Imagining & Sensing:
Understanding and Extending the Vocalist-Voice Relationship Through Biosignal Feedback

Courtney Nicole Reed

Submitted in partial fulfilment of the requirements of the University of London Degree of Doctor of Philosophy

School of Electronic Engineering and Computer Science
Queen Mary University of London

2023

Courtney N. Reed: *Imagining & Sensing*, Understanding and Extending the Vocalist-Voice Relationship Through Biosignal Feedback

Supervised by Prof. Andrew McPherson (primary), Dr. Marcus Pearce

Thesis Submitted: November 16, 2022
Viva Voce: January 26, 2023

Assessment Committee:
Prof. Alexander Refsum Jensenius, *University of Oslo, NO*
Dr. George Fazekas, *Queen Mary University of London, UK*

Abstract

The voice is body and instrument. Third-person interpretation of the voice by listeners, vocal teachers, and digital agents is centred largely around audio feedback. For a vocalist, physical feedback from within the body provides an additional interaction. The vocalist's understanding of their multi-sensory experiences is through tacit knowledge of the body. This knowledge is difficult to articulate, yet awareness and control of the body are innate. In the ever-increasing emergence of technology which quantifies or interprets physiological processes, we must remain conscious also of embodiment and human perception of these processes. Focusing on the vocalist-voice relationship, this thesis expands knowledge of human interaction and how technology influences our perception of our bodies.

To unite these different perspectives in the vocal context, I draw on mixed methods from cognitive science, psychology, music information retrieval, and interactive system design. Objective methods such as vocal audio analysis provide a third-person observation. Subjective practices such as micro-phenomenology capture the experiential, first-person perspectives of the vocalists themselves. Quantitative-qualitative blend provides details not only on novel interaction, but also an understanding of how technology influences existing understanding of the body.

I worked with vocalists to understand how they use their voice through abstract representations, use mental imagery to adapt to altered auditory feedback, and teach fundamental practice to others. Vocalists use multi-modal imagery, for instance understanding physical sensations through auditory sensations. The understanding of the voice exists in a pre-linguistic representation which draws on embodied knowledge and lived experience from outside contexts.

I developed a novel vocal interaction method which uses measurement of laryngeal muscular activations through surface electromyography. Biofeedback was presented to vocalists through sonification. Acting as an indicator of vocal activity for both conscious and unconscious gestures, this feedback allowed vocalists to explore their movement through sound. This formed new perceptions but also questioned existing understanding of the body. The thesis also uncovers ways in which vocalists are in control and controlled by, work with and against their bodies, and feel as a single entity at times and totally separate entities at others.

I conclude this thesis by demonstrating a nuanced account of human interaction and perception of the body through vocal practice, as an example of how technological intervention enables exploration and influence over embodied understanding. This further highlights the need for understanding of the human experience in embodied interaction, rather than solely on digital interpretation, when introducing technology into these relationships.

Acknowledgements

These Acknowledgements were harder to write than any other portion of this thesis. This research is the product of the collective support, encouragement, and love from so many people.

I want to first thank all of the vocalists who lent their time, energy, and investment to this research. You are the new generation of change in how we teach and share our craft with others, and I can see the difference it has made in the voice community already. Thank you for your passion and for your continued work in encouraging others around you to listen to and be kind to their bodies. It goes without saying that the best education and prep for this undertaking came from the wonderful vocal teachers I have had in my life. Thank you to **Carrie, Donna,** and **Randy**, for pushing me into new opportunities and learning with me. Your inspiration is the reason I am here, still singing and feeling the joy of making music above all else.

I want to thank everyone at in C4DM at QMUL, which has become like a second home for me. I never set out to do a PhD when I first came to London, but my time working with **Elaine** during my MSc thesis project inspired me to carry on with my research. Elaine, I am eternally grateful for your encouragement and friendship and for introducing me to this wonderful world of music research. I literally would not be writing this without you and your believing in me. I would also like to thank **Nick, Mathieu, Charis, Paula**, and my wonderful second supervisor, **Marcus**, for their ongoing feedback on this PhD and beyond. Doing a PhD goes beyond just writing the big document - thank you for teaching me to be a well-rounded researcher and person. I've enjoyed working with you and look forward to continuing to do so in the future!

I want to thank my friends in the Augmented Instruments Lab: **Jacob**, for being a calm presence in the next desk over and for all the laughs in London and New Orleans; **Lia**, for being an inspiration in every sense of the word and for always helping me to push myself further and believe in my abilities; **Andrea G.**, for the helpful and insightful discussion and support during some truly tricky performance studies; **Charlotte**, for every wonderful lunchtime break, for helping me to remember to be kind to myself, and for being a source of encouragement always; **Nicole**, for the silliness that comes with making conch shell prototypes and the quiet introspection in a soft-spoken micro-phenomenology interview; **Adán**, for the best hugs, KiCad support, and the most contagious love (and all of the origami rats, which are among my prized possessions). A massive thank you here to **Gala** and **Duna** as well, for being a constant source of fun and friendship, and always cheering me on; **Andrea M.**, for always making me feel included and all the music and invigoration of my performer soul after the long winter of COVID; **Eevee**, for letting us into your mind and the inner workings of some of your amazing NEP7UNO performances, experience querying, and lovely poetry; **Lewis**, for all of the music and good vibes at IKLECTIK and in the performance lab; **Teresa**, for the much-needed heart-to-hearts and coffee drinking; **Franco**, for the laughs over our shared suffering and the plans for future fun in Germany and beyond; **Jack**, for all of the troubleshooting help and great advice in the early days of the PhD; **Giacomo**, for your creativity and for reminding me not to be too serious and to enjoy my journey; **Robbie**, for invaluable guidance with my first research paper and with everything Bela; and **Giulio**, for your sense of humor and saving my ass so many times in the studios I lost count. I tell people about my research lab and they are usually very jealous - rightfully so! Thank you for your friendship, support, and for as much time spent on hard work as having fun. I'd be lucky to work with any of you again, anytime.

There are many others from my time at QMUL who I would also love to thank: **Sophie**: Thank you for the lovely walks around Mile End and for your support, honesty, and helpful advice. I really enjoyed working on the Singing Knit with you and hope we can create more together. **Arooj**: Thank you for keeping me sane and helping me to enjoy the journey with as much silly, spontaneous adventure and fried chicken as possible. **Rida**: Thank you for always being there to listen, for late-night chats and laughs, and for celebrating all of the small things with me. **Aparna**: Thank you for being your goofy, happy self and for always looking after and being there for me.

I would also like to thank my friends at my new home at the Universität des Saarlandes. Thank you for being patient with me while I finished this work. **Paul,** thank you for your encouragement and faith in me, for reviewing a lot of this work, and for already playing so many awesome gigs with me. Special thank you also to **Nihar, Dennis**, and **Alice** for getting me through writing this thesis by distracting me with German puns and Thai food.

I would absolutely have not have accomplished everything I have done during this PhD without the guidance and friendship of my supervisor, **Andrew**. I accidentally ended up in the Augmented Instruments Lab, but you adopted me into the group all the same and I can't imagine having done this thesis otherwise (because I would have likely quit this work 1000 times over, had it not been for your never-ending support). Despite the chaos in the world and in my personal life while doing this research, you were always there for me to vent, to go completely off-topic and talk about orchestra or animals as synthesisers, and to celebrate even the tiniest victories. I appreciate so much how open and honest you are about your experience during your PhD and career and being a fount of personal experience for me when I struggle with imposer syndrome or worry. Doing a PhD can be very lonely and isolating; but, I never felt alone, and I never felt like I couldn't come to you for help. Instead, I felt safe to make errors (all inevitably caught by you almost immediately, especially errors in CAD files), to ramble on about some niche finding I was inspired by, and to try everything in front of me. I really cannot thank you enough for everything.

Above all, I want to thank my chosen family, who were here with me along the way:

Erin: Thank you for your unwavering love over the years and supporting me through all my impulsive endeavors (like deciding to do a PhD). I am looking forward to seeing you and playing more board games, watching cartoons, and eating snacks in celebration. I love you always, my sister and the other pea-in-my-pod.

Clay: Thank you for the video games, the mead, the hugs, and especially the memes. Your humor has been such a light of reassurance and I'm so glad to have you in my life.

Hannes: Thank you for being a good listener and doing your very best to encourage me and keep me on track. I appreciate all of the games we played together, the long conversations about life, and the beer - you kept me from going crazy or just plain giving-up and I won't forget it.

Anja: Thank you for being part of our bubble during the lockdowns and making the most of the little things we did together. All of the baking, dinners, TV binging, and other time we spent together made those dark months less stressful and full of love and laughs.

V: Thank you for being so unapologetically yourself and helping me learn to be the same. I love our time together and our excellent banter, which, like a fine wine, has gotten only better over the years! I can't wait to see what nonsense we get up to next.

Dario: Thank you for being the brother I never had. You inspire me to be silly with reckless abandon and do my best to uphold your mantra to "reduce pain in the world, wherever possible." Thank you for the laughs and hugs and being there whenever I needed.

Inês: Thank you for being such a bright spot in the world. Every time I see you, I always feel better, no matter what nonsense has occurred during the day. I aspire to someday have a whit like yours. You are amazing and make me want to be the best I can be.

Duarte: Thank for just being Duarte - I've enjoyed watching you grow and will be looking forward to being your friend.

Hadeel: Thank you for being the best housemate I could have ever asked for. I absolutely would not have survived the pandemic or this PhD without you, working on your own amazing research beside me. Thank you for the movie nights, the blender, the screaming/crying rota, the hugs, the heart-to-hearts, and for being an inspirational, incredible human being.

JT: For literally everything - for the venting and ranting, the dinners made and coffee brewed, the unrelenting snark, the best hugs, the casual hangs and permanent indent on my couch, and just being there with me every step of this journey. I couldn't have done this without you and I'm looking forward to taking pictures in our cute graduation fits. *Reeeeeeeeeee!*

Emma: For being the absolute best friend anyone could ever have. The lady at Dunkin was half right - we might not be twins, but you are forever my sister. You are my favorite person on earth and I love you so much. And I'm so proud of you for everything you've done lately - you're Thank you for just being yourself and for always thinking of me when you are listening to Beyonce.

And, above all, **Benny**: This PhD is as much an accomplishment of yours as it is mine. You have proofread nearly all of it for me at some point along the way and I'm sure I have spent enough time talking about it that I could probably get you to do my viva for me. We've gotten through a broken leg, a global pandemic, living in separate countries, and probably some other ridiculous things which I have subconsciously blocked out of my mind... plus the completion of this monstrosity of a thesis. You were always there. Thank you for taking care of me and supporting me in a way I have never been supported before, including but not limited to: making sure I eat at appropriate times, giving big hugs, listening to me go off on extensive rants, providing snacks when needed, decorating and DIYing our home to maximum coziness, brewing countless cups of coffee and tea, and for being my best friend. In addition to all of this, you've also grown so much in the last few years and I'm so proud of you for taking new risks and being your authentic self. I've loved being there for the journey and I'm excited for the adventures to come. It's hard for me to say much more because, like the tacit knowledge I will discuss for the next 200 pages, being with you is something wordless and innate. I just know it. I love you.

And finally, to the lovely little friends who greeted me after each difficult day and reminded me that there's no stress in life which cannot be cured with friends, snacks, and naps: **Bear, Goose, Tuna, Beetle, Cricket, Buddy, Pickle, Hugs,** and **Kisses**, I love you very much. *Meep!*

♡

I would also like to thank myself. I love you too.

Licence

Declaration

I, Courtney N. Reed, confirm that the research included within this thesis is my own work or that where it has been carried out in collaboration with, or supported by others, that this is duly acknowledged below and my contribution indicated. Previously published material is also acknowledged throughout.

I attest that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge break any UK law, infringe any third party's copyright or other Intellectual Property Right, or contain any confidential material.

I accept that the College has the right to use plagiarism detection software to check the electronic version of the thesis.

I confirm that this thesis has not been previously submitted for the award of a degree by this or any other university.

November 16, 2022

Courtney N. Reed

"Or music heard so deeply
That it is not heard at all, but you are
The music
While the music lasts."

- T. S. Eliot

Contents

List of Figures

List of Tables

List of Abbreviations

AAF	Altered Auditory Feedback
ANOVA	Analysis of Variance
BAIS	Bucknell Auditory Imagery Scale
DMI	Digital Musical Instrument
DAF	Delayed Audio Feedback
EDI	Electronic Digital Instrument
EMG	Electromyography
Gold-MSI	Goldsmiths Musical Sophistication Index
HCI	Human-Computer Interaction
MIQ-3	Movement Imagery Questionnaire-3
MMIA	Multi-Modal Imagery Association Model
NIME	New Interfaces for Musical Expression
SLHR	Speech, Language, and Hearing Research
sEMG	Surface Electromyography
SMS	Sensorimotor Synchronisation

Chapter 1

Introduction

This research uses the voice to explore how internal interactions are learned and perceived within the body. All of our interactions with our bodies are unique and driven by our own individual life experience (Höök, 2010; Ihde, 1975; Merleau-Ponty, 2014; Tuuri et al., 2017; Wakkary et al., 2018), which often makes understanding the sensations and perceptions of others incredibly difficult (Núñez Pacheco and Loke, 2016). It is often challenging to describe our own tacit and wordless lived experience, even to ourselves (Svanæs, 1997). I investigate the internal understanding and perception which drives our interaction with technology and the world by investigating the interaction which singers have with their voice, as both body and instrument.

1.1 Context

Despite being a part of the physical body, an element of our identity, and an instrument which is culturally widespread (even amongst "non-musicians"), the voice is regarded as highly complex and somewhat mysterious. With physiology being hidden from sight, the vocal apparatus is obscured from the perspective of the listener, who is able to interact with a singer mainly through what they hear. For the singer, the sound and the sensations they feel within their own body provide an ability to interact with their instrument. However, despite being highly personal and often viewed as a representation of the self, the voice is still somewhat removed from the vocalist. Vocalists have refined command over their instrument and body based on internal associations between action and sensory feedback. A vocalist must rely on multimodal aspects of musical imagery — how the singing experience feels, sounds, and appears visually within their mind in order to plan their actions and reactions during performance. Musical imagery is practiced by all musicians and is critical in performance to create a sound that has accurate timing and pitch, in addition to being musically expressive. For the vocalist, musical imagery drives the connection between what they intend and how they create the sound within their body.

However, because the relationship between vocalist and voice is covert and based on internal sensations (Hines, 1983; Jestley, 2011), the singer must be able to translate between bodily, linguistic representations of experience, which are rooted in more abstract mental understandings — although expressed through language and movement, this understanding often is pre-linguistic and instead felt and understood innately, as a result of prior knowledge and experience. Were this another interaction with an instrument or piece of technology, this ambiguity and perception-based relationship would present a massive challenge for communication and design. The interaction with the voice is largely without a physical interface and it is difficult to describe the sensation of singing; however, singers and vocal pedagogy as a whole have been successful at using metaphor to communicate sensory experience between student and teacher. The voice therefore provides us with a unique perspective to examine how we understand, talk about, and teach sensory-based understanding and embodied practice.

1.2 Motivation

The central goal for this PhD is therefore to better understand how we interact with the world around us, including with technology, through the understanding of our bodies. I explore the embodied relationship that vocalists have with their voice, as both their instrument and part of their body. I aim to design and employ novel interaction methods which utilise singers' existing experience and tacit knowledge. I investigate how biofeedback can influence the perception of movement and understanding of the body.

To do this, I study how vocalists are able to control and interact with their bodies through hard-to-articulate or understand, very internal sensory feedback. I have explored these internal associations, their use and formation, and how singers understand, express, and teach this knowledge of their sensory experiences. In understanding how this relationship exists within this specific practice, I discover insights about how humans understand and interact with technology, and how we can encourage connection, communication, and collaboration within our own bodies and the bodies and experiences of others. In this PhD, the voice is explored through several different perspectives, including experiential, physiological, and communicative interactions. Further, I explore how we understand and convey sensory experiences to others in both vocal pedagogy and beyond, and how perception of the body is shaped through experience living in it. The findings of this PhD have implications for vocal and musical pedagogy, human computer interaction (HCI), and design, amongst a number of fun and silly discoveries made while playing and working with the voice.

To explore the internal relationships singers have with their bodies, I also develop sensory measurement tools to measure low-level muscular movement as a way to provide feedback about movement, better connect to the body, explore creative application of subconscious movement, and explore vocalists' perception of their practice and movement. Through externalising internal movement using sonified surface electromyography (sEMG), I aim to give presence to aspects of vocal movement which have become subconscious or automatic. Understanding the rich internal connection with the voice can provide insight into how we understand our own bodies, particularly as we engage in well-defined interaction learned over long periods of time (Cotton et al., 2021b; Reed and McPherson, 2021). Through exploration of the voice, I aim to address larger questions present in HCI of how perception is shaped through experience, and we can communicate sensory-based interaction and practice between different bodies.

This research has a personal motivation for me as a singer. I have worked as a semi-professional vocalist for over ten years, teaching and working within a number of styles and genres. I had the distinct pleasure of getting to test and use much of the knowledge and technology which was developed during this PhD in my own creative and performance time. To me, the connection between the vocalist and voice is one I am very familiar with but one I had difficulty describing or understanding. There are many days where I feel at ends with my body — some days we are instead perfectly in sync, others as if we are completely separate entities. Through understanding better the way that singers relate to their instrument, I hope to further vocal pedagogy and alleviate some of the strain, tension, and often heartbreak that comes when this communication breakdown happens. The voice lesson can be a place for creativity and musical joy, or a hotbed of miscommunication and anxiety when teachers and students do not understand each other. It is my goal through this PhD to better understand how singers, and indeed people as a whole, can better understand each others' experiences and bodies, communicate more efficiently through novel tools, and work with our instruments, rather than against them.

1.3 Research Questions

Vocalists' rich, multi-sensory experience is driven by awareness and understanding of internal sensory feedback and tacit knowledge of the body. Vocal pedagogy uses musical imagery references and metaphor-based teaching and articulation to help describe this experience. The goal of this thesis is to explore how this abstract understanding actually unfolds. Despite its covert nature, vocalists *do* have very strong connections and relationships with this internal body part and instrument. My main research question is:

> **How are vocalists able to control parts of and interact with their bodies through internal sensory feedback, when such feedback and action is hard to articulate or even conceptualise except in abstract representation.**

I break this down into three sub-research questions:

SQ1: How do vocalists use abstract mental representations of their actions, through musical imagery and metaphor, to perform, understand, and speak about their vocal practice?

I explore the way vocalists rely on their mental musical imagery to perform, for instance relying on internal understanding of their practice in situations where their auditory feedback is disrupted and they cannot reply on external feedback. I also explore how vocalists perceive and understand their movements, and how do they articulate and communicate their sensory experiences to others; more specifically, I examine the way that voice teachers describe and explain their practice so that the experience can be understood by others — their students.

SQ2: How can we use biosignals to capture the internal sensory experience of singing by conveying aspects of low-level muscular movement during singing?

In order to pull some of these internal sensory experiences out of the body, I use surface electromyography (sEMG) to detect unconscious internal movements. I explore how effectively sEMG can be used to measure the movements of the laryngeal muscles. As well, I examine how this method of biosensing can reveal subvocal movement, for instance during audiated mental rehearsal, or unconscious behaviours which are rooted in tacit knowledge.

SQ3: How does the real-time sonification of sEMG signals influence perception of movement and create new connections between the vocalist and their body?

Finally, I explore how these biosignals can impact the perspectives of singers when they are moved from an internal feedback source to an external one through sonification. I explore which aspects of movement are revealed to the singer as they become aware of the unconscious movements inside their body through sonified sEMG feedback. As well, I uncover ways in which singers perceive control over and connection to their body when incorporating this biosignal feedback into their practices.

1.4 Contributions

The results of this PhD tell us more about how awareness of movement and control over the voice engages the body, impacts vocalists' perception of their action, and how technology can be incorporated into artistic practices to complement and extend existing practices. This understanding of the body is further extrapolated to the wider HCI community:

This research contributes to existing work within the HCI topics of interactive biosignal sensing, interactive arts and musical instrument design. I create a system for acquisition and amplification of the laryngeal muscles through a new musical system, the VoxEMG board, for dedicated sEMG processing and wearable implementation, and the Singing Knit, a soft-knit wearable for vocal sEMG interactions. Additionally, I provide the experience of a long-term first-person reflection on the incorporation of biosignal feedback in understanding my own body movements and vocal practice, as well as in-depth work with other vocalists on how such feedback can function as a metaphor and allow singers to investigate their movements and embodied relationships. From this, I also propose a strategy for adopting technologies from other practices into traditional arts and other contexts through the use of soft wearables. This culminates in a comprehensive examination of how technology and the feedback we receive in HCI can shape our perception and understanding of our bodies and our actions.

The methods and findings of this PhD are also contextualised within design research and elaborate on the inclusion of our lived experiences and perception of our movement and action when creating new digital technologies. This thesis delves into the relationship between interaction, digital or otherwise, and how it shapes our awareness, understanding, and feeling about our bodies and ourselves. I provide an analysis of how metaphors used in fundamental vocal pedagogy and how metaphorical communication between humans works. This work contributes to the wider interaction design community through this examination of interaction modalities and the communication and teaching of sensory-based experience in interaction. I discuss the ways in which singers understand and describe their experience and propose novel ways in which we can structure interaction with technology to aid in sensory communication between different people. I believe this research to be beneficial to not only understanding our own behaviours and body relationships, but for better understanding each other's experiences and better communicating sensory-based experiences across our individual bodies.

Finally, the study of musical imagery applications will allow for more insight on how imagery and its cultivation and strengthening can be taught in vocal pedagogy and beyond. The thesis provides novel understanding in how multi-modal imagery allows singers to perform tasks and remain accurate in their performance by relying on internal mental understanding, even when external feedback is disrupted. This is particularly useful in cognitive science and in the use of technology for learning applications.

1.5 Thesis structure

The body of this thesis uses an inter-disciplinary approach to address these research questions and aims. This topic covers a wide breadth of different research perspectives, including vocal physiology, interactive arts technology, musical instrument design, cognitive science, HCI and design research, and vocal pedagogy. I begin with a literature review in two chapters — the first introduces the voice, its physiology, and methods which have been used to interact with it in and outside of musical applications. The second delves more into qualitative approaches which have been taken to explore the experiential aspects of musical and vocal interaction. To unite these fields of research, I present a methodology to be used for the thesis, which balances the differing backgrounds to explore specific perspectives of vocal interaction and how this study introduces a valuable paradigm for design and human-computer interaction research. This thesis involved six major studies, which are discussed in Chapter 5 through Chapter 9. A brief summary of the subsequent chapters:[1]

Chapter 2: Vocal Controllers
This chapter examines the objective qualities of the voice, including physiology and how the functions of the body drive vocal production. I outline studies of the vocal physiology and how the voice has been both controlled and used as a controller itself in different musical applications. This section also introduces sEMG as a biosensing method used in the study of vocal interaction.

Chapter 3: Vocal Phenomenology
This chapter provides background on the experiential and perceptual aspects of singing in existing research. I outline relevant background research the domains of cognitive science, vocal pedagogy, and psychology. By uniting these disciplines within this research context, I demonstrate how vocal interaction is rooted in musical imagery and how this interaction is typically perceived by vocalists.

Chapter 4: Methodology
Given the broad related work which formed this thesis, this chapter outlines the methods to be used throughout the research. I here describe and balance the various approaches, both quantitative and qualitative, and discuss their place in HCI and design practices. To this end, I describe how this research in vocal interaction reveals aspects of human interaction that are necessary to consider in HCI and the design of systems which work with the body and our perception of it.

Chapter 5: Vocalists' Use of Auditory Imagery
This chapter outlines the first exploratory study done to investigate these questions - to determine the link between auditory imagery and vocalists' ability to sing under altered auditory feedback (AAF). The study revealed that, while greater auditory imagery ability allowed vocalists to perform more accurately with pitch shifted AAF, the effect on accuracy with delayed AAF was not related. This informed further study related to tactile links and proprioceptive feedback in singing.

Chapter 6: Understanding Vocal Perception
This chapter presents a study about vocalists' perception of their own movement and interaction while singing; working with vocal teachers, I expanded on the current knowledge of how vocalists use abstract language and gesture as metaphorical representations. This chapter highlights how vocalists' understanding and communication of their practice is both pre-linguistic

[1] In these chapters, I will often switch to "we" to discuss the work of myself and my collaborators.

and non domain-specific, allowing for a fluid, adaptable representation of action for both the self the understanding of students.

Chapter 7: Surface Electromyography for Vocal Interaction
To move towards the body and away from audio-based vocal controllers, I designed a system for measuring aspects of physical laryngeal movement while singing through surface electromyography (sEMG). This chapter outlines two studies related to the creation of a novel vocal controller – first, I discuss the design and use of the VoxEMG board for vocal sEMG acquisition and pre-amplification, to observe both subvocal and vocalised singing. Additionally, I present the integration of the system as the Singing Knit wearable for vocal interaction.

Chapter 8: Autoethnographic Interaction & Evaluation
Using the VoxEMG, I conducted a long-term autoethnographic study of my own practice and how I was able to incorporate audio feedback controlled by the sEMG signals from my own laryngeal movement. This first-person use informed the design of the board in further iterations. As well, the feedback allowed me to explore, question, and challenge the embodied relationships with my own voice.

Chapter 9: Exploring Vocal Movement & Perception
This chapter presents a study of other vocalists' practice and perception while working with sonified sEMG signals, similar to my own exploratory practice. Vocalists received the VoxBox, a vocal EMG kit, and worked with sonified feedback of their own movements over an extended period of time. This chapter explores the embodied relationships, challenges, and discoveries made about vocalists' expectations and perception of their bodies during their practice and informs further work in performance and educational contexts.

Chapter 10: Discussion
Finally, this chapter summarises the findings and contributions of this thesis, returning to the research questions outlined in this Introduction chapter and how they have been addressed through these studies. I here outline other implications for HCI derived from this thesis, as well as a few personal reflections before concluding this work.

1.6 Associated Publications & Presentations

This PhD research has been included in international scholarly publications and presentations. As well, I have presented this research at several academic venues, including workshops and invited talks, and also in a few musical performances.

The publications which are derived from the studies conducted in this PhD are provided again at the start of their relevant chapter for easy reference.

1.6.1 Journal Proceedings

Courtney N. Reed, Marcus T. Pearce, and Andrew P. McPherson. Auditory imagery ability and singing accuracy with altered auditory feedback. *Under review in Musicae Scientiae.*

1.6.2 Conference Proceedings

Courtney N. Reed, Paul Strohmeier, and Andrew P. McPherson. 2023. Negotiating Experience and Communicating Information Through Abstract Metaphor. In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23). Association for Computing Machinery, New York, NY, USA, Article 185, 1–16. DOI: 10.1145/3544548.3580700

Courtney N. Reed and Andrew P. McPherson. 2023. The Body as Sound: Unpacking Vocal Embodiment through Auditory Biofeedback. In Proceedings of the Seventeenth International Conference on Tangible, Embedded, and Embodied Interaction (TEI '23), February 26-March 1, 2023, Warsaw, Poland. ACM, New York, NY, USA, 15 pages. DOI: 10.1145/3569009.3572738

Courtney N. Reed, Sophie Skach, Paul Strohmeier, and Andrew P. McPherson. 2022. Singing Knit: Soft Knit Biosensing for Augmenting Vocal Performances. In Augmented Humans 2022 (AHs '22), March 13–15, 2022, Kashiwa, Chiba, Japan. ACM, New York, NY, USA, 20 pages. DOI: 10.1145/3519391.3519412

Courtney N. Reed. 2022. Examining Embodied Sensation and Perception in Singing. In Proceedings of the Sixteenth International Conference on Tangible, Embedded, and Embodied Interaction (TEI '22), February 13-16, 2022, Daejeon, Republic of Korea. ACM, New York, NY, USA, Article 47, 1–7. DOI: 10.1145/3490149.3503581

Courtney N. Reed and Andrew P. McPherson. 2021. Surface Electromyography for Sensing Performance Intention and Musical Imagery in Vocalists. In Proceedings of the Fifteenth International Conference on Tangible, Embedded, and Embodied Interaction (TEI '21), February 14–19, 2021, Salzburg, Austria. ACM, New York, NY, USA, 11 pages. DOI: 10.1145/3430524.3440641

Courtney N. Reed and Andrew P. McPherson. 2020. Surface Electromyography for Direct Vocal Control. In Proceedings of the International Conference on New Interfaces for Musical Expression (NIME '20), July 21-25, 2020, Birmingham, UK, pp. 458–463. DOI: 10.5281/zenodo.4813475

1.6.3 Presentations

Courtney N. Reed. 2022. Sensory Sketching for Singers. In 2022 CHI Workshop on Sketching Across the Senses, April 22, 2022, New Orleans, LA, USA. ACM, New York, NY, USA, 4 pages. [Paper]

Courtney N. Reed. 2022. Communicating Across Bodies in the Voice Lesson. In 2022 CHI Workshop on Tangible Interaction for Well-Being, May 1, 2022, New Orleans, LA, USA. ACM, New York, NY, USA, 4 pages.

Courtney N. Reed, Andrew P. McPherson, and Marcus T. Pearce. 2021. The Role of Auditory Imagery and Altered Auditory Feedback in Singers' Timing Accuracy. In 16th International Conference on Music Perception and Cognition/11th Triennial Conference of the European Society for the Cognitive Sciences of Music (ICMPC16-ESCOM11), July 28-31, 2021, Sheffield, UK. **Best Paper Nomination**.

Courtney N. Reed. 2021. Translating the Body: Abstract Language in the Teaching of Fundamental Vocal Pedagogy. Presentation at the Language, Discourse and Communication Doctoral Lab, May 6, 2022, King's College London, UK.

1.6.4 Performances

The vocal sEMG system designed in this thesis has also been a part of a few musical performances in the research and arts community:

Duet with Andrea Martelloni (augmented guitar), Music from the Augmented Instruments Lab, performance at IKLECTIK Art Lab (London, UK), 23 November, 2021. [Link]

Duet with the Singing Knit with Paul Strohmeier (drum machine), performed at the 2022 Augmented Humans Conference at LMU München (Munich, DE), 14 March, 2022 [Link]

Solo set with the Singing Knit, Music from the Augmented Instruments Lab, performance at IKLECTIK Art Lab (London, UK), 23 April, 2022.

1.6.5 Other Research & Collaborations

In addition, I have also been involved in a few collaborations researching related music technology and HCI topics while doing this PhD:

Micro-phenomenology, as both a psychological discipline and an evaluation method for studying pre-reflective aspects of sensory interaction:

Courtney N. Reed, Charlotte Nordmoen, Andrea Martelloni, Giacomo Lepri, Nicole Robson, Eevee Zayas-Garin, Kelsey Cotton, Lia Mice, and Andrew McPherson. 2022. Exploring Experiences with New Musical Instruments through Micro-phenomenology. International Conference on New Interfaces for Musical Expression (NIME '22). DOI: 10.21428/92fbeb44.b304e4b.

Explainable AI (XAI), working with human perception and understanding to make generative musical models more understandable to non-experts and usable in musicking environments:

Nick Bryan-Kinns, Berker Banar, Corey Ford, **Courtney N. Reed**, Yixiao Zhang, Simon Colton, and Jack Armitage. 2021. Exploring XAI for the Arts: Explaining Latent Space in Generative Music. 1st Workshop on eXplainable AI Approaches for Debugging and Diagnosis (XAI4Debugging@NeurIPS2021), December 14, 2021, Online. [Paper]

Media and Arts Technology, scientific, inter-disciplinary methodologies surrounding research and design in music, fashion, and the fine arts:

Nick Bryan-Kinns and **Courtney N. Reed**. 2023. A Guide to Evaluating the Experience of Media and Arts Technology. *Creating Digitally: Shifting Boundaries: Arts and Technologies - Contemporary Applications and Concepts.* Intelligent Systems Reference Library No. 241. Springer.

Music Emotion Recognition (MER), understanding the role of different musical and performance features in human perception of emotional qualities in music:

Simin Yang, **Courtney N. Reed**, Elaine Chew, and Mathieu Barthet. 2021. Examining Emotion Perception Agreement in Live Music Performance. IEEE Transactions on Affective Computing, Early Access. DOI: 10.1109/TAFFC.2021.3093787

Courtney N. Reed, Simin Yang, Elaine Chew, and Mathieu Barthet. 2019. The Listening Participant and The Living Instrument: A Thematic Analysis of Why Listeners Actively Annotate Emotion Changes Through Live Performance. Workshop-Symposium on Research Methods in Music and Emotion, September 14, 2019, Durham, UK.

Simin Yang, **Courtney N. Reed**, Elaine Chew, and Mathieu Barthet. 2019. Listener-informed Features for Time-varying Emotion Perception in Live Music Performance. 7th Seminar on Cognitively Based Music Informatics Research Workshop (CogMIR), August 8, 2019, New York, USA.

Chapter 2

Vocal Controllers

Designing Technology for Vocal Physiology

To begin this thesis, I will introduce the voice as a part of the body and the way that we discuss it in both a musical and physiological sense. I first present relevant research on the physicality of performance and musical gesture. With this general background, I will then outline the vocal anatomy and the roles of the muscular system to produce our voices. This will involve some literature from medical contexts and the study of the vocal apparatus in action. Then, I will move toward the musical components of singing as they relate to this physiology, for instance defining aspects of voice quality, vocal quality, and emotional quality. This is to give a better sense of some of the vocabulary which singers and musicians use to refer to the voice.

2.1 Embodied Musical Gestures

Leman and Godøy regard the musical experience as "inseparable from the sensations of movement" and that movement during any musical experience is an expression of full engagement with listening and performing (Godøy and Leman, 2010). These movements can include head bobbing or nodding, swaying of the torso, flourishes of the hands, movement of the instrument itself, or even facial expression and eye contact. Gesture as an aspect of movement during musical experience falls into three general categories (Godøy and Jørgensen, 2001; Godøy and Leman, 2010): (1) *effective gestures*—those caused by the physical demands of playing an instrument and generating sound, (2) *accompanist gestures*—those used to strengthen and facilitate sound production, and (3) *figurative gestures*—ancillary movements not related to the sound that may occur as a result of musical interpretation and expression (Buck et al., 2013; Davidson, 2012; Delalande, 1990).

2.1.1 Expressive Gesture in Performance

Figurative expressive gestures — those ancillary gestures which arise from expression and interpretation — can be defined by whether they support the encoding or decoding of the expressive material in the music. Gestures that support encoding, *expression-supporting gestures*, occur during performance by the performer. On the opposite end, gestures that support decoding, *expression-responding gestures*, occur during listening, when the audience interprets the expression in the music (Leman and Maes, 2015; Leman et al., 2017). The transfer of the expression from performer to listener during the musical experience indicates an idea of "mirroring," where the quality of the sound patterns are related to the quality of the resulting movement patterns (Davidson, 2012; Huberth and Fujioka, 2018; Leman et al., 2017); in other words, the body moves with the musical line, whether that be the performer moving the bell of their instrument through execution of the musical passage, or in the similar swaying of the listener's head while they listen from the other side of the hall. This encoding of emotional information is entirely on the performer in the gestures they choose to exhibit, and is found to be intentional and based heavily on the musical context (Lehmann et al., 2007; Leman, 2008; MacRitchie and Zicari, July 23–28, 2012; Mazzola, 2002).

Interestingly, even gestures with little or no effect on the sound produced are still believed by musicians to add emotional context to the sound; in the case of pianists presented by Doğantan-Dack (2011), although it has been found that there are no changes in sound produced depending on anything other than the velocity of the key press itself, pianists truly believe that the way they touch the key can have an impact on the sound's quality and emotion (Bernays and Traube, Augustt, 2011). In the same way, composers will indicate in scores markings that dictate how the pianist should press the keys to add emotional content (Doğantan-Dack, 2011). Doğantan-Dack (2011) goes on to remark that, given this conceptualisation by pianists, the entire interpretation of a musical piece still depends on this kinaesthetic relationship between the touch and the sound produced, regardless of whether or not there is a sonically perceptible effect.

The repeated patterns of gesture which correspond to the phrasing of a musical passage are deemed *phrasing gestures*. Phrasing gestures are thought to accomplish two tasks: (1) to assist with the performer's timing and consistency through the repeated phrase, and (2) to strengthen to the listener the definition and meaning of the phrase (Buck et al., 2013; Colley et al., 2018b; Goebl and Palmer, 2008; Huberth and Fujioka, 2018; MacRitchie et al., 2009). In expressive gesture, the movement can relate to the features of the music and also exist independently of them, thus relating information to the listener about the tension and release within the music (Davidson, 2012). Individual performers are found to exhibit individual phrasing gestures (which is proposed to relate to the individual interpretation of the line, as well as to the specific understanding each performer utilises); however, the patterns of motion are repeated through successive repetitions of the phrase, furthering the connection between body and interpretation during a performance as one cohesive network (Buck et al., 2013; Huberth and Fujioka, 2018; Keller and Appel, 2010; Leman and Maes, 2015). It has been found specifically that body sway is used to solidify timing and anticipation of events, so-called "forward" internal models (Keller and Appel, 2010; Leman and Maes, 2015; MacRitchie and Zicari, July 23–28, 2012). Internal understanding, mental representations, and gesture are further linked through imitation; where the metaphor of a perfect sound may be derived from the playing of a particular teacher or mentor, a musician may further extrapolate on this image by imitating the gesture of a familiar instructor or idol (Leman, 2008). Ultimately these gestures and expressions may spread through a genre as student-teacher relationships pass on this understanding — these mental models form our mental imagery, which is further discussed in regards to cognitive science in Chapter 3, Section 3.1.

Figurative ancillary gestures for expression are studied greatly in their ability to unite the interpretation of a musical passage between duets or groups of musicians. Because each individual will have a different visualisation of the quality and direction of the music to be played, musicians performing together must communicate their intentions via gesture to the others in the group in order for the joint sound to act as one cohesive unit. Musicians are found to utilise gestures such as eye contact, swaying of the torso, and head movement such as nodding as visual cues in this way (Davidson, 2012; Eerola et al., 2018; Keller and Appel, 2010).

Gestures may also take place as a form of mimicry; social psychology has focused on group motor action as a spontaneous reaction to observing the same reaction in others. This is studied commonly through observation of pain: when observing someone to be experiencing injury, it is common to wince without experiencing the pain firsthand (Bavelas et al., 1986). This reaction as a form of non-verbal communication relies heavily on the ability to see the other person and share eye contact. A particular study by Bavelas et al. (1986) reveals in this context that this type of behavior takes place within a social context as a way of relating to others around in order to convey a sense of understanding: "I am like you, I feel as you do." Facial mimicry is observed in musical contexts, with audience members' facial expressions mirroring those of the performers, who's facial expressions are used to emphasise the emotion and facilitate reflection in the audience's perception (Chan et al., 2013).

2.2 Vocal Physiology

Phonation in singing requires the interaction of two systems: (1) the vocal folds in the larynx and vocal tract (which includes the soft tissues in the throat and mouth) and (2) the respiratory muscles. The respiratory system provides airflow through the larynx, causing the vocal folds to vibrate. This base vocal sound is further shaped by the vocal tract (Hardcastle, 1976; Sundberg, 1994) and provides each individual with their unique voice characteristics (Sundberg, 1994). I will use *vocal folds* as a more anatomically accurate descriptor of the vocal mechanism than the better known *vocal chords* through the remainder of this paper, but the terms are synonymous in practice.

It is important to note going forward that vocal physiology is dependent on genetics and, being rooted within the body and acting as a means to self-express, the voice as an instrument is highly personal; vocalists often feel that their singing is a part of their identity and vocal health is a critical daily practice (Achey et al., 2016; O'Bryan, 2015; Prem and Parncutt, 2008). All of these processes are impacted by individual physiology, as well as other factors such as hydration, nutrition, hormones, and fatigue, making singing a rather athletic act which requires fine motor control.

2.2.1 The Larynx

Vocalisation (in speech, singing, or any other vocal sounds) occurs within the larynx. The movement of the two vocal folds and supralaryngeal vocal tract are responsible for pitch, loudness, and quality of the tone produced during singing (Bouhuys et al., 1966). The folds themselves are composed of bundles of muscle, with each bundle responsible for a specific positioning of the fold (Zhang, 2016a). Within the space between the two opposing vocal folds is the glottis, which is composed of membrane and cartilage (Zhang, 2016a). Movement of the larynx (Figure 2.1) is controlled by two groups of muscles: the extrinsic, which indirectly affect the vocal chords as they attach to the cartilage surrounding the folds, and the intrinsic, which control the tension and position of the folds (Table 2.1 and Table 2.2) (Hardcastle, 1976). The extrinsic muscles are more accessible (Figure 2.2), as they lie below the skin and do not require medical intervention to observe.

The glottis normally forms an open space at rest (abducted); tone vocal tone is a result of the folds being brought together to close the glottis (adducted) (Story, 2002). The tone produced is reliant on the subglottic pressure (P_S), the increase of which causes the elastic vocal folds to be pushed apart and pulled back together at regular intervals. The escape of air at this periodicity of push and pull causes vibrations above the glottis—the frequency of the vocal fold movements is the fundamental frequency of the tone produced (Bouhuys et al., 1966; Zhang, 2016a). P_S is essential also in normal speech production to produce a sound source, which is further modified by the vocal tract (García-López and Gavilán Bouzas, 2010). The amplitude, phase, and anterior-posterior symmetries of the vocal folds are not found to vary amongst individuals, while degree of glottic closure (how much the of folds touch during vibration) to achieve different frequencies varies, especially for female singers (Gelfer and Bultemeyer, 1990). This would result in slightly breathier tones for higher pitches in some female singers, as more air passes through the larynx without vibrating the folds.

Changing the tension in the vocal folds will thus change the resultant pitch. It is common to find that untrained singers increase P_S and tension in the vocal folds to increase pitch (Howard, 2009),

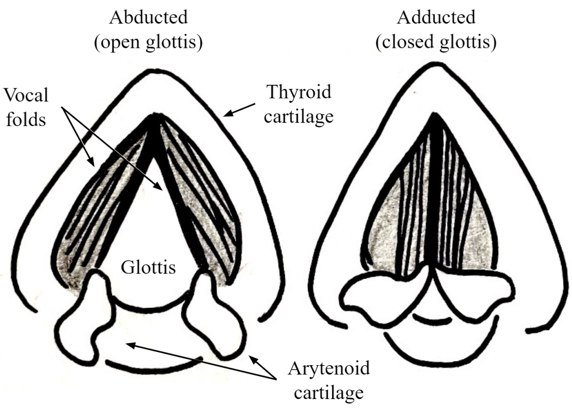


Figure 2.1: Superior view of the larynx when the vocal folds are abducted (left) during respiration and adducted (right) during speech and singing.

Extrinsic Muscles	
Laryngeal Motion	**Muscle**
Elevators (Suprahyoids)	digastricus geniohyoideus mylohyoideus genioglossus hyoglossus stylohyoideus constrictor pharyngis medius
Depressors (Infrahyoids)	sternohyoideus omohyoideus thyrohyoideus sternothyroideus

Table 2.1: Extrinsic laryngeal muscles and corresponding movements.

Intrinsic Muscles	
Laryngeal Motion	**Muscle**
Sphincter muscles (of laryngeal inlet)	aryepiglotticus thyroepiglotticus
Abductor	cricoarytenoideus posterior
Adductor	cricoarytenoideus lateralis arytenoideus transversus arytenoideus obliquus
Tensor	vocalis cricothyroideus
Relaxer	thyroarytenoideus externus

Table 2.2: Intrinsic laryngeal muscles and corresponding movements.

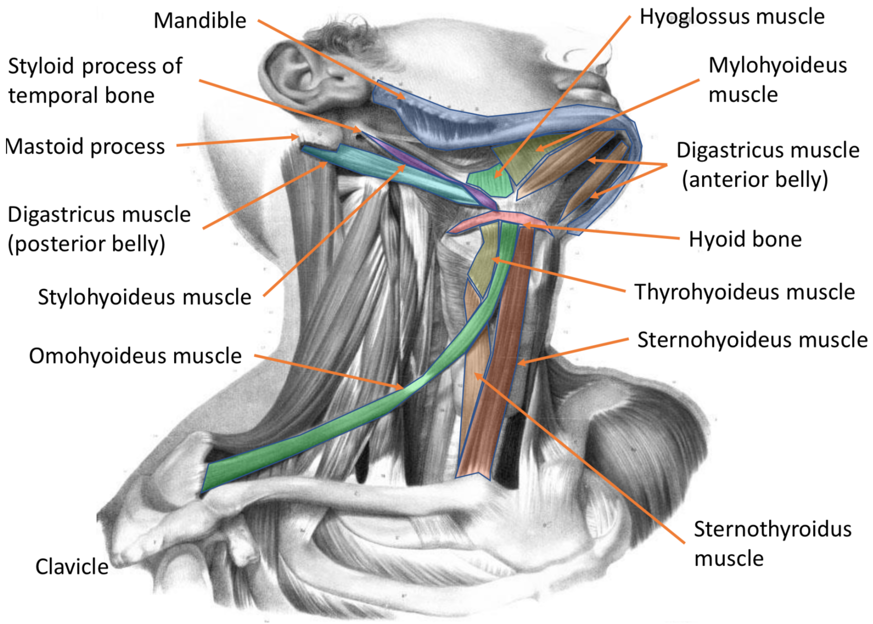


Figure 2.2: Extrinsic muscles of the larynx (adapted from image available in the public domain, retrieved from Flickr (Qasim Zafar): https://flic.kr/p/u6pAzL).

although this can be potentially damaging to the soft flexible tissue and result in vocal nodules (hardened areas on the folds that make it difficult for the folds to vibrate together). Classically trained vocalists are taught to elongate the vocal folds by contracting and lowering the cricothyroid (CT) muscle, which places less strain on the vocal folds and creates more stable P_S in the changing of pitch (Erickson et al., 1983; García-López and Gavilán Bouzas, 2010). In order for the tone and loudness to remain constant, the P_S must also remain constant (Bouhuys et al., 1966).

The supraglottic tract goes from the vocal folds to the mouth and the nose and is responsible for resonance and harmonic boosting of the voice (García-López and Gavilán Bouzas, 2010). The supraglottic tract provides what are commonly referred to as "voice formants." These are groups of boosted harmonics that give clarity and definition, as well as projection to the voice. The first two formants, with the deepest harmonics, are responsible for definition of the sound (or understanding of a word, when spoken), while the higher formants are responsible for tone and individual characteristics of each voice (García-López and Gavilán Bouzas, 2010). Formants and their amplification can be controlled in professionally trained vocalists, who can control the movement of their larynx (as described above) and the soft tissues of the mouth (the tongue and soft palate). Nasal qualities resultant from incorrect positioning of the mouth and jaw, and subsequent resonance in the nasal cavities, are common in untrained singers (Howard, 2009). Additionally, it is believed that classically trained vocalists have control over what is known as the "singer's formant"—a boost in the energy projection of the voice at around $2,500-3,000$ Hz (García-López and Gavilán Bouzas, 2010; Weiss et al., 2001). This particular formant is what allows a soloist to sing with projection over a full orchestra without the use of microphones or amplification. Lowering the larynx position allows for a wider pharynx to achieve this specific formant; this is one of the believed conditions for a singer's formant, although this is not consistently observed to be the case in high-register singers—sopranos and tenors (Pabst and Sundberg, 1993).

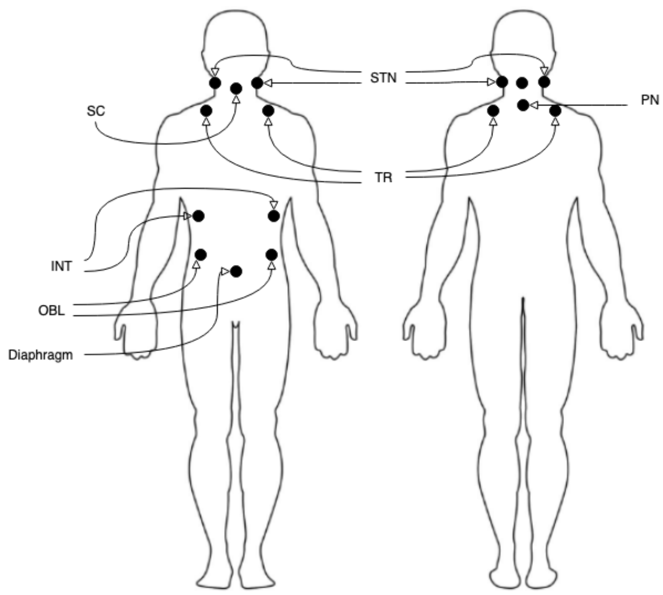


Figure 2.3: Approximate location of the thoracic and neck muscles activated during respiration in singing on the anterior (left) and posterior (right).

Because the thickness and size of the vocal folds and tract, as well as the other muscles and tissues involved in singing, are dependent on genetics, each individual will have a different resultant sound based on this physiology (García-López and Gavilán Bouzas, 2010; Zhang, 2016b). The length of the vocal folds also determines the timbre and range of the voice, and thus ultimately the voice classification. A correlation exists between the anterior-posterior diameter of the glottis and the trachea; measurements of vocal fold lengths to this diameter support the idea that singers in the same voice classification share consistent anatomical features (Roers et al., 2009).

2.2.2 Respiratory Patterns in Singing

In order to keep P_S constant and sustain sung tones, the respiratory muscles must act as a set of bellows to provide support as lung volume decreases and the chest wall's elasticity returns it to normal shape (Bouhuys et al., 1966; Griffin et al., 1995). At rest, inhalation is active while exhalation is passive. While speaking, and even more so during singing, the process of exhalation becomes more critical for maintaining P_S and duration of airflow, and so exhalation becomes more active (García-López and Gavilán Bouzas, 2010). Greater elastic recoil forces in the muscles during the singing of musical passages mean lung volume must be taken into greater account during singing than during speaking, with female singers consuming more air and using wider lung volume ranges than male singers (Thomasson and Sundberg, 1997).

The major muscles in the thorax involved in respiration are the diaphragm and intercostals (INT), with support from the lateral abdominal/obliques (OBL). In the neck, the sternocleidomastoideus (STM), scalenus (SC), upper trapezius (TR) and the muscles of the posterior neck region (PN) are also active during singing (Figure 2.3).

Most singers are taught that the diaphragm is responsible for controlling airflow during exhalation, and that pressure resulting from its tension provides the support; in actuality, the diaphragm relaxes at this point, while the OBL and internal INT muscles control the exhalation process (García-López and Gavilán Bouzas, 2010; Thorpe et al., 2001). The diaphragm is employed mostly during inhalation to re-inflate the lungs (Watson and Hixon, 1985). During singing, the movement of the diaphragm has been found to become dissociated with the movement of the other inspiratory muscles (Bouhuys et al., 1966; Salomoni et al., 2016). The other abdominal muscles (INT and OBL) work to set the posture necessary to carry out proper supported breathing and, along with the rib cage, airflow control (Watson and Hixon, 1985; Watson et al., 1989).

Classical singers are also found to support their singing with greater percentage of abdominal contribution to lung volume and greater asynchrony between movements of the rib cage and abdomen. This is thought to increase the length and pressure generation in the rib cage for voice quality (Salomoni et al., 2016), and provide lifting of the rib cage to activate the abdominal muscles for increased projection and strengthening of the singer's formant (Thorpe et al., 2001). Phasing of the TR, STM, SC, and PN have also been discovered to work in conjunction with the OBL and INT muscles, and are proposed to increase efficiency in breathing by reducing the space between positions required for maximum inhalation and exhalation (Pettersen and Westgaard, 2004, 2005). The activity of all neck muscles is found to increase consistently when singing at higher pitches (Pettersen and Westgaard, 2005).

The "breath pressure" in supported breathing is found to provide different spectral characteristics, higher sound pressure level (SPL), peak airflow (airflow through the glottis), and P_S than the unsupported voice (used often by vocalists without classical training) (Griffin et al., 1995; Thorpe et al., 2001). It has been found that male singers rely more on this breath pressure to support singing in upper ranges, where female singers utilise more movement at the end of the airway in the jaw and tongue to support the resonance and SPL of the singer's formant without increasing P_S (Griffin et al., 1995; Weiss et al., 2001). Supported breathing from the abdomen is core to formal voice training in order to achieve these distinctions in sound quality and power and breath capacity to sing through longer phrases. Posture and alignment of the body is essential to vocalists: standing supported on both feet allows the diaphragm and lungs to expand and maintain this tension for vocal emission (García-López and Gavilán Bouzas, 2010). Keeping the chin down and neck without tension provides more space for lowering the larynx (Pettersen and Westgaard, 2004). It is common for singers to practice in front of a mirror or standing with their backs flush against a wall so they can monitor their posture while practicing. Students in the early stages of learning voice technique may also practice while the instructor repositions their head and shoulders to reduce tension.

2.3 Vocal Gesture

In the context of singing, we can divide these physiological activities into the types of gesture outlined in Section 2.1. The movements associated with the larynx, vocal tract, and respiration as outlined in Section 2.2 are effective gestures and help to generate and shape sound; for instance, intentional movement of the soft tissues in the mouth to change vowel sounds can be classed as effective gestures. Accompanist gestures might comprise actions such as raising the eyebrows to facilitate the production of higher pitches or adjustments to the body's posture to provide space in the chest cavity and release tension. Finally, figurative gestures encompass expressive physical reactions to expressive markings in music, such as changing the facial expression, or body sway.

2.3.1 The Challenge of Vocal Gestures

The challenge in developing digital musical instruments (DMIs) and musical controllers for the voice specifically (and particularly the larynx) come from the facts that 1) much of the voice is not outwardly visible or measurable, for instance in the way moving a limb might be, and 2) that most of the study of the physical vocal activity has been done in a medical context, which is inapplicable or impossible during performance. The covert movement and sensory experience of singing, combined with refined technique and intricate physiological movement, provide a difficult yet intriguing prospect for musical interactions. In order to digitally expand the capabilities of the voice while preserving the emotional communication and technique in singing as a musical craft, the inherent challenges of interfacing with the voice must be tackled: how do you design for something which you can neither see nor touch? Up to now, the voice controllers have revolved around more overt audio analysis and feature extraction. Refocusing design around the physiological interactions previously discussed, independent from the audio produced, can provide a means of direct control and get more to the root of the vocalist's actions and intentions.

The 2003 NIME paper by Michael J. Lyons well summarises the core issues with physiological interaction in voice controllers: "Current ways of interacting with computers neglect most of physiology of human-human interaction and are surely unsuitable for most forms of communication, especially expressive forms such as music" (Lyons et al., 2003). Emotional expression and communication in singing is hypothesised to utilise existing neural pathways from verbal communication for encoding and interpreting emotion in speech (Juslin and Laukka, 2003; Juslin and Västfjäll, 2008), making this interaction especially critical in vocal music. Current vocal interfaces can be broken down into two main categories: controllers which use audio characteristics of the voice to control another instruments, or model-based vocal synthesis controllers which use other forms of interaction, mainly the hands; thus, there is a present gap in the utilisation of vocal technique and clear need to center the vocalist in the design of voice controllers. I will introduce firstly the distinction between direct and indirect control and the different controllers which have been developed for the voice up to this point. Then, I will describe how sEMG provides a way to bridge this gap in voice controllers and to provide a source of direct control and a sense of tangibility to the voice.

2.4 Direct and Indirect Control

The distinction between direct and indirect control is well-defined in HCI: coined by Shneiderman, *direct manipulation* describes user actions which are rapid and mimic real-life interactions with objects in an incremental and easily reversible way (Shneiderman and Maes, 1997b). An example would be a touch screen, which allows users to directly "touch" objects to open them. Indirect manipulation instead involves an intermediary stage where some translation must occur between the user and machine. Instead of touching an application, a user could indirectly open them at the command line.

In direct control, as the name implies, the action of the user directly manipulates the machine; for instance, the touch screen on most modern phones allows the user to directly "touch" an application to open it or type onto a digital keyboard. Direct control mimics real-life interaction and so allows for ease of use and understanding, as well as ease of amendments to the change made. In contrast, indirect control involves an intermediary stage where the action of the user must be translated into the output of the machine; a basic example would be the volume control on most digital devices—a slider or knob which moves left to right yet brings the volume down or up. The user is controlling an object which only loosely mirrors the action being performed and the directional difference between left/right and softer/louder must be translated with some mental effort. This example is a fairly intuitive one, but indirect control can become more complicated in contexts such as working at the command line; the connection between action and result becomes much less clear and thus carries a heavier cognitive load.

In musical interfaces, we thus define this intermediary translation stage which exists in indirect control as feature extraction. Indirect control would be audio-signal-driven sound synthesis (Pöpel and Dannenberg, 2005) where analysis and parametrisation of sound drives interaction; features are extracted from input audio and mapped to other elements of synthesised output. Truly indirect control-based interfaces would include instruments such as the MIDI guitar, which uses audio signal to generate symbolic MIDI data (Verner, 1995) or Max Mathews's electronic violin through which filter parameters are controlled by audio amplitude (Mathews, 1984). Direct control would not involve this parametrisation. A direct control comparison to the MIDI guitar would be the K-Bow, a violin bow controller which generates MIDI data from bow position, acceleration, pressure, and grip (McMillen, 2008). Features for control could also include sensor measurements and raw audio for excitation of digital synthesis, for instance piezo sensing for string plucking (Harrison et al., 2018) and resonance modeling in the Caress instruments (Momeni, 2015). As seen in cases like this, it is important to note that the use of audio signal does not imply indirect control; the distinction is in the presence of the translation as a result of audio analysis and feature extraction for control parameters.

Therefore, the key affordance of direct over indirect control when interacting with the voice is that it does not rely on audio production to provide interaction. Figurative gestures often do not produce sound on their own (Godøy and Leman, 2010), but are essential in regulating emotional communication (Buck et al., 2013; Doğantan-Dack, 2011) as well as group dynamics and synchronisation (Eerola et al., 2018). Such aspects of performance could be used for direct control. Additionally, most audio analysis, especially spectral analysis, introduces some latency into a system. An unpredictable connection between user and interface can also result from imperfect audio analysis; pitch tracking algorithms are not always accurate and can behave in strange ways. Finally, audio-based indirect control implies that there is an acoustic sound which any digital synthesis must compete with.

Direct control involves manipulation of sound based on the action which produces it; for instance, using sensors to directly map on pressure piano keys to manipulate MIDI note velocity. Indirect control involves examining the audio signal produced; in the case of the piano, perhaps applying a filter when the pianist plays above a designated note. An intermediary step is introduced in this translation and requires the pianist to learn these new associations. Furthermore, there are many ways to produce a pitch on a piano (physically reaching in and moving the hammers, electromagnetic activation of the strings, etc.). Thus many facets of creative performance technique that are inaudible or do not produce sound can be lost when only indirect audio-based features are used for control.

In instrument design there are benefits to each type of control. Direct control is critical when the focus of the interface is in capturing performance gesture, particularly in cases where auxiliary gestures (those which do not produce sound but convey emotional meaning) are important in musical communication. Indirect controllers can allow for ease-of-use, greater precision and scaling, and versatility in flexible mapping of many parameters to a few controls.

2.4.1 Combining Direct and Indirect Control

Many digital instrument designers have successfully balanced both types of control in a single interface. This is particularly present in augmented instrument design, where the common design goal is to allow the musician to use their existing technique on an otherwise traditional instrument for digital synthesis aspects, sonic or otherwise.

Such augmented instruments include the Overtone Fiddle (Overholt, 2011), Svampolin (Pardue et al., 2019), and other related hybrid violin controllers (Pardue and McPherson, 2013; Pardue et al., 2014). The augmentation of the violin involves multi-modal tracking of several elements of performance, including upbow and downbow detection with electrodynamic pickups on the bridge and pitch tracking via left hand finger placement on pressure sensors embedded into the fingerboard. The key control element here, pitch data, is based on a fusion of this sensor data as well as parameters extracted from the audio signal. Some augmentations were intended for use in teaching, meaning that "both pedagogically and motivationally, players need to feel like they are playing a real violin;" thus, focus was placed on low-latency response and coupling acoustic sound analysis and gestural-based controls to reinforce sensorimotor mappings in learning (Pardue et al., 2015, 2019).

This work on the violin was partly inspired by the ESitar, an augmented sitar which also uses a variety of sensors for detecting gestures such as hand position, fret placement, and thumb pressure in addition to audio analysis (Kapur et al., 2004). This combination of control helps to determine gestures which may otherwise be undetectable from an audio-only standpoint, such as the performer bending a string for pitch variance. The ESitar also provides coordinated visual representation based on direct gestural control, which is useful in teaching contexts for reinforcement of finger placement as well as in creative performance.

An example of combination control aspects involving the voice can also be found in the "auditory masquing" tools developed by Stahl and Clemens for work in diegetic character voices (Stahl and Clemens, 2010). This work again aims to preserve the personal connection created within the unprocessed voice whilst adding the desired element of digital transformation (in this case, preserving the actors' voices when they are enhanced for character work). The authors' motivation derives from the idea that "the sense of personal connection between increasingly synthetic performers and increasingly diffuse audiences is vital to storytelling and entertainment"—not unlike musical storytelling in vocal composition. The design of the voice interface, a wearable sound reproduction system using predictive forward stimulation, allows for key qualities of the acoustic output (particularly low-end sources which add defining characteristics to the speaking voice) to be retained and improve vocal clarity and individual characteristics of speech when combined with other auditory masquing techniques. This is beneficial to consider in aspects of work with the personal aspects of the voice.

This blend of controls can also be found in new instrument design, such as the Bellyhorn (Verdonk, 2022); Verdonk describes how visible excitation methods can reinforce connections between synthesis elements to preserve human interaction and expression (Verdonk, 2015). Using vocal audio features and direct pressure from body, the bellyhornist can influence the drone produced inside the instrument. Singing loudly or putting one's head further into the horn define the volume of the drone, while lifting the horn influences pitch. The player can also lie on the instrument and change its shape to influence the sound. This combination of control allows the user to create sound through exploration.

2.4.2 Vocal Interfaces

A look through NIME conference proceedings from 2001 to 2022 reveals the voice is a relatively uncommon focus in academic research, with fewer than 20 papers devoted to control aspects of the singing voice (Kleinberger et al., 2022). Although this is a non-exhaustive search and vocal DMIs have been published in other academic venues focused on HCI (e.g, Kilic Afsar et al. (2023), which focused on sharing gestural information and bodily sensations between a vocalist and listeners), NIME demonstrates how neglected vocal interaction is, even in a conference focused solely on musical interaction. Up to 2022, 1,887 papers had been published at NIME, meaning voice makes up about 1% of the NIME literature. On the other hand, from a commercial standpoint, there are thousands of mobile apps which use vocal audio signals for synthesis, identity recognition, and entertainment, such as in karaoke apps. The controllers which do exist can be divided roughly into two categories: those using features of the voice (most are indirectly extracted from audio) to control aspects of other-instrument synthesis, and those using non-vocal direct control to manipulate digital vocal synthesis.

It might be important to note that, although I originally searched through the NIME archives up to 2022, this continues to present a challenge in instrument design, despite the prevalence of voice as an instrument. To the best of my knowledge, commercial applications exclusively use indirect audio processing, largely via the inbuilt microphones on mobile phones. Within research fields, where there is potentially more room for exploration of direct control methods, this still seems to be the case, suggesting that it is not a financially motivated or product- or consumer-driven barrier disrupting the creation of novel vocal interfaces. Rather, the lack of advancement in vocal control appears to come from the difficulty of direct vocal control, remaining almost exclusively in the audio domain. In the last two years, only two additional papers have been published regarding this — my own (Reed and McPherson, 2020), which will elaborate on this exact challenge of vocal interaction and addressing it through sEMG, and another example of blended direct and indirect control in The Body Electric (Cotton et al., 2021b), which I will elaborate on below.

The Voice as a Controller

There are many instances of vocal audio signal features being used to indirectly control synthesis for other instruments. Vowel detection with the Wahwactor allows for control of guitar filtering (a wah-wah pedal) in the guitarist uttering "wah-wah," as a way to reduce the learning demands of using a foot pedal (Loscos and Aussenac, 2005); a similar example is the synthesis of bass guitar using volume, pitch, and timbre extracted from the voice (Janer, 2005). Other devices have been developed commercially, notably the Vocoder and TalkBox for changing instrumental filtering using vocal formants and mouth shape derived from the audio signal. Newer digital controllers include imitone, a voice-to-MIDI controller comparable to the MIDI guitar (Evan Balster and Richard Hogg, 2022), and the OVox plug-in by Waves (Audio, 2020) which uses vocal features to control filtering and modulation. Audio signal has also been translated into tactile physical vibration in musical installations (Holbrow et al., 2014). These devices, developed specifically for vocalists, highlight their ease-of-use and intuitiveness as primary selling points. OVox in particular highlights a similar blend to its controls as seen with augmented instruments, with the specific goal of preserving expression in the original vocal signal for authenticity: "The human voice is the original instrument—the richest in expression and the fullest in color. OVox turns your vocal—any vocal—into an even more limitless musical playground for you to express your creativity."

Articulatory aspects of the voice have also been the subject of a few direct control-driven interfaces. There have been several mouth or vocal tract interfaces developed using facial and mouth tracking, such as the Mouthesizer (Lyons et al., 2003), mapped to a variety of sound synthesis parameters (de Silva et al., 2004; Pöpel et al., 2014) or as MIDI controllers (Orio, 1997). Ultrasound has been used in the case of the Tongue'n'Groove (Vogt et al., 2002) to use tongue contour and motion for controlling other digital instruments as well as a vocal model, although the system was not used for gestural recognition of vocal technique, "but rather to explore how to leverage the fine motor control skills developed by the tongue for expressive music control." Outside of a musical context, ultrasound has also been used to detect speech formants for direct control (Kimura et al., 2019). The installation by Poepel et al. is notable in its direct control aspects, employing a kind of "air-opera" where the participant would be able to pantomime operatic performance without singing by miming vowel shapes with their mouth (Pöpel et al., 2014).

Corsetto (Kilic Afsar et al., 2023) is a wearable corset containing pressure sensing pillows. The pressure sensing drives the actuation of pneumatic fibres to expand and contract the garment as the wearer breathes (Afsar et al., 2021), creating an extension of the sensation of breathing, or that the garment breathes back in response to the wearer's movement (Cotton et al., 2021b; Tsaknaki et al., 2021). With the help of this feedback, the wearer can observe and become attuned to the sensation of their breath in the abdomen and back. By capturing the movements of one singer and presenting them to another singer through haptic actuation, it is also possible to physically feel the movement of another person while singing (Cotton et al., 2021a). This has the potential to create mutual understanding of the motion between two individuals.

A handful of NIME installations have also focused around control aspects related to the voice, most prominently in 2014 with the Vocal Vibrations experience installation (Holbrow et al., 2014), which allowed those visiting to experience vocal composition while considering the physiology and expressive interaction of their own voice in reflection on its various structures; participants were able to feel their own voices more tangibly in transformation of physical vibrations of their own voices to a handheld device, the Oral Response Ball (ORB). The ORB, using a Max/MSP patch, translated the audible vocal signal of the participants' singing and humming into tactile response as a way for those using it to feel their singing in their hands for more detailed and sensitive interaction.

Controllers for Vocal Synthesis

Direct control is more prominent in controllers for vocal synthesis; however, the majority of this direct control relies on gestures unrelated to the voice or vocal performance, such as hand movements (d'Alessandro et al., 2006; Xiao et al., 2019; Yonezawa et al., 2005), manipulation of vocal tracts made of soft materials (Yoshimura and Kazuhiro, 2019), or browser-based (Thapen, 2017) and stylus/tablet control interfaces (Delalez and d'Alessandro, 2017; Feugère and d'Alessandro, 2013) to change vocal models. Digital vocal processing in computer-based audio plug-ins is also popular; for instance, the Dehumaniser (Audio, 2022) provides modulators, scrubbing, spectral shifting, and a variety of filtering to create artificial monster voices or modify existing audio. Some instruments such as the SqueezeVoxen, COWE, and VOMID (Cook, 2005) incorporate direct controls for voice synthesis which are similar to actual singing, including air pressure sensing for breath control and mouthpieces for phoneme measurement, through amalgams of other instruments such as accordions and keyboards. Although removed from organic voice production, some groups such as the Cantor Digitalis (d'Alessandro et al., 2014; Feugère et al., 2015) team have been able to turn this control of voice synthesis into an art form in its own right (Synthesis, 2016).

There are several notable vocal performers working with vocal controllers and extended vocal technique (using the full-range of vocal technique) to augment and explore the virtuosic and cultural aspects of the voice that shape vocal performance. Vocal composer and instrument designer Kristin Norderval notes that some of the main barriers in introducing interactive technology into vocal practices, particularly opera: existing technology does not mesh well with theatrical aspects of performance and need for mobility, there are few systems which do not rely on a visual interface, and systems often do not accommodate both acoustic, unamplified voice and processed sounds (Norderval, 2020). Norderval's work with the HotHand (SourceAudio) and Wave (Genki) rings use gestures of the hands to process the voice, which is heard alongside the unamplified vocal signal. In this work, the focus is on the aesthetics of balance and interplay between the two components. (Norderval, 2020). In another example, Pamela Z famously uses custom MIDI controllers and a variety of processing software to process their own voice in real-time, combining an extensive knowledge of traditional vocal techniques with digital capabilities (Pamela Z, 2020). Pamela Z's work, like Norderval's, incorporates balance between abstract vocal expression and literal or easy-to-grasp vocality. As vocalist Franziska Baumann highlights, everyone has a voice and can make a connection to vocal content quickly in this balance (Baumann, 2021). Baumann's own vocal designs use communicational gestures, for example to create a visual metaphor of limbs that demonstrate interactions between musicians (Baumann, 2023). Similarly, Alex Nowitz explores the Strophonion, built at STEIM (Studio for Electo-Instrumental Music, Amsterdam), to augment his voice using live sampling techniques controlled with its hand controllers (Nowitz, 2019). Through this work, for instance in the piece like Nowitz's *Moving Tongues: Playing Space,* multivocality — various aspects of performer knowledge — are addressed (Nowitz, 2018). Controlled by gesture, the introduction of vocal processing extends the voice, blurs processed and unprocessed voice, and captures the "vocal imaginary," and how these multivocal processes are linked (Nowitz, 2018, 2021).

2.5 Summary

In short, the vocal apparatus requires fine motor control during singing. Although studied extensively in medical contexts, the methods of interfacing with the voice presented by physiological research are not suitable for musical interaction, or most other interactions outside of the lab for that matter. In current musical interactions, controllers which deal with the voice often lack a source of direct control; that is, they operate by extracting features from the audio signal resulting from the body's characteristics, rather than the body itself.

Among this variety of vocal interfaces, there is a clear gap: control of synthesis using direct vocal control. Being interested in utilising the well-developed sensorimotor techniques of vocalists and examining the internal connections singers have with their body, I particularly sought devices which used audio-independent control. Later, in Chapter 4, I will introduce my exploration of surface electromyography (sEMG) for direct physiological vocal sensing in a minimally disruptive way to bridge this gap and provide a method for voice controller design that can be used by the wider music community. To provide a way to work with the vocal physiology in a direct and non-invasive way, this thesis explored and designed a vocal controller using sEMG to better understand vocalists' movement and perceptions at their source — in the body, rather than in the resulting sound.

In order to understand how this gap of direct control will be addressed, keeping the body at the forefront of the vocal interaction, I will now move to the more perceptual aspects of singing and musical interaction, to build a clearer picture of interaction from the musician's point of view.

Chapter 3

Vocal Phenomenology

Imagery & Tacit Knowledge as a Basis for Vocal Interaction

Musicians relate to physical aspects of performance and gesture through more abstract cognitive processes, namely in mental musical imagery. This chapter addresses the phenomenology of the musical and, more specifically, the vocal experience — how musicians really understand their practices. I first discuss musical imagery and how it is formed, measured, and used by vocalists and other musicians. Then, I outline vocal expression and materiality; that is, how the physical quality of and the way we feel and hear the voice shape our interactions, and how this is related to the formation of musical imagery. I conclude by discussing how metaphor is a critical tool for understanding and describing our imagery, and how vocalists are taught and understand their practice through such abstract representations in the voice lesson.

3.1 Mental Imagery

In general, imagery refers to the anticipated outcomes of a particular action in its preparation or intention, and shares neural and behavioral similarity to actual execution (Halpern and Zatorre, 1999; Halpern et al., 2004; Kleber et al., 2007; Kosslyn et al., 2001; Trusheim, 1991). The action and its execution rely on a single underlying mental representation as the imagined action (Zatorre et al., 2007). There are different modalities of imagery as they relate to all the senses; however, the focus in cognitive research is commonly on visual, auditory, and kinaesthetic modalities. Visual representations relate to what is "seen" in the mind and can sometimes be from a first-person (1PP) or third-person (3PP) perspective (Cumming and Williams, 2012; Williams et al., 2012), while auditory imagery refers to what is heard in the "mind's ear." Kinaesthetic modality involves the "feel" of completing the action (Cumming and Williams, 2012; Leman, 2008). It is important to note that the sensory experience in general contains large amounts of overlap and therefore none of the modalities of imagery occur in isolation; additionally, imagery is highly context based and is often accompanied by elements of emotional state, including feelings of pleasantness or tension (Zagacki et al., 1992).

3.1.1 Imagery Theories

There are several theories in psychology which propose how imagery works to help us learn skills, visualise actions and reactions, and retain memory of previous experiences. These have been previously referenced when discussing musical imagery (Jestley, 2011). These are summarised in Table 3.1.

When referring to imagery in this thesis, I will largely focus on Functional Equivalence (Kosslyn et al., 2001); that is, that imagined representations of prospective actions play a preparatory role in

Author(s)		Theory	Summary
Sackett	(Sackett, 1934)	Symbolic Learning Theory	Imagery works as a "blueprint," mapping out the sequential aspects of an action. This image is essentially a motor programme in the nervous system and can be rehearsed to help encode symbolic patterns of action.
Paivio	(Paivio, 1975, 1978)	Dual Code Theory	Imagery exists in two separate channels, as either sensory imagery or verbal imagery. These are processed differently but can be used together or in combination. Information can therefore be stored either as an image or a word; images are better for concrete concepts and words for abstract concepts.
Grouios **Hale**	(Grouios, 1992) (Hale, 1994)	Insight Theory	The perception of a whole situation is encoded in imagery, rather than in parts or as details. Rather than learning over time, the skill is encoded through sudden insight.
Kosslyn **Jeannerod**	(Kosslyn, 1980) (Jeannerod, 1994, 1995, 1999)	Functional Equivalence	The mental practice of an action and its realised execution share the same neural pathways and therefore the two are functionally equivalent.
Pylyshyn	(Pylyshyn, 2003, 1981, 2002)	Tacit Learning Theory	Rather than images being encoded directly in memory, people view and describe their imagery as what they believe they are seeing, based on their implicit knowledge.

Table 3.1: An overview of imagery theories and how they describe the encoding of knowledge and action paths in the brain (adapted from Jestley, 2011). Functional Equivalence, the theory used in this thesis, is highlighted in the blue box.

motor control and have been found to share neural and behavioural similarities with the execution of actions (Cumming and Williams, 2012; Halpern and Zatorre, 1999; Kleber et al., 2007; Trusheim, 1991; Zatorre et al., 2007). Later, in Chapter 4, I will introduce the most common methods for measuring imagery ability. These methods are primarily validated with respect to neural imaging and therefore utilise Functional Equivalence Theory. However, this thesis's investigation and findings are not heavily predicated on the correctness of this theory; regardless of the exact encoding of the vocal imagery, its use is essential to relationship between vocalist and voice. There are several considerations of grounding the thesis in this theory, which I will discuss in Chapter 4 when outlining the methods used for this thesis.

3.1.2 Musical Imagery

Musical imagery therefore encompasses the mental imagining of all of the aspects of the musical experience—the auditory, visual, kinetic, and tactile aspects of sound and sound production—without or prior to action and sound production, when no audible sounds are present (Godøy and Jørgensen, 2001; Keller, 2012; Trusheim, 1991). The sound represented in a mental musical image can be consciously called upon by performers. In the context of performance, kinaesthetic musical imagery might take the form of planning out piano fingerings in one's head or the imagination of the fingers' movements while stationary on a table, while visual imagery could involve 1PP imagination of a particular performance hall and standing on its stage, or 3PP imagination of observing another performer from the front row.

Much research in this musical context focuses on the auditory aspects of musical imagery—that is, the ability to imagine the qualities of a sound without hearing it. This is especially critical for musicians in sound production, where an auditory image can be used to prepare the body and the instrument to properly execute musical passages and adapt technique learned in rehearsal for a performance setting. Musicians notably use auditory imagery in *audiation*. "Audiation is to music what thought is to language," as described by Gordon (1999), and operates in a similar way to allow musical communication: it is how we think of music, construct ideas and understanding, and give meaning to what we hear, as we would do with linguistic communication (Gordon, 2011). Musicians and music educations commonly use the term audiation to refer to mental imagery executions of hearing music covertly within the mind, without overt playing or singing (Brodsky et al., 2008a; Halpern and Overy, 2019; Pfordresher and Halpern, 2013). This mental execution of a musical exercise in silent rehearsal is useful while preparing for performances to explore differing technical and expressive aspects of a piece (Bailes, 2006; Cumming and Williams, 2012; Holmes, 2005; Loimusalo and Huovinen, July 5-9, 2016), as well as to coordinate sensorimotor actions needed to execute these aspects in their performance (Fontana et al., 2015; Keller, 2012; Leman and Maes, 2015; Pfordresher, 2019). Personal experiences including instruction, practice, and time spent in performing environments will form the basis for imagery associations and are therefore highly individual. The use of imagery and its application during practice and performance will also vary based on a musician's specific background and training (Bianco et al., 2018; Fontana et al., 2015; Pfordresher, 2019).

3.1.3 Using Musical Imagery

During performance, imagery is used to anticipate the outcomes of certain actions and is employed by musicians during practice to prepare for upcoming performances (Bailes, 2006; Cumming and Williams, 2012; Holmes, 2005; Leaver et al., 2009). The use of imagery gives musicians the ability to audiate — in other words, hear the music in the mind and examine other elements of performance within silent practice, without actually performing a piece (Loimusalo and Huovinen, July 5-9, 2016). Musicians commonly state that the ability to imagine a performance is essential to expressive qualities in their music (Holmes, 2005; Woody, 2006) and for coordination in sensorimotor action while playing or singing (Fontana et al., 2015; Keller, 2012; Leman and Maes, 2015; Pfordresher, 2019). Deficit to this internal translation between auditory and motor imagery results in inability to predict the outcomes of action and is believed to be the reason vocal pitch-imitation deficit (VPID) singers are unable to match a pitch they hear (Pfordresher and Halpern, 2013; Pfordresher et al., 2015).

The formation of auditory imagery is also used for the purposes of memorisation and expression, where, without instrument in hand, imagery "frees your brain to study it [the piece and its components] in more detail (Holmes, 2005)." Motor and auditory imagery and emotion are closely linked as a singular feeling during performance. Musicians describe these images as existing together in spacial awareness, with an emotional feeling acting as a type of cue or trigger which warrants a corresponding musical response to communicate this feeling (Holmes, 2005; Keller, 2012). Furthermore, musicians with the ability to audiate are able to better discern the qualities of their produced sound compared to their desired sound and to adapt over the course of a performance to incorporate expressive features (Loimusalo and Huovinen, July 5-9, 2016; Trusheim, 1991; Zarate and Zatorre, 2008).

Use of musical imagery and its application during practice and performance will vary based on the specific background and training an individual musician has (Bianco et al., 2018; Fontana et al., 2015; Pfordresher, 2019; Woody, 2006). Personal experience during learning, practice, and performing will form the basis for imagery associations. This individualistic component of imagery is thought to cause difference in expression between different performers and be responsible for the quality of interaction within musical ensembles (Keller, 2012).

Adapting to Altered Feedback

Imagery is particularly necessary when musicians need to produce a desired sound or expressive quality to their performance but have reduced feedback. For musicians, auditory feedback is the strongest source of information about their own playing–after hearing the sound produced, adjustments can be made to achieve more desirable results. However, the available feedback is not always ideal in particular performance venues and conditions. Hearing one's own playing can become difficult, if not impossible at times. Altered auditory feedback (AAF) in live performance environments can happen through sound masking, external noise and speech distractions, poor placement of monitoring speakers, delays, mix issues, and uncontrolled resonances from the space itself. These AAF conditions are difficult to control during the course of a performance and must be met with adaptation and flexibility on the part of the performer. In these settings, musicians rely on their auditory imagery abilities to monitor their performance; through the recall of an auditory image and associated actions used to create the desired sound, performers can anticipate the results of their actions without hearing for certain the sounds they produce.

Building musical imagery during ideal rehearsal conditions allows connections to be made between sound and action and results in better accuracy in subsequent performances (Goebl and Palmer, 2008; MacRitchie and Milne, 2017). Several studies focusing on pianists have found that auditory imagery built during rehearsal with strong auditory feedback allowed the musicians to recall and play a piece accurately with reduced or even no auditory feedback (Brown and Palmer, 2012; Edmonson, 1972; Finney and Palmer, 2003; Highben and Palmer, 2004). The ability to rely on the calculated outcome of an action with auditory imagery extends further to expressive technique, such as dynamics and articulation, which can be performed regardless of the presence of auditory feedback (Bishop et al., 2013). Auditory imagery also assists with motor coordination and expression in duet and group performances, where musicians must adapt to other players and communicate intended sound through expression and gesture (Brown and Palmer, 2012; Davidson, 2012; Highben and Palmer, 2004; Zatorre et al., 2007).

Connections made between sound and action can thus provide later auditory recognition and accuracy in playing (Goebl and Palmer, 2008; MacRitchie and Milne, 2017); when practice is done in methods that use auditory-only listening or combinations of auditory-motor learning, pianists have been shown to have stronger recognition of melodies than in conditions of motor-only learning (Brown and Palmer, 2012). The association of a particular produced sound to a specific action can be referenced later to recognise patterns of playing. Research shows that, when a piece is learned with auditory feedback, pianists are also capable of recalling and playing a piece with accuracy when there is reduced or no auditory feedback; this is due to the application of imagery constructed during the learning process and the confidence in recreating that action without actually perceiving it (Brown and Palmer, 2012; Edmonson, 1972; Finney and Palmer, 2003; Highben and Palmer, 2004).

Pianists have also been shown to be capable of imagining certain expressive features, such as dynamics and articulation (Bishop et al., 2013), which can be performed regardless of whether auditory feedback is available. Motor-learning is found to be directly related to mental imagery abilities in terms of expression; musicians with greater ability to apply mental imagery are able to "fill-in" for missing auditory feedback by imagining their produced sound during the learning process (Brown and Palmer, 2012; Highben and Palmer, 2004; Zatorre et al., 2007). Anticipatory imagery from associations made in familiar melodies also allow musicians to better predict melodic movement; this activity is present in activation of premotor areas of the brain, which play a role in sequence learning (Leaver et al., 2009). Existing ideas of expression in music are also found to help unite collaborative playing in duets, where musicians pull from experience to communicate expression to others in the group (Davidson, 2012).

Training Musical Imagery

It is believed that imagery can be trained to some degree by rehearsing within environments that encourage its formation. This could include practicing within the intended space or in situations which may strengthen the associations (e.g., focusing on the tactile sensations of a performance in settings where AAF is present).

The PETTLEP model (Physical, Environment, Task, Timing, Learning, Emotion, Perspective) of learning (Figure 3.1) is thus encouraged for use by musicians (and those in other arts, as well as in athletics) to rehearse with ideal conditions as close as possible to actual performance in order to form new mental imagery(Cumming and Williams, 2012; Trusheim, 1991; Wright et al., 2014). This model is based on the functional equivalence theory of perception, believing that imagined and executed actions recruit the same centres in the brain; therefore, we can essentially train action through imagery and its construction (Wakefield et al., 2013; Wright et al., 2014). This model mimics the formation of imagery "in the wild" but draws focus on the experiential elements which are believed to make images stronger in practice. With application of these contexts in rehearsal, it is seen that performers can increase their aptitude for building new imagery; additionally, by combining as many sensory elements into the experience as possible, the image itself is further strengthened (Godøy and Jørgensen, 2001; Holmes, 2005; Trusheim, 1991; Wilson, 2006).

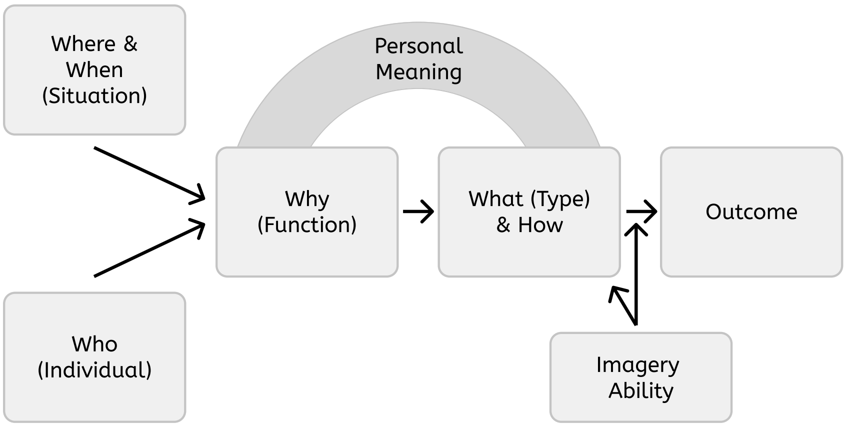


Figure 3.1: The incorporation of imagery tasks and ability in an individual's anticipated action (adapted from the revised PETTLEP model of learning (Cumming and Williams, 2012)).

3.1.4 The Role of Imagery in Singing

As an instrument, the voice provides an interesting case for study of auditory imagery. Despite singing being universally present in all cultures, and a naturally intuitive ability for many members of those cultures, it requires strong imagery, particularly in the connection between internal physical sensation and resulting sound. Singing lacks external tactile feedback compared to other instruments, which involve some overt tactile aspect of control for feedback; for instance, a pianist is able to rely somewhat on the one-to-one mapping of a key to a specific note; if the auditory conditions of a performance are not ideal, the pianist can rely solely on the tactile feedback for accurate performance. In other instrument interfacing, such as in winds, the mapping relies more on auditory imagery: a trumpeter can rely on their valve combinations to some degree, but their embouchure must be adjusted with the correct partial in mind to achieve the correct note.

Auditory imagery is essential in translating the ideal sound to physiological expression (Clark et al., 2011); for vocalists, with an instrument existing fully within the body, this is especially so. As in the case of embouchure adjustment, there is some degree of kinaesthetic imagery (the musician is able to consider "what does this note feel like?" in both a tactile and proprioceptive sense, while a vocalist would be limited to proprioception alone), which is coupled to auditory imagery through sensorimotor links (Cumming and Williams, 2012; Pfordresher et al., 2015). The connection becomes more obscure as the instrument is further removed from direct tactile connections (Hemsley, 1998; Hines, 1983); vocalists are highly dependent on internal kinaesthetic feedback (Larson et al., 2008). The internal sensory experiences of vocalists can be less apparent than tactile feedback received at the fingertips.

Previous work demonstrates that singers may be more susceptible to pitch-shifts than keyboardists, due to the particular perceived response-effect associations and coordination that singers have learned in normal circumstances (Pfordresher and Mantell, 2012). However, self-reported experienced vocalists exhibit similar success to instrumentalists in a study conducted by Zarate and Zatorre, where neural activity was monitored while singers adapted to pitch-altered feedback (Zarate and Zatorre, 2008). While both untrained and trained vocalists utilise the same functional networks during singing, the trained vocalists showed higher levels of activity that allowed them to adjust to pitch-shifting and ignore external stimuli to closely monitor their own pitch. With altered feedback, vocalists exhibit more awareness of their own vocal-motor programs to ensure their notes are being produced correctly in terms of pitch (Zarate and Zatorre, 2008). In the case of AAF, vocalists must rely even more so on internal, often wordless imagery and understanding of their action paths to perform with high accuracy (Salaman, 1989). With this less well-defined feedback and difficult-to-articulate imagery, vocalists depend heavily on abstract metaphor to understand sensory feedback and create and use musical imagery.

3.2 Vocal Expression, Entanglement, and Materiality

Before going into details of how musical imagery is formed and vocal pedagogy, I want to quickly define some elements of vocal quality and musical expression. It is important to consider the perceptual aspects of music and the voice which are important to singers during their practice to understand what is important in interacting with the voice. I will also discuss vocal and instrument entanglement and how the materiality of the voice impacts our interaction with it. It is likewise important to consider that, as the author of this work, my own vocal background is a major influence in this research. I have been trained in a Western classical vocal context, with specialisation in Baroque and Romance opera and art song. This is all to say that context is everything: musical culture forms our perception and understanding of the voice. I will briefly introduce terminology used in physiological research and in Western classical vocal contexts to help solidify this context going forward.

3.2.1 Vocal Terminology

Three qualities — voice, vocal (not the same thing!), and emotional quality — are used to describe aspects of singing. *Voice quality* is influenced by the register the singer is currently using (Prem and Parncutt, 2008). A singer's register is generally either the "chest voice," "mixed voice," or "head voice" (*falsetto* is the more common term for head voice in male singers), referring to the low and high registers of the voice, respectively. The distinction between registers is not well defined but generally refers to the resonance placement of the voice at different pitches; this is particularly noticeable in the lower register, where the lowering of the larynx causes resonance in the chest.

In contrast, *vocal quality* refers to the physiological aspects of the vocal folds and resultant opening and closing of the vocal tract (García-López and Gavilán Bouzas, 2010; Prem and Parncutt, 2008). This is therefore a physiological indicator of the voice, rather than a situational descriptor, in the case of voice quality. Physiology (and prior genetic factors) will determine the singer's vocal range and subsequently designated voice part. From highest to lowest the main voice parts are: soprano, mezzo-soprano, contralto (or just alto), countertenor (or contra tenor), tenor, baritone, and bass. Most female singers make up the first three groups while most male singers the latter three; males whose voice range occupies the same range as the female mezzo-soprano or contralto are referred to as countertenors. An individual's physiology determines the frequencies which compose the higher voice formants that, along with the resonances created by the space in the mouth, provide the timbre and individual characteristics of the voice (Sundberg, 1994). Voice instructors will sometimes define the color or characteristics of an individual voice in comparison to an instrument with similar formants—some voices are defined as "reeds," having a thicker, rounded tone, while some are labelled "flutes" and have a thinner, clearer sound.

Finally, *emotional quality* refers to the emotional and mood quality added to the base timbre of the voice. This is highly individual and can depend on the personal mood or interpretation of the singer (Prem and Parncutt, 2008). As with other musical practices, these qualities can be difficult to describe in language but are sometimes described through colors and visuals (blue, bright, dark) or other physical qualities (mellow, soft, dense) (McAdams and Giordano, 2009; Siedenburg et al., 2016). However, there is still little agreement on how individuals express their perception of musical timbre in other modalities (Löbbers and Fazekas, 2021; Löbbers et al., 2021).

3.2.2 Expressive Singing

The ability of music to convey emotions has long been studied, especially in regard to listener interpretation. Emotion has been found to be present in various musical features, including *timbre* (Farbood, 2012; Farbood and Price, 2014, 2017), *rhythm* (Krumhansl, 2000; Schellenberg et al., 2000), *tempo variability* (Juslin and Madison, 1999; Kamenetsky et al., 1997), *loudness and intensity* (Sloboda and Lehmann, 2001), *pitch and tonality* (Krumhansl, 2000; Lerdahl, 1996; Lerdahl and Krumhansl, 2006), and *harmonic and melodic movement and expectation* (Bigand et al., 1996; Farbood, Aug 25-29, 2008; Farbood and Upham, 2013; Herremans and Chew, 2016; Steinbeis et al., 2006); these are elements manipulated in composition by performers in order to provide emotional context and change to the listener during a performance (Gomez and Danuser, 2007; Hevner, 1936; Krumhansl, 1997, 2002; Sloboda, 1991). The range of emotions experienced during musical performance fall into the same semantic space as emotions experienced during other parts of daily life (Hevner, 1936), with the aforementioned features providing a way for the listener to apply their own personal and individual life experience to what they hear (Sloboda, 1991).

In a way that mirrors speech, sung vocalisations are able to carry emotional context; it is hypothesised that singers communicate by "exploiting" the same neural paths used in encoding and understanding emotion in speech (Juslin and Laukka, 2003; Juslin and Västfjäll, 2008). Vocalists are also able to express strong emotional context in a sung piece by manipulating its musical features, with even young children (aged 5-10) being able to both identify the emotional quality of a sung musical excerpt and convey different emotions in their own singing (Ebie, 2004; Gabrielsson and Juslin, 1996; Kendall and Carterette, 1990). The voice as an instrument can also be regarded as more emotionally expressive due to the context carried by the lyrics of a piece. Ultimately, the vocalist is responsible for relaying their desired interpretation of those lyrics and the composition as a whole to the audience by applying learned performance techniques, including musical imagery, to achieve the necessary sound quality.

3.2.3 Instrument Entanglement

Musical imagery forms through experiences of making music in practice and performance and the ongoing construction of relationships with instruments. This interaction includes, as mentioned at the start of the section, the influence of the musical culture and the qualities of the voice we perceive. Context, our relationships, and our experiences with our bodies and with tools (instruments or otherwise) change how we interact with them. For instance, with advances in musical technology and changing roles in performance and composition, such as AI in music and physiological- or movement-based control, the relationships between musicians and musical tools has also changed (Ihde, 2021). This is of course also true of more "traditional" musical interfaces; with the introduction of new musical technology, we can observe this changing interaction with respect to novel designs. This relationship to instruments and our understanding of interaction with them is driven in a cyclical way, where the feedback we get from the world influences our behaviour (Frauenberger, 2019), which then provides new feedback about the world. Imagery changes over time with these new experiences and is closely related to the post-phenomenological theory of entanglement: entanglement with the world or the tools (for instance, musical instruments) shapes our understanding and the imagery we form through interacting with them. Mental imagery is a key component in interaction in the sense that it is formed through and depends on the user and system or environment interacting with each other (Armstrong, 2007; Essl and O'Modhrain, 2006; Frauenberger, 2019). Through the subjective experience of the user and their application of past experience, mental imagery provides the basis to examine the co-dependency and collaboration of a user with their environment (Höök, 2010; Tuuri et al., 2017; Wakkary et al., 2018). Mental musical imagery forms as musicians participate in rehearsal, performance, instruction, collaboration, and so on, and then use the experience to relate back to their world (Ihde, 1975). There is a cyclical relationship between performer and performance, where they continually shape each other and act, in a way, as a feedback mechanism - the performer learns from the space they are in or from whomever they perform with and the mental image is updated and refined as this interaction goes on. This makes up the theory of enactivism in interaction; that is, the system and the context in which it is used are inseparable and forms the way we learn actions and their consequences (Fenwick, 2000; Li et al., 2010). There is no permanency to the musical imagery and it evolves through the relationship of the musician to their world (Armstrong, 2007; Ihde, 1975).

The instruments themselves also dictate our understanding and the relationships we have in our interaction. The design of musical instruments and other technology therefore shapes how we interact and even more so who we are (Frauenberger, 2019); as Homewood succinctly puts it, "When we design for/with bodies, we actually design bodies themselves." (Homewood et al., 2021). This *more-than-human* shift highlights that, rather than humans being a more privileged or more-in-control position than other agents, humans and non-human machines, sensors, and tools all have agency within the interaction (Andersen et al., 2019; Devendorf and Rosner, 2017; DiSalvo and Lukens, 2011) and are part of each other in their entanglement (Barad, 2003):

3.2.4 Vocal Materiality

Post-human and material-discursive practices (Barad, 2007) focus on this decentring of the human in interaction to a more-than-human interaction (Forlano, 2016; Nordmoen and McPherson, 2022). For instance, the ideology of agential realism (Barad, 2007) focuses on material discursive approaches which balance the role of material and culture in the interaction; actions and interactions depend on materials and environment in which they were designed and used (Nordmoen et al., 2019). The different agents which make up the interaction are part of one another and are inherently entangled, rather than being ontologically separate entities, and all of these factors shape our reality; experiences and realities are "enacted in practices." (Law and Lien, 2012) In the design of instruments, it is therefore possible to shape the musical interaction and indeed the musician through the design itself. This has led to suggestions that, rather than designing instruments and therefore dictating the musical interaction, designers should create tools that provide a context for "musicking" (Waters, 2021). Because of the personal connection and individual background of a musician, the use and perception of an interaction will depend on the individual (Mice and McPherson, 2022; Rodger et al., 2020). In this sense, creating more open-ended contexts or providing ambiguity can create space for exploration and understanding of musical understanding (Erdem and Jensenius, 2020). The musical context exists because of the identity, understanding, and perspectives of the musician; in a sense, the instrument is only an instrument *because* of the musician (Hardjowirogo, 2016). At the same time, the musician's perspectives are a result of the instrument and the interaction with it.

Body Lutherie

The materiality of the voice itself, as a physical aspect of the body and self, also plays a role in musical interactions. Human and body and body and voice are not separate entities but are rather entangled and depend on each other in interaction. In a material-discursive way, I have started to think of this relationship as a kind of "Body Lutherie" (Reed, 2023). Luthiers — musical instrument designers, specifically those who create string instruments like violins — work with wood, a material that is living, ever-changing, and unique. As with other woodworkers, the quality or uniqueness of that material often influences what it becomes or how the luthier works with it. This relationship with the material comes from deeply knowing and understanding the material itself, as an agent with its own role in the design process (Dew and Rosner, 2018; Nordmoen and McPherson, 2022). Although specific tree types may be grown in very specific conditions, there are many uncontrollable aspects of their growth which ultimately influence the wood. The environment, especially with increasing impacts of climate change, location, age, weather, and so on ultimately make up the material; only some of these (age) or portions of these (perhaps hydration and sunlight aspects of weather) can be managed. The other influences must be managed by the luthier in the creation of the instrument; in woodworking, like other crafts, the material often dictates the design. This has also been the case for "digital lutherie," wherein aspects of a digital instrument such as programming environment or language and available parameter controls also influence the digital sounds created (Lepri and McPherson, 2019; Renney et al., 2022), as well as in the dynamics of agency in fabrication (Devendorf and Ryokai, 2015). Much like other materials, the *physical* voice is its own agent in vocal practice; we can only work with the voice we have, and this shapes our interaction.

Vocal Organology

This materiality of the voice goes beyond the body itself to the space in which it resides, both culturally, socially, and physically. Nina Sun Eidsheim's work and critical organology of the voice as a means to consider vocal materiality (Eidsheim, 2017). The voice is not treated or perceived like other musical instruments; as Eidsheim describes, the voice perceived like other instruments would then be described in a material context referring to specific physiological features — respiratory organs, soft tissues, and articulatory bodies in the mouth, perhaps. By thinking of the "voice" as only one of the many functioning parts, rather than a singular entity, can provide a point of critical inquiry that isolates these features to instead focus on its materiality. Eidsheim focuses as well on the space within and outside the body, through which vocal sound must travel to be experienced, felt, and heard by both the vocalist and listener (Eidsheim, 2011). This view of voice as a "vibrational practice" defines music as a multi-sensory materiality (Eidsheim, 2015) resulting from the dynamic relationship between spaces, the performer's and audience's bodies and experience in the world, transitioning from "thinking about music as a knowable aesthetic object to thinking about it as transferable energy" (Eidsheim, 2015). For example, Eidsheim examines Juliana Snapper's underwater opera performances as a case study of the importance of space outside the body to perceive the voice (Eidsheim, 2011); traveling through water instead of air acts as a 'denaturalisation" (Stadnicki, 2016) to examine this materiality.

The experience of the voice also goes beyond its physical properties and materiality. Cultural and social backgrounds also dictate experience. Vocal organology aims to incorporate feminist philosophies to examine such bias and predisposition (Eidsheim, 2011). Our lived bodies are embedded in culture and context. A critical organology approach addresses present socio-cultural power dynamics and gender disparities which form much of our view of the physical body, people, and identity, and therefore vocal vocabulary and discourse. (Eidsheim, 2017). In this thesis, I will focus on the dynamics between vocalist and voice as one part of this complex vibrational vocal relationship. The relationship between vocalist and voice is entangled with culture and perceptions of the body, and this in turn impacts vocal musical imagery and how we understand vocal practice. With this vocal context of expression, entanglement, and materiality, I will introduce vocal metaphor and how the perception of the voice as body and instrument is described and taught within the vocal context. This dynamic relationship, rooted in complex culture and entanglement between body and voice, is most commonly expressed between vocalists through metaphor.

3.3 Metaphor & Underlying Schema

Where it is difficult or impossible to understand or describe these material experiences and imagery in direct language, metaphor allows us to apply what we do understand for explanation and reasoning. We understand our experiences through proprioceptive senses, which make up awareness and control over the body and understanding factors such as its positioning in space, movement, and tension and effort in action (Candau et al., 2017). This knowledge — our innate tacit knowledge — arises from living in our bodies and therefore it difficult to describe (Núñez Pacheco and Loke, 2016; Svanæs, 1997; Svanaes and Solheim, 2016). As well, much of the fine-grained dimensions of our interactions are lost in the rapidness of experience (Petitmengin, 2006); although we receive sensory information, much of the experience in any given moment is not readily available to us. In the case of musical imagery or the sensory experiences which form our imagery, there is often a need for these abstract references. Metaphor can help us to understand and articulate these understandings to others.

In order to use these abstract representations, we must be able to imagine them and understand them through other life experience (Lakoff and Johnson, 1999). These images can take any kind of sensory modality (visual, kinetic, auditory, etc.) or any combination thereof (Godøy and Jørgensen, 2001; Kosslyn, 1980; Ohrenstein, 2003). Imagery allows us to imagine experiences without needing to experience them and we can recall images we have previously encountered to inform future action or understand a new experience (Jeannerod, 1994, 1995, 1999; Kosslyn, 1980; Kosslyn et al., 1995, 2001). Therefore, the relationship we have with imagery is also a constantly evolving one; the experiences and interactions in day-to-day life help us to form mental images and representations and then these images inform further interaction (Chiel and Beer, 1997; Depraz et al., 2003). These associations, with repetition, are made up of schema.

3.3.1 Image Schema

Schema are the underlying ideas and concepts we observe in our interaction (Johnson, 1989; Langacker and Lakoff, 1988). These schema can be extended to more abstract concepts and become more embodied in our understanding (Langston, 2002). We make associations with different qualities based on the tacit knowledge formed around these schema, which become embodied metaphors (Antle et al., 2009; Daudén Roquet and Sas, 2021); for instance, the idea that future is forward and past is backward (Hurtienne et al., 2020). The schema associations we make between different interactions are found cross-linguistically (Gibbs et al., 2004), demonstrating that metaphor is rooted in the body and then informs language.

Perception mapped onto a physical equivalent forms what is referred to in cognitive science as an image schema, learned Gestalt mental patterns that provide structure and understanding to a particular experience (Amant et al., 2006; Dewell, 1994; Hampe and Grady, 2005). Image schema can be either static, dynamic, or action schema (Amant et al., 2006). Static schema are instantaneous descriptions of a particular non-changing relationship, such as the relationship to an object as being near or far away. Dynamic schema are just as the name implies — changing with the relationship and related to recognition of this change; for instance, the act of determining whether the distance between two objects is decreasing. Finally, action schema involve the idea of intention: the understanding that different actions relate to different outcomes, and that an individual must determine a course of action to create a specific change. This type of schema chains together the other two; in a musical context, an example of an action schema could be the understanding of a melodic line and interpreting the individual actions needed to reach each note along that line. Once a technique, emotional expression, or other musical outcome is achieved, musicians are able to refer back this image schema to recreate the experience or understand differences in future experiences. This is done through metaphor; because image schema are part of higher-level cognition, it is necessary to apply manageable and tangible substitutions, in the form of metaphors (Amant et al., 2006; Dewell, 1994; Hampe and Grady, 2005; Howell and Archer, 1984; Lakoff and Johnson, 1980).

3.3.2 Metaphor in HCI

HCI Metaphor traditionally focuses around WIMP (Windows, Icons, Menus and Pointers) organisation of digital actions. For instance, *icons* are small images which represent computer resources as familiar, physical objects, such as a trash bin representing file deletion or folders location in digital memory where documents are stored. The *Desktop Metaphor* was intended to make the computer relatable to and part of a typical office environment (Canfield Smith, 2020), often by recreating a physical desktop in a digital space (Johnson et al., 1989); for instance, windows represent physical sheets of paper. When computers became a part of typical office environments, metaphor made this complex technology, previously was only usable by computer experts, accessible to users in other fields. The ability to interact with a digital desktop as one would with a physical desktop goes as far as to introduce physics properties so that papers will flutter down when dropped onto a pile (Agarawala and Balakrishnan, 2006), leveraging real-world physical interaction to make things as life-like as possible.

Icons provide a source of *Direct Manipulation* by providing interaction through representations of action, rather than command line input (Shneiderman and Maes, 1997a). Canfield Smith describes icons as effective because they have both visual and computer semantics; when dragging files to that trash bin icon, somewhere something is being deleted from the computer's memory (Canfield Smith, 2020). Desktop Metaphor allows complex processes to be expressed as everyday functions, democratising computer use by providing familiarity to non-experts. However, there are also many instances of effective computer UI that does not involve icons, for instance the Sketchpad's representation of functional programming through manipulation of shapes and patterns (Sutherland, 1963), graspable interfaces and gestural communication (Voorhorst et al., 2000), and other post-

WIMP interaction (Beaudouin-Lafon et al., 2001a,b) and tangibles (Hornecker and Buur, 2006). Where Canfield Smith suggests that icons have fallen by the wayside due to a "failure of imagination," in coming up with new Desktop Metaphor objects, Beaudoin-Lafon suggests a breakaway from monolithic representations of physical objects through a variety post-WIMP interaction techniques (Beaudouin-Lafon et al., 2001a). This is because there are notable limits to the Desktop Metaphor, namely that simulating a real-world environment neglects the affordances of the digital world and enforces physical limitations that can easily be overcome with a computer (Smith et al., 1985); for instance, processing thousands of digital documents — many more than can be physically placed on a desktop (Dumais and Jones, 1985).

This disparity between Desktop Metaphor and other representations of computer processes suggest that there is a tension between HCI and real life which causes a breakdown in what can be communicated by traditional HCI metaphor. Hornecker notes in response to physical embodiment through tangibles that social understanding provides grounding; they require learned associations through experience and context (Marshall et al., 2009) and do not work merely because they are physical objects (Hornecker and Buur, 2006). Although Canfield Smith is critical of Apple's movement away from the Desktop Metaphor (Canfield Smith, 2020), users are still able to use their phones through experience. Additionally, humans are able to understand and relate to each other through very abstract metaphor, particularly in cases where there are no external objects to be represented: in regards to sensory experiences, icon representation would be difficult and perhaps too rigid for representing something so difficult to express. This understanding, and perhaps even the definition of Metaphor in HCI as a depiction of reality, is something which should be questioned (Blackwell, 2006); I explore this later in Chapter 6 within the context of vocal pedagogy.

3.3.3 Contemporary Metaphor Theory

Humans use metaphor in other, even abstract ways to exchange information and understanding. Contemporary metaphor theory describes a more flexible and multi-modal representation (Lakoff, 1993): rather than depicting reality, metaphor is the way we map information existing in one domain to another. This mapping can take the form of words, but metaphor is cognitive and conceptual (Lakoff, 1993). Humans are able to understand and use abstract representations based on other life experience (Lakoff and Johnson, 1999); information can become easier to understand (Blackwell, 2006) and communicate (Utsumi, 2005) when expressed in a different modality. This explains why the Desktop Metaphor can be effective: we have taken a complex computer process and expressed the action through a relatable image we understand from the physical world. But, based on this contemporary metaphor theory, "Metaphor" is not just icons and should also include representations of unfamiliar processes through tangibles and multi-modal data representations, for instance visualisations. Through the remainder of the thesis, we therefore refer to this contemporary theory when talking about metaphor as a concept, with the aim of expanding the way we think about metaphor in HCI.

Metaphor mappings are based in understanding of the body through proprioception and tacit knowledge (Svanæs, 2013, 2019), as well as mental *imagery*, which allows us to imagine experiences without needing to actually experience them (Godøy and Jørgensen, 2001). Lived experience helps us to form this mental imagery, enabling us to plan further interaction (Chiel and Beer, 1997; Depraz et al., 2003) and understand new or abstract information (Jeannerod, 1995, 1999; Kosslyn, 1980; Kosslyn et al., 2001). The relationship we have with metaphor is also a constantly evolving one; these mappings are made up of mental *schema*, which are underlying ideas and concepts we observe in our interaction (Johnson, 1989; Langacker and Lakoff, 1988). Schema are built through experience and cultural exposure. These schema, with repeated association, become embodied over time and can be extended to more abstract concepts (Antle et al., 2009; Daudén Roquet and Sas, 2021; Langston, 2002); for instance, the directional schema that future is forward and the past is backward, or that happy is up and sad is down (Hurtienne et al., 2020). Metaphor therefore allows us to understand and express our lived experience and embodiment. The schema associations we make between different interaction can sometimes be cross-linguistic (Gibbs et al., 2004), demonstrating that metaphor is rooted in the body and conceptual mental mappings (Lakoff, 1993), which can then inform language or other multi-modal mappings (Kosslyn, 1980; Ohrenstein, 2003). Metaphor is **not reality**, but rather provides a way for us to apply our lived experience to understand new information and interaction with the world.

Although we might not currently call it metaphor, the focus on phenomenological and embodied perspectives in HCI (Harrison et al., 2007; Höök, 2018; Poulsen and Thøgersen, 2011) has provided a number of approaches to express individual experience and information through abstract representations (Schiphorst, 2011). Somatics and somaesthetic design in particular use metaphor to explore embodied understanding of interaction (Höök et al., 2021; Shusterman, 2008) and keep the body and movement in focus. For instance, body mapping can be used to provide a non-linguistic representation of experiences rooted in cognitive processes, which might otherwise be difficult to explain to others and even to ourselves (Boydell et al., 2020; Daudén Roquet and Sas, 2020). Sensations in the body are represented visually through sketching (Gastaldo et al., 2018; Ståhl et al., 2021), painting (Cochrane et al., 2022), and clay sculpting (Núñez-Pacheco, 2021). Material speculation (Daudén Roquet and Sas, 2020; Wakkary et al., 2015) and temperature stimuli (Daudén Roquet and Sas, 2021) are used to provide tangible presence to sensory experience. Mapping can also help express temporal change within the body (Tennent et al., 2021) and illustrate design choices and strategies to others (Cochrane et al., 2022).

These mapping practices are metaphor, helping participants to find words for reflection and offering a substitute for language (Boydell et al., 2020). By redefining metaphor in HCI within contemporary theory, we can encompass these practices and open the possibility for expressing information in abstract, multi-modal ways, without being limited to physical objects. Metaphor is not monolithic, being more hermeneutic and dependent on the interpretation and understanding of the individual. This approach in expressing information can fulfill the need to acknowledge the plurality of the human experience and understanding, rather than attempting to find a "true" or perfect representation. Humans are relatively good at expressing their tacit knowledge and understanding through metaphor, as seen in the approaches mentioned above.

3.3.4 Metaphor, Embodiment, & Lived Experience

Contemporary metaphor theory also aligns well with principles of third-wave HCI, which centre interaction through and with our bodies, moving away from work- or task-driven, "purposeful" objectives (Norman and Draper , Eds.), to focus on social, cultural values in interaction, human understanding, and being in the world in everyday life (Felipe Duarte et al., 2019; Harrison et al., 2007). Post-phenomenological approaches focus on how our lived experience — our knowledge of being in the world (Tuuri et al., 2017),[1] as embodied through our individual bodies — is intertwined with the body as part of our identity. In opposition to more Cartesian mind-body dimensions, the cognition of the mind and sensory experience of the body are seen as inseparable; embodied knowledge arises through the understanding of and interaction with the world through the *whole* organism. (Husserl, 2014; Klemmer et al., 2006; Merleau-Ponty, 2014; Varela, 2010). Every action

[1] When discussing "being in the world," it is often required to cite Heidegger. It is not possible to separate the writing from the writer, as it were, and I have made a conscious choice to not reference the antisemitism and Nazism entangled with Heidegger's work in this thesis.

depends on a highly internalised working routine formed as a result of this being in the world (Tuuri et al., 2017). Perception, action, and behaviour are linked (Depraz et al., 2003; Thompson and Varela, 2001) and entangled with the feedback we get from the world and from technology (Frauenberger, 2019), which is informed through the sensory information we get from acting and living in our environments (Gallagher, 2005; Gallagher and Zahavi, 2012).

The principles of mental imagery, lived experience, and embodiment are therefore very similar, although these ideas have formed somewhat separately between cognitive science, HCI, and design practice. The key similarity is that the mind and body are indeed inseparable (Armstrong, 2007; Thompson and Varela, 2001; Varela, 1979; Varela et al., 1991) and that, rather than actively moving through each part of a refined action, the execution of a task relies on a broader mental image and intention. Recreation of an action and behaviour depends on personal lived experience and environment and is constantly changing with further knowledge. A vocalist's technique and application of their experience, rooted in mental imagery, become more ingrained through information gathered from the bodily sensations while achieving a desired outcome. This creates a feeling of "one-ness" with the voice; for many musicians, the instrument is entangled with and viewed as an extension of the body (Nijs et al., 2013; Souza, 2017) and part of the self (Bates, 2012; O'Bryan, 2015).

Over time and with experience, the line between body and technology begins to blur and effectively these become one functioning entity (Nijs et al., 2013). This allows a musician to execute complex actions through focus on their big-picture intentions of a performance, with less attention on finer actions such as motor control over their instrument. With experience, musicians learn to match their gestures with desired sounds and, over time, the fundamentals of sound creation can be done without constant or complete attention. This also means that technology and its presence in our interaction also becomes embedded in these action paths; technology becomes a part of our lived experience, and indeed who we are (Frauenberger, 2019; Verbeek, 2015) and the way we think of ourselves (Mice and McPherson, 2022). Experiences are therefore the result of our individual perception, made through our unique bodies and physical environments (Spiel, 2021). Because of individuality and plurality of experiences, it can be difficult to understand the sensory and embodied relationships of another person (Núñez Pacheco and Loke, 2016; Svanæs, 1997); in these cases, metaphor becomes a fundamental tool for sharing sensory-based experience.

3.3.5 Metaphor in Musical Pedagogy

Metaphor is commonly used during classical music training on any instrument, including the voice. Examples of such metaphors include teaching voice students to "throw their sound to the back of the hall" to achieve better projection of sound (some voice teachers even have students move their arms as if they were throwing a ball to help visualise their projection), or describing the breath as being "warm" or "cold" to represent the speed and pressure of the air through a brass instrument. These metaphors can relate both to the physical production of the sound and also to the quality or timbre of the sound, and musicians have a large volume of vocabulary to describe these more abstract concepts (Godøy and Leman, 2010; Trusheim, 1991).

In the context of teaching, students learn heavily through metaphor because it is often easier and more natural or intuitive for the instructor to relate abstract ideas to body movement and musical expression than to give detailed physiologically or anatomically based instructions (Jestley, 2011; Woody, 2010). Musicians will adapt these ideas during the course of their training, building a vocabulary through experience of what works to their individual preference. The formation of schema and resultant metaphor in music occurs in the blending of different modalities to relate understandable structures together; for instance, the relationship between text painting and expression in music. Language and resultant sound are tightly combined in multi-modal metaphors such as text painting in speech, allowing for the expression of such schema through concepts understood outside a musical context; this can include movements through space (descent down a staircase compared to descent down a musical scale), physical gestures (trembling or shivering compared to tremolo textures), or emotional state (Zbikowski, 2009).

The use of metaphor and descriptive words in performance therefore provides the basis of musical-emotional relationships. Reference to emotional state can be easily observed in the expression markings of nearly every composition; *Maestoso* (majestically), *Brillante* (bright and sparking), and *Luminoso* (luminously) are such examples. These ideas relate to a specific sound and expression learned by the musician and reapplied the next time they see a score with the same marking. Studies have also found that planning of expression in performance results in greater emotional quality and efficiency in execution (Woody, 2010). The imagery and metaphors used by musicians demonstrate a high level of emotional and tonal detail, even to the extent of mimicking an exact sound—often that of a teacher or of a widely regarded performer (Trusheim, 1991). Musicians are found to possess what has been described as an "inventory" of sounds by Trusheim, in the sense that they can identify the demands of a particular performance in terms of technicality and expression needed and pull from previous life experience and training to select the correct visualisation of sound in the mind's ear (Trusheim, 1991). The overall quality of the sound and performance depends on the ability for the musician to visualise their goal before execution, and explains why so many musicians use metaphor as part of their regular practice in honing their musical imagery abilities.

The perception of a sound is thus very dependent on the culture and language used to relate to it; similarly, musical image schema are formed in different ways and from different perspectives based on the learning environment. This relates heavily to musical expression, such as characteristics of articulation, melodic and harmonic movement, dynamics, and tonality; in an example relating to tonal distance, Serbian children describe a pitch and a lower octave to be "high and low," while Romani children describe this as "big and small" or "thick and thin" (Antović, 2018). Performers apply both their personal interpretation of the musical score and their image of what the score should sound like to convey the written ideas of the composer. In this way, the resulting interpretation of any piece is entirely up to the performer to convey as they see fit (Ebie, 2004; MacRitchie and Eiholzer, 2012); this explains why songs performed by different artists will ultimately sound different, and why covers of popular works are common in different styles and genres. Despite individual differences, it has been found that musicians (even on different instruments) employ similar manipulations of melodic lines to convey specific expressive performance goals (Gabrielsson and Juslin, 1996).

3.4 Vocal Pedagogy & Perception

The interaction with the body while singing happens almost purely internally; there is no physical interface to touch. Although we can see some movements as we sing, many action paths employed in singing consist of small internal movements. As with other fine-grained actions based on tacit knowledge, it is very hard to find appropriate language to describe what it feels like to sing. This is true of other vocalisation — what does it feel like to speak? (Cochrane et al., 2022; Núñez-Pacheco and Loke, 2020; Petitmengin, 2006). Yet, voice teachers are able to communicate about healthy singing practice using metaphors to represent their own experiences in an understandable way (Dunbar-Wells, 1999). Because of the removed nature of the voice from the musician, vocal pedagogy uses a depth of metaphor-based teaching to relate vocal concepts to students, or through sharing of personal experience, rather than being rooted in physiology (Callaghan, 1998; Miller, 1996). Voice teachers must be able to describe their awareness of their own bodies in a way which is relatable but not so specific so that a student cannot interpret the metaphor within their own experience. This metaphor-based teaching has been used as the basis for vocal pedagogy for hundreds of years. This is particularly the case when working with students who are still developing their own sensory understanding of their movement.

3.4.1 Vocal Metaphor in Practice

Commonly addressed focus areas in voice pedagogy include physiological aspects, such as body alignment, breath management, resonance and phonation, and sound production (Jestley, 2011). The highly individual nature of the voice as an extension of one's body also means singers are found to use metaphors for timbre intuitively, suggesting a kind of synaesthesia between the imagined or desired sound and the body itself (Prem and Parncutt, 2008). Voice instructors employ metaphors as observed in one of two overall categories described by Dunbar-Wells in the context of a voice lesson: (1) trigger metaphors, single word metaphors which are used by a teacher to refer to a grouped set of actions once understood by the students (for instance the word "tuck" being used to describe the alignment of the abdominal muscles and spine while singing and then later as a cue to remind the student to realign their body), and (2) descriptive metaphors, references to a certain aspect, such as the tone of the sound having a characteristic, perhaps "bright" or "nasally (Dunbar-Wells, 1999)." In the thesis later written by Jestley about metaphorical and non-metaphorical imagery in the voice lesson, it was generally found that teachers use a combination of literal explanations and metaphorical imagery to elicit a particular response from their students and examine the sensorimotor coordination and feeling of the action (Jestley, 2011). Metaphor often relies on tacit knowledge (to be explored more in the study presented in Chapter 6); for instance, the image of a tyre expanding around the waist is a common metaphor used to help vocal students focus on the tension in their diaphragm. Although most vocal students would not have experienced this exact sensation, it is not hard to imagine what this might feel like; through tacit knowledge, the singer is able to extend their awareness and understanding of the body and create an idea of un-experienced sensations such as a tyre around the waist, being pulled up by an imaginary string, or wagging an imaginary tail (Svanæs, 2019; Svanaes and Solheim, 2016). It is believed that, with the repeated practice and training using these metaphorical references, singers can produce sensorimotor responses to join and coordinate their perception and physiological actions (Ohrenstein, 2003).

Dunbar-Wells presents a model of metaphor-based voice teaching (Figure 3.2) wherein the teacher's metaphor is translated into a sensory-based image that is internalised and then recreated by the student (Dunbar-Wells, 1997, 1999). In this way, the teacher must be able to first understand and articulate their sensory experience (Emmons and Thomas, 1998; Green and Gallwey, 1986) and then the student must apply their own understanding and experience through imagery to recreate that sensation (Cumming and Williams, 2012; Dunbar-Wells, 2003; Godøy and Jørgensen, 2001). Singing is an interesting context to examine this translation of metaphor because there is little else to help the teacher and student; most of the vocal practice is visually hidden within the body. Vocal pedagogy is an age-old method for teaching the movements and internal sensory experiences which are felt during singing. With the instrument existing fully within the body, the singer must heavily rely on this internal perspective and the relationship they have with their body. For vocal pedagogy, metaphor is therefore considered to be critical to understanding these sensorimotor experiences and relaying them to others (Günter, 1992a,b; Hemsley, 1998; Salaman, 1989). Image-schematic structuring is believed to help "codify" terminology in vocal pedagogy (Wilson Spillane, 1989). Jestley describes how voice teachers consider metaphor to be "the main vocabulary or language for describing what they actually sense while singing" (Jestley, 2011). This is not for a lack of understanding of the interaction, but rather because the act of learning and using the voice occurs within a sensory domain; singers innately think about singing in a sensory and non-verbal way (Hines, 1983).

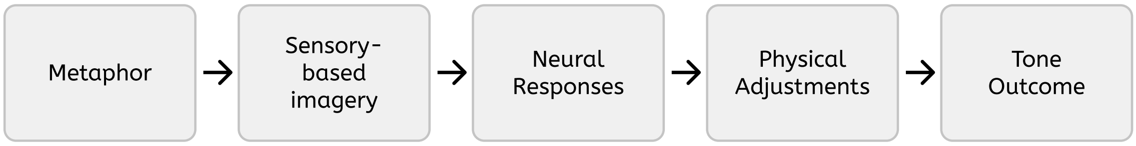


Figure 3.2: The model for translation of a teacher's (communicating agent) metaphor by a vocal student (receiving agent) into physical movement and then sound, proposed by Dunbar-Wells (Dunbar-Wells, 1997, p. 152).

It is worthy to note also that voice teachers are often found to be very unaware of physiological processes or acoustics (Callaghan, 1998), meaning voice education consisting of metaphor is based primarily on this internalised feeling or underlying schema, which is passed down from teacher to student. This is apparent in examination of the contrasts between teachings in classical voice pedagogy, and varies from teacher to teacher, according to their individual experiences and understanding of singing and its mechanisms (Coffin, 1989). The student's own understanding comes from the application of these abstract ideas to their study and performance. Compared with Desktop Metaphor, metaphor as understood through its use in the voice lesson actively *avoids* being a truthful representation of how the body works. Both models of metaphor rely on what I will call the *receiving agent* (e.g., a user or student) to interpret the reference given by a *communicating agent* (machine or teacher).

3.4.2 The Difficulty in Vocal Metaphors and Communication

There is a source of debate in the vocal education community about the validity of metaphor as a teaching method. Some metaphors encountered in the voice lesson are indeed found to be helpful and clear, yet it is not uncommon for singers to run into references that seem arbitrary or confusingly fantastical (Miller, 1998b,c). For instance, this metaphor discussed by Jestley (2011), as referenced from Sell (2005), and now repeated here to again bring attention to its ridiculousness:

> "Sing like you are an ice cream sundae with hot fudge dripping down the sides. . . . Pretend your diaphragm is an ice rink in front of your body, and every time you begin to sing, a little angel comes down from heaven and lands on the rink, twirling as fast as she can. Make your voice sound like that." (as documented originally by Sell (2005), pg. 116).

This kind of extreme descriptive metaphor highlights some of the negative aspects of communicating or teaching in abstract languages. While it might make sense to the teacher, the student would likely not be able to decode the abstract meaning of the teacher's imagery. Many vocal pedagogues speak of this kind of metaphor when discrediting its use in the lesson, arguing that such language is too random, subjective, and "mumbo jumbo" without a solid foundation of the objective physiological aspects underneath (Miller, 1980, 1998a; Vennard, 1967). There is a push in pedagogy for students and teachers alike to question the subjectivity of such metaphors and to seek practical, consistent, and direct connections instead (Daniels, 1983; Miller, 1995). It is important to note these kinds of examples while exploring metaphor and to acknowledge that harm can be done to students when teachers direct without knowledge of or attention to vocalisation as a physical act (Miller, 2001).

Misunderstanding and miscommunication between students and teachers can lead to frustration, confusion, and often physical harm when not properly addressed. Even for non-singers, the voice is part of our identity – it is how we express ourselves and communicate with others. As well, since the quality of the voice depends on individual physiology, everyone's voice is a unique, identifying aspect. Many singers form strong identities through their voices and their craft (as is common with other artistic practices) (Achey et al., 2016; O'Bryan, 2015). While this connection is a powerful tool in terms of musical expression, it can cause detriments to emotional, mental, and physical well-being when the connection to the voice is disrupted. There is a large body of musical research which stresses how metaphorical-based teaching can potentially lead to communication breakdown between teachers and students when not employed correctly (Miller, 1996). This is unfortunately a common occurrence in the voice lesson. This can be personally distressing to students when the miscommunication is not quickly addressed or met with flexibility and support from the teacher during the learning process. These students, who have personal identities rooted in their singing, may push their bodies too far in an attempt to please their teacher or meet their own expectations of themselves. Vocalists are also often trained in ways which perpetuate an overemphasis on sound and replicating particular behaviours, rather than physiology (Eidsheim, 2015; Stadnicki, 2016). With pedagogical approaches that neglect vocal materiality and physical bodies, this can potentially lead to physiological damage and even loss of the ability to sing.

However, there is still a need and even desire for metaphor in vocal pedagogy. Metaphors are critical even in understanding our own experiences (Antle et al., 2009; Daudén Roquet and Sas, 2021) Historically, vocal education has been based around imagery and communication through abstract references (Cleveland, 1989; Hines, 1983), having existed for much longer in time than any of the concrete physiological knowledge of the voice. As well, many vocal teachers are not well educated in physiological structures of singing (Callaghan, 1998) and there is much to suggest that incorporating a more literal, anatomical approach would prove equally confusing to a student (Annett, 1994). Metaphors and image schema extend the understanding we have of the world, making them often more intuitive and easily understood (Hurtienne, 2016; Macaranas et al., 2015). The way metaphors are understood or internalised by others can potentially influence their understanding of movement; when considering the body, it is crucial to ensure that we use metaphors which provide a foundation for individual meaning while focusing on the perception of this communication.

3.5 Summary

To summarise, the voice has a lengthy history of being taught through abstract conceptions which form the basis of a singer's musical imagery. Imagery, the imagined sensory and emotional aspects of an action, co-opts the same neural pathways as execution and allows us to predict and react to the outcomes of our behaviour. In a musical context, this is used for motor planning, adaptation to altered feedback, and the inclusion of emotional content in performance. Vocalists have less overt connections to the feedback they receive while singing; because the instrument is located within the body and the vocalist must interact with hard-to-articulate internal sensations, vocalists have used metaphor as a tool to understand the voice. In vocal perception, abstract representations allow a singer to understand the physiological components of their craft and provide reference and language when discussing or teaching voice. To this end, when the content of vocal metaphors focuses on sensory experiences, it proves to be a useful tool for understanding the body.

However, the connection between the vocalist and voice is still difficult to explain to others, potentially causing confusion and harm when communication of sensory experience is misunderstood. It is very difficult even to comprehend our own understanding of such refined tasks, as they are innate, exist outside of a linguistic explanation, and can be done without conscious thought. In order to get at this understanding the relationship with the body, I ultimately use the surface electromyography methods discussed in the previous chapter to move some of these internal relationships outside of the body; when presented with an external sound, instead of an internal sensation, we can explore the movement in more detail and provide the sound as a sort of metaphor and context through which we can discuss the experience.

Because the vocalist-voice relationship is rooted within this abstract understanding, it provides a interesting case in which to study not only our interaction with our bodies but also with each other and with other tools. Having now outlined the more quantitative aspects of vocal physiology and interaction and the qualitative perception vocalists (and musicians in general) have about their practice, I would like to present a defined methodology, rooted in HCI and design practice, to further study this relationship. These differing fields of research can be united through current understanding of human-human and human-computer interaction; by utilising methods and theories from HCI and design, this thesis will uncover further understanding of the interaction we have with computers, other humans, and ourselves.

Chapter 4

Methodology

The related work presented in the previous two chapters demonstrates a tension between the methods used to observe vocal behaviour and gather data on vocalisation — vocal control and musical interaction methods being highly reliant on external data streams — and those used to teach and provide understanding of the vocal mechanics — pedagogy and metaphorical representations reliant on internal awareness of the vocal apparatus. From a machine perspective, the primary data stream comes from external vocal audio. Vocalists also use audio feedback, but as it relates to internal sensory experiences. The multi-modal entanglement from internal kinaesthetic feedback and external audio feedback allows for an understanding and connection between the vocalist's body and their sound through an abstract, internal perspective. Yet, if we consider the difficulty faced by these seemingly-opposing external vs. internal interaction methods, we see that they are really two sides of the same coin, united by the same challenge: the vocal mechanism is not easy to access, understand, or even really feel, other than through innate, intuitive understanding of our individual bodies. This PhD therefore requires a mixed-method approach to expand on the understanding of the vocalist-voice relationship in order to reconcile the traditional, quantitative, and generalised evaluations of physical movement and auditory feedback with the subjective, qualitative evaluations of highly individual sensory and emotional experiences. Although it will be necessary to develop methods for measuring aspects of vocal performance, it is equally necessary to *not* generalise individual experiences, but rather to create a clearer, more coherent examination of the different perspectives which are explored in this thesis.

4.1 Objective and Subjective Perspectives

I will begin with study of how the vocalist-voice relationship is used in task-based setting and how auditory imagery assists in performance, specifically in tasks where they must rely on these mental representations. Objective data and assessment through statistical analysis of performer accuracy in different singing tasks which force reliance on musical imagery can provide a clearer picture of the impact of this perception and how vocalists use it. Further, when examining metaphorical references in vocal practice, it will be necessary to look for correlations between vocalists' imagery abilities and the kinds of abstract perceptions they use to discuss and teach the voice. As well, it will be necessary to focus on measurable aspects of sensor implementation when designing a wearable for vocal sEMG acquisition; I aim to develop a tool which can detect activation of the laryngeal muscles in both vocalised and subvocalised singing (where imagery would be most heavily used).

However, because this thesis focuses on singer perception, I must also rely extensively on subjective methods to evaluate and explore the interaction between vocalist and voice with the technology designed and introduced in this thesis. Beyond developing an sEMG system for voice, I am not currently aiming to explore objective vocal gesture classification, but rather to sonify this data stream as a driver for interaction with one's own body in a novel way. I want to know how vocalists perceive their action and movement, how this changes their awareness of their practice and behaviour, and how they react to the information conveyed to them through sonification, which is not something that can be overtly measured.

All of this is not to say that data streams from audio analysis of singing or sensors are inherently "objective" and human-centred data is not. As we see in Chapter 2, vocal audio analysis can be interpreted in a number of ways by machines and teachers, and indeed sensors and their data collection are entangled with and dependent on the context in which they are designed and used (Barad, 2007; Nordmoen and McPherson, 2022). The important consideration of this thesis is in how these different data streams — objectively measured and subjectively described — used and where each is best suited. For this, an objective point of view will only be so helpful; there will be no "ground truth" in this interaction and the perception will be dependent on the individual singer and their background experiences, as discussed previously. For this reason, I focus on a mixed methodology and a combination of objective and subjective perspectives to examine individual experiences and interactions between vocalists and their voices.

4.1.1 Defining the Design Space

Using this dual-perspective approach within the existing literature space of several different research fields including phenomenology, HCI, cognitive science, vocal pedagogy, instrument design, I can now contextualise the methodology within a design space for this thesis going forward. Figure 4.1 demonstrates the main concepts, theories, and methods discussed so far. This figure can be referenced through this chapter as I discuss how they align within the methodology of this thesis. This design space will need to provide a way to examine the internal, subjective aspects of the vocalist-voice relationship through the external, observable feedback about the body which we are able to measure. I will endeavor to first solidify the connection between physical vocal physiology, vocal controllers, cognitive science, and vocal pedagogy research. The relationship between mind and body is expressed in many design principles which are necessary to consider when working in the musician's relationship with their instrument. I will define here *embodiment, lived experience, tacit knowledge*, and *entanglement* as they make up understandings of interaction in broader HCI research, and how these perspectives can be used to address the vocalist-voice relationship.

Going forward, I will refer to sensation and perception as related yet distinct concepts. ***Sensation*** is the most difficult to define, as it is an ambiguous and often wordless state of consciousness, but it will be used in this thesis to refer to feelings in or around the body or having bodily awareness through the senses during an interaction. I use ***Perception*** to refer to the sensory feedback received through the bodily senses. This includes feedback from conscious sensations and also what we unconsciously detect through our bodily senses. Both conscious and unconscious feedback form our understanding of how we *believe* we act and interact, whatever that truth may be for us.

Further key definitions of lived experience, embodiment, tacit knowledge, and entanglement are outlined in Chapter 3; in short summary, perception is formed through the base of knowledge we have ***Lived Experience*** through being in the world. Lived experience arises through our ongoing relationship between the world, what we learn from it, and changes the way we interact. ***Embodiment*** theory further outlines that our knowledge arises through our whole organism; the mind and body are not separate, but rather our sensations, perceptions, experience, imagery, and understanding are integrated together (Merleau-Ponty, 2014; Varela et al., 1991). Our individual bodies form our understanding of and interaction with the world (Husserl, 2014; Klemmer et al., 2006; Merleau-Ponty, 2014). This involves ***Tacit Knowledge***, the innate, embodied understanding which arises from the experience of living in our bodies (Svanaes and Solheim, 2016), which is similar to lived experience, but refers often in this thesis to the *un-experienced* extensions of knowledge we have as

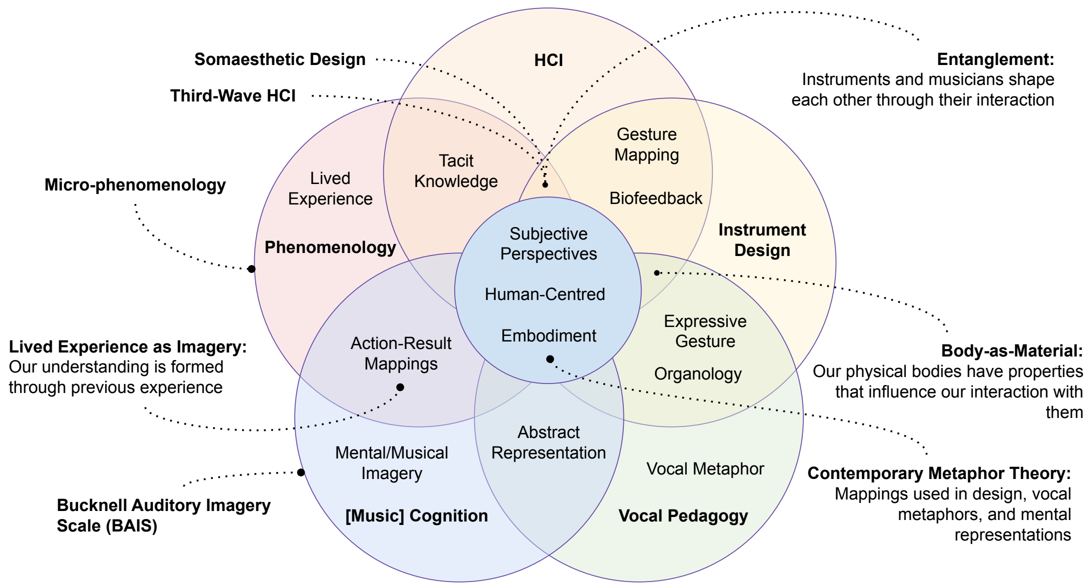


Figure 4.1: Theories, methods, and concepts outlined in Chapter 2 and Chapter 3, demonstrating the relationship between fields in this interdisciplinary thesis.

a result of living in our bodies (Svanæs, 2019); for instance, imagined sensations evoked through the vocal metaphors discussed earlier in Chapter 3 and later in a metaphor study in Chapter 6. The feedback we receive from this interaction further impacts our perception, actions, and approaches living within it in a highly cyclical way (Tuuri et al., 2017). When we are exposed to new interaction with different elements — instruments, technology, and other agents that make up parts of the interaction — those elements act in ***Entanglement*** with this lived experience; for instance, the musical instrument becomes entangled and shapes its relationship with the musician, who responds and acts accordingly to this relationship, particularly with prolonged experience (Nijs et al., 2013; Souza, 2017). As described in Chapter 3, Section 3.2.3, entanglement is not just a relationship between the human and non-human; I use this term specifically because it encapsulates how the human and non-human agents "co-constitute one another," shaping each other and enacting inseparable realities (Barad, 2007).

4.1.2 Designing for Interaction with the Voice

The main difficulty in developing technology which works with this embodied interaction comes from the fact that systems require some way of measuring gesture and action for input. When the machine's interpretation relies only on the results of a movement or interaction, rather than the lived experience and mental imagery driving that gesture, we lose its meaning and the intention in that action (Tuuri et al., 2017). Considering the covert ways in which most interactions with the voice occur, this is especially the case.

As seen in Chapter 2, the body's role in vocal sound production is often not fully considered when there is a preference on audio feature extraction alone. Many different vocal gestures may produce the same or very similar sounds, with nuances either going unnoticed or undetectable by current audio processing algorithms. We have little way of seeing or sensing how the vocalist moves and what technical knowledge and experience is driving their actions. In an expressive sense, much of the emotional intent in performance may also go unnoticed as recognition and classification of emotion in music audio analysis is often difficult and struggles to align with listener or musician perception (Alluri and Toiviainen, 2010; Aucouturier and Bigand, 2012), or even produce agreement on expressive intention and emotion among those listeners (Cowie et al., 2012; Yang et al., 2021). In a more serious sense, this interpretation or mapping of audio in an arbitrary way can lead to unnatural movement or miscommunication of gesture (Tuuri et al., 2017). Given the entanglement between musician and instrument, technology introduced to the musical practice has the capability to shape the body itself. With too much reliance on measured audio feedback to regulate vocal behaviour, the role of the physical body can be shaped in a way which leads to unsustainable behaviour, potentially ignoring or encouraging strenuous or damaging activity. It can be difficult to tell if someone is using unhealthy technique, as the resulting sound may still be deemed "desirable," to an interaction system, but the action may impact the health of the voice or cause long-lasting damage (Watson, 2009). Audio-based interactions without insight into the body may encourage singers with less-developed imagery and technical control to engage in unsustainable vocal practice.

This thesis therefore focuses on the internal aspects of vocalisation, both through imagery use, physiological movements, and internal sensations. There are four specific approaches and methods I have used throughout this thesis: I take a ***Somaesthetic focus*** to keep the vocal body at the centre of this thesis, specifically referencing my own lived experiences as a vocalist and my role as the designer and researcher in these studies, and ***Micro-phenomenology inspired co-investigation*** with the vocalists I worked with. I explore imagery use first through a performance study with other vocalists to frame the role of internal awareness and imagery in performance, focusing on how the vocalists are able to audiate and rely on previous interactions in performance tasks. I use ***self-assessed imagery questionnaires***, the Bucknell Auditory Imagery Scale (BAIS) and Movement Imagery Questionnaire-3 (MIQ-3) to examine the effects of auditory imagery on tonal and temporal accuracy in performance. With this practical, contextual understanding in place, I then provide a novel method of vocal interaction which returns to the bodily origin of the sound. Most of the current methods of looking at physiology related to singing, while effective and highly informative, would be largely unsuitable for performance or creative contexts. Due to the internal nature of the vocal mechanisms, medical equipment is often needed to observe voice physiology, making this potentially an invasive or expensive task. For this reason, we propose ***surface electromyography (sEMG)*** for sensing vocalists' performance intention through internal muscular movements.

4.2 Somaesthetic & Micro-phenomenological Approach

To address the embodied relationships that vocalists have with their voices, it is important to keep bodily sensations and physiology in focus. We must also acknowledge the role of the body in the design; by providing attention to their movement in this way, we may change the action or perception of the vocalists' practice and bodies (Homewood et al., 2021; Mice and McPherson, 2022). There are several strategies in HCI which have focused on the human's role and, more specifically, the role of the body, in interaction. This attention in design choices is rooted in phenomenology and works to make sense of the underlying understanding of both the designer and the intended user (Frauenberger et al., 2010). I specifically focus on Somaesthetic design and micro-phenomenology as disciplines to address the body's role in this interaction and design with the intended focus on individual lived experience and bodily sensations. These interactions are further explored through thematic analysis of the data collected on other singers' perspectives and autoethnographic portrayals of my own experience during this thesis:

4.2.1 Somaesthetic Design Principles

Somatics addresses the body and our connection to our internal sensations and tacit knowledge (Hanna, 1995) by incorporating felt experiences and *somas* — Shusterman defines soma as our "living, purposive, sentient, perceptive body or bodily subjectivity" (Shusterman, 2012) — into the design process (Höök et al., 2016). This is not an easy task, and one that requires the designer slow down enough to become sensitive to and engaged with their own soma in order to design from this place. This design practice has been used to explore embodied understanding of interaction (Höök et al., 2021; Shusterman, 2008). As well, tacit knowledge can be uncovered through awareness of movement or changes introduced into typical behaviour during a task (Höök, 2018); this can further inform the design process by using existing embodied relationships as a source for creativity and reflection on one's movement (Cotton et al., 2021b; Martinez Avila et al., 2020).

I referenced my own soma and my lived experience as a semi-professional vocalist throughout this study. The design and use of the vocal sEMG platform I developed in this thesis were made with consideration to my own body; how I understood my own vocal practice, my changing experiences when interacting with my voice in new ways, and the abstract understanding I use myself during my practice. This is detailed in Chapter 7 and Chapter 8. At the core of the work in this thesis are the 7 tenets of the Soma Design Manifesto (Höök, 2018):

1. We design for living better lives - not for dying
2. We design to move the passion in other and ourselves
3. We are movement, through and through
4. We design with ourselves - through empathy and compassion
5. We design slowly
6. We cultivate our aesthetic appreciation
7. We disrupt the habitual and engage with the familiar

Again, because tacit knowledge can be difficult to articulate, different soma-based approaches have been used to provide designers and practitioners with a way to convey the sensory experiences which emerge (Schiphorst, 2011). The sensations in the body can be explained or represented in visual ways, such as through sketching (Gastaldo et al., 2018; Ståhl et al., 2021), painting (Cochrane et al., 2022), and clay sculpting (Núñez-Pacheco, 2021). In a similar way, material speculation can be used to provide tangible presence to sensory experience (Daudén Roquet and Sas, 2020). The practice of making body maps or other tangible objects can help participants to find words for reflection, or offer a substitute for language in a different modality of interaction (Boydell et al., 2020). Mapping can help to express qualities of feeling and change within the body (Cochrane et al., 2022; Núñez-Pacheco, 2021); mapping soma trajectories at various points in an experience to highlight how these sensations develop in a temporal manner (Tennent et al., 2021). Additionally, body mapping and material speculation are useful for soma design to illustrate design choices and strategies to others and to transfer elements of experience into physical components (Cochrane et al., 2022).

Often, the design takes an approach of disrupting habitual practice to bring attention to elements of interaction we normally would not be aware of (Höök, 2010) and reveal hidden knowledge (Homewood et al., 2020). For instance, in the use of sonification to relay information about muscular movement to musicians, we adopt a somaesthetic approach (Höök et al., 2021; Shusterman, 2008) wherein something about the typical practice is disrupted to allow the singers to explore their tacit knowledge and bodily awareness in their practice, which they might not normally examine (Höök, 2018). As well, I will use the connection to the sonification to explore how vocalists might describe their movement through sound. This "sound sketching" in both hearing the body in a novel way and using sound to depict information about movement can provide an additional form of metaphor and communication about the vocal practice (Wirfs-Brock et al., 2021). As well, using these new perspectives can draw and stabilise attention to the body (Mah et al., 2021); through this modality, I will explore vocalists' perception, awareness of the interaction, and their sensitivity to different elements of the vocal practice.

4.2.2 Micro-phenomenology

Disciplines such as micro-phenomenology, developed by Petitmengin (Petitmengin, 2006) have been used to explore pre-reflective dimensions of experience (Petitmengin, 2021; Prpa et al., 2020). The discipline is derived from neuroscience practices (Varela and Shear, 1999) to investigate dimensions of sensory perceptions (Petitmengin and Bitbol, 2009; Petitmengin et al., 2013). The discipline builds on Varela's work in centring the human in experience (Thompson and Varela, 2001; Varela, 2010, 1979), Vermersch (Vermersch, 2006) and Depraz's creation and employment of an interview method for investigating professional practices (Depraz et al., 2003), and the work of Petitmengin in structuring lived experiences (Petitmengin, 2006, 2021).

Interview Methodology

Micro-phenomenology explores pre-reflective dynamics (Petitmengin, 2006; Vermersch, 2009); that is, perceptions which occurred during the initial experience but are inaccessible upon reflection. During daily life, sensory perceptions occur rapidly, and we are not immediately aware of all of the elements of our experiences. Micro-phenomenology helps uncover the hidden dimensions of experience by zooming in on a precise moment during an interview (Petitmengin, 2006, 2022). A trained micro-phenomenologist helps bring the interviewee back into the experience through an evocation. The interviewee becomes a co-investigator in recounting and refining their experience and dimensions. The interviewee is invited to explore the diachronic structure (the entire experience chronologically) and synchronic elements (dimensions in a clear-cut moment) of the experience. Synchronic details form the "landscape" of an experience, while diachronic details depict that landscape's evolution (Petitmengin et al., 2018).

The micro-phenomenological interview works to evoke an experience and bring an interviewee back into their interactions (Prpa et al., 2020). The data provided by the interviewee is from a second-person perspective, which can be thought of as narration; the interviewer conveys the interviewee's experience offering balance to the first-person subjective and third-person objective components (Reed et al., 2022a; Varela and Shear, 1999). Because the experience captured is pre-reflective, the discipline allows investigation of tacit knowledge and can bring attention to the individuality of experience in interaction. The discipline has been used within HCI and music interaction, more specifically, to focus interviewees on their experiences (Prpa et al., 2020) and the explanation of *what* happened, rather than *why*. It also aids the development of a vocabulary for embodied dimensions of knowledge and has provided space for interviewees to re-live their experiences and gain new insight on their interactions (Reed et al., 2022a).

I describe this work as "micro-phenomenology inspired" as I am still in the process of completing my micro-phenomenology training. At the time of writing, I have completed interview and analysis training in the micro-phenomenology discipline during the course of this PhD and engaged in about

200 hours of interview practice. So, I will use micro-phenomenological inspired methods to uncover important aspects of interaction. This includes the co-investigation singers' perspectives and relationships formed when working with the novel interaction methods developed in this PhD. Although the experiences will again be subjective, micro-phenomenology has been used to uncover generic structures of experience amongst and within participants (Petitmengin, 2006; Petitmengin et al., 2018). I will hope to uncover new perspectives about singers' pre-reflective experiences through its use in this thesis.

Considerations and Limitations

As with other subjective evaluations, we will uncover very individual aspects of experience and it is worthwhile to mention that the sensory details discussed should not be taken as any kind of generalisation for the whole human population. I will endeavor to explore commonalities and generalisation in the *structure* of the evoked experiences and how these experiences occur. This structure which provides validation to the interview method; although the details of the experience may differ in subsequent interviews, interviewees convey consistent structure as they become aware of their experience through the interview procedure (Petitmengin and Bitbol, 2009).

There is always the potential, as with any research method, for bias. Participant bias can be mitigated through the ambiguous way in which the interview questions are constructed. The method's core tenet is to guide but not lead; that is, the interviewer must be carefully trained in precise methods of evoking the experience to allow the interviewee to unfold the experience and report on it. Questions such as "When X happens, what happens/what do you experience?" provide direction for the interviewee but do not give detail on an expected reply. In cases where experimenter bias might arise, for instance in unintentionally leading questions, the interviewer is trained to recognise and remove the question-answer pair from further analysis. Although participants may still come up with exaggerated or manufactured sensory experiences in order to please the interviewer, the structure should remain true to the evocation as questions do not imply particular facets of experience and instead provide space for the evocation.

It is important to note some current inaccessibility of micro-phenomenology training (Reed et al., 2022a). The training works like an apprenticeship or luthiery and can pose considerable financial (800-1600€) and time costs (80+ hours of training and supervised interviews) to the researcher. These are high due to the intensity of the course and a limited number of micro-phenomenologist supervisors. While the time requirement is necessary to learn and correctly apply the discipline, it may be hard for some researchers to commit to the training due to work or care responsibilities. The financial cost poses inaccessibility to researchers without funding or available support from their institutions.

Even though micro-phenomenology focuses on pre-reflective experience, describing embodiment is still difficult to do. It is challenging to provide language for elements of tacit knowledge which do not have a language to begin with. Further, when working with experts, it is possible that the awareness of finer details and insights into practice may be complicated to describe. Micro-phenomenologists might find that some people are potentially less responsive to this method or find it challenging to reach evocation states. The responsiveness of an interviewee or their ability to describe experiential states might depend on their engagement in other meditative practices or cultural and social differences. Additionally, the achievement of the evocative state depends on the interviewer's experience and time studying the discipline.

4.2.3 Reflexive Thematic Analysis

I will also frequently use reflexive thematic analysis to explore vocalists' interaction and perception in this thesis. The method, as proposed and extensively researched by Braun & Clarke, focuses on salient themes which arise in participation (Braun and Clarke, 2006, 2012); typically, the method is useful in analysing interview, video, and other more open-ended feedback from participants where answers would not fall into a categorical or numerical structure. Rather, the method allows for examination of key aspects of the examined paradigm; in the case of this thesis, notable perceptions of vocal practice and interaction.

Interview Methodology

I will conduct a number of semi-structured interviews during which I will ask vocalists guided questions that are pre-determined before the time of interview. Unlike micro-phenomenology, this method is reflective; interviewees look back on their experience and we discuss the examined elements from a more general viewpoint to better understand high-level concepts. I will use a bottom-up inductive strategy for the thematic analysis; that is, rather than going into the analysis with an existing hypothesis, I will approach the collected data openly and refine what I learn into themes of understanding. Thematic analysis works firstly through generating a series of "codes;" the codes are notable pieces of feedback from participants. The codes are reviewed and then categories through an iterative process into larger and larger topics, and then finally into relevant themes.

Reflexive thematic analysis at its heart goes back to the lived experience, context, and understanding of the person doing the analysis. In my somaesthetic approach, I reference my own experiences reflexively to provide understanding and narrative to the experiences of the vocalists.

Considerations and Limitations

Again, the findings of subjective methods should be taken as representative of the individuals who were interviewed and not that of the general population. This method is appropriate in the examination of subjective experience because it allows for the individuals' data to inform the exploration. There is also no assumption of homogeneity in the responses (Braun and Clarke, 2020a); thematic analysis does not rely on a set of rules about how many codes must be found or how many must exist to create a theme. Indeed, themes of a single code can exist; if the data point provides a pertinent contribution to understanding the participant, it is included. In a more traditional research mindset, this is perhaps problematic because thematic analysis cannot provide a generalisable, reproducible result in practice. However, this provides the benefit that outliers in data are still contributed to the larger picture and individuality in experience and perception is accounted for and acknowledged, rather than being discarded or argued.

It is possible that participant bias is introduced as interviewees may attempt to help the researcher; however, the questions are decided and reviewed beforehand and participants are continuously reminded that there is no right or wrong answer (and sometimes, there might be no answer at all). As the interviewer, it is my responsibility to structure the interview as a dialogue; allowing space for interviewees to elaborate their point and provide as much information as they like. By sticking to the structured interview questions and not pressing for more details which may not exist, the interviewee can mitigate leading or reacting to responses in a way that might produce experimenter bias.

4.2.4 Long-Term Autobiographical Design

Finally, this thesis will involve an autoethnographic approach and autobiographical design. As the designer, I applied my own experience as a classically trained mezzo-soprano to the development of the vocal sEMG interaction. I sing regularly at the semi-professional level in both choral and solo settings. At the time of the writing of this thesis, I have been developing and living with the system for nearly three years and so have a long-term view of its use in a daily context. I tested the system around my use and the observations the sonification had on my own vocal practice informed much of the resulting research and studies done with other singers. Although autobiographical design is only recently becoming more reported on in HCI research (Erickson, 1996; Neustaedter and Sengers, 2012), I believe that this form of design and perspective is beneficial to focusing on incorporating lived-experience and embodiment in interaction (Goodman et al., 2011; Höök et al., 2015; Neustaedter and Sengers, 2012). It is a well-suited method for this exploration of intention with the voice, as I was able to provide a context-specific knowledge to the creation and wear of the vocal sEMG system. The autobiographical design also allowed for development in its intended environment and immediate implementation with my own well-established vocal practices and imagery, which could be called upon in active reflection during my use to meet any challenges (Schön, 1984).

A key consideration of autobiographical design is that documentation of my experience does not provide a universal view of others' perspectives. Living with the system through its development and designing for use by the designer, rather than a hypothetical imagined user, the design became highly specific to my individual use — this somatic approach provides more of a perspective of my body and individual experience than the system as a whole. The use of the design grew up around this long term relationship and evolved as it did in a way that would unlikely be the same for others (Erickson, 1996), who might approach the system from a completely different perspective. However, this type of long-term personal use allowed for its evolution through customisation (Erickson, 1996), and means that the sEMG system in its current state actually describes some of my own perspectives and intentions in using it. The act of tinkering with the system to meet a more tailored use as time went on allowed me to create a relationship with the system, which has been previously observed in similar long-term interactions (Erickson, 1996; Gaver, 2006; Wakkary et al., 2018). Similarly, when used by other singers adapting this system and growing through it (Zimmerman et al., 2007), I would expect the journalling and self-study methods to reveal factors unique to those individuals. This could either lead to generic features which are useful to many or implementation of features which allow any potential user to customise the setup around their own practice, which inherently would reflect the user in the system.

In the end, the goal remains the same — to better understand the relationship between vocalist and voice as it exists for the singers I will work with in this study. Through this approach, I will demonstrate how interactive system design, even beyond the singer-specific context in this thesis, can impact users' perception of their actions and their bodies and change the way we interact. With this in mind, I will now address gestural and physiological measurement technology and how I will explore the vocalist-voice relationship through surface electromyography.

4.3 Measuring Musical Gestures

Most of musical gesture can be observed from an outside perspective, so technologies surrounding this area of study are mainly non-invasive, as compared to voice physiology previously mentioned. There is similarly worth in the development of non-disruptive, portable, and inexpensive devices as to capture unbiased performance — that is, performance data which is true to the body's movement, unaltered by recording devices, image capture, or other observer perspectives. Outside of a medical context, it is necessary to use other devices for sensing movement during musical performance. Common tools used in the design of NIMEs include motion capture, popular for larger groupings of the limbs and whole body movements (Dahl, 2014; Nymoen et al., 2011; Santos et al., 2021; Skogstad et al., 2010), biosignals such as electromyography (Reed and McPherson, 2020; Tanaka and Ortiz, 2017), heart rate, and respiration (measured through pressure and stretch sensing) (Cotton et al., 2021a,b; Tsaknaki et al., 2021), other sensing methods for acceleration (Lee, 2021; Mice and McPherson, 2022), inertia (Skarha et al., 2021), rotation (Kirkegaard et al., 2020), position detection (sometimes co-opting existing technology for other gesture recognition, such as Leap motion controllers (Granieri and Dooley, 2019; Han and Gold, 2014), Wii remotes (Tsoukalas et al., 2018), or Kinect controllers (Fernandez et al., 2017)), optical tracking (McPherson, 2013), magnetic resistance (Morreale et al., 2019; Temprano and McPherson, 2021), ultrasound (Ciglar, 2010; Vogt et al., 2002), pressure (Cazzani, 2015; Essl et al., 2010), and many other examples (Wanderley et al., 2006).[1] I acknowledge the breadth of methods here because any of them would be useful for tracking physiological aspects, body movement, and large and small gestures alike while singing. However, I chose to focus specifically on surface electromyography; referencing these methods demonstrates the particular perspective sEMG offers into the body's movement to address the challenges of working with the voice, which I will elaborate on in Section 4.3.2.

4.3.1 Measuring Aspects of Voice Physiology

Most measurements of the vocal physiology have originated within medical research. Frequently used methods of examining the detailed physiological activation in vocalisation (particularly in regards to the larynx) involve invasive procedures, such as endoscopy (Roy et al., 2013) or cannulation (Herbst et al., 2015; Nacci et al., 2017), or extensive MRI (Echternach et al., 2012; Hertegård, 2005). There are also several commonly used tools for noninvasive examination of voice physiology, including strain gauging sensor bands, ultrasound (Hamlet and Reid, 1972; Kimura et al., 2019), electroglottographs (which measure the airflow through and contact in the glottis), and surface electromyography (sEMG) (O' Keeffe et al., 2022). In this thesis, I will focus on sEMG, which will be discussed later in Section Section 4.3.2. Industry standards for research of voice physiology include the use of other tools such as:

- *Respitrace Respibands (Ambulatory Monitoring, Inc., Ardsley, NY)*

  A respiratory inductance plethysmograph with elastic band transducers to detect ribcage and abdominal motions (Griffin et al., 1995; Prisk et al., 2002; Salomoni et al., 2016).

- *Physiometer PHY 400 (Premed, Norway)*

  An sEMG interface to record surface signal of the upper trapezius (TR), intercostal (INT), sternocleidomastoideus (STM), scalenus (SC), and posterior neck region (PN) muscles, and lateral abdominals (OBL) (Pettersen and Westgaard, 2004, 2005).

- *RES-117 Strain Gauge*

  A strain gauge sensor for measuring thorax circumference change during respiration (Pettersen and Westgaard, 2004, 2005).

[1] Methods used in musical applications have been documented in the SensorWiki: https://sensorwiki.org/doku.php

- *MC21 (Glottal Enterprises, Syracuse, NY)*

  An electroglottograph (EGG) to measure vocal fold contact area (Griffin et al., 1995). Two electroglottographs (dual channel electroglottography) can be applied to accurately observe vertical larynx position (Pehlivan and Denizoğlu, 2009)

- *Laryngoaltimeter*

  A portable device utilising condenser microphones positioned above the glottis for measurement of vertical laryngeal position and control (Pehlivan and Denizoğlu, 2009).

- *Digital Phagometer*

  This has not been applied in singing studies; however, a combined piezoelectric sensor and digital event counter/recorder have been used to track vertical motion of the larynx in swallowing (Pehlivan et al., 1996). The upward-downward motion in inhalation may similarly be observable.

With the current state-of-the-art in vocal physiology measurement defined, I will now turn to movement and gesture in music and how elements of the body are measured in a performance context.

4.3.2 Surface Electromyography

Electromyography (EMG) involves measuring the electrical impulses produced by the nervous system which cause muscular contraction. EMG is a common method used to observe the musculature in medicine and involves inserting an electrode via a needle into the muscle tissue to collect data on electrical signals passing through the muscle which trigger muscular contraction (Chowdhury et al., 2013). These signals convey information such as the intensity of contraction and work done by a muscle (Becker et al., 2018; Chowdhury et al., 2013; Tanaka and Knapp, 2017; Tsubouchi and Suzuki, 2010). More recent studies focus attention to surface level measurements of activity in the vocal tract and respiratory muscles, including the use of *surface* EMG (sEMG) electrodes placed on the skin (Kapur et al., 2018; O' Keeffe et al., 2022) and sensory bands placed around the muscle. This type of device not only avoids invasive procedures for the participants but also limits the construction cost and increases portability of measurement tools. Such portable tools are more easily used in natural performance settings, allowing the researcher to observe singing practices from a physiological sense while limiting bias caused by discomfort or disruption to the musician.

It is important to note that the raw sEMG signal is generally very low in amplitude, with signals ranging between 0-10 mV peak-to-peak (De Luca, 2002). sEMG is also notoriously noisy, as there are range of other bodily activities such as skin movement and cardiac activity which can interfere (Chowdhury et al., 2013), and the usable frequency of EMG signals lies mainly between 50-150Hz, making EMG susceptible to power line interference (Chowdhury et al., 2013; De Luca, 2002). This provides an opportunity to use ambiguity to explore movement and perception, described below, and brings design challenges which will be explored in the study presented in Chapter 7.

4.3.3 sEMG Systems & Tools

The most state-of-the art sEMG acquisition and filtering, the Trigno sEMG system (Delsys Inc., Natick, MA), allows up to 12 channels of data. This system has been used to gather activation of the inferior and superior infrahyoids, superior and inferior sternocleidomastoid, and trapezius muscles in vocal performance (O' Keeffe et al., 2022). However, the system is medical-grade and is financially inaccessible to the vast majority of instrument designers and academics working outside of speech-language pathology and related medical contexts. Other systems such as the BioMuse (Lusted and Knapp, 1988), MYO arm bands (Martin, 2018; Nymoen et al., 2015), and the BITalino EMG system (Tanaka, 2019) have been used for musical interaction; however, instrument designers have also turned to more of a DIY approach with sEMG in order to focus on flexibility into specifically musical interactions and extend beyond measurement of the forearms (the ever-popular but now discontinued MYO bands were only able to be used on the limbs).

The development of sEMG as a method for the study of musical practice and interfaces for expressive control has lead to the development of highly specified processors for this type of data. sEMG data is very noisy due to activity in surrounding muscles and other bodily processes occurring in nearby tissues; for instance, measuring muscular activity on the chest is particularly difficult given the interruption by the heartbeat (Chowdhury et al., 2013). Similarly, measurement at the neck and throat will be disrupted by the major arteries carrying blood to the brain with high blood pressure. Work by the Embodied Audiovisual Interaction (EAVI) Group at Goldsmiths University of London has addressed some of these difficulties. Currently, the group is in development of a low-cost dedicated board for sEMG human-computer interaction in music and instrument making (Di Donato et al., 2019), as many sEMG sensing devices are limited to DIY, as done in this paper, or medical-grade (and therefore highly expensive) equipment. The EAVI EMG board has been developed for dedicated collection and filtering of sEMG signals and is currently being improved for use in variety of interactive controllers and measurement systems (Di Donato et al., 2019).

4.3.4 sEMG in Musical Interfaces

sEMG has already been employed in a number of human interaction studies and tangible wearable systems to extend natural body language. In other computer science applications sEMG has been used previously to convey information to the user or another system about the force and direction of a gesture (Becker et al., 2018; Koike et al., 2006; Lim et al., 2020; Theiss et al., 2016), or as an alternative to other sensors for emotional communication through movement (Costanza et al., 2005; Hartman et al., 2018; Koike et al., 2006; Lim et al., 2020; Woodward et al., 2018). Converted to an auditory signal, sEMG is capable of providing discernible feedback about complex muscular activity through frequency and rhythmic conversion (Tsubouchi and Suzuki, 2010).

Within a musical performance context, the sEMG work of Tanaka and colleagues (Di Donato et al., 2019; Donnarumma et al., 2013; Tanaka and Knapp, 2017; Tanaka and Ortiz, 2017) and researchers at RITMO at the University of Oslo (Erdem and Jensenius, 2020; Jensenius et al., 2017; Martin et al., 2017, 2018; Nymoen et al., 2015) has extensively furthered understanding of interaction with this biosignal. sEMG has been used to detect different gestures, particularly movements of the head and limbs (Donnarumma et al., 2013; Tanaka and Ortiz, 2017), for control over digital audio synthesis (Tanaka and Knapp, 2017; Tanaka and Ortiz, 2017). The MYO armbands (Thalmic Labs, discontinued) in particular have been used for wearable controllers which incorporate both rotational and sEMG sensing of the arms and fingers in performance (Nymoen et al., 2015) and composition (Martin et al., 2018). In studies involving the use of EMG for detection of performance gesture, it has been found that, although it is difficult to classify some gestures, amateur users were able to quickly adapt their movements according to audio feedback and create new gestures to achieve desired sound (Jensenius et al., 2017; Karolus et al., 2020; Martin et al., 2017). This interaction allows for users to explore how their movements impact audio synthesis and can be useful for spontaneous composition using the body (Erdem and Jensenius, 2020). sEMG can be used to reflect and represent changes in a user's intention informed by learning with a system. sEMG in this sort of creative context does not only capture the user's existing knowledge, but also provides insight into the user's adaptation and interaction as they gain more feedback about their movement (Tanaka, 2015).

4.3.5 Affordances with sEMG

As mentioned, sEMG is an effective method for measuring performance gestures. sEMG also affords a few interesting perspectives that benefit its use addressing the gap of direct vocal controllers; because sEMG measures the neural impulses which lead to movement, systems designed using sEMG make use of subvocalisation and the "negative latency" factor. As well, the ambiguity in sEMG interaction is beneficial to exploration of gesture, cause and reaction, and musical intention during performance.

Subvocalisation

Subvocalisation, or the "inner voice," is the covert activation of muscles which takes place while reading or speaking to oneself in internal thought, and does not rely on any actual speech articulation (e.g., in mouthing words) to occur (Baddeley et al., 1981). Engaging in imagined action draws on the same neural structures used in actual execution of the task (Brodsky et al., 2008b; Kosslyn et al., 2001; Smith et al., 1995). The occurrence of subvocalisation in mental rehearsal when reading music (notational audiation) is well established. This has been the case of activity found by positioning throat contact microphones and sEMG electrodes around the thyroid cartilage (the Adam's apple) in reading melodic lines, as well as sEMG readings on the limbs of professional drummers reading percussion notation (Brodsky et al., 2008b). Mental action and executed action share the same neural pathways; because of this, the muscles are still excited by neural impulses when imagination occurs (Kosslyn et al., 2001). It has been found previously that the subvocalisation which occurs during use of this inner voice is linked to internal understanding of both auditory and kinetic relationships (Aleman and van't Wout, 2004). Thus, during subvocalisation slight activity occurs in the muscles used in speech articulation occur; although these are not outwardly discernible, even to oneself, the nerve impulses that occur are detectable and can be captured by electrodes positioned on the muscle. More on this neural connection will be discussed in Chapter 3, Section 4.3.6.

In recent work at the Media Lab at MIT, Kapur et al. have designed a wearable interface for silent speech — AlterEgo is a non-invasive mask that discerns low-level neuromuscular activity on the face and jaw to detect subvocal words (Kapur et al., 2018). Nerve impulses across the face and neck are measured on the surface of the skin during subvocalisation. The design of the mask focuses on the laryngeal, hyoid, buccal, mental, oral, and infraorbital regions as muscle articulators used in speech production (Hardcastle, 1976). These regions would also be of interest in the study of sung vocalisation, particularly the laryngeal and hyoid regions as they relate to the tension and movement of the larynx. Electrodes are placed at points in each of these regions to capture nerve impulses through these muscles. The signal is then input to a recognition model to classify the movements into words, which are then converted via Text-to-Speech back to the user. The current recognition model is trained on a limited set of vocabulary involving arithmetic and computation, with further expansions to the dataset planned. sEMG taken off the facial and jaw muscles during subvocalisation has been successfully used in the recognition and classification of internally spoken words (Garrity, 1977; Kapur et al., 2018; Meltzner et al., 2008).

The "Negative Latency"

The most notable difference in using sEMG, compared to other biosignals, is that the impulses being measured are present as a precursor to visible movement. The neural impulse causes the muscular tension and then the contraction with movement. sEMG therefore provides information about the intention of a wearer just before the gesture is made (Tanaka, 2015). As described by Tanaka, this results in a sort of "negative latency," where sEMG signals can indicate the occurrence of a motion a few milliseconds before it occurs (Tanaka and Ortiz, 2017). This is because of where sEMG lies in the action path: "A classical sensor, then, is at the 'output' of a gesture while the EMG is a signal that is the 'input' to a gesture (Tanaka, 2015; Tanaka and Ortiz, 2017)." Pressure, flexion, and force sensing would measure the results of an action; motion or audio capture of a vocal performance would likewise only allow for analysis of movement or sound which has already been produced, even in the case of real-time systems.

Ambiguity and Exploration

Due to sEMG's noisiness and the fact that we are not always consciously moving all of our muscles, there are inherent aspects of controllability and disorganisation when using sEMG-based control (Erdem and Jensenius, 2020). This, however, can be beneficial to studying embodied relationships in design. This ambiguity allows for users to develop their own relationship with a system, learn through play and experience, and act intuitively towards design without any pre-existing ideas of how the system should operate (Gaver et al., 2003; Sengers and Gaver, 2006). sEMG is highly responsive to movement, which would allow musical systems to take advantage of the very refined and subtle movements of experienced performers. In the opposite sense, these systems can also incorporate unconscious movement into interaction, allowing for users to explore their existing relationship with their bodies and aspects of technique which may have been adapted into the subconscious (Erdem and Jensenius, 2020). In this sense, sonic exploration with sEMG allows for intention and also effort and restraint to be used for control (Tanaka, 2015). Performers can adapt the force and exertion of their movements in addition to the overall shape of the movement in interaction. Restraint, conscious loosening of muscular tension, or explicit "letting-go" of control over muscles is a critical part of the embodied musical experience (Tanaka, 2015), making sEMG-based sensing well-suited to the existing highly-defined control musicians have over their instruments and bodies in performance.

The sEMG can then be used to drive a sonification, which allows a vocalist to "hear" aspects of their movement which otherwise would have occurred subconsciously. This practice mimics the normal action path which singers use – hearing the voice and then making small adjustments to technique, in a cyclical entanglement. However, this forces them to focus on the movement of a single muscle, isolated from a larger gesture, presented back to them in a non-vocal sonification (Igarashi et al., 2010; Tsubouchi and Suzuki, 2010). As well, multi-modal feedback from visualisation and sonification of gesture can strengthen imagery formed during rehearsal (Goebl and Palmer, 2008; MacRitchie and Milne, 2017). Sonification can lead to better reproduction accuracy in movement patterns (Effenberg, 2004) and allow for voice teachers to better relate to students (Igarashi et al., 2010) through this sensory feedback, in addition to relaying experiential knowledge. Sonification of biofeedback is found to be engaging and rewarding in rehabilitation through the emotional cues provided within music (Kantan and Dahl, 2019; Matsubara et al., 2013), which may further strengthen the learning process in terms of skill practice. sEMG also provides a way to experience bodily sensations externally, either visually or aurally as presented here, and potentially with other senses, which is known to reinforce learning (Bernal et al., 2015) and provide understandable information about the amount of and change in muscular activity (Tsubouchi and Suzuki, 2010). This kind of interaction is beneficial in that it provides a way for users to not only explore their conscious movements, but to also become aware of the unconscious action paths which have become internalised through prolonged experience.

4.3.6 Self-Assessment Questionnaires

The thesis used a variety of pre-existing questionnaires to examine imagery ability and musical experience.

Measuring Musical Sophistication

The Goldsmiths Music Sophistication Index (Gold-MSI) (Müllensiefen et al., 2013, 2014) was used to assess general musical sophistication and examine musical experience throughout the studies in this thesis. The Gold-MSI is possibly the most widely used questionnaire for self-assessment of musical experience. It has been validated and developed through an implementation by BBC LabUK's *How Musical Are You?* test, providing a large survey of the general population (more than 190,000 people between 2011 and 2012).

With respect to the general population — all the way from professional musicians to casual listeners — the Gold-MSI's evaluation does not give emphasis to any musical genres. It provides assessment in five subscales: 1) Active Musical Engagement (time and resources spent engaging with musical activities), 2) Perceptual Abilities (abilities to understand musical features and listening skills), 3) Formal Musical Training, 4) Singing Abilities (the ability to sing in tune, accurately, and be aware of one's own singing), and 5) Sophisticated Emotional Engagement with Music (being able to talk about music and the expression of a given piece).

This questionnaire is used when working with a large variety of musicians in the studies presented in Chapter 5 and Chapter 6 to provide objective information as a comparative assessment of the participants' varied experiences. As with other self-assessed questionnaires, it is possible that participants judged their abilities too harshly or too leniently. This is discussed with respect to the specific studies in the aforementioned chapters. However, the questionnaire currently provides the most robust and validated metrics for music sophistication and is the preferred method for this evaluation across the various research fields covered in this thesis.

Measuring Imagery Ability & Use

Mental imagery has been found to be used very frequently used by musicians in daily life. Experience Sampling Methods (ESM) have been used to explore how musical imagery occurs in every day life; this method utilises participants' self reflections to report on their experience at certain times of day when signaled by the researchers, and is a common methodology in communications research (Kubey et al., 1996). Sloboda et al. first provide application of the ESM in inquiry about an individual's music-related activities during the day, including the type of music in question, location, mood, related activity, and control of the music (Sloboda et al., 2001). This method is further applied to musical imagery in the context of imagined music: in a study by Bailes, it was found that the average music student spent 44% of their time listening to music, 32% imagining music, and 3% both listening and imagining together (Bailes, 2006). Within the time spent imagining music, the two most common reasons for doing so involved having heard the piece recently and preparation of the piece for performance. Because of the prevalence of auditory feedback in musical performance, auditory imagery ability is of particular interest to imagery researchers:

Auditory Imagery with the BAIS: Self-reported auditory imagery ability has been effective at predicting neural and behavioural outcomes; the Bucknell Auditory Imagery Scale (BAIS) was developed to determine auditory imagery based on two sub-scales for vividness (BAIS-V) and control (BAIS-C) of imagery (Halpern, 2015). The questionnaire offers questions for assessing each subscale with three types of sound source: musical, environmental, and spoken voice. For vividness, the participant is instructed to imagine a particular sound; for example, one question asks the participant to consider attending a choir rehearsal and to imagine "the sound of an all-children's choir singing the first verse of a song." The participant then scores their image on a Likert scale from 1 to 7, where 1 means 'no image is present at all,' 4 means 'fairly vivid,' and 7 means the image is 'as vivid as the actual sound.' In the case of control, the participant is given a pair of sounds: the same sound as imagined before in the BAIS-V and a specific change to that sound. For example, after imagining again the children's choir singing the first verse of a song, and "an all-adults' choir now sings the second verse of the song." The participant would indicate their control over this change to the image with a separate 7-point scale, where 1 means 'no image present at all,' 4 means 'could change the image, but with effort,' and 7 means 'extremely easy to change the image.'

Self-reported imagery ability on the BAIS has been found to correlate significantly with behavioural aspects of musicality. Higher BAIS-V scores correlate with better pitch imitation ability (Pfordresher and Halpern, 2013), recall and recognition of transposed, reversed, or serially shifted melodies (Greenspon et al., 2018), and the occurrence of "earworms," or involuntary musical imagery (INMI) (Floridou et al., 2014). BAIS-C scores also correlated positively with better prediction of melodic line movement (Gelding et al., 2015) and anticipation of tempo changes (Halpern, 2015). BAIS scores have also been associated with some aspects of neural processing of music, particularly with asymmetrical activity in the right side of the brain (Zatorre and Halpern, 2005). Neural activation was greater in participants with higher BAIS-V scores in the right anterior superior temporal gyrus (secondary auditory cortex) and right dorsolateral prefrontal cortex (involved in working memory) during encoding of imagined melodies (Herholz et al., 2012). Higher overall BAIS was positively correlated with activity in the right secondary auditory cortex and right intraparietal sulcus in mental reversal of a melody (Zatorre et al., 2010). Furthermore, higher BAIS-V scores correlated positively with grey matter volume in the left inferior parietal lobule and left supplementary motor area (Lima et al., 2015). These areas have also been implicated in functional studies of musical imagery (Foster et al., 2013; Halpern and Zatorre, 1999; Zatorre et al., 2010). It has therefore been suggested that higher BAIS scores allow an individual to perform musical tasks with accuracy, access more neural attention to memory and auditory processes, and ultimately exercise more control over mental images of a sound.

Kinaesthetic & Visual Imagery with the MIQ-3: The BAIS is perhaps the most researched questionnaire in imagery self-assessment, but focuses on only a single imagery modality. Other similar questionnaires have been developed for multimodal imagery use. The Multi-Modal Imagery Association Model (MMIA) focuses on tactile and auditory associations to describe more complex sensorimotor imagery (Pfordresher et al., 2015). As well, the Movement Imagery Questionnaire-3 (MIQ-3) assesses both internal and external visual imagery and kinaesthetic imagery (Cumming and Williams, 2012).

In this thesis, I use the MIQ-3 as a companion to the BAIS to address visual and kinaesthetic imagery, which are most commonly discussed and studied in previous research. It is of course likely that gustatory and olfactory imagery play a role in music performance. This is seen in an image involving potpourri smelling by a participant that will be described in Chapter 6. I however limit the scope of this thesis to auditory, visual, and kinaesthetic imagery as gustatory and olfactory imagery are not well-researched at present and require separate attention outside the possibility of this thesis. This version of the MIQ-3, which combines the Movement Imagery Questionnaire (Hall and Pongrac, 1983) and Movement Imagery Questionnaire-Revised (Hall and Martin, 1997), has also been extensively validated with respect to psychometric properties (Williams et al., 2012) and is commonplace in investigating movement imagery in sports science and dance research (Cumming and Williams, 2012).

Like the BAIS, the MIQ-3 also uses 7-point Likert scales for rating imagery ability, for instance rating visual images between “Very hard to see” and “Very easy to see” and kinaesthetic images between “Very hard to feel” and “Very easy to feel.” Like the BAIS, the questionnaire examines the ability to imagine a changing movement and body orientation, covering vividness and control but within a single scaled score. First, the participant is given a starting position and an action to perform. For instance, the first question of the MIQ-3 asks participants to use the Starting Position “Stand with your feet and legs together and your arms at your sides.” and then an Action to “Raise your right knee as high as possible so that you are starting on your left leg with your right leg flexed (bent) at the knee. Now lower your right leg so you are once again standing on two feet. The action is performed **slowly**.” Then, the participant is given a Mental Task to “Assume the starting position. Attempt to **feel** yourself making the movement just observed without actually doing it. Now rate the ease/difficulty with which you were able to do this mental task.” This question in particular examines kinaesthetic imagery; other questions examine both external and internal visual imagery by directing the participant to either “Attempt to **see** yourself making the movement just observed from an **external perspective**” or “from an **internal perspective**” during the Mental Task.

Considerations & Potential Bias: As described above with reference to the Gold-MSI, it is of course possible that participant bias may be introduced. Participants self-assessments may be too harsh or too lenient, depending on individuals. However, these questionnaires are among the most utilised assessments of imagery. The analysis of imagery ability in this thesis mirrors findings from other studies (see Chapter 5, for example), further adding to the validation of the metrics. It is also important to note that I here adopt the Functional Equivalence theory of imagery encoding (please refer to Section 3.1 in Chapter 3) as, in particular, the BAIS is validated on neural correlates with imagery ability in particular musical tasks. Functional Equivalence is not the only method proposed for imagery encoding; the results of this thesis, like the questionnaires themselves, depend on this particular grounding and this is worthwhile to note. The questionnaires operate under directed tasks, which assume that deliberate imagery production scored in their use is equatable with contextual and unprompted imagery use, in this case in singing. Both have been validated in-context and correlated to particular behavioural skills, but this is important to note. With consideration of these potential biases, these remain among the most studied in previous imagery research, especially the BAIS in musical interaction, and provide robust context to the findings of this thesis.

4.4 Apparatus and Materials

In addition to existing materials, this thesis involved an extensive amount of protoyping and the design of a dedicated system for vocal EMG amplification, the VoxEMG, as well as two wearable probes, the Singing Knit and the VoxBox, for sEMG interaction in both an autoethnographic and first-person research with other vocalists, respectively. These research artifacts were created for this thesis and the design process is outlined in Chapter 7. The implementation and use of the probes are fully discussed in their respective chapters: the Singing Knit in Chapter 7, the VoxEMG in Chapter 8, and the VoxBox in Chapter 9.

A complete list of other technical equipment used to facilitate each study can be found in Appendix D.

4.4.1 Ethical Approval

The studies conducted in this thesis received ethical approval from the Queen Mary Ethics of Research Committee. The result of ethics proposals determined that the proposed work did not present any ethical concerns and deemed very low-risk studies. The relevant ethics code for each study is as follows:

> **QMREC2125**, *Granted: 26/06/2019*, Effect of musical training and musical learning experience in singers' performance in non-ideal listening conditions (ref: Chapter 5)
>
> **QMERC2505a**, *Granted: 19/10/2020*, The roles of metaphor in teaching kinetic aspects of sustainable and desirable technique in a one-to-one voice lesson (ref: Chapter 6)
>
> **QMERC20.592**, *Granted: 07/12/2021*, Interacting with Voice Physiology Through Sonification of Laryngeal Movement Measured with Surface Electromyography (sEMG) (ref: Chapter 9)

It is worthwhile to note that I did not receive external ethical approval for the autoethnographic work I carried out during my own design and use of the VoxEMG system and Singing Knit (Chapter 8), as it is not required for self-study. I am however very passionate about ethical concerns in autoethnographic contexts to ensure that we adhere to a standard of care for ourselves as both researcher and participant. For the autoethnographic work I carried out, I documented aspects of my experience as a participant, for instance addressing issues of discomfort and skin irritation when working with sEMG sensors. I agreed with myself to adhere to a code of conduct, derived from ethical approval granted for the other studies in this thesis: *I am free to stop the research at any time, without repercussion, for any reason. If anything does not feel natural or I feel unusual discomfort in my voice, I will stop immediately. In exploring I will ensure not do anything out of my comfort zone that can be harmful to my voice, including but not limited to straining or pushing, screaming, and holding my head or neck in uncomfortable or locked positions.* This information later formed part of the Participant Information Sheet for QMERC20.592, given to other vocalists utilising the VoxEMG system. By acknowledging my own interaction and potential risks, I was able to ensure my own care in autoethnographic work and also carry this forward to working with other participants.

Chapter 5

Vocalists' Use of Auditory Imagery

Using Imagery to Adapt to Altered Auditory Feedback

Based on the existing literature surrounding imagery cognition, the main focus of the beginning of the study was on auditory imagery and feedback as the driving mechanism behind singers' crafts. This also arose from my own experience and what I believed myself to be doing while performing: while singing, we listen to the resultant sound and are able to determine when something is right or wrong based largely if not completely on that feedback. In order to explore the relationship between singer and voice and how singers understand their sound, I thus focused on auditory feedback.

The initial study for this PhD aimed to explore audiation and adaptability in vocal performance. The use of altered auditory feedback (AAF) conditions through tasks involving forced audiation in this study provides information about how well musicians are able to perform a specific performance task. We explored the relationship between auditory imagery ability and maintenance of tonal and temporal accuracy when singing and audiating with AAF, including upwards pitch shifts and delayed audio feedback (DAF). These tasks required the singers to rely on training and their "imagery toolboxes" while both singing and audiating. Results also determined whether there is any significant difference between vocalists and instrumentalists, as well as formally trained and non-trained musicians, in terms of musical imagery ability, and whether participants' auditory imagery skills determined by the BAIS correlate to measures of performance accuracy. Musicians sang and audiated through a self-selected piece under conditions of pitch shifts and DAF and with speech distraction. The results indicate that musicians with higher self-indicated auditory imagery scores on the Bucknell Auditory Imagery Scale (BAIS) produced a tonal reference that was less disrupted by pitch shifts and speech distraction than musicians with lower BAIS scores. On the other hand, there is no effect of BAIS on temporal deviation when singing with DAF. Musical imagery ability as measured by the BAIS was not related to musical training but did correlate significantly with years of performance experience. The significant effect of auditory imagery ability on tonal reference deviation remained even after partialling out the effect of performance experience. The results indicate that auditory imagery ability plays a key a role in maintaining an internal tonal centre during singing but has, at most, a weak effect on temporal consistency. This work outlines directions for future studies to explore the multi-faceted role of auditory imagery ability in singers' accuracy and expression.

Portions of this chapter have been submitted for publication in:

Courtney N. Reed, Marcus T. Pearce, and Andrew P. McPherson. Auditory imagery ability and singing accuracy with altered auditory feedback. *Under review in Musicae Scientiae.*

5.1 Method

This study examined how imagery is beneficial in tasks requiring audiation and adaptation to AAF. The hypothesis of the study was that, with greater auditory imagery ability, vocalists would be more able to adapt to hearing their voice in pitch-shifted or delayed feedback and continue singing accurately by relying on internal references. We use an approach focusing on auditory imagery specifically with respect to AAF, centring the exploration within the extensive work of Halpern and Pfordresher in auditory imagery in using and validating the BAIS (see Chapter 4, Section 4.3.6) in specifically singing tasks specifically (Greenspon et al., 2018; Halpern, 2015; Pfordresher and Halpern, 2013). Similar singing studies examining auditory imagery ability in performance have used a variety of methods to prompt imagery use, as discussed at length in Chapter 3, Section 3.1.3. This includes sensorimotor synchronisation tasks (Colley et al., 2018b) mental time and pitch transformations (Foster et al., 2013), pitch imitation tasks (Greenspon et al., 2018), and AAF tasks such as singing with DAF (Pfordresher and Palmer, 2002). Our method, based on the methods employed by Pfordresher and Palmer (2002), Pruitt et al. (2019), and Herholz et al. (2012), specifically focuses on realistic, performance-based tasks where vocalists sing familiar songs with strong, existing mental imagery (Herholz et al., 2012; Zagacki et al., 1992), rather than completing an isolated exercise like imitation or synchronisation. Using these familiar pieces, we investigate as well how audiation is impacted by auditory imagery ability.

5.1.1 Participants

A total of 16 participants (7 male, 9 female) with a mean age of 28 years (range: 22 – 37 years) were recruited for the study (Table A.1). Participants were recruited through an open call online and via emailing lists; in order to participate, participants were required to "be able to sing confidently and do so with reasonable pitch accuracy" in an unaccompanied setting. A short questionnaire provided at sign-up collected information regarding the participants' demographic and experience data. Participants were asked to provide their primary instruments and whether or not they had undertaken any formal musical training (outside of compulsory schooling) on that instrument. Nine participants identified themselves as being primarily vocalists, six with formal voice training and three without. The remaining seven indicated a primary instrument other than voice (two pianists, two guitarists, one flautist, one dhol player, and one electronic digital instrumentalist), six with formal training on that instrument and one without. There was some overlap in backgrounds, with some participants being both vocalists and having substantial performance experience ($>$ 5 years) on another instrument (Table A.1).

To ensure a wide range of principal instruments and musical abilities, participants were invited from a range of different musical groups in the Greater London area, including theatre and choral groups at QMUL, music science research centres at QMUL and Goldsmiths University of London, and professional and semi-professional choruses. Participants represented a variety of nationalities, were all fluent English speakers, and lived in the UK at the time of the study. Participants provided written informed consent for the collection, inclusion, and publication of their data, including anonymised questionnaire results and audio-video recordings. Participants were paid for their time following the completion of the study.

5.1.2 Materials

After sign-up and initial screening, participants were invited to bring "a solo piece/excerpt that you enjoy singing and can perform accurately without accompaniment." The piece could be in any style

Figure 5.1: The recording and monitoring setup used for completing the tasks and receiving visual cues for audiation.

(a) Indication to sing aloud.

(b) Indication to audiate.

Figure 5.2: Visual stimuli displayed on the monitor during the Toggle and Toggled & Voice Distraction tasks.

and be 2-3 minutes in length. In order to ensure the strongest imagery association possible for the duration of the study, it was important that participants sang a piece they knew very well and felt confident to perform unaccompanied. It is also known that imagery has strong emotional associations (Zagacki et al., 1992), so we hypothesised that bringing a preferred song would further add to the strength of the present image, while also creating an enjoyable environment for the participants. This also helped to avoid the significant time costs or potential anxiety associated with having all participants learn or sight-sing an unfamiliar piece specifically for the study.

Song choices were agreed upon before the study took place. The reference key and tempo for each runthrough of the piece were agreed on at the time of study (Table A.2). This was done to ensure participant's comfort in repeating the piece in a suitable range, as well as starting consistency between the iterations.

The participants completed the Gold-MSI (Müllensiefen et al., 2013, 2014) to assess general musical background and demographics and the BAIS questionnaire to assess auditory imagery.

5.1.3 Apparatus

The study was completed within an isolated, acoustically- treated recording room at QMUL (Figure 5.1). Directions and auditory stimuli were given with the use of a set of RCF ART 412-A speakers in the room or with a Beyerdynamic DT100 closed, circumaural studio headset, depending on the given task. The headphones offer approximately 20 dBA of noise attenuation and were chosen to provide isolation for AAF stimuli from the unaltered voice. Participants were recorded with an AKG C414B-XLII condenser microphone via a Yamaha MG16XU mixing console and Universal Audio 4-710d Tone-Blending mic preamp. The recording was made into Logic Pro X running on a Mac (Mac OS 10.14.1) in a neighbouring recording and mixing studio. The latency between the onset of a recorded audio signal and the return via the headphones was measured to be 1.63ms. Visual stimuli created in MAX/MSP (Cycling'74, San Francisco, CA) were presented on a 24" BenQ LCD monitor placed approximately 2' away from the participant.

5.1.4 Procedure

Participants began by singing their chosen piece unaccompanied and without any additional feedback or stimuli.

All participants demonstrated qualitatively confident ability to sing their piece accurately (assessed by myself) and so were able to proceed with the remainder of the study. The remainder of the study involved three tasks done in six AAF conditions.

Tasks.

Three tasks were performed, in the following order:

1. **Normal**. Sing the piece as written, from start to finish.
2. **Toggled**. Perform from start to finish, alternating between singing aloud and audiating as visually instructed.
3. **Toggled & Voice Distraction**. Perform as in the Toggled task, with an additional stimulus of external dialogue supplied during the audiated sections.

In the Toggled and Toggled & Voice Distraction tasks, the participants were instructed to either "sing" aloud or "wait" while continuing to audiate, as indicated by the monitor (Figure 5.2). The screen changed from one state to another on a random interval of 5-15 seconds. The toggling within these tasks was designed to force the participants to rely on their musical imagery abilities through audiation. In the Toggled & Voice Distraction task, a BBC Radio podcast ("Fermat's Last Theorem," from In Our Time, Melvin Bragg, BBC Radio 4, 25 October 2012) was played while the participants were instructed to audiate and was switched off as they were instructed by the monitor to resume singing aloud. This voiced dialogue distraction was included to observe how well participants were able to focus on their audiation while a conversation happened in the room, as audible speech is found consistently to be more disruptive to task performance, including to audiation in reading, than other sound distractions (Vasilev et al., 2018; Venetjoki et al., 2006).

Auditory Feedback Conditions.

In each task, participants sang their piece in six different feedback conditions (2 control, 2 DAF, 2 upwards pitch shifts), for 18 total performances of the piece:

- **Normal Feedback (NF)** (control, room feedback)
- **Headphone Feedback (HF)** (control, setup latency)
- **200 ms Delay**
- **600 ms Delay**
- **+ 1/4 Tone Pitch Shift**
- **+ Whole Tone Pitch Shift**

DAF timing was determined based on related literature from previous musical imagery study, as well as Speech, Language, and Hearing research (SLHR). 200 ms has been found to be the most disruptive to speech production (Atkinson, 1953; Black, 1951; Fairbanks, 1955; Sasisekaran, 2012;

Zimmermann et al., 1988),functionally interrupting the action-effect path in sensorimotor coordination (Howell, 2004; Howell et al., 1983). The 600 ms delay was chosen as it provides a qualitatively distinct delay; longer delays have been found to inhibit monitoring of vocal parameters and internal beat subdivisions (Bartlette et al., 2006; Finney and Warren, 2002; Lee, 1950; Pfordresher and Palmer, 2002; Zarate and Zatorre, 2008), and potentially to disrupt longer phrases (Howell et al., 1983).

The pitch shifts used were both upward pitch shifts; this was decided based on the fact that most singers drift downward in pitch over time (Howard, 2007; Mauch et al., 2014), with the hypothesis that upwards pitch shifting would provide a more unnatural stimuli and not confound results by exacerbating the typical detuning during an a capella performance. The quarter tone pitch shift was intended to provide a sense of chorusing, being closely situated to the sung pitch. It was anticipated that the singers might try to adjust over time, and drift upwards with what they heard to remain "in-tune" (Howard, 2007). The whole tone shift was meant to provide a distinctly out-of-key sensation, rather than a slight detuning, forcing the singers to ignore the stimuli outright.

All AAF stimuli were played back via the headphones with added gain in order to mask the unaltered voice (Malloy et al., 2022). The gain was set to a level that was "comfortable, but you should not be able to hear your voice outside of the headphones." The stimuli were ultimately presented at approximately 80-82 dB SPL; this does not limit any bone conduction of the unaltered voice, which should be noted, but minimises the presence of veridical feedback when AAF was presented.

DAF was introduced with Logic Pro X's built-in Sample Delay plug-in. Pitch shifting was added with in the built-in Pitch Shifter plug-in, set for 0.0 ms delay and with Latency Comp enabled. The I/O buffer size in Logic was additionally set to the smallest size (32 samples) for further latency reduction. Participants listened to direct monitoring on the track using this plug-in to sing along to with the AAF. Using an oscilloscope, the round trip latency (RTL) between input through the microphone and output through the headphones was measured to be approximately 7 ms of additional latency, and so would not provide any noticeable latency beyond the intended DAF. The raw audio signal from the performance was recorded on a separate track without the plug-in and bounced in place to ensure there was no latency introduced in the audio files used for analysis.

Experiment.

The NF task was always performed first for each task as a control. This also allowed participants to become familiar with the task before addition of extra AAF demands. Following this, the remaining conditions were randomised.[1]

The ideal performance for this study would involve the performer maintaining a consistent tonal centre and tempo throughout the piece, effectively ignoring AAF and singing as they had done under the HF conditions. In the pitch shifted conditions, a performer would be able to ignore the AAF and continue singing with their internal tonal centre, maintaining the key throughout the piece and not trying to adjust to what they hear. In delay conditions, the participant would not react to the AAF and maintain a continuous and consistent pulse. In the Toggled and Toggled & Voice Distraction tasks, the participant should switch from audiating to singing in the same tempo and key as where they left off, without late or early entries.

Participants were therefore instructed with the following guidelines: "Sing the piece as you did during the first run-through, ignoring any auditory feedback you hear in the headphones. Sing the piece as you would normally, keeping in key and staying in time. If you make a mistake, keep going;

[1]Some participants did not complete all task-condition combinations due to external constraints. See Table A.3 for details.

if you find you are off-key or are changing tempo, stay consistent and continue to the end of the piece with your new key or new tempo. It is important to not stop and to continue on singing as well as you can." Participants were reminded of this between performances to ensure that their goal was to maintain consistent tonal and temporal centre in their performance.

For each condition, the participants were allowed to talk into the mic to hear how the AAF would sound before they began singing. This was done to give an idea of what to expect and to reduce the effects of surprise. Toggled tasks were described to be performed "as if someone has just muted you for a short time." Participants were reminded to do their best to maintain their pitch and tempo centres while audiating and to come back in when instructed to sing aloud again. The screen changes for Toggled tasks were shown to the participants beforehand, again to reduce the effects of surprise when completing the sung performances. Once participants were comfortable with the performance's conditions, the sung performance began.

At the start of each task, the starting pitch and the melody of the first two bars of the piece were played for consistent tonal reference. This reference was generated in Logic Pro X and played on the default Steinway Grand Piano software instrument patch. The reference tempo was provided as a count-in with Logic's digital metronome. The full study took one hour for each participant, including completion of questionnaires and approximately 45 minutes of singing.

5.2 Analyses

We did not expect the participants would be able to ignore AAF perfectly, so participant "accuracy" was measured in terms of how little deviation occurred through the piece. In the cases where participants were unable to maintain these tonal and temporal centres, the accuracy measures chosen for the analysis address the extent of the deviation. If the tonal or temporal centre was lost, especially after audiating for a passage, we expected participants to continue consistently and finish the performance with the new tonal or temporal centre. The hypothesis was that, with a stronger internal representation of sound driven by auditory imagery, participants should be able to better ignore both pitch shifts and DAF and maintain tonal and temporal consistency.

Therefore, we used three calculations to measure accuracy in each of the 18 task-condition combinations: one for tonal deviation and two for temporal deviation. It is important to note that these accuracy measures focus objectively on quantifiable measures of the performance in timing and tonality and that the "best" performances according to these metrics were not always the most musically expressive or given by the participants who were subjectively more talented singers.

The recorded audio was analysed within Sonic Visualiser (Cannam et al., 2010) to determine the pitching and timing of each performance.

5.2.1 Pitch

A pitch track and notes were extracted from the recording using the pYIN algorithm implemented in Tony (Mauch and Dixon, 2014; Mauch et al., 2015, 2014) and were checked and corrected by hand if necessary. A reference pitch track created from the score was then aligned to provide a reference[2] for the expected notes in each performance. This reference track, along with fundamental frequency $f0$ output by Tony were further converted to musical pitch p (Dai et al., 2015). Here, p is measured in semitones and is equivalent to the note's MIDI pitch number, where A4 (440Hz) is 69, using

[2] In the Toggled and Toggled & Voice Distraction tasks, the pitches omitted during the audiated sections were excluded from this reference pitch track.

Equation 5.1.

$$p = 69 + 12 \log_2 \frac{f_0}{440\text{Hz}} \tag{5.1}$$

We aimed to determine how well participants were able to sing correct intervals, maintain intonation, and limit tonal drift from a key centre.

We used Kennedy and Kennedy (1980)'s definition of intonation as "the act of singing or playing in tune," which requires the existence of a tonal reference. Given that participants were singing unaccompanied, the reference is internal (Mauch et al., 2014) and cannot be measured directly; therefore, estimations of intonation given this internal reference were calculated as proposed by Dai et al. (2015), using Tonal Reference Deviation (TRD), also measured in semitones.

Tonal Reference Deviation.

To calculate TRD, an internal tonal reference curve must be estimated. Assuming the reference changes over time in unaccompanied singing, this estimate approximates the local, internal tuning reference, with respect to the neighboring pitches. TRD captures this trajectory of the reference pitch over time with score normalisation by adjusting the pitches that are actually sung to the expected pitches, as indicated in the score (Dai et al., 2015). Because the internal reference is based on pitch memory (Mauch et al., 2014), a sliding window can be used to estimate the magnitude of the pitch trajectory over time. The window represents deviation along the natural drift, which occurs during a capella singing, by judging the each note against the notes directly before and after. Pitch error is calculated as the difference between the tonal estimates and the tonal reference curve. The TRD is therefore the fluctuation of this reference pitch and is calculated as the standard deviation of the reference curve. Please refer to Dai et al. (2015) for further detail.

First, the sung pitches from the performance, p_i, are adjusted by removing the anticipated pitches, s_i, given by the score, $t_i^* = p_i - s_i$, and then adjusted further by subtracting the mean $t_i = t_i^* - \bar{t}^*$. The smoothed trajectory of these tonal estimates, t_i, is used to find the reference pitch, c_i , from the weighted mean of the tonal estimates within a triangular window around the note (Equation 5.2).

$$c_i = \sum_{k=-n}^{n} w_k t_{i+k} \tag{5.2}$$

In this equation, $\sum_{k=-n}^{n} w_k = 1$ and the triangular window $W^{T,N} = \left\{ w_k^{T,N} \right\}$ is used. Here, N refers to the length of the window and k is the index of all notes from $-n$ to n. We used a window size of $N = 5$, which gives a window consisting of the two notes on either side of the i_{th} note and gives more weight to the notes nearer while giving no weight to the i_{th} note itself, proportional to 1, 2, 0, 2, 1, as in Equation 5.3.

$$w_k^{T,N} = \begin{cases} \frac{2N+2-4|k|}{N^2-1} & 1 \leq |k| \leq \frac{N-1}{2} \\ 0 & \text{otherwise} \end{cases} \tag{5.3}$$

Pitch error is therefore calculated as the difference between the tonal estimates, t^i,and the tonal reference curve, c_i. The TRD is the fluctuation of this reference pitch and is also calculated as the standard deviation of c_i.

We therefore expect participants with stronger auditory imagery will have lower TRD, which indicates less drift or inconsistency in the tonal centre. As this measure captures the reference pitch, small errors in tuning will not be penalised as much as wholly inaccurate intervals. In the same way, if a participant abruptly loses their key, the measure penalises the initial error as compared to the score and the deviation from the original tonal centre, but this error is subsequently smoothed by the overall trajectory of the deviation. By use of a sliding window and accounting for the notes preceding and following, the TRD will be lower if the participant is singing consistently in a new key than if the pitches are moving around between tonal centres.

5.2.2 Tempo

It is found to be generally impossible for people to maintain a consistent timing, even if they try (Repp, 2002), but the degree to which drift occurs varies between individuals. Previous research identifies the importance of a central clock variance in drifts found in isochronic tapping with different limbs (Collier and Ogden, 2004). Collier and Ogden (2004) suggest that musical experience or inclination, among other unknown factors, influences this clocking and therefore the natural tendency to drift. There is less understanding about how this drift occurs over time, compared to tonal drift. Humans use contextual information for determining timing, which suggests temporal variation should be examined successively across a series of onsets (Madison, 2004). In related work studying the effect of DAF on timing in musical performance, the coefficient of variation (CV) is used as a measure of timing variability (Pfordresher and Palmer, 2002). CV is calculated as the standard deviation of the inter-onset interval (IOI) divided by the mean IOI. We include this measure to indicate drift while singing. In Toggled and Toggled & Voice Distraction tasks, we are unable to determine the internal timing while participants audiate. In order to examine timing variation while audiating, we calculate the absolute average number of missed beats (MBs), adjusted for the length of the audiated section, as a representation of the singer's internal temporal reference.

Coefficient of Variation: Error and Drift.

At the start of analysis, note events and beat onsets were annotated by hand in each audio file within Sonic Visualiser, based on the placement given by the score. The distance between each note and its predecessor (omitting the first beat of an performance) is calculated as beats per minute (BPM) and converted to milliseconds, allowing the distance between beats to be expressed as an IOI. For Toggled and Toggled & Voice Distraction tasks, the beat before the change in the visual display is marked and the audiated sections are assumed to continue at the same tempo from the last beat previously vocalised. This gives a clear image of the timing between each beat through each performance.

Accuracy in this case is again defined as the ability to maintain consistency when presented with AAF. It is normal for musicians to have inconsistencies in timing at a local level, especially when performing solo; without the reference of other players, this drift would not be noticeable. Because CV is calculated as the standard deviation of the IOI divided by the mean IOI, it acts as a measure of the dispersion around the mean tempo and is used to depict the general temporal deviation. Similar to TRD, in cases where the participant would change to a new tempo, for instance after an audiated section, the deviation from the original tempo would be penalised but not compounded if the participant remained consistent to the new temporal centre for the remainder of the piece.

Missed Beats: Drift During Audiation.

Timing accuracy in Toggled and Toggled & Voice Distraction tasks also uses MBs to examine audiated sections. As discussed previously, the tempo through an audiated section is assumed to be the same BPM as the last beat sung before the start of the audiation. The number and length of audiated sections varied depending on the frequency of the random toggle and the duration of the piece in each performance. Therefore, MBs were averaged according to the length and frequency of the audiated sections.

If auditory imagery and temporal accuracy are in fact related, a singer with greater imagery ability would likely have more accurate timing and consistent beat keeping. They would be expected to be able to keep more consistent tempo and miss fewer beats on average through these audiated sections.

5.3 Results

The Gold-MSI indicated that all participants were engaged in active performance for at least one year prior to the study (M = 10.9 years) and all had some training in musical theory, including self-study (M = 6.6 years). Mean ratings on the BAIS-V ranged from 4.14 to 6.57 (M = 5.19, SD = 0.7). Ratings on the BAIS-C ranged from 3.79 to 6.43 with (M = 5.23, SD = 0.8).

5.3.1 Group BAIS Results

Preliminary analysis was conducted to establish any group differences in terms of musical imagery ability as measured by BAIS-V and BAIS-C scores. The groups were divided based on indications made at the time of sign up and again on the Gold-MSI of (1) being primarily a vocalist or an instrumentalist and (2) having been formally trained on the principal instrument or not.

Vocalists vs. Instrumentalists.

A two-sample t-test was performed to determine whether the groups differed significantly in either BAIS subscale. There was no significant difference in the BAIS-V scores of the vocalists (M = 5.18, SD= 0.81) and the instrumentalists (M = 5.2, SD = 0.56), t(14) = − 0.06, p = .95. This was also true of BAIS-C scores between the vocalists (M = 5.23, SD = 0.99) and the instrumentalists (M = 5.22, SD = 0.57), t(14) = 0.01, p = .99).

Formal Training.

For BAIS-V score, there was no significant difference between the participants with formal training (M = 5.26, SD = 0.7) and those without (M = 5.0, SD = 0.69), t(14) = 0.63, p = .54). This was also the case for BAIS-C, t(4) = 1.38, p = .24, although variance between the participants with formal training (M = 5.42, SD = 0.64) and those without (M = 4.64, SD = 1.07) was unequal.

Given the lack of any statistical differences in BAIS scores between vocalists and instrumentalists, or participants with and without formal training, no further distinction was made between participants on these grounds. All participants were examined as a single cohort of confident singers regularly performing vocals in some capacity.

Aggregating BAIS Scores.

Additionally, participants' BAIS scores from the two subscales had a strong positive correlation (r = .76, R^2 = .57, p < .001, Figure 5.3). This is consistent with previous research using the questionnaire (Halpern, 2015; Pfordresher and Halpern, 2013) and indicates that individuals with ability to produce a more vivid auditory image may also have greater control over that image. Given this strong relationship between the BAIS subscales, BAIS-V and BAIS-C scores were averaged for

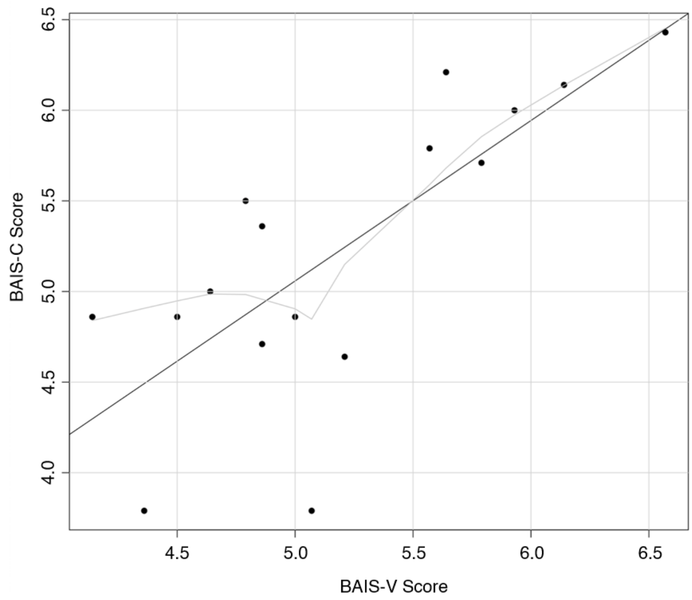


Figure 5.3: Positive correlation between participant scores on the BAIS subscales linear (black) and loess (grey) regression

each participant and this aggregate BAIS score (referred to simply as "BAIS score") is used in all further analyses. Participants were therefore divided into two groups by median split (median BAIS score = 5.02) as being either high or low BAIS scorers (Details on participant demographics by BAIS group can be found in Table A.4).

5.3.2 Potential Covariates in Music Selection

Before examining anything else, we first determined whether the participants' self-selected pieces might introduce covariates into the analyses.

Music Complexity.

Several elements of complexity were used to compare the different participant-chosen pieces; these measures were calculated using the MIDI toolbox functions for MATLAB (Eerola and Toiviainen, 2004) and included the melodic range in semitones (ambitus), melodic complexity (complebm, derived from the expectancy-based model of melodic complexity by Eerola and North (2000) and Schaffrath (1995)), durational variability of note events (nPVI, by Grabe and Low (1995) and Patel and Daniele (2002)), note density as number of notes per beat (notedensity), and tonal stability of notes in a melody (tonality, by Krumhansl (1990)).

We found that correlations between these features in each song and the respective participant's BAIS score, years of study, and years of performance experience were all found to be weak ($r < \pm 0.39$). Ambitus and years of study ($r = -0.48$), note density and melodic complexity ($r = .66$), BAIS score and years of performance ($r = .6$), and durational variability of note events with note density ($r = -.52$), melodic complexity ($r = -0.44$), and BAIS Score ($r = 0.43$) were found to be moderately correlated (Figure A.1). However, these potential correlations were all found to be non-significant

($p > .05$). As the complexity of the pieces in these dimensions did not correlate with any of our participant measures, we assume that the subject-selected music does introduce confounding factors in our primary analyses.

Interaction with Delays.

Additionally, potential interactions with DAF were examined; given previous research on timing accuracy with DAF (Pfordresher and Palmer, 2002), it is possible that the event timing of some of the chosen pieces might result in DAF occurring at a binary subdivision of the beat, thus creating a less distracting delay for some participants For instance, if the IOI between the notable beats of the measure was 400 ms, the 200 ms delay condition would result in DAF occurring at a subdivision of the beat and could potentially benefit a participant's accuracy in timekeeping. In order to determine whether this was an applicable factor in any of the participant-selected pieces, the IOI between notable beats in each piece was determined using the reference tempo in BPM (provided at the start of each performance) and the meter of the piece. The IOI of notable beats, as well as the measure length, in milliseconds is presented in Table A.2.

In both cases, the length of the beat and the measure for all pieces would not provide reasonable binary subdivisions with the two DAF conditions. For this reason, we assume interactions with delays are not likely to be a confounding factor in comparison of participant experience and performance outcomes in the DAF conditions.

5.3.3 Primary Analysis: Effects of Auditory Imagery on Accuracy

Having addressed potential covariates, we began the primary analysis to determine whether there were any effects of auditory imagery on the examined accuracy measures.

As all AAF conditions were delivered via headphones, we first determined whether using HF had significant impact on accuracy. We compared the NF performances with the HF performances in a Welch Two-Sample t-test. There was no statistical difference between these performances for any of the accuracy measures. We thus determine there is likely no confounding factor in delivering AAF via headphones. The HF performances are used as a control for the other AAF performances.

We conducted two analyses using different baseline performances. First, participants' AAF performances were examined against their own individual control performance scores in each task. This was conducted to determine the effect of BAIS on individuals' ability to perform consistently between AAF and control conditions. A subsidiary motivation for comparing within-individual accuracy was to account for individual differences in control performances and with respect to the participant-chosen pieces. In each task, a participant's accuracy scores for in their AAF performances were normalised with respect to their score for their control performance. The resulting adjusted score thus functions as a measure of how much better or worse the participant was in a certain AAF condition compared to their initial performance: an adjusted score of 0 indicates the participant achieved the same error as their control performance (consistent performance), a positive score more error, and a negative score less error.

Second, a group-adjusted analysis was performed. This analysis was conducted to account for some participants having more error than others in their control performances. The group adjustment was achieved by normalising AAF performance scores to the average control score of all the participants in each task (Table 5.1). The adjusted score thus determines how much better or worse the participant did in an AAF performance than the average control performance. An adjusted score of 0 means the performance was as accurate as the average, a negative score less error than average, and a positive score more error than average. There were no significant correlations between participants' BAIS scores and their accuracy score for any measures in control HF performances. We therefore assume the average control performance for the group is not affected by BAIS score.

Task	**Group Accuracy**		
	TRD	**CV**	**MBs**
Normal	0.47	9.74	N/A
Toggled	0.46	9.09	0.82
Toggled & Voice Distraction	0.43	8.99	1.10

Table 5.1: Averaged scores in each measure of accuracy across the group.

Individual-Adjusted Tonal Deviation.

We conducted a 2x3x4 mixed analysis of variance (ANOVA) with *task* (3 levels) and *condition* being the repeated-measures (within-subjects) independent variables, and *group* the between-subjects independent variable. TRD scores were found to be non-normally distributed and were transformed with Ordered Quantile Normalising Transformation (ORQ) to meet parametric assumptions (Peterson and Cavanaugh, 2019).

There was a significant two-way interaction found between BAIS group and task, F(2,144) = 3.304, p = .040, and between BAIS group and condition, F(3,144) = 3.628, p = .015. There was no significant three-way interaction found between the variables.

A planned simple two-way fit for each BAIS group showed that the effect of condition on TRD was significant for the low BAIS group, F(3,144) = 3.70, p = .013. Bonferroni-adjusted pairwise comparisons show this was a result of the Whole Tone Pitch Shift condition's significant difference between the two groups, t(162) = −2.97, p < .001. Here, low BAIS participants had higher TRD compared to their control performances than high BAIS participants (Figure 5.4). The effect of the Toggled & Voice Distraction task differed significantly between the groups, t(160) = 3.69, p < .001, with low BAIS participants again having higher TRD compared to their control performances (Figure 5.5).

Full-factorial results for the individual-adjusted TRD scores can be found in Table A.5.

Individual-Adjusted Temporal Deviation.

Individual-adjusted CV scores were found to be normally distributed and so were not further transformed. A similar 2x3x4 repeated measures ANOVA was performed.

There were significant main effects of condition (F(3,144) = 7.321, p < .001) and BAIS group, F(1,144) = 7.323, p = .008, but no significant interactions between the independent variables.

Planned two-way analyses between each BAIS group showed a significant effect of condition on CV for the low BAIS group only, F(3,144) = 8.22, p < .001. Bonferroni-adjusted pairwise comparisons for each condition indicated significant differences between the groups for both the 200 ms Delay, t(160) = 2.85, p = .005, and the 600 ms Delay, t(160) = 2.34, p = .021.

Interestingly, while high BAIS singers tend to perform similarly under the AAF conditions compared to their control performances (CVs cluster around 0, meaning there is little difference in accuracy between control and AAF performances), some low BAIS singers had *lower* CV scores in these performances (expressed by the negative adjusted CV scores, Figure 5.6). Thus, while the high BAIS singers had AAF performances with similar error to their individual baselines, some low BAIS scorers managed to achieve lower timing error in the DAF conditions compared to their

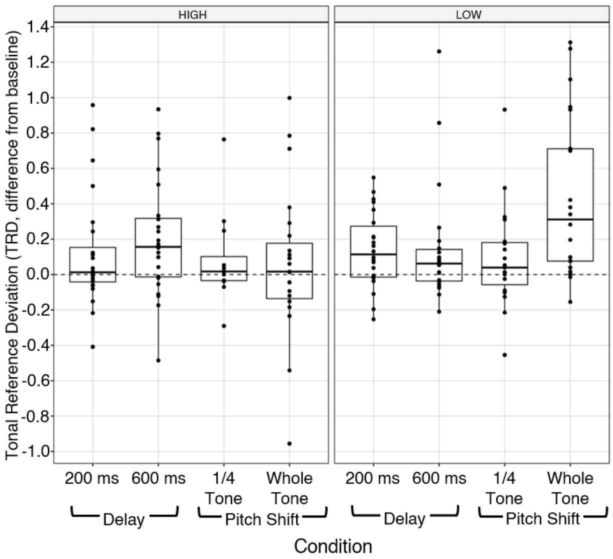


Figure 5.4: Tonal Deviation: Individual-adjusted TRD score (AAF conditions by BAIS group).

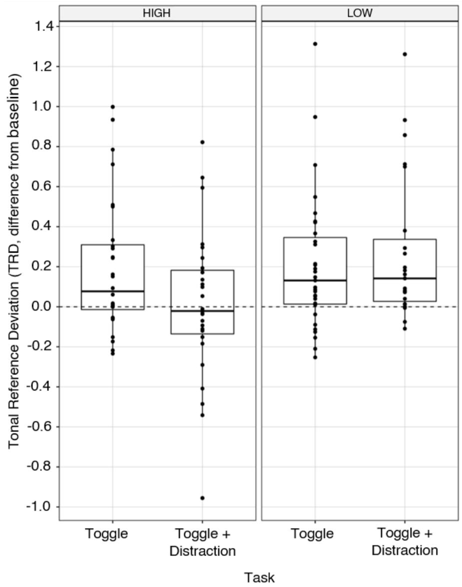


Figure 5.5: Tonal Deviation: Individual-adjusted TRD score (audiated tasks by BAIS group).

control performances. As the independent variables had no significant interactions, we assume that this decrease in error is not an effect of BAIS.

Full-factorial results for the individual-adjusted CV scores can be found in Table A.6.

For the audiated sections in the Toggled and Toggled & Voice Distraction tasks, the repeated-measures ANOVA analysis was performed with individual-adjusted MBs as the dependent variable. No significant main or interaction effects were found (Figure 5.7). In general, the two groups appear to have similar accuracy in the AAF conditions compared to the baseline, where the vast majority of participants drifted less than one beat difference through an audiated section compared to their control performances. Although condition was found to be significant for the other individually-adjusted accuracy scores, the MBs are largely consistent with control performances, with no readily visible differences in either group.

Full-factorial results for the individual-adjusted MBs scores can be found in Table A.7.

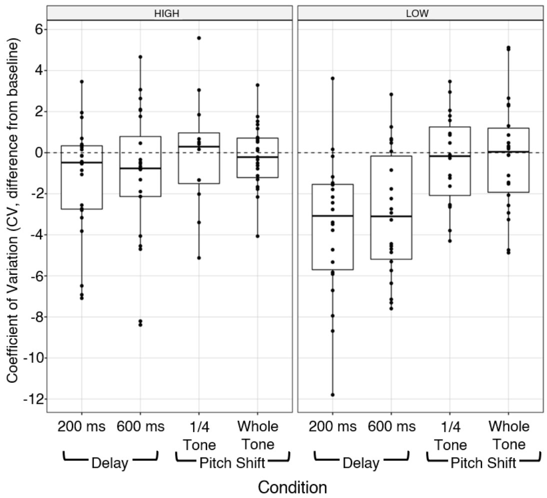


Figure 5.6: Temporal Deviation: Individual-adjusted CV score (AAF conditions by BAIS group).

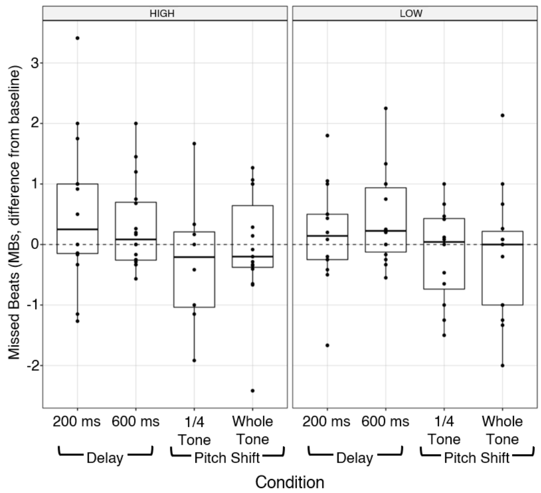


Figure 5.7: Temporal Deviation: Individual-adjusted MBs scores (AAF conditions by BAIS group).

Group-Adjusted Tonal Deviation.

Group-adjusted TRD scores were again transformed with ORQ transformation and used as the dependent variable for a repeated measures ANOVA. There were no significant interactions found between the independent variables.

Planned analyses for interactions by group were carried out to determine any other interactions. A two-way analysis by BAIS group showed a significant effect of condition on TRD for the low BAIS group, $F(3,144) = 2.78$, $p = .043$.

This was further explored through Bonferroni-adjusted pairwise comparisons between the BAIS groups, which revealed a significant difference between the high and low BAIS groups in the Whole Tone Pitch Shift condition only, $t(160) = -2.00$, $p = .047$, with low BAIS scorers having higher TRD scores than the group average control (Figure 5.8).

Full-factorial results for the group-adjusted TRD scores can be found in Table A.8.

Group-Adjusted Temporal Deviation.

Group-adjusted CV scores were normally distributed and so were not further transformed. While there no significant interactions between the effects, there was a significant main effect of condition, $F(3,144) = 12.721$, $p < .0001$.

Further two-way comparisons by BAIS group revealed that condition was a significant effect on CV for both the high ($F(3,144) = 3.36$, $p = .02$) and low BAIS ($F(3,144) = 10.6$, $p < 0.0001$) groups (Figure 5.9). Both groups managed to achieve slightly better than average control performance in the DAF conditions, mirroring the strange result seen in the individual-adjusted analysis of CV scores.

Group-adjusted MBs were also investigated; the 2x3x4 ANOVA revealed no significant interactions between effects on the number of MBs. Two-way interaction by group revealed a significant

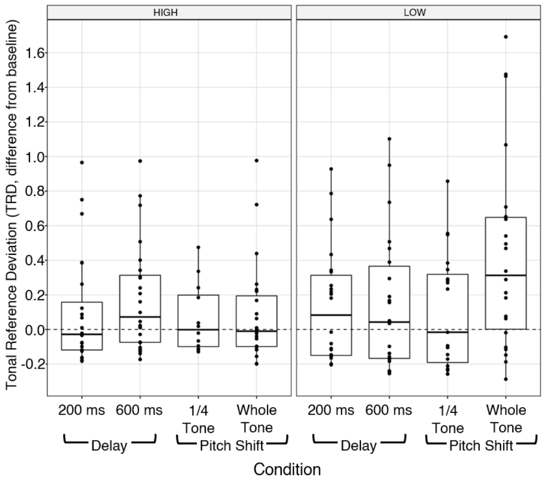


Figure 5.8: Tonal Deviation: Group-adjusted TRD score (AAF conditions by BAIS group).

effect of task for just the low BAIS group, (F(1,93) = 5.57, p = .02) (Figure 5.10).

Full-factorial results are listed in Appendix A for both CV (Table A.9) and MBs (Table A.10).

5.3.4 Musical Experience

A follow-up analysis was performed to determine whether there were any correlations between participants' musical experience (other than formal training as described earlier) and self-assessed imagery ability on the BAIS. Information about participants' years of performance experience on their primary instrument and years of music theory study gathered from the Gold-MSI were regressed against participants' BAIS scores.

Years of theory study did not correlate with BAIS, r = .32, F(1,14) = 1.62, p = .22; however, active performance experience in years was found to correlate positively with BAIS, r = .6, F(1,14) = 7.81, p = .014. Examining the High and Low BAIS groups and these musical experience measures, we find this correlation remains: using a two-sample t-test, we found the groups did not differ significantly in terms of years of music theory, t(13) = -1.33, p = 0.1, but were significantly different in terms of years of performance experience, t(13) = -3.6, p = 0.002.

As BAIS score and performance experience in years (PEY) were observed to correlate in this way, a follow-up full correlation and partial correlation analysis was conducted to determine which of these factors was driving the observed performance accuracy. This was done for the two significant differences observed between the BAIS groups: the Whole Tone Pitch Shift condition as an effect on TRD, and the 200 ms Delay condition as an effect on CV.

TRD had a moderate negative correlation with aggregate BAIS, r(43) = -.44, p = .003, but was not significantly correlated with PEY, r(43) = -.12, p = 0.18. The partial correlation between aggregate BAIS score and per-participant TRD when controlling for PEY in the former case remains significant, r = -.45, F(1,45) = -3.25, p = .002. When controlling for BAIS score, the partial correlation between TRD and PEY remains insignificant, r = -.15, F(1,45) = -0.99, p = .33. This supports the originally observed effect, where a higher BAIS score (indicating better ability to produce and control auditory images) correlates with less tonal deviation when presented with pitch shifts even when performance experience was accounted for, whereas the relationship between accuracy and

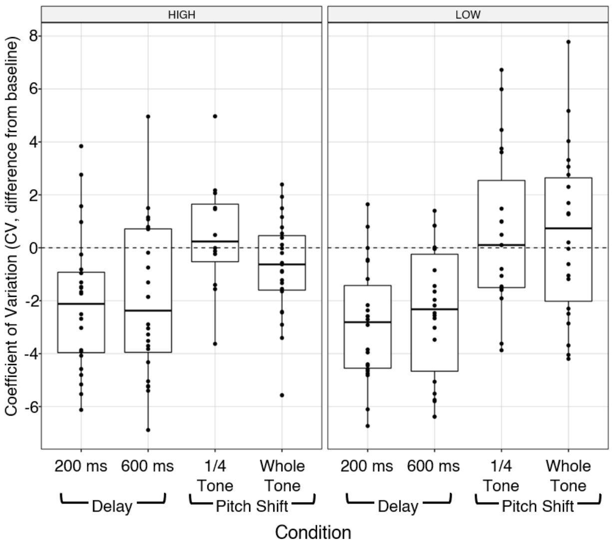


Figure 5.9: Temporal Deviation: Group-adjusted CV score (AAF conditions by BAIS group).

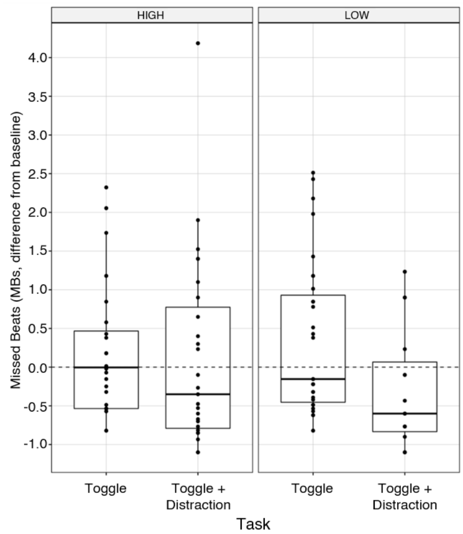


Figure 5.10: Temporal Deviation: Group-adjusted MBs score (AAF conditions by BAIS group).

performance experience is fully accounted for by inter-participant differences in auditory imagery assessed in the partial correlation analysis.

CV did not have a significant correlation with BAIS, r(44) = .29, p = .053, nor with PEY, r(44) = -.28, p = .06. Controlling for PEY, the partial correlation between BAIS and CV in the 200 ms Delay condition was still insignificant, r = .28, F(1,46) = 1.92, p = .06. Controlling for BAIS, the correlation between PEY and CV in this condition also remained insignificant, r = -.27, F(1,46) = -1.86, p = .07. This is fitting with the previously observed effects; while there were significant (and unexpected) effects of the experimental conditions on CV, neither this relationship nor CV itself were affected by BAIS. This will be a major component in the Discussion of this research.

5.4 Discussion

This study aimed to explore connections between musicians' auditory imagery ability and their ability to maintain accuracy when introduced to AAF during a performance. Separate analyses using individual-adjusted and group-adjusted accuracy scores were performed to determine a singer's ability to perform under AAF. The individual-adjusted analyses examined how well a singer was able to maintain consistent accuracy between their control and AAF performances. The group-adjusted analyses examined singers' accuracy when compared with the other participants with similar auditory imagery abilities.

The results show that pitch deviation under pitch shifts correlates negatively with BAIS score, even when controlling for years of performance experience, whereas this is not the case for temporal deviation under DAF.

5.4.1 Tonal Deviation

When examining differences between the groups for tonal deviation, there were significant interactions between BAIS group and task and BAIS group and condition on individual-adjusted TRD. The Whole Tone Pitch Shift condition (Figure 5.4) and the Toggled & Voice Distraction task (Figure 5.5) were both found to have a significantly greater effect on individual-adjusted TRD for low BAIS singers than high BAIS singers. Low and high BAIS singers were also significantly different in terms of group-adjusted TRD in the Whole Tone Pitch shift condition.

In all cases, the low BAIS group had higher TRD, suggesting more pitch error in these performances. Participants in the high BAIS group appear to be able to better maintain tonal accuracy compared to their control performances, demonstrated by adjusted TRD scores close to 0. The data suggests auditory imagery assessed by the BAIS has an impact on the ability for an individual to maintain their tonal reference accuracy across different challenges in performance. This reinforces the hypothesis for stronger imagery ability to provide a stable internal reference which is preserved and can be used to adapt to non-ideal performance settings.

The examination of group-adjusted TRD (Figure 5.8) indicated that condition was a main effect. About half of low BAIS scorers had upwards of 0.4 semitones greater deviation in the Whole Tone Pitch Shift. This suggests that, while auditory imagery may contribute to an individual's ability to keep consistent tonal error across performances with different AAF, stronger auditory imagery does not provide the ability to outperform others in these cases. This is relevant as previous work has shown large effects of BAIS in pitch-poor singers (Greenspon et al., 2018; Pfordresher and Halpern, 2013; Pfordresher et al., 2015). When examining confident and actively performing singers, there appears to be little effect of BAIS score on tonal accuracy compared to other singers.

The significant difference in tonal reference deviation between the high and low BAIS groups for the Whole Tone Pitch Shift condition indicates that this type of AAF is more disruptive to singers with weaker auditory images. These singers may find they have a harder time performing to the best of their abilities with larger pitch shifts. This is important because similar auditory distractions can be common when singing with others. For instance, in an a capella choral setting, listening to others within a voice section who are unable to pitch match and begin to sing incorrect intervals could lead to overall drift of the group. Choral singers in particular are trained to adjust to what they hear from others in their section and across the choir for timbral and tonal matching to ensure consistent sound and blending between voices (Howard, 2007); if enough singers within a voice section are more distracted by a "pitch shifting" caused by the intonation of others, these singers may find it harder to blend or to stay within a tonal centre.

However, this impact of pitch shift on low BAIS musicians is significant only when shifted a whole tone up, but not when shifted only a quarter tone; this perhaps indicates that there may be a threshold above which the difference between the internal reference and the off-pitch signal makes maintaining intonation difficult for singers with lower imagery ability. SLHR examining the influence of fundamental frequency shift on speech indicates that larger pitch shifts lead to more incomplete compensation (Daliri et al., 2020; MacDonald et al., 2010), which might explain why the quarter tone shift was easier to adapt to. Where this smaller shift might feel more like a chorusing or, as one participant described it, a "robot" filter on the voice, the audible feedback still feels like part of the vocal expectation. The larger whole tone likely creates a separation from the expected frequency output strong enough to disconnect the action-effect relationship. As well, the present study did not examine how the direction of pitch shifting impacted the direction of drift, which is important to assess in follow-up work. Typically, choirs drift downward in intonation, so this would benefit from examination in an empirical setting.

Previous work with singers has found that drifts and compensation to auditory feedback is essential to staying in tune (Howard, 2007); in contrast, SLHR also demonstrates that people compensate for pitch shifting by drifting in the opposite direction (Franken et al., 2018). Franken et al. (2018) summarise that there is a mixture of opposition and matching when compensating for tonal discrepancies, likely as a result of the motor system's attempt to minimise the difference between the predicted and perceived sensory feedback. In some cases, the speaker might not be able to understand where the discrepancy originates (either from the self or at a later point in the feedback), leading to variance in coping strategies (Franken et al., 2018). However, in this study, all participants were aware that the feedback was upward pitch-shifted and had time to experiment with the stimuli before singing through their selection. It is possible that the low BAIS musicians acted as "self-blamers" and behaved as though the pitch shift originated in their singing, rather than the feedback. Such self-blamers tend to rate themselves as being "less musical" (Franken et al., 2022) perhaps parallel to the low BAIS scorers' indication of fewer years of performance experience. With less experience and lower auditory imagery ability, they were perhaps unable to reference action-effect paths in existing imagery, leading them to adjust their physical behaviour when hearing unexpected auditory feedback and producing greater TRD than their high BAIS counterparts, who were able to ignore pitch-shifted feedback by relying on internal predictions (Jones and Keough, 2008; Zarate and Zatorre, 2008).

Tasks of singing with forced audiation and dialogue distractions may also have effect on an individual's internal tonal model, as the interaction between BAIS group and task was found to be significant. There was a significant difference between the individual-adjusted TRD of the low and high BAIS groups when performing under the Toggle & Voice Distraction task. We designed this to be the most challenging of the tasks, given that the audiation was further disrupted by the dialogue happening in the background. This further supports the idea that stronger imagery may provide more consistency in maintaining a tonal reference when faced with difficult tasks, including audiation.

5.4.2 Temporal Deviation

Coefficient of Timing Variation.

In terms of temporal deviation, the results suggest a less straightforward relationship with BAIS. There was a significant effect of BAIS group and condition on individual-adjusted CV, although these effects did not interact. In the low BAIS group, the 200 ms Delay and 600 ms Delay conditions had significant effects on individual-adjusted CV. Unexpectedly, low BAIS group singers had *lower* CV in these conditions when compared with their control performances (Figure 5.6).

As expected, high BAIS singers were more consistent with their control performances, with adjusted CVs closer to 0. BAIS group and condition were found to independently have a significant effect on CV. As there was no interaction between the independent variables, the effect of condition was not related to auditory imagery ability. This is reflected in the analysis group-adjusted CV, where only condition was found to be a significant main effect (Figure 5.9). In the group-adjusted analysis, *both* groups performed slightly better than their control performances in the 200 ms Delay and 600 ms Delay conditions. Pitch Shifted conditions give adjusted CV values near the average for control performances. This suggests that pitch shifts had little impact on the CV across the group, which seems plausible. However, SLHR provides a link between DAF and sensorimotor coordination and understanding auditory feedback in speech (Franken et al., 2018; Malloy et al., 2022). This has several implications for understanding the relationship between imagery ability and temporal deviation, which we discuss with respect to vocalisation in singing and speech.

Firstly, it is important to note that, although there were no correlations between BAIS and CV score in the control performances for each task, there were some participants who had high timing variation in the control. This would have resulted in "better" performances in the AAF conditions. We therefore suggest that there may be a separate aspect of auditory imagery ability corresponding to temporal imagery. SLHR suggests that delays, especially longer ones, give a feeling of compounding, and subjectively feel harder to overcome (Malloy et al., 2022). It is suggested that such delays are disruptive at a neurological or cognitive level as a result of increased attention and awareness to the mismatch between expectation and actuality (Malloy et al., 2022), which would provide little room for coping using auditory imagery alone. DAF has fundamentally different effects on speech than frequency-altered feedback; in cases of pitch-shifted feedback, people are found to make changes in their speech to better match expectations of the sounds they produce, but which do not result in transpositions of the speech or stutters, as found with DAF (Burnett et al., 1998; Purcell and Munhall, 2006). This may highlight why imagery appears to be more linked with pitch shifted feedback in this study as singers are more successfully in negotiating expectations with reality. DAF, operating perhaps more heavily on sensorimotor feedback (Malloy et al., 2022), might not be so easily overcome by auditory imagery ability alone.

Given that there is no temporal aspect within the BAIS, it is difficult to say whether an individual's vividness or control of an auditory image includes any aspect of timing ability. An additional measurement may be needed to examine temporal aspects of auditory images. This suggests exploration for further testing of time-based auditory images. For instance, prior to the establishment of the BAIS, another study by Halpern (1988) sees participants recalling the tempo of a familiar piece and then setting a metronome to the correct tempo; while it is clear this task relies on an existing auditory image, this aspect of auditory imagery may vary independently from the subscales making up the current version of the BAIS.

Furthermore, it is established that internalised pulse is reliant on both auditory and kinaesthetic or tactile imagery. Examination of auditory imagery alone may not be enough to accurately determine someone's ability for timekeeping. Auditory-motor interaction utilises the predictive role of the motor system to make judgements about timing (Cannon and Patel, 2021; Proksch et al., 2020), and the motor system is involved in beat prediction even when no movement occurs (Gordon et al., 2018). Sensorimotor synchronisation (SMS) ability is found to be positively correlated with auditory imagery ability, as well as years of playing musical instruments (Pecenka and Keller, 2009) The Multi-Modal Imagery Association Model (MMIA), which links auditory and kinaesthetic imagery through sensorimotor associations to describe the translation between complex systems used in singing (Pfordresher et al., 2015), may be most appropriate for closer study of timing consistency or body-based representations. Auditory imagery, closely linked with bodily sensations and sensorimotor representations in multimodal imagery, allows for accurate mapping of planned motor images to achieve desired auditory outcomes, which are also applicable for examining planned timing of note events and mental tempo representation (Pfordresher and Palmer, 2002), as well as in SMS in expressive music (Colley et al., 2018a). Previous work suggests that this relationship may vary depending on the task, between individuals who might exhibit preference for either somatosensory feedback or auditory feedback (Lametti et al., 2012), or depending on which subsystem of the vocal mechanism — involving articulatory or laryngeal control — is more active (Weerathunge et al., 2022).

The idea that tempo-specific image production and tempo memory recall are widespread in the general public is also well established. Non-musical individuals are found to have accurate absolute tempo references preserved in long-term memory (Levitin and Cook, 1996). This is a widespread ability and is believed to be unrelated to musicality and training, as with other aspects of auditory imagery; rather, it is believed that absolute tempo is related to tactus, an internal rhythmic period representation or some other body-based reference (Gratton et al., 2016). The urge to move to a musical work and the "feel" or "groove" within the body is driven by internal representation of temporal regularity (Hosken, 2018; Senn et al., 2019; Vuust et al., 2018). It is possible that participants did better in the DAF performances than in the controls because they introduced other elements of external timekeeping which do not rely on an auditory image, such as foot tapping, to help during these challenging tasks. In a similar sense, visual stimuli of timekeeping were found to help cope with DAF in speech (Malloy et al., 2022); some participants conducted beat patterns in front of their bodies, employing both kinetic and visual references to help reduce delay.

As well, it is critical to note how performance goals changed through the study as new challenges were introduced, and how this might have caused participants to perform with less variation in their tempo when confronted with DAF. In the initial control performances, the participants were tasked with singing as they normally would; in solo singing, vocalists might not keep explicit timing in favour of expression (Müller et al., 2010) or rely more on salient perceptual onsets rather than strict metronomical timing (Coath et al., 2009), thus resulting in "higher" initial CV. This further indicates that temporal drift over time requires further study in the way in which tonal reference drift has been studied, to produce alternate measures that more adequately capture "accuracy." CV is limited in that it defines accuracy measured in strict millisecond timing; a piece sung a capella will be judged more harshly by CV of IOI than by both singer and listener. It is also suggested that humans are naturally lax in determining isochrony, as perception of repetition and timing are formed through natural stimuli, which rarely exhibit precise timing (Madison and Merker, 2002). Madison (2004) suggests that internal periodicity operations are present and people will perceive regularity in timing even when there is noticeable drift (Madison, 2004; Madison and Merker, 2002). Subsequently, when needing to perform with DAF, these singers may have been less stringent in the control and only became more aware of their timing when it was the focus of the performance or necessary to work with DAF. As mentioned before, they reverted to other methods of time keeping which may have allowed for less variation between beat onsets for the sake of "ignoring any auditory feedback you hear... keeping in key and staying in time." It would be worthwhile to conduct the experiment in an accompanied context, to see how other external timing cues may impact drift and variation when performing with DAF.

In order to demonstrate these factors, and why it might not be so unexpected to see these results, we can examine Participant 9's performances and individual-adjusted scores: In the Normal task performances, Participant 9 had a baseline CV of 15.43 in the Control condition but achieved a CV of 3.64 in the 200 ms Delay condition – this was the lowest individual-adjusted score of any other participant at −11.79. This particular participant's background and behaviour during the study can help us to understand what may be occurring during the performances. In the control performances, the participant's timing was very free; at the start of the study, the participant spoke of really enjoying singing the song and placed lengthened emphasis on certain words in the text. The chosen song itself also had frequent syncopated timing. Once singing with DAF, it was clear the participant's focus changed. Audio-video recordings of the performances revealed that the participant began heavily dancing along with their singing, incorporating full-body sway, arm movements, and foot tapping in time with their pulse. Longer held notes became more punctuated, with the participant using pronounced glottal stops to separate any melisma into individual notes. The participant also described enjoying the challenge of singing with the delays, making it clear that this became the focus of the performance: not to recreate the initial performance but to instead get through the delays with consistent timing. This participant is also a highly experienced performer. Their regular performances of experimental and electro-pop music would have also provided ample previous experience with performing under AAF, which may have also assisted in this performance. In this case, we observe how body-based time keeping and focus on the task itself over performance aesthetics may have led to less variation when compared with baseline performance.

Missed Beats in Audiated Sections.

There were no notable effects of BAIS group, task, or condition on the number of beats a participant missed in an audiated section, which further supports the conclusion that BAIS has limited relationship with temporal deviation (Figures Figure 5.7 & Figure 5.10). There is reasonable consistency amongst performances for both groups: the greatest timing drift was about ±2 beats compared to the control performance. The majority of participants averaged less than 1 MB in any audiated section.

In this sense, rhythmic stability appears mostly unaffected when switching between audiation and singing aloud and, while audiating, participants in both groups are able to remain at a tempo consistent with that followed when singing aloud. This also suggests more awareness of the current tempo and ability to adjust, as participants did not default back to a remembered tempo once they began audiating, but rather continued mostly at the same tempo where they left off; otherwise, the number of MBs would be expected to be much higher with adjustments made during audiation.

An interesting comment by a participant who struggled with the DAF conditions was "I could feel I was going too slow, but it became so hard to stop." There were other similar comments made, which further indicates that there is an existing auditory image for tempo recall that can be relied on during AAF (and perhaps also a tactile image in that participants can "feel" they are moving too slow, potentially in the duration of sung words and physical movement of the vocal mechanism and mouth). A similar effect has been found in SLHR with DAF, which notably produces reduced speaking rate as a compensation (Fairbanks, 1955; Finney and Warren, 2002). Given the sensorimotor components in speech production, it is likely that DAF disrupts the neural feedback loop active in monitoring whether expectation of actions matches what is heard. This is believed to be the cause of stutters (Howell, 2004; Zimmermann et al., 1988) or serial exchanges of syllables in speech, also found when DAF is introduced (Malloy et al., 2022). A speaker — or in this case, a singer — might feel as though they cannot keep up because of the delay in feedback, causing a slowed movement. This might further be amplified by the change in performance goals, as outlined previously by Participant 9's experience, to reduce stylistic variation in tempo when confronted with more demanding DAF.

5.4.3 Imagery Acquisition and Experience

Based on these analyses, it is plausible that the ability to produce vivid auditory images and control them does assist performers in dealing with AAF and non-ideal performance situations. Imagery may help them to maintain internal references and achieve consistency between control and AAF performances. Comparison of individual performances indicate that imagery recall may assist in the maintenance of consistent performance when challenges arise, as seen in previous studies (Brown and Palmer, 2012; Edmonson, 1972; Finney and Palmer, 2003; Goebl and Palmer, 2008; Highben and Palmer, 2004; MacRitchie and Milne, 2017). This is therefore a useful ability for successful performance and a valuable asset to those pursuing performance. It remains unclear however how auditory imagery ability can be developed or trained in musical practice.

As previously mentioned, existing research on the use of the BAIS for self-indicating imagery ability determined that this ability may only mildly correlate with training. Previous study indicates that reliance on external auditory feedback decreases with training, particularly classical vocal training (Bottalico et al., 2016, 2017). In this study, it appears there there is no correlation with *formal* training; however, there is a notable relationship between years of performance experience and score on the BAIS. It might therefore be suggested that, rather than having a particular musical education background or training, the act of making music and performing is the more desirable component to training imagery ability. Although singers might have hefty performance loads in a formal conservatory environment, this study suggests that “experience” in performing in any context is valuable. There were several participants in the study who, despite having no formal musical training, maintain very active professional and semi-professional musical careers – this is indeed common among performing musicians.

Partial correlation analyses showed that BAIS was negatively associated with per-participant pitch shifting (TRD) effects, even when controlling for years of performance experience. This supports the association of musical imagery with better handling of altered pitch feedback. This was not the case for temporal AAF (CV) effects, which also supports the idea that temporal deviation observed in this study, and the internal tempo model itself, may have been influenced by other factors, such as performance goals and interactions with other abilities, namely kinetic imagery and body-driven representation. Perhaps, rather than years of experience, different kinds of performance experience, such as that of Participant 9, may be more influential in this adaptation. Further studies which replicate this experiment but also take into account individual performance backgrounds, musical styles, and other imagery aspects will be beneficial to fully understanding the relationship between these factors. However, we stand by the overall assessment that auditory imagery alone (at least as measured by the BAIS) does not have significant impact on temporal deviation when performing with DAF.

The main finding of the study is that auditory imagery, as assessed by the BAIS, is related to the ability to sing accurately and maintain tonal centre when performing with pitch shifted AAF. This appears to be driven directly by imagery ability, rather than years of performance experience. There remain two possible relationships between BAIS and performance experience; that high BAIS-scoring individuals are likely to become more successful performers or that performance experience itself develops auditory imagery ability. Any number of issues can occur during live performance, including issues with acoustics, feedback, monitoring, and often very prominent noise distraction from audience members; we suggest that, to thrive in this kind of environment and perform successfully, one would need to have more ability to maintain access to and adapt an auditory image.

It remains to be seen whether training can hone auditory imagery ability specifically and strengthen the image of ideal performance, for instance by rehearsing under stressful conditions with background noise and talking or practicing with artificially introduced AAF. In sporting performance, athletes utilise the PETTLEP model for imagery construction during training, allowing them to build multi-modal images of their techniques (Wakefield et al., 2013); this has also been shown to benefit musicians in their performance (Wright et al., 2014) and could be further considered for increasing one’s BAIS score and overall imagery ability. In addition to training imagery abilities through this type of experience, it seems likely that long-term performers self-select – only those with strong auditory imagery survive their gigs year after year.

5.4.4 Further Considerations

There are several considerations to the current study which should be considered in the interpretation of these results and the planning of further studies.

Participant-Selected Music.

The decision to allow participants to bring their own piece and potential confounding factors arising from this decision were considered during the design of the study. Although we have investigated these factors, it is possible that the individual pieces impacted the performance accuracy in an unknown way. Different pieces may have qualitatively different effects on and relationships with DAF and pitch shifting; it remains possible that the random toggled sections aligned in such a way as to be more in-time in some pieces rather than others, or the length of a delay's impact on pitch overlap created unexpected tonalities in some pieces but not others (Pfordresher and Palmer, 2002).

However, as discussed previously, we believe that allowing participants to choose their music offers significant benefits to experimental investigations of singing, as well as to the exploration of auditory imagery. There is a fundamental trade-off between undertaking a life-like task with a performance of familiar pieces and controlling all possible aspects of these interactions. Furthermore, having a musical context can provide a better representation of how performers act and how these effects manifest in an actual performance setting than using isolated performances such as tapping or pitch matching, although these provide more straightforward analyses.

Additionally, the long duration and challenging nature of the study and the cohort of singers recruited warranted the decision to provide an enjoyable and challenging activity for the participants; by using a comfortable and well-known piece, we hoped to reduce noise introduced by lack of familiarity. Sight-singing can be quite stressful for some singers who do not practice it as a specific skill, and even more so for participants who cannot read music, which was not a requirement of the study or a skill being selected for. In order to examine performance, we believe participants choosing their own piece to be an effective part of the method and hope that the evaluations used here can be useful for others wishing to include participant-selected stimuli in research.

Sample Size.

The small sample size should be noted in order to avoid making undue generalisations from these results. The 2x3x4 ANOVAs performed as planned analyses in this study were chosen during the planning stages in order to work with an anticipated small sample, given the constraints on participant recruitment and rather intense nature of the task; where larger sample sizes are available, other statistical tests may be more suitable. For instance, to more accurately determine the interaction between all of the factors at play, including BAIS, performer experience and background, condition, task, and complexity of the musical selection, mixed-effects regression and multiple regressions could be employed but these would require larger sample sizes.

A median split was used to divide the continuous BAIS scores into two groups for the between-group variable in a three-way ANOVA, providing a tractable and robust way to examine the different factors and interactions of interest. Although this method is statistically acceptable given that the predictors were found to be uncorrelated, it is possible that this introduced some generalisation and loss of statistical power by forcing participants into a "high" or "low" dichotomy (McClelland et al., 2015). In this study, none of the participants had a BAIS score lower than 4 and were all on the "high" end of the BAIS scale itself; therefore, this split effectively divides a group of people who are all reasonably able to produce and control auditory images. At the same time, it is important to note that the equality between the participants above and below the median in a median split is more likely to suppress the effect size and produce a conservative analysis (Iacobucci et al., 2015). Given that there are significant reactions between these two groups when using a median split, we would also anticipate observing this effect of BAIS on the ability to maintain consistency with AAF when using other measures such as multiple regression. It is therefore important to note that, with the small sample size, we intend for this work to provide the basis for future studies to examine the different facets of this work with larger sample sizes, for confirmation of the findings and validation of the previous work referenced here.

5.4.5 Future Study

This partially exploratory study supports existing knowledge on auditory imagery's relationship to maintaining tonal references and suggests intriguing opportunities for better understanding its relationship to temporal maintenance.

To the best of our knowledge, there is currently no model being used to express the temporal drift of singers over time in unaccompanied singing in the way that is done for tonal reference. It would be worthwhile to determine how this drift in timing occurs normally, both for the study of accuracy in different conditions as well as for the study of performance dynamics in improvisation and choral settings. Further links and collaboration between SHLR and singing research will further help to determine how this compensation might work. For instance, DAF produces slower rates of singing; as noted by Pfordresher and Palmer (2002), operating at IOIs at binary subdivisions of the given delay might provide a way to compensate or work around what is heard in musical performance. It could be possible that, by slowing down, participants subconsciously reduced the discrepancy between expectation and actuality in this sensorimotor interaction loop. Future work might explore the trajectory of timing as it functions to create a more manageable delay, thus providing a way to model of temporal drift similar to tonal drift.

Expanding on tonal drift, it would be useful as well to compare the findings of this work with other appropriate measures of pitch error estimation used in previous studies. This might include the difference between expected and sung f_0 (Watts et al., 1994), absolute pitch relative to a tuning reference (Vurma and Ross, 2006), relative pitch interval accuracy (Dalla Bella and Berkowska, 2009), mean deviation from target pitch, and consistency in producing repeated pitch (Pfordresher et al., 2010). Because we are working with unaccompanied singing of longer musical passages, we considered that using TRD over other measures provides the most appropriate measure of internal representations of pitch reference, rather than a focus on each localised instance of pitch error [3]. Other accuracy measures may reveal further details about tonal deviation, for instance in consistency of tuning.

In the same vein, it would be useful to develop the understanding of time-based auditory images and their reliance on both internal and external timekeeping. The link between auditory and kinaesthetic imagery is becoming more understood, and we hope that this research can provide the basis for specific study of this relationship; for instance, in the effect of employing different timekeeping methods, strengthening kinaesthetic imagery, and working with sensory-based perception of movement on timing accuracy. It would be particularly worthwhile to apply blended imagery models, such as the MMIA, in similar studies to test participants' ability to create timing-specific auditory images as an indicator of imagery ability; this finding potentially offers a place for questions related to temporal elements of auditory imagery assessed with the BAIS questionnaire.

In further iterations of this study, it would be useful to include a condition that incorporates further sound masking using an additional noise source, as done in (Parrell and Niziolek, 2021; Pfordresher and Mantell, 2012). This would further ensure that the unaltered voice is masked while AAF stimuli are provided, although it would remain difficult to mask any bone conduction of the vocal sound in these conditions. Additionally, using just noticeable difference (JND) tasks would further help to contextualise the TRD and CV measures by providing further reference to where pitch and delay discrimination, as well as any additional noise stimuli, became perceptible to the participants. As well, additional tasks not related to AAF, such as pitch-matching, mental

[3] Although only TRD is discussed, we also performed analyses introduced with TRD in Dai et al. (2015), including calculated Mean Average Pitch Error (MAPE) and Mean Average Interval Error (MAIE) measures. MAPE and MAIE produce identical patterns of results. We believe TRD to be the most appropriate of the three for our exploration of tonal stability, as it represents the drift of a singer's internal tonal centre.

transformations, and sensorimotor synchronisation tasks might be incorporated with respect to participant-selected pieces, to evaluate isolated actions within a known piece.

5.5 Conclusion

In this study, we examined the ability of musicians to maintain tonal and temporal accuracy while singing a familiar piece of music under combinations of AAF conditions and different forced audiation and voice distraction tasks.

We find that singers with greater auditory imagery ability on the BAIS are able to produce more consistent accuracy in performances across different conditions. High BAIS singers have similar TRD between their own control and AAF performances, while low BAIS singers have greater TRD in their AAF performances. This is significant when singers are presented with feedback pitch shifted a whole tone up. As well, in a forced audiation task where a voice distraction was also present, the high BAIS singers again had more consistent performance. Stronger imagery is therefore likely to provide a stronger reference for singers to call upon when confronted with AAF or required to audiate.

On the other hand, we find there are no significant interactions between BAIS score and temporal accuracy; both high and low BAIS singers were significantly affected by DAF, but surprisingly this resulted in less temporal deviation when compared to average control performances. This suggests other effects. Singers may not prioritise timing in unaccompanied singing until confronted with feedback which focuses on maintenance of a temporal centre. We believe these results support multi-modal imagery theories: the ability to maintain temporal reference and timing consistency is likely not a factor of auditory imagery alone, but rather an interaction with other factors, namely kinaesthetic imagery and body-based tactus. More investigation is necessary into the effects of multi-modal imagery ability on temporal deviation.

Finally, we find that auditory imagery ability did not correlate with formal training, but rather with more years of active performance experience. We hypothesise this may result from performance-driven learning to adapt to non-ideal listening conditions. This suggests that auditory imagery can be improved and trained through performance practice, providing future direction for music education.

Chapter 6

Understanding Vocal Perception

Metaphor-Based Communication of Sensory Experience

The next part of this research was focused on understanding more about how singers understand, describe, and perceive their own movements. Voice teachers rely on abstract metaphor to express information about unseen and intangible processes inside the body, for instance in the forms of abstract language and gesture, to convey their own understanding and sensations to the student, who is expected to translate this instruction into their own action and movement. An implicit assumption in metaphor use is that it requires grounding in a familiar concept, prominently seen in the popular Desktop Metaphor outlined in Chapter 3, Section 3.3.2. In human-to-human communication, however, abstract metaphors, without such grounding, are often used with great success. To understand when and why metaphors work, we present a case study of metaphor use in voice teaching. I worked with vocal teachers about their teaching, not only reflecting on how those teachers understand their own practice, but also how they adapt this knowledge for others. We see through this study the basis through which singers become acquainted with their bodies.

This study explored how metaphor is used and adapted from different teaching and performance experience and internalised in vocal practice. The metaphors examined focus on the training of fundamental physiological components of healthy singing, including posture, breathing, sound production, projection and vowel formation. We present a thematic analysis of metaphor use by 12 voice teachers. We found that metaphor works not because of strong grounding in the familiar, but because of its ambiguity and flexibility, allowing shared understanding between individual lived experiences. Our findings can be used as an analysis tool, for better understanding why metaphor works in HCI, as well as a design resource for thinking about metaphor use in new designs. As well, the study uncovered other interesting cases of metaphor use specific to the voice lesson; for instance, that the images used to teach directly mirror a physical movement or work against it, how technique is encouraged through distraction, and how metaphors can be used to help students focus on different sensations in their body.

Portions of this chapter have been published in:

Courtney N. Reed, Paul Strohmeier, and Andrew P. McPherson. 2023. Negotiating Experience and Communicating Information Through Abstract Metaphor. In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23). Association for Computing Machinery, New York, NY, USA, Article 185, 1–16. DOI: 10.1145/3544548.3580700

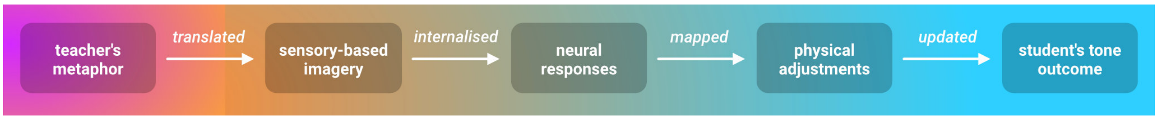


Figure 6.1: The top down process of metaphor provided by the teacher (communicating agent, left) being translated into the student's (receiving agent) understanding through sensory-based imagery. It is then internalised and mapped to physical adjustments, updating how the student executes their singing (right).

Before moving into the study, I again want to present the model of metaphorical communication and translation of a teacher's (the communicating agent) sensory-based knowledge is translated into the student's (receiving agent) knowledge (Figure 6.1). I will conclude the chapter by discussing how the understanding of metaphor in the voice lesson is beneficial not only to the understanding of vocal practice and the vocalist-voice relationship, but also to HCI in general, as an alternative approach to metaphor in communicating between humans and technological agents.

6.1 Method

We conducted a series of interviews to explore how, in human-to-human communication in the voice lesson, information can be shared between individuals using metaphor-based communication and teaching strategies. We focus on beginner vocal technique in this study because lessons at an early stage in the singing career typically focus more on building a relationship with the body, physical movements, and behavior needed for healthy singing, rather than the nuances of any particular musical style. We are thus able to work with teachers in a mix of teaching environments and musical genres. The end goal of each teacher is largely the same: the student should have awareness of the movement of breath in and out of the body and the appropriate level of tension and resistance to pass air through the larynx and control the resulting sound. The teacher must therefore convey information about behavior for the student to recreate, negotiating their different lived experiences and physical bodies. This is done through metaphor.

Voice teachers were interviewed about the metaphors they use to teach vocal fundamentals to beginner students. We focused on beginner technique because lessons at an early stage in the singing career typically focus more on building the relationship with the body, physical movements, and behaviour needed for healthy singing, rather than the nuances of any particular musical style. We are thus able to work with teachers in a mix of teaching environments and musical genres. The vocal practices discussed in the interview build the foundation of all singing practice, regardless of style or genre, and included:

- **Supported breathing**
  Providing muscular support for air flow, using the lower abdominal muscles to breathe in a controlled way.
- **Posture**
  Aligning the body in such a way as to control tension and provide space for the breathing apparatus.
- **Sound production**
  Creating sound and changing pitch with the vocal folds and changing the tension in the laryngeal muscles.

- **Sound shaping**
  Using different resonant spaces with the soft tissues of the mouth and throat to change the quality of the sound.

Despite the variety of metaphorical language, gesture, and imagery used to teach these fundamentals, the end goal of the teacher is largely the same: the student should have awareness of the movement of breath in and out of the body and the appropriate level of tension and resistance to pass air through the larynx and control the resulting sound.

6.1.1 Materials

To investigate potential reasons *why* we use particular metaphors in describing our experiences, teachers' musical backgrounds and use of different methods of mental imagery were examined. We used a set of self-reporting questionnaires, including the Bucknell Auditory Imagery Scale (BAIS) for auditory imagery ability (Halpern, 2015), the Movement Imagery Questionnaire-3 (MIQ-3) for visual and kinetic imagery ability (Williams et al., 2012), and the Goldsmiths Musical Sophistication Index (Gold-MSI) to assess general musical experience and basic demographics (Müllensiefen et al., 2013, 2014). The aim was to determine if a particular kind of metaphor was preferred or used because a teacher had more aptitude for that mode of mental imagery (e.g., someone who is able to produce auditory images well might use auditory metaphor to explain their experience to someone else).

6.1.2 Participants

Twelve voice teachers (10 female, 2 male) participated in the study. The participants were recruited through my own musical groups and social media groups specifically focused on vocal education. The teachers were diverse in terms of personal background, teaching practice, and vocal genre. All of the participants currently teach in English but have a variety of national backgrounds, including the UK (3), USA (2), Greece (1), Sweden (1), Singapore (1), Austria (1), and three dual nationals: Portugal/UK, Hong Kong/Malaysia, and Italy/USA. The participants range in age from 25-72 years (M = 38.67, SD = 15.2). The general musical sophistication on the Gold-MSI of all the participants was well above the population average musicality of 81.58 (Müllensiefen et al., 2013), ranging from 107-151 (M = 125.9, SD = 12.75). The teachers work in a variety of styles, with an even split in the group between classical (art song, opera, choral singing) and contemporary vocal styles (theatre, pop).

6.1.3 Interview Procedure

The interviews were conducted virtually either on Zoom or Skype and were audio-video recorded. The teachers were interviewed about the metaphors they encountered as students and those they use in their own teaching. The interview consisted of three parts, beginning with a brief introduction to the teachers' musical styles, lesson settings, and other background information. This was followed by a discussion of metaphors related to the aforementioned vocal fundamentals that they had experienced as students, and then discussion of their own teaching methods, including their thoughts on the role of metaphor in the voice lesson. The interview questions in detail can be found in Section B.1.

6.2 Analysis

The interviews were conducted virtually either on Zoom or Skype and were audio-video recorded. The interviews were transcribed down to the level of utterances and facial expressions. Physical hand gestures and body movements were also notated with the conversation text. Analysis consisted of a qualitative thematic analysis of the interview data to explore the metaphors used by the teachers. Further statistical analysis was performed to determine whether there were any significant differences in the teachers' metaphor usage according to their vocal style and their imagery skills.

6.2.1 Imagery Modalities

The teachers referenced a variety of auditory, kinetic, and visual metaphors. In total, there were 54 auditory images referenced, 270 kinetic images, and 237 visual images. In order to determine if teachers used metaphors of a certain modality because they have higher aptitude for that modality, we used a Spearman's ranked correlation test.

6.2.2 Thematic Analysis of Metaphors

The interviews were transcribed automatically using Otter[1] and were manually corrected while adding additional information from the video capture, down to the level of utterances and facial expressions. Physical hand gestures and body movements were also notated with the conversation text. This was done manually for familiarisation with the data and acknowledgment of the multi-modality of the metaphors (Trainor and Bundon, 2020). We used a reflexive inductive thematic analysis approach (Braun and Clarke, 2006, 2020b) to examine the metaphors the participants used in their teaching practices; that is, initial coding was done over a period of two months by myself, providing further insight and introspection within my own history of receiving voice lessons and working in vocal education. The analysis examined both vocalist explanations of metaphor and their use case and my own experience in these roles. The process first involved a look at the modality and method of delivery — how the teachers presented the information using imagery by creating imagery which extended tacit knowledge to imagine new sensations or by borrowing imagery from other lived experience. We then further iterated through the data to derive four main themes which outline how metaphor is used to communicate sensory-based experience more generally.

6.3 Results

We first examined general strategies for how auditory, visual, and kinaesthetic imagery were employed by voice teachers. I here outline the organisation of strategies to provide examples of the metaphors extracted from the interviews. Then, I present the results of the Spearman's ranked correlation test, which examined possible links between teachers' imagery ability and the kinds of metaphor used and the results of the thematic analysis that examined *how* the different metaphor strategies work in general.

6.3.1 Identified Imagery Strategies in the Voice Lesson

Twelve (12) general strategies of communicating sensory-based experiences using auditory, visual, and kinetic imagery, and their total counted observances (CO), were identified (Table 6.1). I would

[1]otter.ai

like to note here that the use of a CO or similar is not an aspect of the reflexive thematic analysis and it is not necessary to provide quantitative data for thematic analysis; here, I use it to demonstrate the balance of metaphor modalities used by these particular teachers and because the count was later used to determine if there was any link between teacher imagery ability and metaphors used. This is not to suggest that one form is better or "more common" than another, as this thematic analysis deals with the specific group of individuals and their accounts. From this specific analysis, I will derive a more general picture in the Discussion.

There were found to be two major categories of metaphor, either *Creating Imagery* and new associations for unfamiliar sensory experiences, or *Borrowing Existing Imagery* which would have been formed from outside life. Teachers created new images or invoked new experiences for their students (e.g., having students create new sounds or try out new behaviours to reference later in performance) and also employed the use of non-singing based metaphorical references from outside life (e.g., utilising a non-musical action the student would be specifically familiar with to reference in performance). These metaphors co-opt the students' existing mental imagery already formed outside the singing context for explanation of target sensory experiences. The resulting strategies were:

Creating Imagery:

- **Auditory Images**: Sound Quality, Practiced Familiarity
- **Kinetic Images**: Vocal Sensations, Physical Touch, Accompanying Gestures, and Alignment & Positioning
- **Visual Images**: Aesthetic Visuals, Observed Movement, Modelled Behaviour, Directional Metaphors, Body Shapes & Spaces

Borrowing Existing Imagery:

- **Metaphors from Other Life**: Heard Sounds, Felt Kinetics and Sensations, Seen Objects & Environments

6.3.2 Creating Vocal Imagery

The main way teachers communicate vocal technique is by using metaphors to build imagery around a task. This often uses the student's tacit knowledge of their body by having them imagine scenarios they have not experienced before. For example, a teacher may ask a student to "*Fill up your belly like a balloon.*" (Pt 01), as a way of getting the student to become aware of the air movement into the lungs and the resistance in the diaphragm to draw in the air and control the "filling" of that space. In this case, the diaphragm and the belly do not fill with air, but this visualisation of expansion can help to bring attention to the lower abdomen and the breath. The image of a balloon as a reference to breathing and air movement in and out of the body is one of the most common metaphors in vocal pedagogy. Additionally, students may be asked to compare different actions to contrast different internal sensations. By having a student imagine a sound, action, or visual as they perform different tasks, the teacher is able to build an association between technique and a mental image.

Auditory Images

The teacher provides a sound reference or describes a desired *Sound Quality* for the student to recreate. This often involves using different exercises for *Practiced Familiarity.* Students learn to

		Strategy	CO	Description
CREATING IMAGERY	Auditory Images	Sound Quality	14	Describes a sound quality or differences between sounds as students explore their range of action
		Practiced Familiarity	21	Focuses on a specific sound or sound movement (e.g., interval distances, musical passages) to reference
	Kinetic Images	Vocal Sensations	28	Describes a sensation similar to what should be felt while performing the target task
		Physical Touch	53	Uses touch on the body or the surrounding space to bring focus to a particular body part and its sensations
		Accompanying Gestures	138	Uses an accompanying movement (e.g., hand gesture) is used to represent internal movement
		Alignment & Position	20	Uses the tension and relaxation feelings of different body positions or postures as a reference
	Visual Images	Aesthetic Visuals	23	Uses an aesthetic description, often inducing an emotional quality, to provide high-level reference
		Observed Movement	130	Uses accompanying movement to externally visualise internal movement
		Modelled Behaviour	12	Creates a visual template for reference, often with body positions and postures of the teacher
		Directional Metaphors	19	Describes spatial orientation or movement within a space
		Body Shapes & Spaces	43	Uses visual references to shape and position parts of the body which are not able to easily be seen
BORROWING IMAGERY	Metaphors From Other Life	Heard Sounds	19	Uses a known sound reference, often not related to the target task (e.g., non-vocal sounds)
		Felt Kinetics & Sensations	31	Uses a known/experienced sensation from daily life, which is recreated or referenced in the target context
		Seen Objects & Environments	10	References a physical object or environment and asks the student to visualise it while completing a task

Table 6.1: An overview and description of the main strategies of metaphor used by the voice teachers, with counted observance (CO) for each strategy mentioned in the interviews.

employ their voice in directed exercises so they are able to match to what they hear and reproduce different or specific sounds. This will allow students to associate different movements and resulting feelings inside their body with different sound expectations.

Sound Quality. The teacher describes the quality of sound the student should aim for and introduce a warm-up or exercises for the student to practice. The teacher will often have a student cycle through different timbres or vowels within an exercise so that they can hear the difference between sounds. This is particularly useful in getting students acquainted with the sounds of their upper registers compared to lower notes. Some singers, especially those with lower voices, may have a markedly different sounding voice when singing high notes, so it is helpful to encourage students to learn the different sounds they are capable of making:

> Pt 12: "When I get them wanting to go for a bigger sound... I get them to move between the two, going on an *Aah* sound, going like that and hearing those different things."

> Pt 11: "I do a lot of whooping *Whoop whoop whoop!* *a high whooping in the head voice, their hand comes up and makes small circular motions by the side of the head near the eyes*."

Practiced Familiarity. The teacher focuses on the experience of specific repeated tasks, for instance to "*train people's ears to recognise triads, and intervals,*" (Pt 17):

> Pt 17: "I will copy lines from that song, and I will adapt the warm ups so that they're actually singing the song when they're warming up... bits of the song which might be challenging for them. And then after that, it's easier for them when we actually go to the song bit of the lesson."

Although these do not seem like metaphors, we view them as such because they extract the particular sensation which would be felt while singing (e.g., in the second example from Pt 11, the whooping would get the student to experience the sensation of using their head voice and lifting the soft palate) without directly discussing the technique or movement or working within context. These exercises provide references for unfamiliar sensations which should be experienced during healthy singing; by introducing them out of context, the teacher is able to provide an experience and association of that feeling with associations and identifications made during warm-up.

Kinetic Images

The teacher introduces new *Vocal Sensations* and movements to provide focus and familiarity for the student. These movements will be reproduced in singing practice; although *"you can't see what you're doing in your voice, you have to feel it and remember it,"* (Pt 17). The teacher may also use *Physical Touch*, *Accompanying Gestures*, and *Alignment & Positions* to focus on internal sensations or provide external references.

Vocal Sensations. The teacher will introduce a sensation to the student which is the same or similar to what the student should feel while singing.

> Pt 17: "What you're doing with a guitar, or a piano or a saxophone, it's the same thing with, with learning your voice, is just that you can't see what you're doing in your voice, you have to feel it and remember it."

> Pt 01: "I kind of get them to sing a really low note, really high note and feel how loose the low note feels. How tight a high note can feel." Pt 01: "I kind of get them to sing a really low note, really high note and feel how loose the low note feels. How tight a high note can feel. And then I get them to understand that if they're doing a note the note really high, but their voice feels loose, they're flat."

Physical Touch. The student or teacher physically touches the body or surrounding objects to bring awareness. This can introduce new sensations of pressure, stretch, and pull, outside the body, utilising the sensitivity of the skin to help bring awareness to movement, even when the hands are eventually removed. This can help control the mouth, confirm the body is in correct posture, or to experience the breath.

> Pt 01: "For for breathing and understanding kinaesthetically, what's happening, I put my hands literally at the back, like on my waist. And always check that this part of the body is moving."

> Pt 10: "...put their hand on the back of their neck *places hand flat on the back of their neck*. Okay, now, do this *they jut their chin and hyper-extend the neck forward.* How does the back of your neck feel? Oh, it's hard. It's tense. Exactly... We don't want any tension which will go into our voice."

> Pt 24: "With your thumb or like here doing like these things *places the heel of their hand on their cheeks and pulls the face down* to like, try to force your jaw apart."

Accompanying Gestures. Students are asked to perform external movements to accompany internal movement which might be more difficult to sense or understand as they how their body feels while singing. This uses the natural expression of the body, particularly the limbs, to provide an understandable sense of motion.

> Pt 04: "To demonstrate a phrase and how it grows, develops and then it ebbs away... hold out the beginning phrase *holds their hands apart in front of their body* and pretend there's bubble gum here *touches their right hand to their left palm* and then draw it... Like an elastic band."

> Pt 22: "Feel like you're lighter, feel like you're floating to the to the other side of this *their arm moves up and over, across their body*, don't take a straight line float, over top of this. And try to you know, try to make sure you land softly."

Alignment & Positions. Students practice postures and positions. These metaphors often have students play with the boundaries comfort by hyper-extending and then releasing tension back into a more natural and relaxed pose.

> Pt 04: "You're gonna pretend that there's there's a egg between your scapula *points to the centre between the shoulders* as well, your shoulders back and just keep that egg there *rolls their shoulders back and ducks their chin slightly*. Like squeeze it back *they move between a tensed squeezing of the shoulders and a relaxed position*."

> Pt 08: "...come up on our tippy toes, and then go back. And when our heels just almost come down, stay there... if someone would come poke at you, you would be steady."

> Pt 24: "The marionette is super useful... they can imagine like, 'oh, like I do have a baby string coming up top of my head. It does make my like spine feel a little bit like more tall'."

Visual Images

Visual metaphors often accompany other references provided in the voice lesson. Visual images can be abstract emotional *Aesthetic Visuals* or ones that focus on how the body should look through *Observed Movement* and *Modelled Behaviour*. *Directional Metaphors* and *Body Shapes & Spaces* also address movement of the vocal apparatus and produced sound through space and within the body.

Aesthetic Visuals. These metaphors are the most abstract and tend to invoke a visual image as well as an emotion. They are high-level representations of multiple other kinds of metaphor or can be used as cues to trigger different actions for the student.

> Pt 10: "imagine there's a huge cave with an underground lake, and this lake is perfectly still, so much that it looks like a mirror... and that's in your chest."

> Pt 24: "Imagine that like, your sound is like sailing through like the clouds... like you're on a plane just like *hands in a straight, forward motion*, the plane is sailing in a straight line forward easily going, like you're already in in the air and you got it it's up high."

Observing Movement. Similar to Accompanying Gestures, the student is tasked with performing an accompanying action (e.g., a hand movement) which pair internal sensations with a visual reference.

> Pt 08: "Having the sense of that airflow going *mimes a circular hand movement, drawing their arm out from their body*, like always going, even use your arm to make the, you know, to have a sense of direction *moves gesture bigger and out in one direction away from her body*, an ongoing thing."

> Pt 12: "I will get them to actually have their hands and step upwards and forwards *hands move up and out in a step-wise motion* not pushing, but having it so they're going *Yeah, yeah, yeah, yeah, yeah!* *increasingly louder and more forward in the resonance as the hands step up*. To help that they've got to let that sound go up and go forwards."

Modelled Behaviour. The teacher models a behaviour or shows an visual reference from another singer as a template, having the student recreate what they see in their own body.

> Pt 01: "I show them pictures of Adele while she's singing with her mouth super open, [you] could count her teeth."

> Pt 19: "...basically show them and make hand movements myself, I would stand upright myself, making hand movements, breathe in deep. And then show them the way and I'm supporting that by the right movements."

Directional Metaphors. Directional metaphors depict spatial orientation or movement of the body or the voice within a space to help students visualise internal movement without seeing an external cue.

> Pt 10: "Look at that poster in front of you pick, you know that person's face or that hand and try and imagine all your sound is focusing on that once like a laser *hands move from the side of the face to converge together in front*"

> Pt 17: "I like to talk about axes a lot, especially with my older students. So there's the X, there's the Y, and there's the Z. And the Z is like how forward or backward it is... And generally, it seems to be that people don't like to talk about this word 'placement.' So I just avoid that completely and I just talk about the Z axis, how front is it? How, like, swallowed, is it? And then Y would the pitch... and then with X, I talked about the different shapes like within your mouth when you're producing different sounds different vowel sounds."

Body Shapes & Spaces. Visual references are provided to describe parts of the body and their positions, again attempting to visualise internal movement.

> Pt 10: "I asked them to imagine the shape of an umbrella handle *outlines an umbrella shape with their hand, straight up with the back and then the slight curve to the hand,* so the back of umbrellas, the back of the neck, and it has a slight curve, and then I sometimes do have them use the hand."

> Pt 24: "imagine they have a ping pong ball in their mouth *forms a small round shape in the palm, placed next to the mouth*."

6.3.3 Borrowing Existing Imagery

The other frequently used type of metaphor teachers often use plays on students' familiarity with other experiences in day-to-day life. Existing imagery is used to describe an aspect of vocal technique, connecting an existing association instead of creating a wholly new one to vocal practice. This is a common part of contemporary Estill voice method, but even teachers working within classical contexts include this as part of their lessons. Again, we can split these references into auditory *Heard Sounds*, *Felt Kinetics & Sensations*, and visually *Seen Objects & Environments*; all revolve around something which has likely been previously experienced by the student. The students' existing non-vocal imagery is co-opted into the vocal space, rather than being created specifically with respect to singing.

Heard Sounds. The teacher uses particular non-vocal or completely non-musical sound. This can include sound descriptions, such as "*twang*," (Pts 01 and 12) or sounding "*reedy*," (Pt 24) comparing the voice to other instruments (a guitar or a woodwind instrument, respectively). Other descriptions come from the speaking voice or use character voices, mimicking different accents or using other familiar sounds, such as animal sounds:

> Pt 10: "Speaking in a funny voice which uses this resonant space to become acquainted with the feeling: a 'witch voice' or 'whiny voice'."

> Pt 11: "Yapping, whining. Lots of *Yeeeowwww* *vocalises a siren high in the head voice, like a cat's yowl*, of treading on the cat's tail, pull your cat's whiskers, all sorts of brutality *laughs*."

> Pt 01: "It's not like they've done singing for a long time, but they have been alive for a long time. So, the use of sounds, it's not like singing sounds are like coming from some alien place..."

Felt Kinetics & Sensations. These references are aimed at getting the students to feel natural, making sure the students do not overthink their behaviour, and that they assume appropriate, comfortable positions.

> Pt 08: "...Maybe we don't need to think about the breath so much. Maybe it's just happening."

> Pt 11: "I teach very much from a postural technique, not a 'oh my God, I've got to think about how I'm standing.' But again, trying to get people to be natural."

Teachers help students become aware of different sensations which might be felt in other actions. Some of these imagined sensations can be quite visceral, with the teacher aiming to invoke a very specific or intense feeling to reference:

> Pt 17: "When you start to yawn, the first few microseconds of your yawn is the optimum head dominant sound space. So, I get them to catch that feeling."

> Pt 01: "I tell them is try and to imagine someone is stepping on your foot, and you go, *Oh!*, because you do it. So, I kind of bring them into the situation, get them to do the reaction that would do the same, you know, in the same situation in real life."

> Pt 19: "We should imagine [we] were like seasick on the ship. And then we would go to the to a dock and open our mouth like that... So that was a funny image I had in mind. And that often comes to mind... to sing with an open mouth to project..."

Seen Objects & Environments. The teacher references a physical object or an environment to strengthen the sensory experience. This can also provide context to generate movement.

> Pt 11: "...imagine that you see your best friend about to step in front of a car, you would make the noise *HEY!*... it would be free and it would be supported."

Visualisation of a physical object can also be used to convey space or distance to the student, for instance to give reference to the space within the mouth needed for good resonance space:

> Pt 10: "Imagine you're *Hnnng* biting into an apple *mimes biting, exaggerated and with bared teeth...* it really gets that nasal resonance."

> Pt 04:"...you're gonna hold out your hand, *extends their hand away from their body, puts one finger up as if it's a candle* there's a candle here, you're gonna blow on it *takes a deep breath and mimics blowing out a candle*."

6.3.4 Correlation of Imagery Modalities with Ability

In total, there were 54 auditory images used (19 borrowed from other life, 35 created in-context), 270 kinetic images (31 and 239, respectively), and 237 (10 and 227). A summary of the questionnaire results for self-assessment of the teachers' imagery abilities is found in Table 6.2. Full questionnaire results for each participant can be found in Table B.1.

There were no significant correlations found between the teachers' imagery abilities and the types of metaphor used (Table 6.3). This suggests that teachers did not use different metaphor modalities simply due to their ability to produce that modality mental image; rather, metaphors were chosen for other reasons, which became more apparent through thematic analysis.

	MIQ-3			BAIS	
	Internal Visual	**External Visual**	**Kinaesthetic**	**Auditory Vividness**	**Auditory Control**
Range	4.75 - 6.75	5.00 - 6.75	3.25 - 6.75	4.00 - 6.71	4.14 - 6.86
M	5.88	6.06	5.50	5.52	5.54
SD	0.73	0.48	1.04	0.94	0.99

Table 6.2: Participant self-assessed scores on the MIQ-3 and BAIS for internal and external visual imagery, kinaesthetic imagery, and vividness and control of auditory imagery. Scores on each scale range from 1-7.

Imagery Type	**Scale Measure**	S	ρ	p
Visual	Internal Visual	179.06	.37	.23
	External Visual	382.97	-.34	.28
Kinetic	Kinaesthetic	246.74	.14	.67
Auditory	Auditory Vividness	271.85	.05	.88
	Auditory Control	215.25	.25	.44

Table 6.3: Spearman's Ranked Correlation testing between participant imagery skill scores and the CO of different modality metaphors used. Rho (ρ) represents the strength of the correlation, while the p value represents the statistical significance of the result.

6.3.5 Thematic Analysis of Metaphors

Within the different metaphor strategies extracted, as defined above, we defined four major themes that capture how metaphor works to convey information on sensory-based experiences and transfer knowledge. Experiences are communicated through abstract metaphors by *Requiring No Existing, Domain-Specific Knowledge*, *Working Independently of Language*, *Providing Ambiguity for Individuality*, and *Intentionally Limiting What is Communicated*. The themes together describe a process of how metaphor is created and used to negotiate information between the lived experience of the teacher and student.

6.4 Discussion

Wilson and Keil summarise the interaction singers have with their sensations as existing "outside the sphere of conscious thought" (Wilson and Keil, 1998), as in tacit knowledge. This is applicable beyond just singing interaction, and is likely the case with most of the actions we perform in day-to-day life (Dourish, 1980; Svanæs, 2013; Varela et al., 1991). The thematic analysis of the metaphors used by the voice teachers reveals a number of ways in which tacit knowledge is communicated. We see teachers' perspectives of their own understanding of vocal fundamentals tied into how they would articulate this understanding to students. Metaphor functions because it utilises pre-linguistic and non domain-specific knowledge for communication. Through metaphor, we are able to explore and teach tacit knowledge and use ambiguity and flexibility in interpretation.

The examination of metaphors used by voice teachers in their lessons reveals why abstract metaphor is useful and necessary to communicate information, particularly about subjective embodied experience. These themes together demonstrate how mutual understanding of subjective, sensory experience is shared between individuals and further inform ways in which information can be communicated between humans and technological agents.

6.4.1 Requiring No Existing, Domain-Specific Knowledge

Metaphor focused on experience outside of the specific context, using tacit knowledge independent of singing; the teacher does not assume the student would already understand the specific vocal practice and rather relies on existing relationships with the body. For instance Participant (P) 12 uses an abstract reference of imaginary "*alien noses*" on the lower abdomen to teach her students proper breathing: *"...we talk about having a belt full of alien noses* *her hands come around her abdomen* *and when they're breathing in, the breathing [comes in here]... we really want to get that connection."* The breath is obviously not drawn in through the abdomen, but this metaphor directs awareness for a sensation which is likely not familiar — the singer must learn attention to breathing and careful control of tension in the diaphragm muscle. By creating an abstract image of the breathing outside the stomach, one can become aware of the tension in and movement of the muscle. As in this case, metaphor can be created when there are limited experiences to pull from, providing an evocative (and in this case, also humorous) image to communicate information. Although it may be hard to imagine what a specialised and new skill like diaphragmatic breathing feels like, we can, through proprioceptive senses and tacit knowledge, imagine what it might feel like if our noses were on our stomachs (Svanæs, 2019; Svanaes and Solheim, 2016).

Other metaphor directly references previous experiences from real-life; for instance, P10 mentioned her own teacher had used a metaphor of smelling potpourri to elicit the same breathing behaviour. This reference, although open-ended and not specifically discussing any particular sensations, allowed P10 to concentrate on the feeling of this breath and what she described as a "*buzz*" in her sinuses. She was able to apply her tacit knowledge to explore this new feeling in her face, but she did not understand exactly why the metaphor worked so well, expressing surprise that she could instantly understand: *"The first time my teacher said this. Oh, oh, my god. How do I feel it there? How does that work?"* The metaphor was not explained by P10's teacher, nor is it explained by P10 when she uses it in her own, present-day voice lessons. The metaphor requires little explanation because the student will already have the embodied knowledge and imagery intact from similar lived experience to make this connection.

For P10, the goal is to create awareness of existing bodily knowledge by focusing on these sensations. This focus then becomes rooted in imagery (Depraz et al., 2003; Kosslyn et al., 2001): *"When you're learning something new - first you want to make things very conscious. Then the next step is to automatise it."* This aspect of metaphor use is highly applicable and already acknowledged in somaesthetic practice and design, particularly in exercises which explore body perception through representations in other sensory modalities (Cochrane et al., 2022; Daudén Roquet and Sas, 2021; Núñez-Pacheco, 2021). This act of making things conscious implies that the awareness of the sensation already exists in some form, but that careful thought and introspective reflection are needed to bring it to the forefront (Höök et al., 2021). The attention on details of experience also allows smaller individual gestures to be pulled from larger, embodied action paths (Grouios, 1992; Hale, 1982, 1994). Focus on the experience helps us to understand the low-level movement and technique behind it (Höök, 2010).

The key benefit is that the metaphors require no existing knowledge of the task to work. None of the metaphors provided rely on understanding of particular concepts or even in a singular modality; in fact, while we might expect singers to heavily use metaphors based in auditory schema, auditory metaphors were barely used compared to visual and kinetic references. This indicates that metaphor goes against the notion of domain-specific knowledge; there is no precursor to understanding the metaphors. With a focus on tacit knowledge of the body, the metaphor can be used within different lived experience. In both an educational and design context, metaphor is a way of leveling the existing knowledge of the different parties in order to communicate. This is beneficial for example in design, which is never a truly solo activity; when working with stakeholders, other team members, or potential users with varying background, metaphor is useful in that it does not require a pre-existing understanding. The designer is not in a privileged position when sharing design and its process through metaphor. Consolidating knowledge into a generalised and non-precise communication allows for different perspectives to be understood on a level playing field.

6.4.2 Working Independently of Language

Another critical consideration is in the understanding that metaphor is not believed to be rooted in language, but rather in the body (Gibbs et al., 2004). While metaphor may be expressed often through language (e.g., linguistic metaphor), the underlying schema is rooted in understanding of the body (Lakoff, 1993). Cross-cultural examination of metaphor suggests that patterns of metaphorical understanding are common across languages and that the power of the underlying schema is rooted in embodiment, creating some mutual understanding across human backgrounds (Gibbs et al., 2004). Although all of the voice teachers interviewed instruct primarily in English, the majority of the teachers in this study are dual- or even trilingual. With the understanding that some teachers may use or were taught with metaphors expressed in other languages as students, the interview prompted teachers to provide details on how the metaphor might be expressed in the original language. None of the teachers provided metaphors in another language; when asked about their own vocal training in other languages, the responses were along the lines of *"Well, I guess it's the same [in the native language],"* (P1). P10 elaborated further *"I know how it feels for me when I do this, so I just express that feeling in English."*

This directly demonstrates how tacit knowledge can supersede language; although our communication is often expressed through language, the understanding is rooted somewhere in a wordless knowledge of the body. In a design space, this highlights how metaphor can assist understanding between individuals. The language is less important than the underlying information. To use another metaphor, *the map is not the terrain:* Metaphor is not the experience itself, but rather a tool for articulating and representing that experience in an understandable way. This of course can take the form of language, but representations through other modalities might also be applicable and there are multiple formats which can be used to reach the same end information (Beaudouin-Lafon et al., 2001b; Voorhorst et al., 2000). This also highlights the importance of wordless characterisation of experience in tools such as body mapping. With the internalised awareness of our experience, it is possible for us to divulge details in color, shape, texture, and other non-linguistic based communication (Cochrane et al., 2022). It is therefore advantageous to pay attention to these unifying, cross-cultural schema when we find them; by providing wordless modalities of expression, we open understanding to other humans. This enables us to communicate experience without the need to find the perfect, specific vocabulary, or explain nuanced details.

6.4.3 Providing Ambiguity for Individuality

The use of non-precise communication provides a way of sharing experience. Metaphor relies heavily on the ambiguity of the references, which are intentionally vague or not explained. Ambiguity is used to benefit the student's mapping to inherent knowledge of their body and outside experiences to new tasks. *Metaphors are not concepts*; rather, the underlying schema are mapped to new information and feedback. The teachers do not aim to describe, but rather to approximate sensations based on their own understanding. Often, these metaphors are open-ended and allow the student to interpret them with minimal guidance. Because metaphors do not have meaning in themselves, but rather convey meaning, there is no singular perfect metaphor to use when describing a sensation or experience. Desktop Metaphor aimed to mimic real-world objects to allow office workers to apply their knowledge to computers; however, the *Save* icon remains a floppy disk, with many computer users having grown up in the time after its existence. This prompted a well-known joke about someone having "3D-printed the 'Save' icon" when seeing one in real-life for the first time (Farokhmanesh, 2017) and also suggests that metaphor works not because it represents something physical, but rather that meaning can be interpreted based on use and lived experience.

This is consistent from a phenomenological or hermeneutic perspective. Different teachers, experiencing movement in their individual bodies and with different backgrounds, will all have different perception of performing the same task (Dourish, 1980; Lakoff and Johnson, 1999; Thompson and Varela, 2001). The understanding of own experiences comes from teachers' lived experience and first-person reflection (Höök, 2010; Höök et al., 2018; Neustaedter and Sengers, 2012). Ambiguity in the metaphors provides a means for translation of this understanding to their students. For instance, P4, P5, P11, and P12 all use the word "*spinning*," paired with a circular hand movement to represent movement of the air continuously out of the body over a period of time. The idea of spinning air is not particularly nuanced in its description, but this is a popular metaphor in reference to airflow. I have also heard this term a number of times in my own voice lessons as a student. The lack of detail means that students can apply their own interpretations to match the sensations they experience while performing breathing exercises. All metaphors by definition have this characteristic of subjectivity, similar to the resulting sensory experience.

Using tacit knowledge, this ambiguity becomes personally applicable. Instead of directing a student on what a sensory experience should consist of, the teacher uses the metaphor to induce relevant sensations in their students (Larssen et al., 2007). For instance, the use of the "*marionette*," or the "*puppet on a string*" as a metaphor for posturing is used by P4, P10, P12, P22, and P24. By providing this metaphor, the teacher can help the student to bring awareness to the posture starting from the head down. Although not having experienced the sensation first-hand, the imagination of the lifting of the neck, back, and arms is possible through tacit knowledge. The student is then free to interpret their own sensations of lengthening the spine and aligning their neck, without any other prompting or explanations from the teacher. In this way, the teacher can cause a sensorimotor reaction for the student without needing to verbally describe their own specific internal sensations of alignment. The metaphor is connected with the sensory feedback and becomes a way for the student to understand the new experience.

Ambiguity in this communication, as in design, allows for personal investment and relationships with our interactions (Gaver et al., 2003). We must allow the learner or user to internalise information in their own sensory language by finding balance between describing, showing, and guiding through metaphor. The material taught in the lesson is shaped by the individual and becomes a part of self-understanding (Wakkary et al., 2018). Rather than creating a one-sensation-fits-all model or forcing a sensation from their own perspective, teachers understand that metaphors must be adapted if they do not work effectively. For instance, P4 works as an Estill Master Trainer — Estill voice theory relies heavily on body sensation, incorporation of sensory experience from other life, and kinetic representation of vocal technique. As well, this teacher was quite passionate about their use of kinetic metaphors and representation during the interview process: *"And for me, you know, this like 'sing more or think more blue, think more orange.' It never worked for me. I was like, What is he saying? What does it mean?... You need to practice the sensations and connect them to the physiology..."*. Although the kinetic sensations are the focus of and compose most of the specific pedagogy she uses, she further elaborated that she takes cues off her students and tailors the metaphor around their understanding. When working with students who have difficulty using kinetic representations or are not used to maintaining attention to physical sensations, she will switch to other references provided by the student and negotiate between the two: *"If singing on a pink fluffy cloud does it for you and is your metaphor to trigger you into that [behaviour] every time? Do it."* She uses other trigger words from the students' own recounting of a behaviour, like *"tuck"* for posture alignment, to reference the process back to the student in their own words and then reconnecting it to new kinetic sensations in her own pedagogy. There can be no single-user approach to sensation or communication of experience (Spiel, 2021). By taking into consideration the modality in which we represent these interactions, how people relate to their environment and their body, and how this is impacted by past experience and other societal factors, the student or user ideally will be able to come to their own interpretation, rooted in their own internal awareness of their body.

P17 describes that *"It's a feeling that they can hold on to, and they can return to [it] and they know how to return to [it]... you know where to go to get to that that place... to record this feeling by using your your eyes and using your sense of touch."* We may not have a detailed, verbal way of explaining what we do, but we know how to get there, what it feels like, and how to return to it later through recall of the mental images we form in interaction. As in the case of P10 and their potpourri, the sensation came without understanding why. We see many cases of teachers not exactly understanding why a metaphor works but knowing that it does through the awareness of a sensation. Understanding the differences in individual perception to metaphor can lead to increased communication during design. As done in body mapping and other somatic practices, the study of our reaction to another person's sensory experience can reveal nuances of our own interaction (Núñez-Pacheco and Loke, 2020). Ambiguity could also be used as a tool for discussing meaning in HCI (Gaver et al., 2003) to evoke a variety of perspectives and understanding between individuals in the design process. Through use of ambiguity, we can not only transfer knowledge between different understandings, but create personal connection to this knowledge rooted in individual bodies.

6.4.4 Intentionally Limiting What is Communicated

We see that information is conveyed by making conscious decisions on what components are necessary to create understanding and what should be deliberately withheld to avoid confusion. The vast majority of voice teachers do not have explicit anatomical knowledge about how their body works (Callaghan, 1998). In fact, many teachers do not desire to have more knowledge of voice physiology or teach it to students (Jestley, 2011), seeing it as a distraction and "information overload" (P10) for most students. During the interview, the final question asked teachers whether they felt it was more beneficial to use explicit information or stay strictly in the metaphorical domain, or some combination of both. None of the teachers felt that physiology should be taught to students; although a few (3 out of the 12) had researched it themselves, they all felt the teacher's responsibility to convey the information to the student in a metaphor-based and sensory-focused way, aiming for function within the body itself and providing a way to understand technical terminology through images (Wilson and Keil, 1998). P11 stated that there must be some degree of trust in what we feel: "The body knows how to do this!..." The body knows best, and we do not need to know how everything works anatomically to understand how it should feel. P5 commented that *I think we we do run the risk of saying too much... you spend most of your time trying to dampen it down, [to] get students to not work so hard on trying to understand it all. Just function better.*

We see here that teachers make conscious decisions on what information is necessary in communication. Vocal pedagogy existed before any refined understanding of anatomy and many accomplished singers have no understanding of the physiological processes of creating sound. It is important that we as HCI researchers and designers acknowledge this as well; decisions must be made about what level of understanding is necessary. As well, we do run the risk of over-engineering systems and providing too many details to the end user. In many cases, understanding and association of action and result can be achieved without fully explaining the interaction. To some degree, we must trust the body and the interpretation of others. When we are able to focus on language which conveys only what is necessary, we can provide clear interaction paradigms which do not aim to explain or describe, but rather to guide another individual to an understanding contextualised for their own perception.

6.5 A Model of Metaphor-Based Communication

he examination of metaphors used by voice teachers in their lessons reveals why abstract metaphor is useful and necessary to communicate information, particularly about subjective embodied experience. The results of our thematic analysis demonstrate how mutual understanding of subjective, sensory experience is shared between individuals and further inform ways in which information can be communicated between humans and technological agents. Based on these facets of metaphor learned from the voice teachers, we propose a refined look at how metaphor is negotiated between two human agents:

> Metaphor negotiates information between a communicating agent and a receiving agent using an abstract reference. The communicating agent (e.g., a teacher or designer) encodes information into metaphor, based on their lived experience. The metaphor is then decoded and internalised by the receiving agent (e.g., a student or end-user) into their own lived experience. This process is flexible, iterative, and manipulated by both parties.

The flexible, iterative nature of this model distinguishes it from our traditional views of metaphor in HCI. See also the visual representation in Figure 6.2.

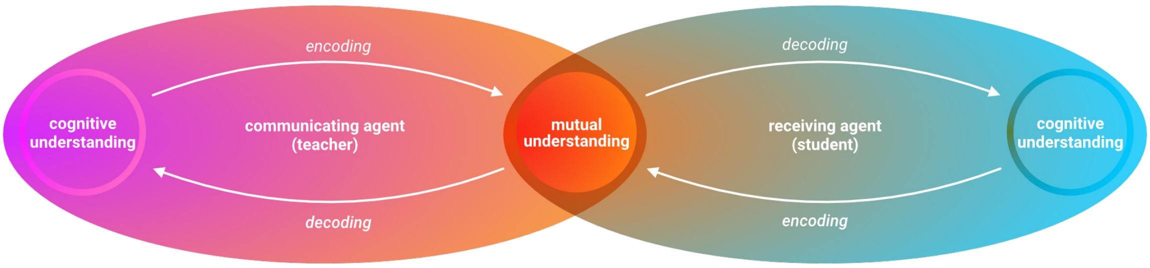


Figure 6.2: A model of metaphor communication, derived from this study, demonstrating how metaphor negotiates information between individual lived experiences. The larger bubbles represent the lived experience of the communicating agent (left, pink) and the receiving agent (right, blue). Metaphor is represented by the arrows, which unite elements of lived experience in mutual understanding by the two parties.

In human-to-human communication, information is modulated between both the communicating and receiving agents; both teacher and student must adapt and revise the metaphor together to achieve mutual understanding. Although the responsibility of the adaptation is perhaps more on the communicating agent — the teacher, in this case — they are able to provide flexibility and

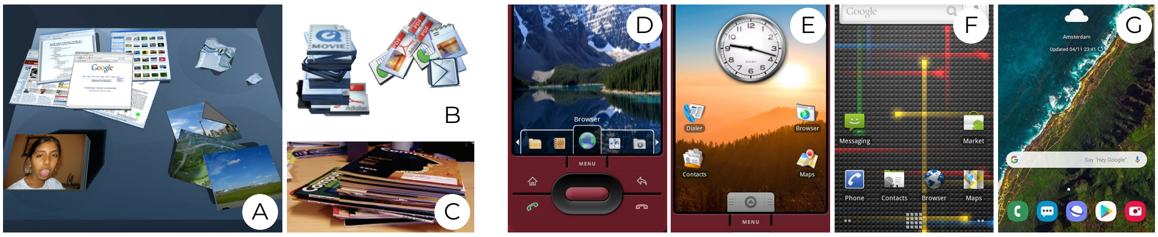


Figure 6.3: Some designs, both inside and outside of academia, expand on the Desktop Metaphor (left); for example, BumpTop's virtual 3D desktops (a) and literal stacking and piling metaphors (b, c) (Agarawala and Balakrishnan, 2006). Android (right) used metaphors much more loosely: The app menu was originally accessed through a physical button (d), then a drawer metaphor (e), and then by clicking an iconic representation of the menu (f), before abandoning visual metaphor all together and using a swipe gesture instead (g).

update their reference to reach shared understanding. The metaphor is grounded in pre-linguistic tacit knowledge and leverages ambiguity to help the decoding of the reference. In fact, we see the metaphors themselves are very specific, providing evocative references, while being unrelated to the task at hand. The metaphors convey information by relying on embodied experience without the need for domain-specific knowledge or specific linguistic representations; rather, the communicating agent (e.g., a teacher or a designer) focuses communicating core components and uses ambiguity in the reference to translate their refined knowledge of the mechanics into the basis of shared knowledge and individual lived experience. This allows the communicating agent to provide a reference which can be understood by the receiving agent, who can further refine its exact meaning within their own understanding. The communicating agent does not assume the interpretation will be the same, focusing on flexible presentation of only what is essential to achieve understanding between the two's lived experiences.

Human agents can together adapt and revise this model, with the metaphor acting as a conduit for exchanging knowledge. Currently, technological communicating agents struggle to produce this flexibility and reactivity, using non-dynamic references and hard-coded metaphor. These are dictated by the designer, who must make a decision on the "best" way to convey this information. The onus is on the receiving agent to interpret and decode this reference appropriately. This means the computer's expression is, in some way, assumed as "truth" that the user should know. Comparatively, as seen with the adaptation of metaphor by P4 in Section 6.4.3, the onus is rather on the communicating agent to put together the receiving agent's understanding and come up with cues and references which function well. Although objective information might be communicated effectively, as seen in icons and Desktop Metaphor, subjective information is lost in assuming individual experience will lead to the same interpretation. Rigid conceptual metaphors have a limitation in how far the concept can be taken (Smith et al., 1985). Data represented in this way might fall into some of the pitfalls of the Quantified Self paradigm, leading human receiving agents to misinterpret aspects about themselves or allow the technological communicating agent to overgeneralise their experience (Prpa et al., 2020).

6.5.1 Expanding the Concept of Metaphor in HCI

If we think of metaphor in its contemporary sense as a flexible mapping of information between one modality and another, we can explore a variety of applications in a new light. Metaphor works not because it is a depiction of reality, but because of lived experience. This leads to an interesting challenge in HCI; while being ambiguous, we must also specifically appeal to individual lived experience. Traditional models of human input of parameter space resulting in machine output as calculation and "truth" are changing rapidly. In cases of sensory experience, we no longer assume there is a ground truth between different bodies. As well, technological agency means that the designer may not always be dictating the metaphor or information communication. We see as well a switch in the roles of human and technology agents in metaphor communication, for instance in the growing popularity of generative art through tools like Dall-E and Midjourney. In these cases, AI must interpret often abstract and ambiguous human input in a non-objective way and come up with a similarly ambiguous result. This, vocal pedagogy metaphor, is a bi-directional and iterative interaction wherein the abstract representation is exchanged and updated until the computer and human agree on the result.

Existing practices such as body mapping also depict metaphor's role in helping us to understand aspects of subjective experience. The next step is in conveying this information in a relatable way. Metaphor can be used to explore this negotiation of mutual understanding. This has been the focus of recent work in sensory translation (Tholander et al., 2008; Wirfs-Brock et al., 2022) and data narratives. For instance, research has focused on how audiences can be supported in learning and responding to data in expressed in narrative sonification (Wirfs-Brock et al., 2021) and visualisation (Dove and Jones, 2012), which act as metaphor to express data. Material experiences have focused on similar interactions, for instance between Friske et al.'s "maker" and "interpreter" in interpreting and re-making personal data (Friske et al., 2020); this dynamic might be thought of as a specific instance of physical and sonic materials as metaphor, being used to negotiate understanding between the maker as the communicating agent and the interpreter as the receiving agent. This practice further fits into this revised model of metaphor, allowing narratives based in individual experiences to entangle and co-exist. Taking additional steps to focus on conceptual mappings and using ambiguity and non-domain specific knowledge, as done by the teachers here, can potentially increase understanding in these instances.

With this model, we can return to address our understanding of metaphor in HCI. The proposed model answers the *why* components of the functionality of contemporary metaphors, as outlined by Lakoff and Johnson and Barr et al. In line with the existing taxonomy, we can address strategies to create and evaluate metaphor in HCI, regardless of their specific type. This model also provides a missing connection to several existing taxonomies of metaphor; although we do not endeavor to explore or unify the existing classifications of metaphor and theories about their categorisation, our contribution to this space is a general model of the functionality of metaphor and how information transfer for metaphor as a whole. Together, the research space then outlines *what* constitutes a metaphor and the kinds of metaphor entailments and mappings that can be used, and now also *how* and *why* metaphors are able to provide these connections.

In terms of practical implementation, this model aligns well with other existing taxonomies of data representation and helps to describe some strategies suggested for creating metaphors in design. In their models of data visualisation, Zhao and Vande Moere include three key components for implementation: 1) the metaphor must be easy to identify, 2) the metaphor must have both 'motor' and 'cultural' affordance which provides structure for interaction, and 3) the metaphor must be intuitive and function without prior training (Zhao and Moere, 2008). Our proposed model of metaphor-based information transfer captures these components as they function together: the metaphor should not require any pre-existing domain-specific knowledge because the schema it relies on exist already in embodied understanding, shaped by cultural and functional experience from other instances in life. Our model of metaphor-based information transfer also aligns with Nesbitt's MS-Taxonomy, demonstrating that there are shared mapping characteristics independent

Figure 6.4: The Heart Sounds Bench (a,c) allows one or two people to sit together (b,c) and, with the help of stethoscopes connected to the inside of the bench (d), listen to a sonification of their heartbeats (Howell et al., 2019). Photos used with permission of Howell et al.

of modality(Nesbitt, 2006); being based on cognitive schema, the resulting metaphor can be adapted and reused, and indeed re-articulated through different modalities as needed for communication between agents with different lived experience.

6.5.2 Operationalising Metaphor in Design

We wish to invite designers and researchers to take a step back and re-evaluate what we believe to know about metaphor. In doing so, we provide a theoretical contribution (Whetten, 1989) by updating the accepted relationship between agents and metaphors (c.f., (Whetten, 1989)), and by suggesting that Metaphor is a continuous, mutually iterative, and reflective process, rather than a unidirectional and discrete phenomena, as suggested, for example, by Dunbar-Wells in the vocal pedagogy context (Dunbar-Wells, 1997).

Evaluating communication strategies in traditional interfaces

These updated, evolving relationships between agents and metaphor in the proposed model can help us think about metaphor use in traditional HCI. To highlight this, let us reflect back on the development of Android phones and the idea of 3D naturalistic desktop interfaces: Both emerged in the mid to late 2000s, and, while Android phones are still with us, the idea of a 3D virtual desktop operating system appears somewhat absurd from today's point of view. Focusing only on the use of metaphor, this might be surprising, as the 3D desktop is a direct continuation of an existing, successful metaphor with a clear, easy-to-understand analogy (Agarawala and Balakrishnan, 2006; Chapuis and Roussel, 2005). Android phones, on the other hand, do not use such a clear metaphor; in fact, over the multiple iterations of the Android software, the metaphors used have often radically changed (Amadeo, 2016).

Both observations can be explained with our model. For example, 3D and physics-based desktop systems attempted to improve interaction by adding naturalistic detail (Figure 6.3a) and leveraging specific knowledge of the world, such as the behavior of stacks of documents (Figure 6.3b and c). As we have shown, this added fidelity and incorporation of domain specific knowledge is not necessarily something that improves the usage of metaphor. By removing the ambiguity of less naturalistic implementations present in desktop methods, the ambiguity required for individual sense-making is lost. Therefore, looking at the usage of metaphor alone, our model suggests why these trends did not catch on.

Looking at the Android UI — for instance, the metaphor for accessing the app drawer — we see that, over time, it has used a physical button (Figure 6.3d), a literal drawer metaphor requiring sliding to open (Figure 6.3e), a virtual button visualised as an iconic representation of the app drawer (Figure 6.3f), or no visual metaphor at all (Figure 6.3g). Without further reflection, such inconsistent metaphors might appear problematic. However, our proposed model suggests that a metaphor's content is secondary to its function of fostering mutual understanding. Designers are free to adapt the metaphor for accessing the app drawer as the underlying understanding (the existence of an app drawer that can be accessed from the home screen) is never disrupted. This same underlying understanding makes it effortless for users to adapt to changing metaphors as Android versions change. This example also highlights that the iterative process described in the proposed model can exist on multiple time-scales, in this case spanning more than a decade.

Evaluating design probes with a metaphor-based communication model

In addition to re-examining the design process and metaphor communication, this model helps to understand existing designs. We believe this model will be especially beneficial to those focused on subjective experience and sensory-emotional communication. Designers and researchers may find it helpful to apply this new model of metaphor-based communication when examining such probes and the design process:

Howell et al.'s Heart Sounds Bench allows people to sit together and, with the help of stethoscopes connected to the inside of the bench, listen to the sound of their heartbeats (Figure 6.4). Their unfiltered and noisy heartbeat sounds can be heard by the others sitting on and around the bench, an auditory metaphor of the heartbeat representing life and connection between people (Howell et al., 2019).

Aligned with our model of metaphor communication, we see how the sonification *relies on embodied experience*: the visceral discomfort that can arise when hearing one's own heart beat and vulnerability in sharing an intimate source of data with others. As well, there are cultural and emotional components — love, excitement, nervousness — associated with heartbeat sounds. The sonification *works independently of language*, relying on conceptual understanding of the heartbeat as a source of life energy, and *communicates core components while limiting unnecessary detail.* Other aspects of the heartbeat, such as its rate, might have provided objective information about someone on the bench; rather, the sonification *uses intentional ambiguity to enable individual sense-making* of that person and their emotional state.

Such designs are successful at communicating subjective experience about the self and soma. In this very brief examination, we have outlined how aspects of the design might be using metaphorical affordances to convey the goals of the designers. Often, decisions made in design are made based on implicit understanding and feeling; using this model in Research through Design practice helps examine these communication practices and underpin their substance and the *why* in interaction approaches. In this way, this model helps to provide a language for describing communication strategies and how information is negotiated between agents. Approaches and underlying schema and concepts can then be identified and abstracted beyond their individual instances for use by other designers. The model as used here provides a tool to investigate strong concepts within intermediate-level knowledge(Höök and Löwgren, 2012). Future work examining data representation strategies and how they evolve over the design process will likely add to the basis of strong concepts which can be used by other designers. Focus on the interactive aspects of designs and *why* designs effectively communicate can help to further ground the specifics of individual experience, knowledge, and understanding within higher level theory of communication.

6.5.3 When Metaphorical Communication Fails

If metaphor works by evoking a sensation, there must be instances where the representation does not work in the way it should, which should also be discussed before concluding this chapter. There is a risk that the personal experience is not compatible with the representation or that the metaphor is too subjective to be relatable. We see that there is no significant correlation between the modality of metaphors used and the imagery skills of the teachers. In fact, some of the opposite is seen; for example, Pt 4's self-assessed kinetic imagery score was the lowest of the group, under the midpoint of the imagery scale, at 3.25, suggesting that they may not be able to use kinetic imagery easily (or at least may not feel they are able to use it easily enough to self-assess a high skill level). This particular teacher is a well-known contemporary teacher and a certified Estill Master Trainer. Estill voice theory relies heavily on body sensation, incorporation of sensory experience from other life, and kinetic representation of vocal technique. As well, this teacher was quite passionate about their use of kinetic metaphors and representation during the interview process:

> And for me, you know, this like 'sing more, think more blue, think more orange.' It never worked for me. I was like, What is he saying? What does it mean?... You need to practice the sensations and connect them to the physiology... because metaphor is one thing, but then like, really going in, a feeling can also like, facilitate a different certain, you know, technical thing."

In this participant, we see a common thread that may explain more of the emotional side of the experiential human interaction on which metaphor and abstract language rely and the context in which these representations exist: teachers will use imagery which aligns more closely with their personal beliefs and feelings. This particular teacher had, as a student, wanted to know more about their body in a logical explanation, but instead was offered very abstract visual representations, such as the colour metaphors described above. The translation from the teacher's abstract visual metaphor (color descriptions of sound timbre are quite common Aesthetic Visuals) did not elicit the appropriate kind of response or create an understanding for Pt 4, which caused a great deal of frustration. Although this participant wanted more concrete, literal explanations of what was going on, their teacher continued to attempt what seems to be a largely visual approach:

> "Definitely no physiology. Which I have now like, Why didn't anyone tell me?... I've realised maybe later that I learned, I wanted to know what's going on. That's how I am, you know... I want to know, what is what are we trying to do? It's so much easier for me to get what it is. No, it was only like, either them trying to show me or, you know, telling me 'a little bit more this, a little bit more that'."

It is unclear where the misunderstanding happened—these kinds of very abstract metaphor are often highly misunderstood in vocal pedagogy and are a great source of debate against using metaphor at all in teaching (Miller, 1996, 1998a). What is apparent is that the teacher reacted to this misunderstanding by completely devoting themselves to learning and teaching in a kinetic sense. There is an enormous deal of frustration discussed by the teachers about metaphors they were unable to understand as students. We assume that, within our own bodies, we should be able to understand and control them the way that we want. This is often a point of frustration in learning any new skill: we can imagine doing something but can't seem to quite get it at first, at least not until we learn the sensations and connect them to our movement and behaviour (Höök, 2010).

There are other strong negative feelings besides frustration for students when there is misunderstanding and misinterpretation of one's own body. Nine out of the 12 teachers (Pts 01, 05, 08, 10, 11, 17, 18, 22, and 24) at some point during the interview described a negative experience they had with a voice teacher. The participants describe being tense, worried, and even scared of their own teachers or their lesson environment. The experiences were remembered in detail years later—this was true even for Participant 22, who still remembered the confusion and the feeling of fighting against their own body some 50 years later. The voice is a very intimate interaction tool between a singer and their own body, one which is also a large part of personal identity (O'Bryan, 2015). When a student finds it difficult to understand or relate to their own sensations based on what their teachers tell them during the lesson, this can result in frustration. We see that the teachers use their own experiences of negative feelings to make decisions about the way they teach:

> Pt 10: "I think singing is different from any other musical instrument, because it's, you're exposed in a way you're not exposed when you're playing an instrument, you have nothing to hide behind."

> Pt 11: "After the first session with him, I thought, oh my god, you know, I'm, I'm going to have to stop teaching. You know, I've got to get come to grips with this... when I came home and cried. But within three weeks of that my whole technique had changed. It was it was phenomenal. And I had better breath and far less tension... And that is how I teach now."

6.5.4 When Subjectivity & Individuality are Ignored

Beyond the emotional negatives of metaphors which fail to translate, there is potential bodily harm in the attempt to force a sensation or experience. Despite singing being a largely sound-focused practice, there is a very large disparity between the types of imagery used by the teachers in this study: compared with 270 kinetic-based images and 237 visual-based images, there were only 54 auditory metaphors provided. Although descriptions of sound qualities and creating aural familiarity are helpful in sound production, they are not as prominent in the teaching of vocal fundamentals as visual and kinetic metaphors. This may be largely explained by the general purpose of the of the non-auditory metaphors in externalising internal movement and sensations. Auditory images may be more difficult for a teacher to convey as they usually require an external reference to be given first in the example of a sound source. Sometimes, teachers have students experiment with the sounds they are capable of creating. This creates action-sound associations and allows students to develop auditory and kinetic imagery to produce a sound. Beyond this, the auditory metaphors lean too heavily into mimicry or modelling behaviour, which diminishes the body's role in vocalisation and multi-modal interaction in favour of aural interaction (Eidsheim, 2015, 2017). It is possible that the group of teachers did not list many auditory references because voice pedagogy as a whole is moving away from sound modelling due to the risk of physical harm to the student.

Sound modelling often sees teachers asking students to recreate their own tonal qualities or volume, or that of other notable vocalists. There is an obvious problem in this as a teaching method: teachers are often older and more experienced, with fully formed vocal musculature. Often, teachers are a different sex from their students, meaning their interaction with their vocal physiology is completely different. By asking students to recreate a sound from a body which is different from their own, the risk for damage to the body, especially the vocal folds, is extremely increased. Pushing students to perform in ways which are not healthy or not physiologically possible can lead to permanent damage — something that many of the teachers interviewed had experienced firsthand:

> Pt 01: "My partner had like had polyps because of this, because of a teacher saying, 'push, push, push push'... a girl in my master's program, she had scarring because of this."

> Pt 05: “I had to unlearn everything you taught me, unlearn all of that. Because I modelled my voice. And of course, now I experienced that with my students as elitism.”

> Pt 24: “Adult voices don’t really settle until ages 25 to 30... I feel like college programs kind of expect your voice to be ready at age like 19.”

The comment from Pt 05 (also mentioned in the interview very similarly by Pt 10) regarding elitism in sound modelling describes a common problem in interaction in general: there is an expectation that everyone will be able to interact and react in the same way or achieve some kind of “ideal” result from this interaction (Spiel, 2021). This ideal voice quality, particularly in opera and in musical theatre, is often extremely Eurocentric:

> Pt 05: “When they come to me and they say, I want to sound [like this]... it’s really quite a dangerous fixation for students, young students that are impressionable... Who are you modeling your sound on? What have you got in your head? Is it useful? Or is it you modeling on something that’s not useful? For example, you know, if you’re a black [female] singer, and you’re listening to the voice in your head that has been taught from a white male voice, is that helpful to you?”

This aspiration to sound like someone else leads many young singers to attempt to sing in ranges outside what their bodies will allow, push or strain to achieve greater volume at pitches where their voice is weaker, and lead many singers to quit their practice because they do not have an “ideal” voice for the style. The voice is a unique part of individual physiology, and there can be no universal interaction paradigm. When bodies are involved, we must assume that the sensory experience is different for each person. The interviews reveal that the voice teachers are very passionate about moving away from these practice, which may explain why there is a marked difference in the number of auditory references given compared to visual and kinetic metaphors. As the latter two modalities require the teacher to summarise or approximate their sensations, it is not expected for students to recreate the teacher’s experience; rather, metaphors allow for the information to become internalised within the individual body.

6.6 Conclusion

We present a study of metaphor as it is used to negotiate understanding between communicating and receiving agents — specifically, voice teachers and their students in the context of the voice lesson. From this work with voice teachers, we see how vocal pedagogy has evolved to teach about a practice and experience which goes beyond linguistic-based understanding. Singing requires us to use parts of our body which we largely do not see or understand except through internal sensations and our tacit knowledge of our body. However, by using existing relationships and experiences from other life, voice teachers are able to elicit imagery and sensory experiences for their students.

We present an overview of strategies for metaphorical communication in the voice lesson as well as a thematic analysis to demonstrate that metaphor functions not because of its grounding in reality, but rather in embodied experience. Due to its flexibility, ambiguity, and non-domain specific context, metaphor is able to unite individual lived experience. Based on these findings, we demonstrate a model of metaphor which addresses the history of HCI metaphors and brings them into modern HCI perspectives of individuality and plurality of experience. Through this model, we demonstrate why metaphor works and how its role in human-to-human interaction can inform better communication and understanding between human and technological agents.

Chapter 7

Surface Electromyography for Vocal Interaction

Externalising the Movement of the Laryngeal Muscles

As first discussed in Chapter 2, Section 2.4.2, digital musical interfaces based on the voice have typically used indirect control, in which features extracted from audio signals control the parameters of sound generation, for example in audio to MIDI controllers. This is likely because of the very features that make vocal interaction difficult to articulate or understand: the muscles are hidden within the body and are not well understood by an average person (or even average singer) and audience members engage through the voice by listening, rather than feeling as the singer does. By contrast, focusing on the musculature of the singing voice allows direct muscular control, or alternatively, combined direct and indirect control in an augmented vocal instrument. In this way we aimed to both preserve the intimate relationship a vocalist has with their instrument and while expanding control over it and its sonic capabilities. Subsequently, we developed a new method of direct control from physiological aspects of singing through surface electromyography. This has been implemented in a dedicated PCB and in a wearable for vocal interaction to study the vocalist-voice relationship and provide new creative outlets.

Portions of this chapter have been published in:

Courtney N. Reed, Sophie Skach, Paul Strohmeier, and Andrew P. McPherson. 2022. Singing Knit: Soft Knit Biosensing for Augmenting Vocal Performances. In Augmented Humans 2022 (AHs '22), March 13–15, 2022, Kashiwa, Chiba, Japan. ACM, New York, NY, USA, 20 pages. DOI: 10.1145/3519391.3519412

Courtney N. Reed and Andrew P. McPherson. 2020. Surface Electromyography for Direct Vocal Control. Proceedings of the International Conference on New Interfaces for Musical Expression (NIME '20), July 21-25, 2020, Birmingham, UK, pp. 458–463. DOI: 10.5281/zenodo.4813475

7.1 System Design

sEMG is the process of measuring electrical neuron activation of the muscles across the surface of the skin in a non-invasive way. sEMG as a biosignal is useful in exploring the vocal mechanism without examining vocal audio signal, thus addressing the control gap. The following section provides a method for sEMG in measuring aspects of vocal performance for direct control. I will first discuss the design process, sEMG signal acquisition, filtering, and integration with the Bela platform (McPherson, 2017), before demonstrating how sEMG can be used to directly measure both vocalised and subvocalised singing.

7.1.1 Sensing

The system uses three 10 mm electrodes — during prototyping and demoing of the system, I used reusable gold-plated silver cup electrodes (Medimaxtech, New Malden, UK), each with a 120 cm wired connection. The end- and mid-muscles electrodes are placed across the muscle being sensed, while a reference electrode is placed on nearby non-muscular tissue, typically a bony or cartilaginous part of the body (Figure 7.1). This third electrode provides reference to a stable electrical conductance on the body — on a non-muscular tissue, this electrode provides a control for the other two. Processing involves two stages, beginning with an analogue preamplification circuit to acquire electrode signal. The circuit is powered by two 9 V batteries, allowing for external powering and portability, as well as noise reduction from grid power sources. During its use in this thesis, the circuit was prototyped and used in tandem with the Bela board (McPherson and Zappi, 2015), an open-source embedded computing platform which allows for ultra-low latency for signal processing (Figure 7.2). We use the Bela for its low latency and to make use of the audio processing which can be done on-board.

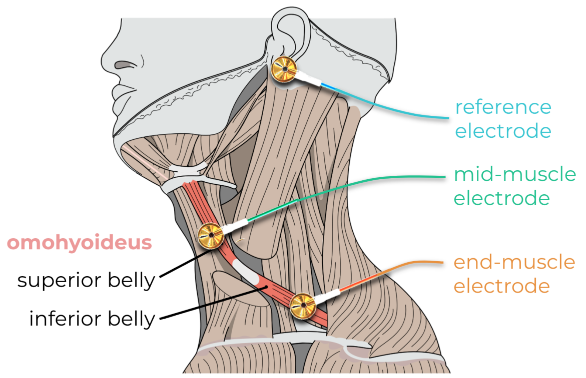


Figure 7.1: Placement of the three electrodes for sensing activation of the omohyoid.[1]

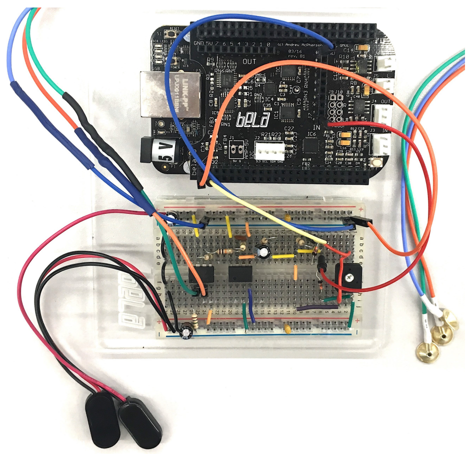


Figure 7.2: Prototyping the initial circuit for laryngeal sEMG with the Bela.

[1] Image adapted from *Musculi coli*, Olek Remesz (CC Attribution-ShareAlike); electrode images: Pulse Medical

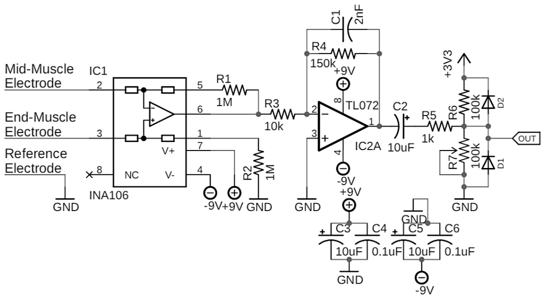


Figure 7.3: Version 1: The original circuit for sEMG signal acquisition and preamplifier schematic using three electrodes.

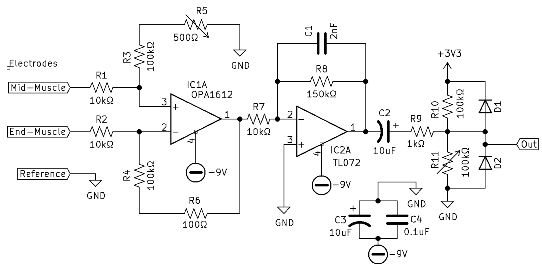


Figure 7.4: Version 2: The INA106 is also replaced with the OPA1612 and trimmable resistances for further noise reduction.

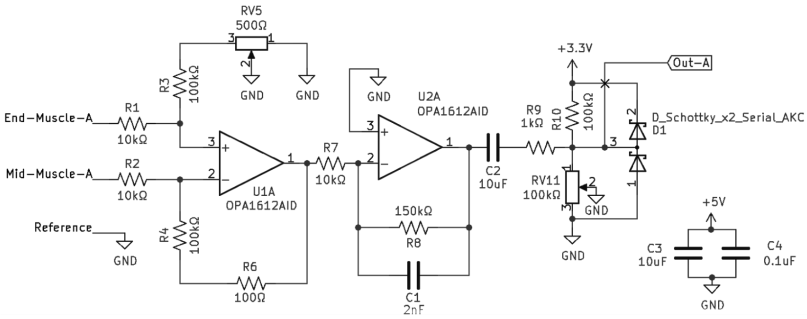


Figure 7.5: Version 3: The currently used circuit used for the VoxEMG, which also replaced the TL072 with the OPA1612. It is important in reproduction of this circuit that high-precision (ideally 0.1%) metal film resistors be used for R1-R4 to ensure the resistance in series with both electrodes is the exactly the same.

7.2 Signal Processing

The signal acquisition and processing stage of this design is based on the open-source Gold Package EMG Circuit v7.1 (Advancer Technologies).[2] In all versions of the circuit, a differential amplifier IC is used to amplify any small voltage difference between the two muscle electrodes. These areas, which otherwise have equal electrical potential, will differ as the muscle is activated and contracts. Differential amplification also reduces noisiness through common mode rejection. In the initial design of the circuit (Version 1, Figure 7.3), the gain of this stage is set to 110. The signal is then passed to an inverting amplifier and a first-order low-pass filter with a 530.5 Hz cutoff to restrict the signal to an appropriate range for sEMG (below 500 Hz, typically most usable frequencies between 50-150 Hz). The signal is then rectified, converted to DC for use by a microcontroller, and further amplified with gain set via a trim-pot by the user. Note that schematics here depict the acquisition for a single-muscle setup; in order to capture two muscles at once, the same signal path can be simply duplicated using the other side of the two op-amps used here. If a second muscle is being measured, the existing reference electrode can be used.

[2] https://advancertechnologies.com/p/muscle-sensor-emg-circuit-kit-bronze.html

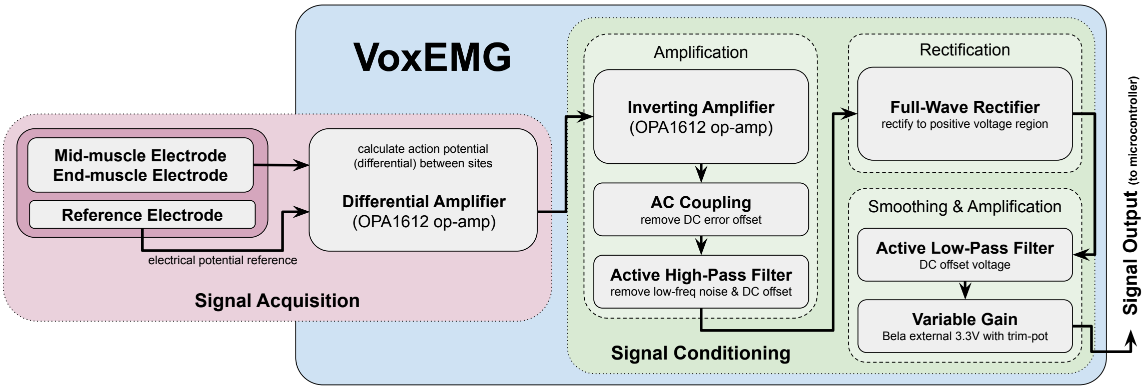


Figure 7.6: Signal flow through the VoxEMG board (Version 3 of the vocal sEMG circuit designed in this thesis).

Further iterations of the circuit improved the signal-to-noise ratio by using different ICs. This was done by switching both ICs to the Texas Instruments OPA1612, which was chosen for its extremely low voltage noise density $(1.1nV/\sqrt{(Hz})$, over subsequent iterations to increase the signal-to-noise ratio when dealing with lower voltages: In the second iteration (Version 2, Figure 7.4), the Common Mode Rejection (CMR) of the circuit averages approximately −37.96 dB across the frequency band for sEMG signals. In the current version (Version 3, Figure 7.5), this is further reduced to −54.06 dB. A variable voltage divider using external 3.3 V power from Bela (or a different microcontroller of choice) is used for gain control and prepares the raw signal with DC offset voltage for use with the platform. I have further designed a dedicated PCB for the final V3 circuit, the VoxEMG board, for incorporation of the circuit into wearable devices; see Section 7.3 for further detail.

Figure 7.6 depicts the signal flow and processing in the V3 circuit, along with processing done digitally, to demonstrate how the circuit can be used to capture vocal sEMG. The output voltage of the VoxEMG board is further processed on-board within the Bela IDE. First, the DC offset is removed after it is passed to the Bela using a high-pass DC-blocking filter. Then, the signal is further filtered digitally with a notch filter to remove any remaining power-line interference.[3] The resulting sEMG voltage signal can then be collected or used in data sonification, as done through the studies with vocalists in this study. See further Section 8.3 for an example of this data sonification.

7.2.1 Proof-of-Concept

The goal of the design was examines the viability of capturing vocalised and subvocalised sEMG using the amplification circuit derived from the Advancer Technologies design and the physical components used for AlterEgo (Kapur et al., 2018), which was the most recent example of vocal and subvocal sEMG sensing at the time of the design. We determined viability through a directed task to produce both vocalised and subvocalised signals. An example of the potential use of this controller for direct control with vocal musculature can be seen in the movement of the omohyoideus muscle when singing descending pitches. The omohyoid is an extrinsic laryngeal muscle which lowers the larynx (Figure 7.1); the main function of this muscle is thus to generate lower fundamental

[3] Bela processing code can be found in the VoxEMG repo: https://github.com/courtcourtaney/voxEMG/tree/master/examples/basic-input

frequencies (Hardcastle, 1976). This muscle passes beneath the sternocleidomastoideus, one of the neck muscles. This provides a concise example of the functionality of metaphorical teaching used in vocal pedagogy without understanding of detailed physiology: vocalists are trained to keep the chin down and the neck in a loose posturing, thus relaxing the neck and "making room for the breath." Such behaviour would keep pressure off the sternocleidomastoideus and allow for the movement of the omohyoideus and other surrounding muscles without increased strain or fatigue resulting from tension in the neck.

In a short self-study using V1 of the circuit (Figure 7.3), I performed a series of short chromatic exercises in the lower-register of my voice. The mid-muscle electrode was placed on the upper portion of my right omohyoid (the superior belly) in the middle of my neck adjacent to the thyroid cartilage, while the end-muscle electrode is placed at the inferior belly close to my scapula (Figure 7.1). The reference electrode was placed on my right earlobe. A conductive adhesive paste (Ten20 Conductive Paste, Weaver and Company, Aurora, CO, USA) was used to secure the cup electrodes and reduce skin impedance; the electrodes were further secured using an adhesive non-woven fabric tape (Hypafix, BSN Medical GmbH, Hamburg, Germany) which was placed over the electrodes to ensure they are in close contact with the skin and do not move around. I then sang chromatics descending from G3 to observe the activation of the omohyoid. I am a mezzo-soprano — the typical mezzo range extends to F3 and this is very near the bottom of my range, meaning my omohyoid would be lowering my larynx as much as possible; although not very musical, the functionality of the sEMG system can be seen in this extreme. In the exercise here, I took a breath before the first two notes but not before the third to observe any contrasts in sEMG as a result of breathing. To synchronise my activity with the data, I also indicated the start and end of note events through pressing and releasing a button connected to Bela. The button presses were timestamped for synchronisation and confirmation of voltages were observed in the GUI during the singing exercise.

The neural activation and contraction of the omohyoid in singing the first three semitones of this downward chromatic sequence can be observed visually in the signal (Figure 7.7). Markers indicate points where different actions occurred in the signal recording; the different pitches are noted above. sEMG signal is not continuous, but rather the sum of discrete neuron impulses (Tanaka and Ortiz, 2017) which can be seen in the voltage spikes during this reading. The inhalation taken before G3 is sung is first visible; with each successive downward movement, the amplitude of the signal voltage increases—this is perhaps due to the greater downward laryngeal movement needed to achieve lower pitches at the bottom end of the voice range. The two inhalations are also visible in this case, as a slight lowering of the larynx also occurs during deep breathing where more space is created in the vocal tract (Hardcastle, 1976). As seen when working with the vocal teachers, this type of breathing is a core facet of vocal pedagogy and allows the vocalist to shape the vocal tract for rounded, warm tones and provide airflow support for vocal fold vibration.

The same exercise was repeated in a mental rehearsal to determine the presence of subvocalisation of the same muscular activation (Figure 7.8). Imagining and executing an activity will result in similar neural activation; in this case, mental rehearsal of a vocal exercise will excite the parts of the brain necessary to perform that exercise (Halpern and Zatorre, 1999; Kleber et al., 2007), resulting in low-level activation which we expect will be able to be detected by the electrodes, as done with AlterEgo (Kapur et al., 2018). Breathing was repeated following the same pattern as done in the previous vocalised trial.

Although it is clear that the subvocal signal has smaller amplitude and lies more closely in the range of electrical noise in the system, the same markers can still be seen. The difference between the notes and greater downward motion of the larynx is less visible, perhaps a muscular distinction which is lost in mental rehearsal. However, despite not actually producing sound, the gesture required for lowering the larynx can still be observed; thus, we see how sEMG measurements of vocal musculature

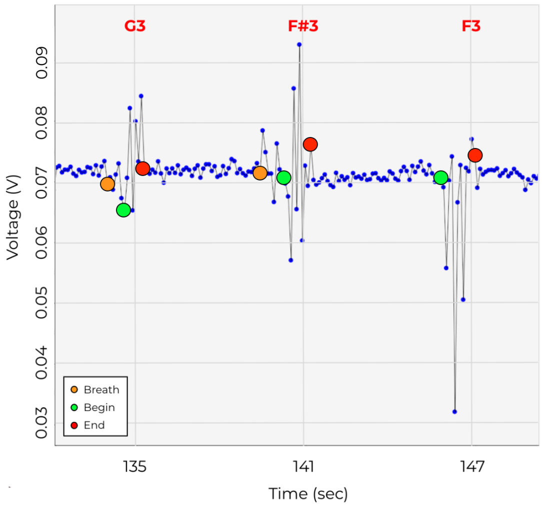


Figure 7.7: Muscular activation during singing.

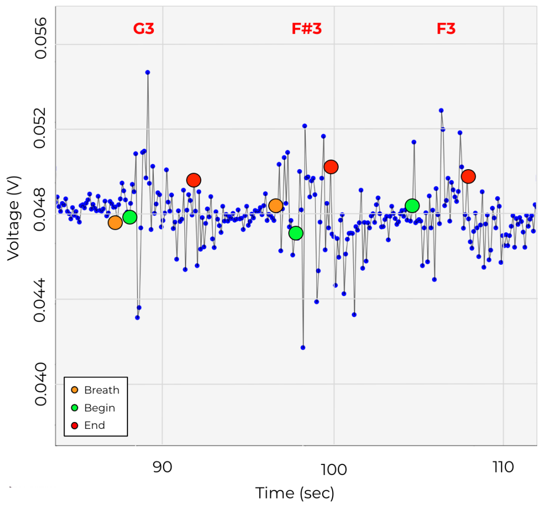


Figure 7.8: Muscular activation during subvocalisation.

can provide an audio-independent method of direct control.

Compared to conventional audio analysis, sEMG provides a more introspective look at what the vocalist is doing or intending to do in their performance and the beginnings of a gestural vocabulary, much like we see in other instruments. In areas where audio analysis of voice may struggle in accuracy or ambiguity, such as in pitch recognition, sEMG data can provide means of augmented support, as done with instruments such as the ESitar (Kapur et al., 2004). The presence of sEMG signal during mental rehearsal and subvocalised singing on its own provides a basis for a wealth of studies regarding musical imagery and rehearsal and learning practices used by vocalists. This system has been further employed to gather and sonify the laryngeal muscles' activation while singing — Chapter 8 outlines autoethnographic study of my own use with the device through its development in a long-term interaction and relationship with my sonified laryngeal muscles, and how this changed my own outlook on my performance techniques and body. Further, in Chapter 9, other singers lived with the system for an extended period and documented their evolving relationship with it in both their typical genres and as a tool for exploring fundamental vocal technique.

7.3 Designing a Vocal Wearable

The VoxEMG was used to incorporate the vocal sEMG circuit into a wearable device, the Singing Knit. The goal of this design was to create a system which could be employed in a real-world setting, adhering to the demands of musical performance. Rather than attaching the three electrodes on the muscle site with conductive paste and tape (Figure 7.9) for each use of the system, we wanted to create a more permanent wearable setup to help address the issues with the systems' use. Namely, we wanted to provide a wearer with the ability to quickly setup the electrodes without needing to locate the muscles, provide flexibility of placement, and to resolve comfort and aesthetic issues within the original system. In order to create such a wearable for vocal sEMG, we took inspiration from similar multi-electrode systems developed within HCI.

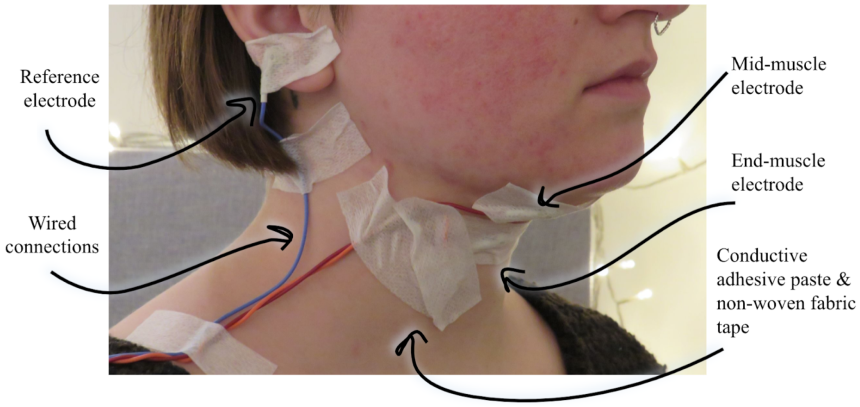


Figure 7.9: A photo of myself wearing the typical sEMG setup, with lots of fabric tape being used (even here, one can see the fabric tape by the labeled mid-muscle electrode is starting to peel off).

Multi-electrode systems have become common platforms for HCI (Duente et al., 2017; Knibbe et al., 2021). Biosignals collected at multiple body locations are used in bio-feedback systems for behavior regulation (Kim et al., 2018) and improved understanding of our own bodies (Karolus et al., 2021a). They are also used to reconstruct gestures (McIntosh et al., 2016) and poses (Knibbe et al., 2021) provide a supplementary channel to influence artistic performances (Kilian et al., 2021; Reed and McPherson, 2021; Stelarc, 2020). Multi-electrode muscle stimulation can create complex physical motion as output in interactive systems (Lopes et al., 2015).

As with the vocal sEMG system designed during this thesis, other multi-electrode systems face many of the same problems when applied in real world settings. Securing multiple electrodes to the body often results in skin irritations and general discomfort (Knibbe et al., 2021; Meltzner et al., 2008).[4] As users move, electrodes may come loose or cables might be caught, leading to reduced functionality (Knibbe et al., 2021). Deploying such systems is difficult, often requiring help from an additional person to be able to wear them correctly (Hassib et al., 2017; Knibbe et al., 2021). Finally, many of these systems do not have an aesthetic which might allow them to be worn without the user drawing attention to themselves (Knibbe et al., 2021; Koelle et al., 2019a,b).

7.3.1 Multi-electrode Wearables

The oldest use of multi-electrode systems is most likely in surface electromyography. sEMG is used to measure neural muscular activations from the surface of the skin using electrodes. This has found a broad range of applications related to HCI (Karolus et al., 2021b). For example, bio-feedback systems have been designed to support physical therapy (Igarashi et al., 2010; Lim et al., 2020; Tsubouchi and Suzuki, 2010) or to simply provide users with additional insights on the actions they are performing (Karolus et al., 2021a). Other uses include sEMG for identifying gestures (McIntosh et al., 2016), for example to be used as explicit computer input (Becker et al., 2018; Costanza et al., 2005; Koike et al., 2006; Theiss et al., 2016) or robotic control (Khokhar et al., 2010).

Multi-electrode systems are not constrained to measuring body activity, but through electrical muscle stimulation (EMS) might also induce body activity. In the HCI literature, such systems have been suggested for improved learning (Tamaki et al., 2010), communicating information to

[4] A good, likely unintentional, example of this irritation can be seen in Meltzner et al. (2008) Figure 1B.

users through their own movement (Lopes et al., 2015), or simulating virtual objects (Lopes et al., 2017). More recently, a number of systems have interleaved sensing and actuation in such multi-electrode devices. For example, Nishida et al. presented a system designed to support users sharing kinaesthetic experiences (Nishida et al., 2015), while Knibbe et al. designed a device which can be used to record and play back movements (Knibbe et al., 2017).

While often not explicitly discussed, applying such devices is one of the main hurdles in their deployment. For example, the system used by Hassab et al. to explore embodied emotion required ~ 60 minutes to deploy (Hassib et al., 2017). Similar limitations were also reported with initial prototypes used by Knibbe et al. (Knibbe et al., 2017). Generic multi-electrode systems have been proposed, including the MYO as a prominent (discontinued) product, or Zap++ (Duente et al., 2017); however, these are not flexible enough to support the requirements of a vocal sEMG wearable or most of the systems described in this literature review. In practice, most custom applications require custom layouts (Nittala et al., 2021).

Case-studies of creating devices based on such custom designs do exist, typically incorporating multiple electrodes into a single garment (Knibbe et al., 2021). However, just as many of the multi-electrode devices do not generalise well. Here too it is unclear how to generalise to other applications or body parts. To add to the existing body of knowledge, we therefore designed a throat-worn multi-electrode device: the Singing Knit.

Soft-sensors for sEMG have been increasing in popularity for wearable designs and the use of body and gestural information in interaction. Textile electrodes (Acar et al., 2019) and embroidered electrodes (Shafti et al., 2017) have been used for flexibility and comfort in wearables for communication of biosignals for health application. Additionally, soft sEMG electrodes have been employed in body extensions for social interactions, where they are used in combination with other sensors on the body to convey and exaggerate gestures for emotional communication settings (Hartman et al., 2015, 2018).

7.3.2 Knit Structures & Soft Wearables

Finding materials and manufacturing methods to create elastic, flexible, and soft devices in place of traditional rigid electronic components has a long tradition. An early example of such work in the textile area was presented by Orth and Post over 25 years ago in the form of embroidered capacitive touch sensors (Post and Orth, 1997). With work like the *Kit-Of-No-Parts* (Perner-Wilson et al., 2010) and many other accessible projects, the High-Low tech group at the MIT Media Lab highlighted how anyone might create electrically functional garments. For the last 10 years, the world has seen hobbyists and researchers alike, for example Limor (Fried, 2012) and Becky Stern (Stern, 2009), augment their garments with sensors and actuators (Skach et al., 2019).

More recent work has explored how to deeply integrate electrical functionality into textiles and garments. Rather than attaching functional objects to a soft structure, the soft structure itself might become functional. This might be achieved by the use of functional dyes (Honnet et al., 2020), or by weaving or knitting fabric or clothing consisting of multi-material yarns (Huang et al., 2008; Pointner et al., 2020). Such multi-material fabrication enables creating devices with integrated sensors (Luo et al., 2021; McDonald et al., 2020) as well as detailed characterisation of interaction between textile design and sensor performance (Wang et al., 2014; Wijesiriwardana et al., 2003). For example, depending on knitting configuration used, one might measure pressure (Preindl et al., 2020) or stretch (Atalay et al., 2017; Liang et al., 2021).

When the electrical device is so intimately linked with the garment, the design of the textiles structure itself becomes relevant for its function (Greinke et al., 2021b). Custom design of textile structure was explored by Hofmann et al., who designed programmatic ways of controlling (Hofmann et al., 2019) and designing for digital knitting machines (Hofmann et al., 2020), supporting designers to implement custom material properties. This enables creating custom garments, where material properties such as texture or elasticity can be fine-tuned dynamically (Jones et al., 2021; Kaspar et al., 2019). Such material properties are essential for chronic deployment of wearable systems (Huang et al., 2021b) both with regards to maintaining electrical functionality (Huang et al., 2021a) and user comfort (Knibbe et al., 2021).

While most wearable multi-electrode systems are deployed on the limbs, Singing Knit is worn around the throat. This area is especially sensitive to pressure, so finding a garment which is both snug enough to maintain good electrode contact, while loose enough to not provide an experience of strangulation, is challenging. In the case study of Singing Knit, we show how we address this issue with a machine knit, and how this knit also guides the integration of electrical conductors in a way which minimises strain.

7.3.3 Wearables & sEMG in Performance

Art and Performance has been a strong driving factor in the development of both wearable and multi-electrode systems. Prominent early examples are the electrical fashion of Diane Dew (Jonas, 1967) and the performance art of Stelarc (Stelarc, 2020). Within HCI research, art (particularly music) has also played a strong role in the development of these research areas. Amongst the first embroidered electronics developed by Orth was a jacket that served as a music-controller (Orth et al., 1998) and a set of soft musical instruments (Orth, 2009). Similar themes can be found in recent explorations of New Instruments, exploring the use of soft materials for performing (Donneaud et al., 2017; Freed, 2008), and wearable devices for creating (Stewart, 2019) or conducting (Greinke et al., 2021a) music. Smart textile wearables have become an increasingly popular for the implementation of integrated audio, such as textile speakers, in daily wearables (Nabil et al., 2021; Perner-Wilson and Satomi, 2012; Preindl et al., 2020).

sEMG is often motivated around music performance, as discussed in detail in Chapter 2, Section 4.3.4. Additionally, EMS (Tamaki et al., 2010) and sEMG (Karolus et al., 2018) are often used with the intention of supporting learning of musical instruments or improving the expressiveness of existing instruments (Karolus et al., 2020; Kilian et al., 2021). In terms of deployability, the result is that sEMG interactions are often limited to wear on the arms. This would not be applicable for measuring the muscular activation of the laryngeal muscles, for instance the system presented here or that of Kapur et al. (2018). With the Singing Knit, we design specifically for wearability in a performance context.

7.3.4 Design Goals for the Singing Knit

With the Singing Knit we intend to improve various aspects of our system, improving comfort, functionality, deployability, and aesthetics:

1) Comfort:

The rigid cup electrodes must be attached securely to ensure consistent measurement, often for hours at a time. As a result, there can be irritation, especially to the sensitive skin of the neck from prolonged contact with the conductive paste, which can cause dry skin, as well as from the adhesive medical fabric tape, which causes mechanical stress to the skin as the performer moves, pulling on sensitive skin and vellus hair. For a professional musician, who might have a multiple hour long performance, or perform multiple events per week, this unnecessary discomfort is not sustainable. For a performer to wearing the sensors in a typical vocal performance, which often lasts for hours and over multiple performances per week, this is unsustainable and can cause unnecessary discomfort — even during some of the prototyping of the system I found that my skin became dry and irritated with the prolonged use. **We therefore aim to use conductive fabrics as an alternative to rigid, discrete electrodes, creating a garment with the comfort of a normal piece of clothing.**

2) Functionality:

Taped electrodes are not only unpleasant, but – much like commercial fabric electrodes – also fragile, frequently peeling off during prolonged wear. This is especially of concern in musical performance settings, where the problem is further exacerbated due to the sweating caused by the exertion of performing and the high temperatures a performer might encounter while working under stage lighting. Gravity also works against the electrode placement in some of the examined muscles, particularly those under the jaw. It is therefore common for the electrodes to fall off or need to be re-positioned by the wearer. This requires time to do accurately and is not feasible in a performance context.

Additionally, our relies on an amplification circuit which is not wearable. This means that the performer must restrict their movement to accommodate the wiring of the system. Disturbing the connections can lead to either disconnection of an element of the electronics or introduction of noise into the system. Longer cables might improve the range of motion of the performer, but lead to signal degradation. This is a further risk to the functionality of the system in the context of live performance. **We therefore aim to design a wearable which is robust, even under extreme conditions. The system should support performers to move freely while performing, without worry of signal degradation or electrode failures.**

3) Deployability:

The problem of lacking robustness is further amplified by how difficult it is to apply the system. Attaching an electrode requires searching for the correct placement through self-palpation on the muscle, then placing the electrode at the correct location before taping it down. Sometimes it is of advantage to mark the location with a pen, to ensure that the electrode is not inadvertently moved while affixing it. At a minimum, this process requires experience and training as well as access to a well-lit mirror. Ideally, it requires assistance from another person.

Lengthy setup times prevents changes mid-set, and add an additional level of complexity to activities such as sound-checking and general setup for performances. It essentially eliminates the ability for any spontaneous performances, and is prohibitive for quick sets. These problems are especially problematic for solo performers. **We therefore aim to design a system which is both easy to put on and take off and keeps the electrodes in the correct position. Additionally, it should be easy for the wearer to easily confirm the placement of the electrodes is correct.**

4) Aesthetics:

In performances where I have used the sEMG system (Martelloni and Reed, 2021) the electrodes are clearly visible. In the context of these performances, revealing the technology in this way might be desirable as the performance is not only about the music but also about new technologies for musical expression. However, this may not be the case for all performers. Based on context the display of technology and the medical appearance of the electrodes my not be desirable. Many musical artists use very elaborate costuming in their performance and this system should aim to act synergetically to such efforts, much as an additional piece of clothing would (Norderval, 2020). **Our goal is thus to further align the use of biosignals and physiological sensors within artistic, specifically musical, applications and create a garment implementation which can be incorporated into all types of existing artistic culture and practices.**

7.4 Design Process

During this design, I collaborated with Dr. Sophie Skach, who is an expert in textile and fashion design and works extensively with e-textile implementation.[5] The first part of the design process was familiarising Sophie with the required physiological background, and for Sophie to share initial ideas of what a textile solution might. Then, creation of the knit collar involved four main design activities: selecting a knit to use for the body of the collar, selecting an appropriate conductive fabric for electrodes, determining a way to host the sEMG signal acquisition on the wearable itself, and designing the reference electrode.

7.4.1 Sharing Competencies

Singing Knit is intended to measure the electrical activity of the 4 pairs of largest extrinsic laryngeal muscles (8 muscles total, on either side of the neck, Figure 7.10). Larger muscles provide stronger, cleaner signals and are easier to identify through self-palpation of the throat, making placement of the electrodes easier. The muscles include symmetrical muscular pairs: two suprahyoid pairs, the mylohyoid and digastric (anterior belly) muscles and two infrahyoid pairs, the omohyoid and sternohyoid muscles. The suprahyoid (above the hyoid bone) muscles lie beneath the chin and work to elevate the larynx. They are also active in moving the floor of the mouth and the tongue. The infrahyoid muscles (below the hyoid bone), depress the larynx, giving a lower fundamental frequency to the voice and assisting in articulation. After familiarising herself with the above background, Sophie suggested that a knit structure might be a useful candidate to explore, and created several sketches (Figure 7.11).

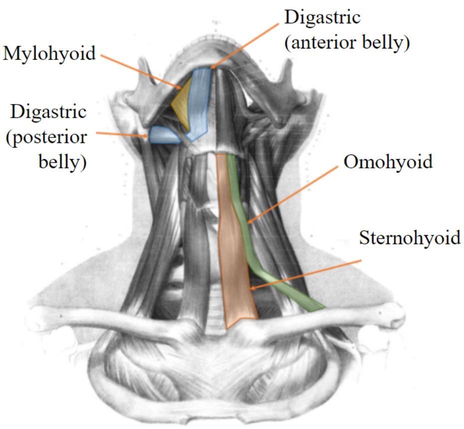


Figure 7.10: The laryngeal muscles measured by the collar (adapted from image available in the public domain, Flickr)[6]

	Muscle	General Function
Suprahyoid Muscles	Mylohyoid	Elevates the floor of the mouth, active in consonant articulation and jaw movement
	Digastric (anterior belly)	Raises the hyoid bone and the larynx, increases supraglottal presssure to generate higher pitches, moves the tongue up and forward
Infrahyoid Muscles	Sternohyoid	Lowers the larynx for producing lower pitches, tilts the hyoid for articulations
	Omohyoid	Similiar to the functions of the sternohyoid, acts as a laryngeal depressor for lower pitches

Table 7.1: Functions of the selected laryngeal muscles in speech and singing (Hardcastle, 1976).

[5] https://www.sophieskach.com/

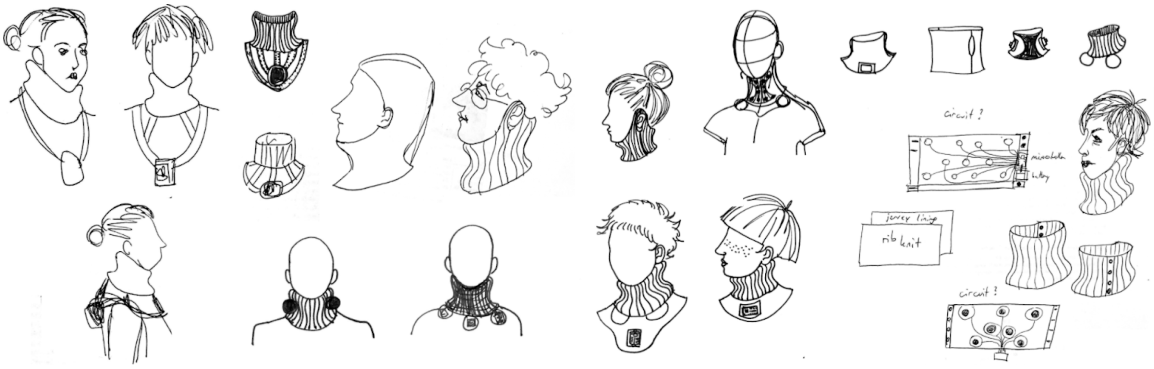


Figure 7.11: Sketching potential structures for the collar, matching the garment's structure to fit the muscles being measured (reproduced here with permission from Sophie Skach).

7.4.2 Selecting Knit

We use a knit collar to provide the right structure for the wearable; using a knit garment provides a balance between comfort and sturdiness to house the conductive electrodes. The base knit body for the collar can be constructed from different knitting techniques, each resulting in slightly different textile properties regarding elasticity, robustness, and volume. An appropriate structure to be worn around the neck, providing enough elasticity to be pulled above the head, as well as sitting tight around the neck are double bedded ribs. This knit structure is often found in high collared turtleneck jumpers or ribbed hems - the places on a garment where maximum stretch is required when put on and enough grip and fit when worn. The throat is a vulnerable part of the body, meaning that the fit of the device cannot be too tight, although beneficial to the contact between the skin and the electrodes. Unlike other sEMG wearables (Knibbe et al., 2021), the fit cannot simply be designed to be as snug as possible. We therefore opt for the elasticity of a knit garment compared to other woven fabric materials, which do not offer the same flexibility. The garment must be tight enough to ensure contact but also flexible to provide comfort, protect the neck, and not restrict head movement.

Several double-bedded knit structures were created and examined for different amounts of stretch and shape related robustness. Amongst these structures created are two Milano Rib variations (Figure 7.12a and b), a 2x3 rib (Figure 7.12c), and a two variations on a 1x1 or full bed rib stitches (Figure 7.12d and e). The difference between these rib variations are characteristics in their stretchability, rolling of fabric sides, and volume. These differences have, amongst other factors, determined the use cases of different double bedded knit structures. For example, Milano Rib has commonly been used as a structure for Jaquard patterns; 1x1 ribs for smoother knit surfaces; and larger ribs like 2x3, 3x4, etc. for soft, voluminous accessories like bonnets. The technical differences of these structures consist of the arrangements of needles on the knitting machine: the numbers in the rib types indicate how many needles are knitted in each needle bed, in an alternating manner. The Milano Rib is in this sense a 1x1 rib with added rows on only one side, making it unique from the other rib structures.

[6]https://flic.kr/p/uKQ8eb

Figure 7.12: Samples of different knit structures for the collar body. Three Milano rib variations (a, b) were compared to a 2x3 rib (c) and two full bed rib stitches (d, e).

Yarn & Gauge

Several knit swatches were created to test both the feel and stretch of the knit. The yarn composition used to create the knit probes is a Merino extra-fine Nm 30/2 yarn - a high quality sheep's wool yarn.[7] Two strands of the yarn were used to produce the samples presented here. The thickness and weight, as well as twist of the yarn further determine the gauge range that can be produced. All probes, as well as the final collar piece, were knitted on an industrial mechanical hand-flat double bedded knitting machine (Dubied) in gauge E7.[8]

Stretch Behaviour

The stretchability was the most critical element to examine for each knit. As the collar must be held tight enough to hold the electrodes in place, yet not as tight as to cause discomfort to the wearer. With the Milano Ribs displaying the most robust surface, they offered the least movement when pulled. Additionally, adjusting the stitch length of the structure with the alternating rows, this structure proved most error-prone in fabrication, with either being too loose or too tight. The other two knits, as expected, outperformed the Milano Ribs. They offered a more elastic fit overall, although the different rib sizes showed differences. Generally, the larger the ribs are, the more loosely they fit and the quicker they wear out. After examining the different swatches in terms of their stretchability, rolling, and optics, the most suitable option in all factors was to use a full rib,

[7] Labelling yarn in a Metric (Nm) system reports the weight and twist of it. In our case, it is double twisted and requires 30 meters to weigh 1 gram.

[8] Referring to the distribution of knitting needles per inch (in our case, 7 needles spread across one inch on the knitting machine.)

Figure 7.13: Samples of different rib structures added to the full-knit to provide indication of electrode sites.

Figure 7.14: The ribbing added to the collar body; the relief stripes' end points provide visual and tactile representation of the electrode positions.

or 1x1 rib stitch. Most importantly, the knit provides a suitable balance in its stretch, meaning the swatch feels secure when wrapped around the neck but in a way similar to a turtleneck sweater.

Relief Stripes to Highlight Muscle Locations

Although affixing electrodes directly onto the skin using adhesive paste is not ideal in terms of practical wear and movement, the visual cues this approach provides are valuable for exact positioning and locating the muscles. To preserve this characteristic, an additional structural feature was embedded: a needle transfer pattern, where selected stitches are transferred from one needle bed to the other, excluding the now empty needle from use. This creates a relief structure, or inset (Figure 7.13), that is visible on one side of the knit fabric. When integrated at the exact position of the electrodes, the wearer has instant visual and haptic information on their positions and can move the collar in place. Figure 7.14 shows these effects on the final collar (left).

7.4.3 Replacing Traditional Electrodes

Several textile materials were examined as replacements for the cup-electrodes used in the original design. The materials examined were chosen to match the collar's body; they also consist of flexible knits. We compared 5 knit conductive materials - two single-bed jersey fabrics and two double-bedded metal mesh fabrics, and one conductive foam. The conductive foam features a knitted backing and so the feel of the pad is similar to the jersey fabrics. We created a set of electrode swatches (Figure 7.15) for testing the materials. The materials were compared against the same rigid gold-plated silver cup electrodes (MediMaxTech, New Malden, UK) used in our previous setups (Reed and McPherson, 2020, 2021). For each material, a 1 cm^2 "electrode" was used for appropriate comparison to the same size cup electrode. The conductive fabric electrodes were stitched into a small knit body, similar to how they would be attached to the full collar, with the appropriate spacing to be worn and tested on the left omohyoid muscle on the lower neck. A highly conductive silvered copper yarn (Karl Grimm) (Satomi and Perner-Wilson, 2007) was used to provide a connection to the fabric pads. The swatch was secured to the neck using an elastic band, as if they were part of the full collar wrapped around the neck. The swatches were worn to test if the conduction was appropriate to replace the traditional electrodes, and to test comfort and feel against the skin.

The resistance across the 1 cm^2 pad was measured using a multimeter. All the materials tested had similar resistance to the cup electrode (0.6 Ω), with the exception of the foam. Overall, the density of the foam lead to a higher resistance (1.4 Ω). The two double-bedded knits had a similar resistance to the cup electrode, 0.6 Ω for the tin and 0.7 Ω for the stainless-steel fabric (Table 7.2).

To test the electrodes in context, the swatches were used in a singing task, where I checked the activation of the omohyoid by singing a low note in my vocal register (G3), similar to the original proof-of-concept. The signal was passed through the VoxEMG board and the amplitude of the muscular activation captured was measured with an oscilloscope. I performed this task a number of times, singing the pitch with a tuner to ensure as much consistency as possible. The average peak-to-peak amplitude of the signal conveyed is listed in Table 7.2. All materials conveyed a usable signal in an amplitude range similar to the rigid electrode, with the exception of the metal mesh fabrics. This difference for the meshes may be the result of the difficulty in wearing in this small electrode-sized piece, which is discussed in the following sections about the qualitative properties of the fabrics.

Single Jerseys

The two single jersey (one-bedded knit) fabrics examined, a gold-colored zebra jersey (Hitek) (Materials, 2021) and a grey stretch conductive fabric (LessEMF) (LessEMF, 2020), are both made from silver-plated nylon yarn but of different weights. In comparison, the grey jersey was the heavier fabric (150g/m^2). It also has a slippery surface due to a slightly higher percentage of elastane than the zebra fabric (128g/m^2). Due to the elastane content and slippery feel of the grey jersey, the fabric can have a variable resistance when integrated into the knit collar body, as some of the material comes in and out of contact with the skin, or shift during stretching of the base material. The zebra jersey is most easily attached to the main collar body with a smooth surface, yet non-slippery grip; particularly because it is striped with non-conductive fabric, it is possible for a larger area of material to be secured while keeping the conductive surface restricted to a 1 cm^2 space. The non-conductive fabric stripes can be stitched into the base material, leaving an even contact across the conductive space in between.

Double Bedded Knits

The double-bedded knits examined were a light grey, tin-coated copper canopy mesh (Less EMF) and a dark grey stainless-steel mesh (Less EMF). Two different ribs were cut in shape and compared: one interlock 1x1 rib, and one 2x1 rib. Both are significantly heavier fabrics than the single jerseys (both around 190g/m^2). Although these are most similar to the cup electrodes in terms of their conductivity, the wear on the skin makes them unsuitable to replace the electrodes. When cut into the needed electrode size, the fabric becomes uncomfortable, especially on the sensitive skin of the neck. Additionally, on integration into the knit body, the mesh structure remains too rigid, leaving space between the electrode and the body of the collar, rather than the two materials feeling as one garment. Because of this space, there can be too much movement between the conductive fabric and the knit, causing the resistance to be variable depending on how tight the collar is. With the scratchy electrodes, the collar cannot be held too tight around the neck, which increases this variability in the resistance.

Foam

The thickest and heaviest material tested in our samples is RayPad conductive polyurethane foam (Less EMF) with nickel and copper elements. Although not a fully knit material, the foam pad is covered by a knitted material and offers additional tactile information to the wearer through pressure on the skin. Its disadvantage, however, is that it is compressed slightly when stitched to the base fabric. This can cause a discrepancy in resistance when further compressed against the skin during wear. Additionally, the thickness of the pad creates extra space between the collar and the skin. The body of the collar needs to be tight to keep it secure in the case of using foam, this results in the foam being always “pressed.” This creates an interesting tactile interaction, where the wearer can easily feel pressure at the electrode site and sense where the connection to the fabric is being made. This may offer an interesting affordance to provide attention to the wearer about certain parts of the body. In this sense, a similar non-conductive foam may be beneficial to give the wearer more information about the presence of the garment on their body.

The two single jerseys provided the best compromise between the conductive materials. They are only slightly more resistive than the cup electrode (0.8 Ω for both) and are able to convey a usable signal in a similar amplitude range to the rigid electrodes because they maintain consistent resistances. They have the most comfortable feel as well, and are less distinguishable from the rest of the collar than the other three materials. The zebra jersey provides the best balance of qualities amongst those examined. It is the easiest to integrate and secure into the knit material and provides suitable contact with the skin which does not shift during wear, due to having less elastane than the grey jersey.

New electrodes were cut and stitched into the knit body of the collar. The collar was knit into the desired shape and size, so no waste material was created, and no post-fabrication cutting was required. The electrodes were then stitched into the collar at the dedicated points; on the outward-facing side of the garment, the ribbing was alternated to mark the location of each pair of electrodes (Figure 7.16).

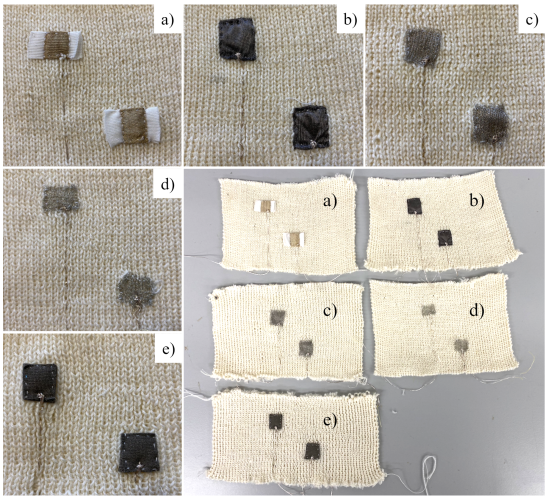


Figure 7.15: Swatches with stitched, soft fabric electrode probes for comparing materials. In order: a) silver-plated nylon zebra jersey, b) silver-plated jersey, c) tin-coated canopy mesh fabric, d) stainless steel mesh fabric, e) RayPad foam cushion.

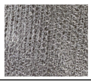
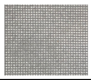
		Resistance (Ω) (1 cm^2 sample)	Amplitude (mV)
	Gold-plated silver cup electrode (MediMaxTech)	0.6	320
	Silver-plated nylon zebra jersey (HITEK)	0.8	220
	Silver-plated nylon jersey (LessEMF)	0.8	220
	Tin-coated canopy mesh fabric (LessEMF)	0.6	80
	Stainless steel mesh fabric (LessEMF)	0.7	80
	RayPad foam cushion (LessEMF)	1.4	260

Table 7.2: The conductive materials examined, approximate resistances and the average amplitudes of the output muscular activation signal through the VoxEMG board in the test singing exercise for each material.

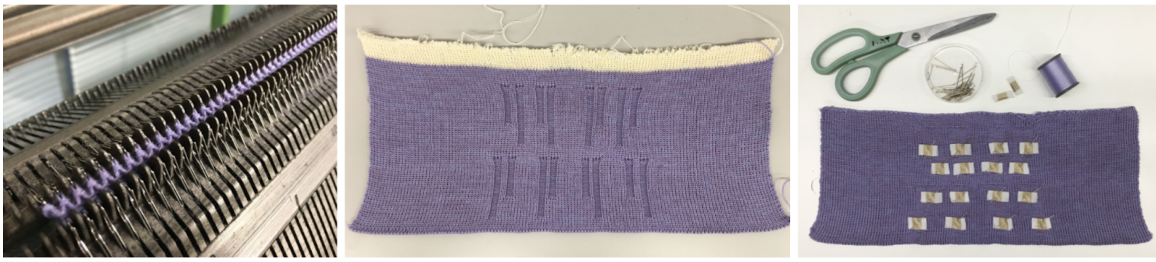


Figure 7.16: Creating the final collar body (left) and attaching the fabric electrodes (right). The final iteration uses the alternated relief stripes to mark the electrode locations externally (middle).

7.4.4 On-Board Signal Acquisition

From the final iteration of the sEMG circuit (Figure 7.5), I designed a custom open source PCB implementation of the amplifier circuit, specifically for integration into wearables for on-board processing — the VoxEMG. It was decided that the PCB should be versatile, to use it in this work and also provide a platform for others wishing to incorporate sEMG into their wearables. Therefore, we offer different connection types for the signal input (both traditional connectors or fabric integrations), and ensure that the PCB is as small as possible, to be easily fitted into a garment without adding bulk or distraction in the design. The resulting VoxEMG PCB measures 3.6 x 3.15

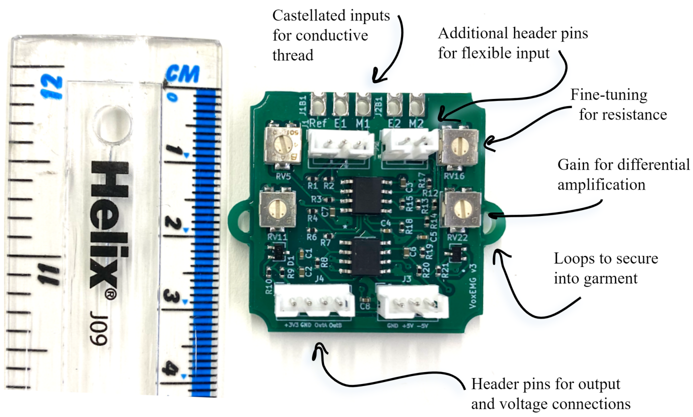


Figure 7.17: The VoxEMG board and some of its features for precision tuning and flexible I/O for wearable integration.

cm, ensuring compact integration for prototyping (Figure 7.17).[9]

The board offers fine precision resistance tuning and features header connectors for integration into other electronics, as well as castellated holes for connection with textiles. The castellated inputs were adapted from designs for the Bela E-textile Capelet (Stewart, 2020; Temprano and Stewart, 2019) for Innovate UK. Additionally, loops were added to either side of the board so that it could be affixed into a garment.

The VoxEMG boards were secured to the back of the collar using loops in the PCBs, added for affixing the board to a fabric element. By stitching the loops to the collar between rows of knit underneath of the board itself, some allowance for the rigid PCB is made to keep the collar stretchable. If the garment is stretched (Figure 7.18), the board will not be pulled from the collar and the knit underneath will be able to expand fully. The wired connections were then stitched between the board and the electrode pads; as seen in Figure 7.18, the conductive thread is able to be wrapped around the castellated hole inputs at the top of the VoxEMG boards.

The conductive thread was then stitched by hand along the knit itself (Figure 7.19); rather than stitching straight across from the electrodes to the boards, this allows the conductive thread to move with the knit. In a sense, the knit has been restitched and so the traces take on the same elastic properties as the rest of the knit. This ensures the conductive thread will not break if the collar is stretched or pulled to the knit body's maximum displacement.

The traces were added after the creation of the knit because we used an industrial handflat machine to create the collar body. Adding the conductive thread, in the manner described, ensures that it does not impact the overall knit structure, which would result from additional conductive yarn added by means of conventional stitching or sewing. Rather, the thread is integrated with the garment seamlessly and maintains the characteristics of the full rib. In addition, this method is easier to employ during the prototyping process, as it separates the components in a way where it is easier to correct or modify the design or layout of the traces as needed, without re-knitting the entire collar.

[9] https://github.com/courtcourtaney/voxEMG

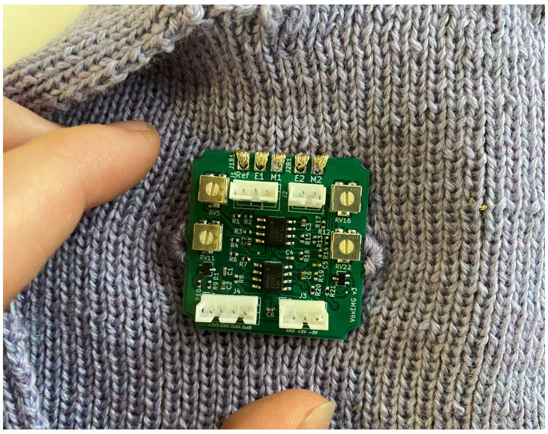


Figure 7.18: The VoxEMG board secured into the back of the collar.

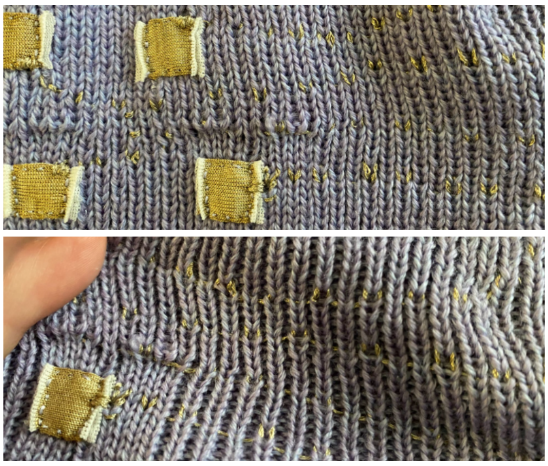


Figure 7.19: The conductive thread stitched along the knit (top). When stretched, the conductive traces stretch with the knit (bottom).

7.4.5 Wearable Reference Electrode

The last piece of design was in the construction of an aesthetically appropriate and easy-to-affix reference electrode. Previously, the electrode had been glued to the earlobe using medical tape. Rather than just securing a fabric electrode to the ear as a replacement, we continued the pursuit of wearables and stage-appropriate costuming replacements and constructed a reference electrode earring.

The reference electrode is made using generic a clip-on earring. The electrode was constructed first by creating a "knit" cable, to give the connection a feeling of being a part of the wearable collar (Posch, 2017a,b). This was done by threading an insulated wire through a piece of cording (Posch, 2019). The end of the wire is stripped and folded over to make a hoop, through which conductive thread is tied and wound (Figure 7.20a). The clip-on earring came with a soft silicon pad to be inserted into the back side of the clip and hold the earring against the wearer's ear without discomfort (Figure 7.20b). The conductive thread was used to stitch another 1 cm^2 fabric electrode, made of the same zebra fabric as the collar's electrodes, directly into the soft silicon (Figure 7.20c). This provides connection to the pad and insulation for the connection otherwise. The other end of the "knit" cable was connected to an alligator clip, which could then be secured to a conductive thread knot, which was stitched into the edge of the collar (Figure 7.20d), leading to the reference inputs of each of the VoxEMG boards (they are able to use the same reference). An alligator clip was used to ensure that the electrode could be easily attached after putting on the knit collar.

The final reference electrode earring is pictured in (Figure 7.20e). The ends of the cording were burned to prevent fraying. It can be seen that the electrode provides appropriate aesthetics as a piece of costuming, blending in with the other earrings I wear normally (Figure 7.20f, Figure 7.20g). The clip-on ensures that anyone could wear it on their earlobe; as well, a different style of clip-on earring could be used, or indeed other types of ear jewelry, such as a cuff to be worn on the upper helix of the ear.

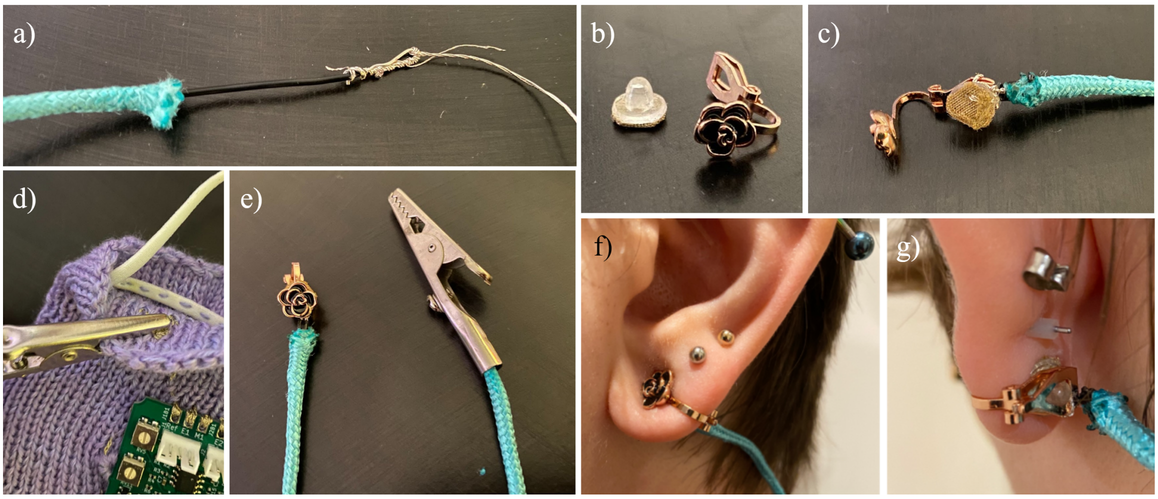


Figure 7.20: The construction of the earring reference electrode, starting with the cabling (a), stitching the conductive fabric to the earring pad (b-c), and securing the electrode to the collar and reference inputs (d) for a complete wearable (e). I wear the electrode along with my normal jewelry (f, g).

7.5 Final Design

The final version of the wearable collar is pictured in Figure 7.21, with the soft electrodes stitched into the flat, neck-facing side in the measured positions, and the alternated ribbing patterns in the outward-facing side, providing the location of the electrodes beneath. Additional elastic has been added to the top of the inside to ensure that the knit remains in place around the jaw when worn. As well, the ear straps hold the collar up and prevent it from folding over or moving around while on the neck (Figure 7.21). A set of buttons and elastic loops at the ends of the open piece close the two ends together after the collar is around the wearer's neck. This makes it easy for the wearer to put the collar on by themselves. Once on, the ribbing in the front of the collar allows the wearer to check that the electrodes align with the muscles, which can be felt beneath the fabric. For the top pair of VoxEMG boards on the back of the collar, the cables were tucked into the folds, which were stitched down at the top and bottom except for a small hole through which the cables could be passed through.

The final version of Singing Knit (Figure 7.22) was tested by myself for suitability in recording signals as done previously with rigid cup electrodes. Testing was done first qualitatively by myself in aspects of wearability, and also in a quantitative comparison between the signal conveyed by the traditional electrode setup and the fabric electrodes in the collar. For a qualitative evaluation, I wore the collar for two hours while moving around my home to assess the feel in lengthy wear. While wearing the collar, I was able to move normally and the knit stayed in position. They felt that the knit was comfortable and could have been even a bit tighter, if needed, as the fabric stretched easily. The ribbing was especially beneficial for preventing slippage of Singing Knit, as the wearer is easily able to check whether the knit is in the right place on the neck and reassure themselves of the placements. The ear straps however were able to keep the collar in place and prevented any shifting. In further musical interaction, I was able to recreate qualitative examples of use with the collar, for instance in visualising particular vocal muscle movements as done in (Reed and McPherson, 2020)

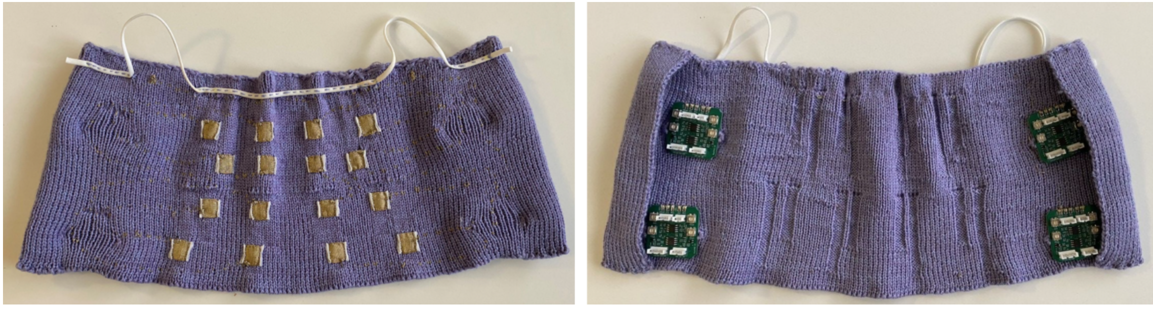


Figure 7.21: Amendments made to the collar: additional elastic is added to the inside of the collar to hold it up around the chin and provide support with straps around the ears (left). The excess fabric is folded over on the back to ensure a tight fit (right); these ends will be joined once the collar is on the wearer.

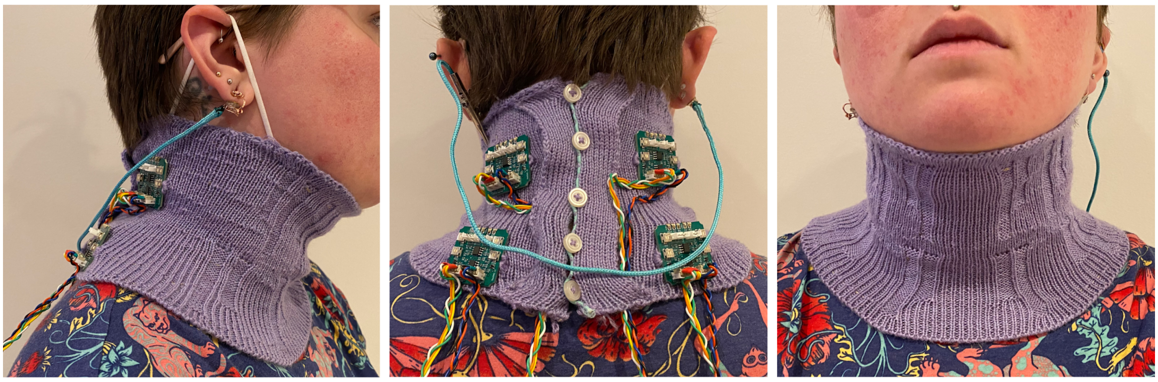


Figure 7.22: Wearing the completed knit collar, showing the knit's form on the neck compared to the original sEMG setup.

or sonic interactions with the vocal muscles as in (Reed and McPherson, 2021).

Additionally, the output signal of the stitched fabric electrodes were compared with that of the original, rigid electrode setup. The goal was to determine whether the collar was able to detect the activation of the muscle faithfully compared to the rigid electrode setup. As determined during material selection, the resistance and amplitude of the signal conveyed through the materials were comparable. To check the final design, I performed the same exercise again, measuring the signal of the omohyoid activation when singing a low note in my vocal register (G3), first with the rigid electrode setup and then wearing the final collar. The onsets of the signals generated during this exercise were captured and examined using an oscilloscope (Tektronix MSO 2024B). The signals are pictured in Figure 7.23; the traditional electrode had a peak-to-peak voltage of 340 mV, where the fabric electrodes for the same muscle gave a voltage of 300 mV. The period for both was 10 ms. As the action will never be exactly the same, there is allowance for some small difference in the signal; we see that the fabric electrodes are able to produce consistent results to that of the rigid electrode setup. In the use case here, the electrodes need to be able to measure signals which can trigger and be mapped to elements of sound design; in this case, the signal can be accurately detected and would be sufficient for sound design. It should be noted however that there is slightly

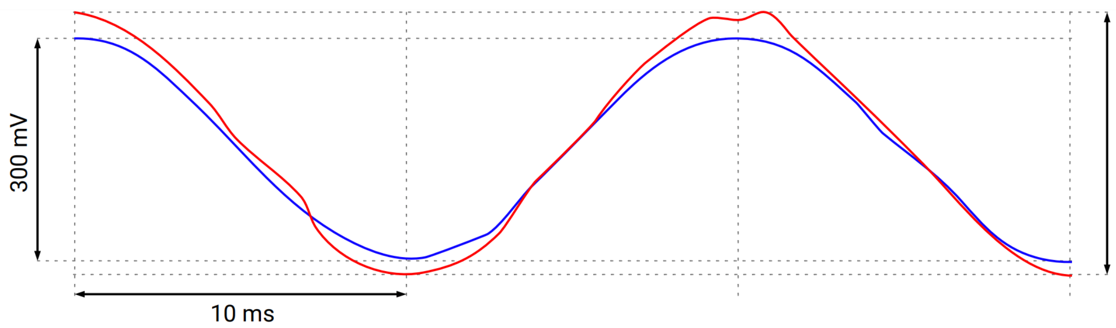


Figure 7.23: The vectorised signals from rigid electrode setup (red) and the fabric electrodes (blue) stitched into the collar. The signals measured are from the activation of the omohyoid during the singing exercise, as captured by the oscilloscope.

higher variability with fabric electrodes, which is to be expected. However, due to the live feedback one receives during performing, this variability did not prove to have any negative impact.

7.6 Evaluation of Design Goals

The purpose of Singing Knit was to explore how an existing multi-electrode design by Reed and McPherson (Reed and McPherson, 2020) might be modified to improve Comfort, Functionality, Deployability, and Aesthetics.

With regard to **Comfort**, the decision to use a soft knit and fabric electrodes provided a more suitable platform for the technology. We encounter knit garments constantly, making the feel of the collar something familiar and comparable to wearing another piece of clothing, namely a turtleneck sweater. By using a full rib stitch, we have captured a balance between the garment being tight enough to hold the electrodes against the skin, but not so tight as to cause discomfort to the sensitive skin of the neck or a sensation of binding around the throat. This provides an improved affordance over other, non-stretchable fabrics, which would need to be secured as tightly as possible. Additionally, the electrodes themselves were tested for feel; in using electrode pads made from a similarly knit jersey fabric, the conductive materials blend into the garment and are not noticeable within the knit body. In a performance setting, a singer would be able to move freely without the discomfort of medical tape constraining movements of the throat and neck. This prevents the irritation caused by prolonged wear.

In terms of **Functionality**, the custom open-source modular sEMG PCB significantly increased the robustness of the system. Integrating the sEMG sensing into the Singing Knit prevents noise or failures previously caused by disconnecting cables from the tethered station. The ability to directly sew the VoxEMG into the knit minimises cable length, while sewing the cables along the knit (Figure 7.19) ensures that all connections have the appropriate level of elasticity. In using a stretchable knit, the collar creates a tight, secure fit; additionally, the use of elastic banding in the key trouble area for the electrode placement, the muscles under the chin, prevents the electrodes from falling off or the collar from slipping. This removes the requirement of using medical tape to connect electrodes, consequently the problem of electrodes disconnecting due to movement or sweat is also avoided.

Regarding **Deployability**, the use of the button-back closure means that the collar can be fit tightly and easily by the wearer alone. Further, if the wearer wishes to check the positioning of the electrodes, they can do so using the ribbing texture on the outside of the collar. The novel use of the ribbing for checking the structure on the other side of the garment allows the wearer to quickly confirm the electrodes are correct by feeling the location of the muscles through palpation with the fingers, and making sure the ribbing aligns with them. Should they not align, the Singing Knit can easily be adjusted, without requiring to disconnect and re-attach individual electrodes. This means that setup, and quick adjustments can easily be done on stage and without the help of another person.

Regarding **Aesthetics**, we have shown how the rather medical appearance of the original system could be transformed into something which superficially looks like an everyday garment. As long as the structural the design decisions, for instance the knit type, are maintained, there is lots of opportunity for others to further customise the design to fir their respective needs. The design can even be incorporated into a full garment, such as a knit top or other piece of costuming (see also (Kaspar et al., 2019)). At the same time, the design also supports the singer to further engage with their body. The collar provides external information through the ribbing construction about the anatomy of the muscles underneath.

7.6.1 Adapting Multi-Electrode Systems

Through the adaptation process of an existing rigid sensing technology to a soft-sensing wearable, we found that the use of knit fabric to provide a sense of familiarity and flexibility (in both a literal stretchy way, and a non-literal adaptable way) which we believe will be useful for other systems too. Adapting the design followed a number of steps, which we believe might generalise to other design processes: The design of Singing Knit started with an exchange between disciplines, with the core team exchanging their technical knowledge of sEMG systems with textile design knowledge of a new team-member with a textiles and fashion background. Together, context-specific needs to be addressed through soft-sensing were identified (e.g., ease of putting on, quickly setting up, and wearing the collar for extended periods of time in vocal performance). Then it was determined which aspects of the existing system should be preserved and which should become part of the textile, to balance the sensing capabilities but increase accessibility for the target context (e.g., providing connection to and awareness of the muscles while placing the sEMG outside of a medical context). Then testing was performed, both for determining suitable replacement for the traditional cup electrodes as well as for (in our case, the conductive fabrics), and for determining the structure needed to properly house the sensors on body and function in-context (the knits). Only then could the Singing Knit be constructed, before finally testing and comparing systems, both quantitatively in terms of sensing suitability and qualitatively in terms of aesthetic and feel, with the original sensing method.

It should be highlighted that these steps are highly iterative. For instance, the body of the collar itself began as a selection for elasticity and comfort - how to get the collar on and off the body and how to ensure that it would stay upright. After examining the suitable fabric electrodes, it was apparent that the knit structure would need to be tighter toward the front of the garment to ensure there was little-to-no movement between the different materials (to reduce the variability of the resistance). Finally, the decisions on structure changed again when re-examining what was beneficial about the traditional electrodes; in seeing and feeling where the measured points were, there was feedback regarding correct placement of the electrodes. Consequently, the ribbing of the Singing Knit was introduced to provide tactile feedback to the wearer on its positioning relative to their muscles. This adds a sense of transparency and understanding to a very covert, internalised bodily function. Through the iterative process and adaptation in choosing materials, we were not only able to preserve the function of the original sensing technology and address some impracticalities, but also to add to and extend its positive features.

7.6.2 Soft Knits for Different Bodies

The knit collar was designed as a bespoke garment for a single user. In different wearers, the muscular placement will be different, as will be the overall size and diameter of the collar itself. There is both negative and positive to using this individual design. On the negative side, the custom-made collar means that others could not effectively wear it. In addition to points of discomfort in wearing a collar which is not sized correctly, the electrodes, if not placed correctly, can end up on two different muscles or have the potential to pick up cross-talk from other nearby muscles with different activations and action potential differences. The design can be custom fitted for an individual's anatomy, but this requires time and skill to correctly size and knit a collar for each wearer. An advantage of the rigid multi-electrode setup is that the wearer can choose different positions each time they apply the electrodes and one setup can be used for different physiologies just by adjusting the electrode placement — for this reason, when working with other vocalists later in Chapter 9, we return to the traditional electrode setup, which is more time and cost-effective for the exploratory study. Applying some more of the original multi-electrode design into the wearable would be beneficial.

On the other hand, there are benefits to bespoke design and the attention given to individual bodies in design. Creating custom collars allows for individual differences, however subtle, to be taken into account and for the best possible fit to be achieved. Also, a prominent theme in vocal wearables, is the exploration of the connection of the singer to their body and their body perception (Cotton et al., 2021a,b; Reed and McPherson, 2021; Tsaknaki et al., 2021). Therefore, achieving a "one-size-fits-all" collar might not be desirable, in fact it might negatively impact this theme, in masking the breadth of human body-diversity (Spiel, 2021). In pragmatic terms, creating bespoke vocal garments might be more reasonable than assumed, as it is already common for many elements of costuming in vocal performance settings, namely opera and musical theatre, to be custom-made for the performer. It is therefore not unreasonable to design individual wearable e-textiles for performers, given the benefits to the wearer's comfort and the biosignals captured. In longer-use cases outside of the study with other singers in Chapter 9, it would be reasonable and worthwhile to undertake the time and effort to create individual wearable collars for different singers. Taking this approach to similar biosensing tasks may help to preserve the connection between our individual bodies, actions, and perceptions, and the resulting use of the technology involved in our interactions.

7.6.3 Future Work

There are several additional design features which would greatly benefit the wear and use of the vocal collar. Although attention to individual bodies is critical, there may be ways to address unique physiology through adjustable sensor positions and re-sizable or "hackable" collars. This is a critical next step in the design to make the collar usable to different wearers. However, both with the creation of new collars and in traditional electrode placement, the wearer (or a technician setting up the sEMG) must have the ability to manually feel the muscles within their body to get an accurate placement and affix the electrodes correctly. In this case, I had received instruction from a speech language pathologist on how to identify the muscles in my own body, which was the model for designing and testing this collar. This however requires practice; although we feel it is beneficial in being acquainted with one's body to know the placement of muscles and how they feel when active or tensed, it is not practical for every user to receive this type of instruction or to begin every performance or setup of the device with muscle identification.

A potential solution to address flexible design and finding accurate electrode sites could involve the inclusion of more electrode pads, potentially placed in an array to occupy multiple potential measurement sites, such as done in SkillSleeves (Knibbe et al., 2021). This might provide flexibility for different wearers' anatomy and exploration of their physiology, allowing the user to switch between placements and adjust until a suitable and clear signal is achieved, without removing the garment or needing a totally custom-fitted garment.

Additionally, previous research evaluates the first-person experience of the wearer interacting with sonified sEMG signals from their own muscular activations. Further research might be done to determine how the tactile aspects of the collar play into the interaction, perhaps providing additional feedback in feelings of stretch and tightness, or constraining movement to certain boundaries. This sensation of externalised tension may be useful in vocal pedagogy or other applications to help train posture and alignment in the wearer.

7.7 Conclusion

This thesis then provides a practical verified system which can be used to gather sEMG signals from the vocal musculature. We therefore provide a method for direct control using the voice which operates independently from audio signal analysis. This thesis has found sEMG an appropriate and minimally invasive way to measure and externalise (e.g., as a visualisation in this chapter, and further as sonification in the studies that will be presented in Chapter 8 and Chapter 9) vocal musculature activation in both vocal and subvocal contexts. Through this direct control, sEMG can be applied for new vocal interfaces and augmentations and provide new directions for research in vocal gesture.

The creation of the Singing Knit collar for vocal interaction involved the adaptation and converting of design using traditional sensors into embedded, knit design. We identified areas for improvement in the existing system, which uses technology derived from another research context, to integrate the technology better into a performance setup. We used these areas to motivate the design of a wearable collar specifically for this context. In doing so, we have converted an application and design for vocal sEMG using traditional sensors into one that utilises embedded soft sensing technology. This also provides an outline a template for converting similar existing rigid sensing methods and present a number of design aspects which will be valuable to creating other wearables.

Chapter 8

Autoethnographic Interaction & Evaluation

Laryngeal Sonification through First-Person Perspectives

While designing and prototyping both the circuit which would become the VoxEMG and the Singing Knit, I engaged in a long term autoethnographic study of my use. Knowing the system in a very intimate way and developing it within my own vocal practice (particularly while testing the setup), allowed me some novel thinking of my own singing. I documented my experience while working with the system through a sort of journalling process while recording and using it; in addition to play within the moment, this allowed me to examine some unexpected interactions, affirmations of my practice, and the balance between control and being controlled while working with my body in this new way.

This chapter will detail three different explorations while working with some kind of sonification of my vocal muscles: a series of improvisations done during testing and prototyping, reflections on use and performance with the Singing Knit, and a micro-phenomenological interview I did with another micro-phenomenologist. In the end, we see some notable points that arise from working with the muscles in this way through sonification: that there is an entanglement between the technology and the body and a push-and-pull effect between my intention and expectation and the system and its response, and that this can evoke different perception and awareness of even very subconscious movement.

Portions of this chapter have been published in:

Courtney N. Reed and Andrew P. McPherson. 2021. Surface Electromyography for Sensing Performance Intention and Musical Imagery in Vocalists. In Fifteenth International Conference on Tangible, Embedded, and Embodied Interaction (TEI '21), February 14–19, 2021, Salzburg, Austria. ACM, New York, NY, USA, 11 pages. DOI: 10.1145/3430524.3440641

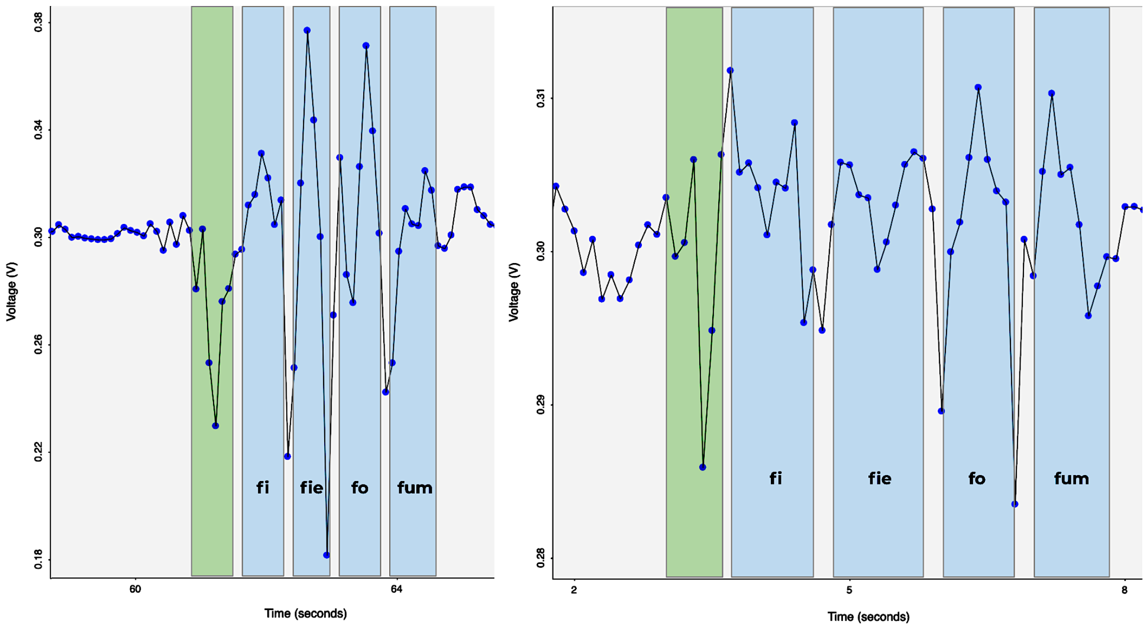


Figure 8.1: An example of a third person perspective: activation of the omohyoid muscle while singing (left) and imagining (right) a major arpeggio, as displayed in the Bela GUI. The activation during the breath is highlighted in green, with the sung syllables in blue.

8.1 Interaction Perspectives

Before I begin to describe my experience with the sEMG sensor, it is important that I reconcile the differences between first-person and third-person perspectives with this system and its usage. sEMG can be used in both ways and it depends on who is the spectator of the information it provides — it can most certainly be used in a third-person context where information about the user's muscular activations are analysed and used to convey objective information such as force and grasp, as utilised in previous research for controllers using sEMG (Costanza et al., 2005; Kapur et al., 2018; Lim et al., 2020; Meltzner et al., 2008; Theiss et al., 2016) and earlier in Chapter 7. Here, predictions are made about body movement or the muscular contractions are used to interact with digital object, but the gesture is not necessarily adapted based on the machine's interaction. A movement classifier to control a robotic hand, for instance, determines a gesture and mimics the user's action and the user acts through the machine (Khokhar et al., 2010). sEMG also allows for useful first-person interaction, where this reciprocal interaction between human and machine is at the forefront. This is more common in musical (Tanaka, 2015; Tanaka and Knapp, 2017; Tanaka and Ortiz, 2017) and emotional interaction (Hartman et al., 2018), where the user is the spectator of their own gestures and learns by observing how the system responds and learning how it acts as an individual agent.

The same can be said for this particular vocal sEMG system. I can use it to simply measure gesture in an objective way, and this is beneficial if I want to see how my imagined action mirrors anything I actually do (Reed and McPherson, 2020). For example, if I look at activations in the omohyoid moving while I sing a major arpeggio, and compare it with my imagining of that same exercise, I can see that this activation occurs in a similar pattern (Figure 8.1). In this exercise, I sing a simple major arpeggio (C4-E4-G4-C5) on the syllables 'fi, fie, fo, fum.' The muscle activates as I articulate each note (or imagine articulating each note). Note that the subvocalised exercise (right) lies within a much smaller voltage range (approximately 0.285 - 0.31 V, 0.025 V peak-to-peak) than the vocalised exercise (left, 0.18 - 0.38 V, 0.2 V peak-to-peak) but similar activation of the muscle occurs in subvocalisation as in vocalisation. This is by no means a quantitative analysis of comparison between vocalised and subvocalised activity, but rather a means to see that activity does indeed occur in the same exercise both sung and imagined, as utilised in previous studies using the muscles of the face, neck, and jaw (Kapur et al., 2018), and potentially to compare it to other use cases, for instance that in the previous Chapter 7. This type of information could be further used for gesture classification or force measurement or a multitude of other analyses; in this sense, it is information about my performance but not about any sort of interaction with the system. The system examines me, and in this sense, it views my performance from a third-person perspective to measure aspects about my movement or imagery.

On the other hand, in my exploration of the system, I was not necessarily looking to determine any gestures or to extract any sort of data. However, I do get to observe how the machine responds to my input and how I can adjust accordingly to change the result if I wish. We are co-dependent in a very entangled way and the interaction is based within a first-person perspective, where I am in control of my responses to the system as my collaborator or accompanist. It is in this type of first-person perspective which we see the benefits of sEMG and musical imagery as tools for embodied interactions; the output is not always predictable or expected, so manipulation of existing knowledge, surprise, and reaction are used to relate to the machine in an improvised performance (Essl and O'Modhrain, 2006; Ihde, 1975). I am able to monitor and change my expectations and work with the system to create something meaningful to me and rooted in both my existing experience and the learning I did while using the system, which I will now elaborate on.

The system was designed and used extensively through iterative testing by myself, with a focus on my own internalised metaphors, imagery, and personal experience as a vocalist. This is therefore a highly specific use case. Keeping this in mind, while my own interaction with the system is unique and the specifics of my interaction would not necessarily be translatable to other vocalists, there are aspects of this interaction which we feel can be ubiquitous. For instance, the ways in which I learned to interact with my body and adapt some of my technique, as well as some perhaps unconventional or unnatural behaviours I attempted, could be distinguishable interaction patterns which are common amongst others using the system. Even in non-musical contexts, the methods of exploration and balance between user action and system reaction may be experienced similarly by individual users.

8.2 Improvising with the Design

During the prototyping and testing process, particularly in the later stages where the system was working functionally but I was experimenting to see what kinds of interactions I could observe with the better CMR, I switched from viewing the signal within the Bela IDE to a sonification. This was done firstly to see if there were notable changes to what I could hear in real-time while singing (rather than viewing a visual plot at a slight delay), and also to provide a presence to the laryngeal muscles in a musical way through sound. I used an improvisatory approach to explore whether the movement of some of the smaller muscles in the suprahyoid region (e.g., muscles we could newly measure with reduced noise in the circuit) was apparent. Within the Bela context, I used the sEMG signal with SuperCollider, a computer language and platform for audio synthesis and algorithmic control. Here, I used my laryngeal muscles to control aspects of digital vocal processing.

I focused on a few different muscles, namely the digastric, mylohyoid, and geniohyoid muscles (Figure 8.2); these muscles, located under the chin, are active during laryngeal elevation and move

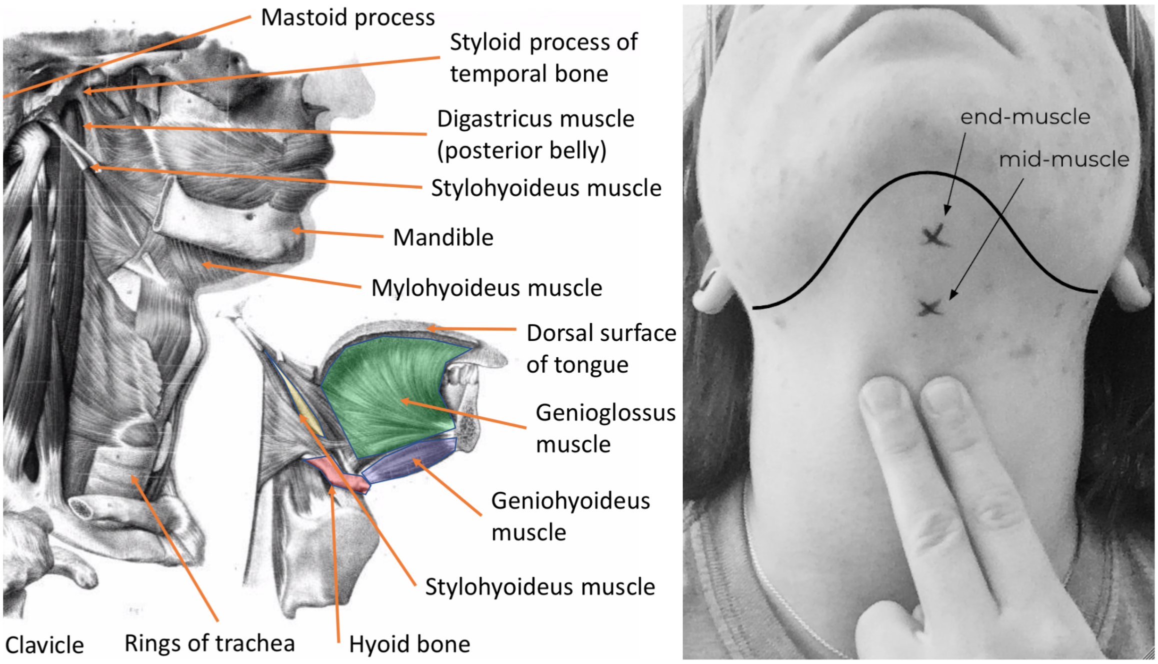


Figure 8.2: The suprahyoid muscles (adapted from image available in the public domain, retrieved from Flickr: https://flic.kr/p/u6pB3Q), and electrode placement on the suprahyoid region (right, finger placement indicates the position of the hyoid bone on the wearer).

with the tongue. The tongue is important in singing firstly for lyric articulation, but also can be moved in a way that creates different kinds of resonant spaces in the mouth. Changing the position of the tongue helps to get different vowels and place the sound in different portions of the face, with either a rounded open sound or a pinched nasally sound in either extreme.

I used these sEMG signals in a SuperCollider script to control the frequency of a sine wave carrier used in a ring modulator. I used an exponential mapping in this case and, rather than just the typical metallic chorused sound of a ring modulator, I found that using lower frequencies provided a sort of fluttering or "breathiness" to the sound of my voice, which I quite enjoyed.[1] It felt more visceral and bodily to me, rather than a companion to the audio I was modulating. This is what I focused on for much of the interaction during my improvisation. With increases in suprahyoid activity, I was able to increase the frequency of the sine wave which modulated my direct voice input. As well, I included some other synthesis effects to my direct voice input such as a pitch follower and chorusing to create an environment in which I could play around with the sEMG signals.[2]

8.2.1 Re-Learning and Reacting with the Body

Firstly, the decision to use the suprahyoid region was made because of the flexibility of interaction I learned through testing and playing with the sEMG sensor. In testing the system, I found as well that the larger infrahyoid muscles in the lower neck, particularly the omohyoid muscle, provided sources of reliable activations, although these are laryngeal depressors and so are most active at the

[1] An excerpt of an improvisation using this setup: http://instrumentslab.org/data/courtney/TEI2021_demo.mp3
[2] An explanation video in our presentation at the 2021 TEI Conference: https://youtu.be/XujlMjoBG04?t=461

bottom of the vocal range. On the other hand, because the suprahyoid muscles move not only to elevate the larynx but also with the chin, jaw, and tongue movement, I felt I would not have to rely on pitch production or singing within a certain range, and so might be more flexible in this case; however, restricting movement by focusing on muscles which perform a more defined action would also be worth further consideration when it comes to observing how learning occurs through interaction with a system.

With the suprahyoid, I know I can choose between a number of techniques which actively utilise these muscles, ranging from vowel formant shaping and text articulation to some unorthodox practices in singing, namely stretching my neck all the way out or squishing it towards my throat (ie. the "double-chin"). These sorts of movements are completely undesirable in singing, as they create lots of tension and make the voice sound constricted; in students who are soprano voices like myself, it is a common tendency in an untrained voice to "reach with the chin" when trying to sing higher notes, and several voice teachers worked very hard to hammer this out of me. However, when it came to actually playing with the system, I realised I could get a lot of activity out of these activities and so I started to work them into my practice, leaning into some of the more constricted sounds for a tense effect. For me, this was a very useful practice tool–I had to recall motions which had been trained out of me and try to work with them while still keeping my "well-trained" open and supported sound. An untrained singer might have less awareness of these tensions in the chin, and providing a source of feedback to it would likely call attention to it.

In an improvisation sense, I found I behaved normally when it came to the pitches I was singing. Being very comfortable with the sorts of fast melisma-style phrases you'd find in opera (where a singer strings many notes alone to one lyric), I would have previously focused my vocal contributions in improvisation on moving the melodic line, toying with dissonances, and mimicking any other instruments I was performing with. In the process of coding the synthesis in SuperCollider, I added just a small sine wave chorus in the background so I had some context to sing with. I play a game with myself where I try to sing complex melodies over the drone of the microwave while heating up my lunch, so I imagined doing something similar in this improvisation; however, I found it was more interesting to focus on changing my vowel and resonant space over a singular pitch instead. I became a part of this background drone, trying to harmonise with it, but not doing very much over top of it, in favor of playing with formant placement instead.

8.2.2 From Unconscious to Conscious

As well, some movements of the suprahyoid I had not noticed until I started getting feedback from the activation during testing of the sEMG system. sEMG in this way can provide a basis for *re*action to a system and adaptation to the unexpected. My tongue moves down and out of the way when I breathe so I can take in more air and then moves again as I start to articulate any text. I had observed this while refining and testing the circuit, but I did not realise how prevalent it was until I began to sing with it. As someone who has been singing for quite some time, the act of supported breathing is very natural and I do not have to think how to get enough air and sustain a note for a long period of time. The beginning of phrases were modulated almost always because of this movement of my tongue after breathing. Certain consonants require more movement and so have stronger activation; for instance, plosives made with the lips, such as 'p' and 'b,' minimally involve the tongue compared to sharp consonants such as 't' and 'k.' This awareness of my breathing made its way into the improvisation as well, resulting in sharp or exaggerated breaths, exhaling without producing pitch, and using different consonants.

Being aware of both conscious and unconscious movements, especially those which do not produce any sound (such as the extreme chin movements I describe previously), are extremely useful when considering vocal technique and voice pedagogy. Many younger or inexperienced singers move in all sorts of ways which are incompatible with singing in a healthy and sustained way, but without the help of a voice teacher to physically point them out, they can go unnoticed. As well, the act of a teacher correcting posture or head positioning or some other type of movement during an exercise does not allow a student to examine their behaviour in the moment, but rather by comparing and contrasting as a before and after correction. Sometimes tension can be introduced through, for example, a jutting chin, and be heard in the sound; however, these types of unsustainable movement and overexertion are more likely to be felt in the long-term (Watson, 2009). Myoelectric potential captured through sEMG in beginner artisans is found to be unstable and power used unnecessarily, compared to experts, making this a valuable indicator of control or lack thereof (Hiyama et al., 2010). If this type of information were relayed through digital modification of a sound, as done here with muscular tension in my chin movements, this type of soundless movement could be relayed back to the student, who would potentially then notice their straying posture and correct themselves, as done successfully with other auditory biofeedback (Dozza et al., 2007; Okada and Hirai, 2019), leading to in-the-moment learning of action-sound relationships. In the same way, this reinforcement could provide additional information to educators.

In these couple examples of my interaction with the sEMG sensing, what was most striking was that, by drawing attention to one aspect of my technique, I actually found myself singing less, at least in a conventionally attractive way, and just paying attention to my movement. The tongue is constantly moving during most vocal exercises and diction and text clarity are common focuses in rehearsal; however, the movements of the tongue and chin need to also remain relaxed to keep resonant spaces and projection. I know how to open this space in my mouth and throat for a proper operatic or choral sound, and I don't usually think about it. When I gave presence to this aspect of my voice, I found myself putting it at the forefront of the interaction, often at the expense of any clarity or support in my pitch. Even though I designed this system to provide an alternative to audio-only voice interaction, I was surprised to find that I quickly abandoned some practices which I have been refining for years; in fact, I felt it was because I know so well what not to do while singing that I could actually supplement a number of new ways to appropriate these "poor" techniques.

8.3 Performing with the Singing Knit

With the development of the Singing Knit and planning future studies with other singers besides myself, I designed a different kind of sound for the vocal interaction. This was done within Pure Data and, rather than manipulating the vocal signal itself, used the sEMG signal to control a soundscape in which a singer could explore their action and movement without changing something about the vocal signal.[3] The synthesis is based on a physical model of a rubber duck by Christian Heinrichs (DC:A).[4] The model calculates the differential of the input signal which simulates the sound of the airflow through a rubber duck — the greater the change in squeezing pressure (based on the input signal), the greater the filtering (in this case, the squeak). For the vocal model, the differential of the sEMG signal is calculated and mapped to the cutoff of a highpass filter applied to a white noise generator. This causes a sort of *whooshing* when the muscle contracts and there is a large change in the signal. With another noisy drone, the result was an ambient filtered noise, which in my intention and design sounded a bit like a crackly or crunchy wind with changing intensity. The goal was to create a non-musical backdrop to represent the body, stimulating a sort of wind or breath with a

[3] The PureData patch can be found here: https://github.com/courtcourtaney/voxEMG/tree/master/examples/VoxBox

[4] Christian Heinrich's sound design work, http://dca.mt/, and the Rubber Duckie model: https://learn.bela.io/tutorials/pure-data/synthesis/rubber-duckie/

Figure 8.3: Performing a duet with Paul at the 2022 Augmented Humans Conference in Munich.

slightly scratchy quality to indicate tension within the muscles during contraction. Before passing this sound along to other singers to work with, I explored this soundscape in a few recordings and performances with the Singing Knit.[5]. In these cases, I also focused on the suprahyoid muscles.

8.3.1 The Body as a Partner

These performances come over a year after the previously discussed example (Figure 8.3); at this stage, the entanglement between the way I view my performance and the response from the system has been developing over the course of yearly three years. With such an extensive relationship, I find that working with the system has become extremely similar to singing without it. The Singing Knit allows me to focus on the interaction, rather than the device. As hoped in the design goals, it is more seamlessly a part of my performance and I do not worry about it as much as I used to; at least, not anymore than I would worry about another piece of gear or clothing when doing a gig. This is a level of connection which feels "normal," both from the comfort of having used the system for an extended duration and a better integration into a typical performance routine. The biggest difference at this stage is that I can trust the equipment in the same way I trust my body (Figure 8.4).

When I first began the design and testing of the sEMG, I was not always convinced it would work (and often, it would not). After furthering the design, solidifying some of the components through the VoxEMG and the Singing Knit, and spending a lot of time with it, I know more of its behaviour. In this sense, I do not think of it as a part of my body, but rather more like a symbiant. We are mutually entangled and depend on each other to a certain degree, as was the case in earlier use; however, there is now a sensation of freedom and trust between the two. Within the demos included here, I allow myself to be more musical and react to the improvisational setting, instead of performing with an aim to elicit the sonification. I am perhaps no longer trying to "prove" myself and the system, and this allows me to be comfortable with its response, rather than trying to force it to behave in a particular way. I have identified particular elements which I know elicit a response

[5]A demo of this sonification, presented at AHs 2022: https://youtu.be/grvRBR5DjRs

Figure 8.4: Performing with the Singing Knit at IKLECTIK Art Lab in London.

from the sonification, for instance changing the sharpness of my inhalation and switching into a head register and lifting within my mouth for higher notes. I feel that I have a sort of language with which to react to the device at this point — we respond to each other and I have tools to allow me creative freedom. While not a direct translation, I can expect some kind of response from the system. To me, this mirrors the way I feel about my body while singing in general.

The way I understand my body is not the same way I understand a person — we do not have a direct, spoken communication but I know and trust that we understand each other. As with the vocal sEMG, I may not get exactly what I want from my body, but I have a control and reciprocal relationship with it. If I want to run a marathon, I have to spend time with my body and make sure to take care of it during a long period of training (I did do this during my PhD, so I was able to make this comparison to other physical training). I can learn what does and does not feel right. Some days, I wake up with a "better" connection and I am able to act and react more to how I feel. In a musical sense, we do the same: singers, including myself, are notorious for monitoring their vocal health, behaving in a particular way leading up to performances, and practicing difficult technique until we *feel* it. Something switches in during this extended time period, wherein I have spent enough time with my body to know that I can trust it. When I approach the start line or the first bar, I don't worry so much about the 20-mile slump or the ridiculous melisma in measure 196 as I did when I first encountered them. With any luck, I monitor my physical sensations, the results of my activity, and trust that my body will behave the way we practiced.

8.4 Exploring Disconnect with Micro-phenomenology

Even though most of the interaction I have with my vocal sEMG goes as above, there are still moments which go wrong. The sensation of fighting against the body or not having our expectations honored by our action can cause a feeling of disconnect. Sometimes these things are within our control: perhaps we pushed ourselves too much the previous day by staying up all night to write a thesis and now are in a battle of will with our drooping eyelids. At other times, changes to our bodies happen without our consent: hormones, hydration, illness, and other physiological factors can impact our day-to-day relationship with our bodies and our activity. For singers, these factors can make or break a performance (I have twice heard of singers who purposefully scheduled their concerts around their menstrual cycles to ensure their laryngeal mucosal lining would be unaffected by hormones). Some days the body just does not cooperate with what we want, and this can lead to not only disconnect from ourselves but also feelings of guilt, frustration, and anxiety.

I explored a moment of communication breakdown between myself and the vocal sEMG in a musical performance with a micro-phenomenologist, Charlotte Nordmoen. I will present here the analysis of the experiential details uncovered in that interview, along with a reflection on my experience and the investigation we did of the moment of disconnect.

8.4.1 Exploring a Moment of Tension

I sang with the same vocal sEMG soundscape previously discussed during a duet with Andrea Martelloni, who performed with an augmented guitar in percussive fingerstyle.[6] This was the first time I had performed with the vocal sEMG with anyone else. During the course of the duet, Andrea and I played together and then took typical turns where one of us would take the lead and the other might fall to the background in accompaniment or remain tacet for a few phrases, giving the other a chance to shine. During the performance, there was a moment where, although I was attempting to create space for Andrea to play solo, the sonfication of the sEMG continued, seemingly without me (Figure 8.5).

During a micro-phenomenological inspired interview with Charlotte, we explored this moment where I realised the sonification was acting without my intention (Reed et al., 2022a). She summarises it:

> The experience begins with Courtney taking a step back from the duet for Andrea to have a solo moment. Courtney has her head turned slightly towards Andrea and notices him getting into the performance. At one point, Courtney hears the sound design from the EMG system. The sound is like a "wave," starting quietly and suddenly hitting her. She panics; she is not singing, so the system should not trigger. She goes through a mental checklist of why the system is triggering and realises that her tiny movements while listening to Andrea are the cause. Courtney immediately freezes and changes her breathing to eliminate all movement.
>
> We explore the moment of realising the interruption by the sound design in greater detail. Courtney describes the sensation as "almost like something grabbing and kind of pulling open [..] pulling my attention and so it starts to spread between these two places [..] the feeling that I have kind of inside where I'm hearing in my ears, it starts to be actually pulled away and towards these two different sound sources [..] it's the sensation of now I'm completely like ripped, ripped apart or divided in half and I'm no longer focusing a lot on either of them - there's too much of a divide and I'm hearing both things at the same time."

The interview uncovered a visceral physicality of having the attention pulled away towards something "going wrong" with the sEMG. When describing the sensation of hearing the "wrong" sound, I described the locality of the sound as being "inside the ear" and its effect, rather than the sound itself. There is a state of disconnect when the instruments behaved in this unintended way. Despite

[6] The performance can be found here: https://youtu.be/axn_wQM_I_c?t=3426

Figure 8.5: Andrea and myself enjoying a duet, despite the messiness of communication breakdown.

being generated by biosignals from my own body, but this disconnect resulted in a feeling that the sEMG had become uncontrollable and like a separate entity. Even though the body is controlling the DMI, the end feeling is that the DMI ends up controlling the body.

8.4.2 The Body as a Constraint

In addition to the summarisation of this disconnect being uncontrollable, I distinctly feel the sensation of being controlled, or indeed limited by my body. I am limited by my body's needs, its constraints, and its shape. During the evocation of the experience, I remember feeling frustrated and guilty at interrupting Andrea but being unsure of how to help it. It was easy to blame my system design at first, but upon realising that it was in fact my own movement causing the disruption, I was forced to constrain my movements to stop the sonification from pinging around. Although this is one situation where I was able to respond and provide control, the reflection caused me to think of other instances where we are beyond control and how this might create negative association. For instance, while studying in conservatory, I nearly gave up singing while working with a teacher who wanted my voice to sound a particular way and could not accept when I struggled to achieve this unattainable goal, getting frustrated with me instead of helping me to find my own voice.

I am reminded of one of the teachers interviewed about teaching vocal metaphor: when discussing kinetic imagery and visualisation, she commented on how it was strange that, although she could imagine particular movements, her arthritis and general physical capability would limit her from ever doing that activity:

> "I asked you about the visualising, because if I can imagine, I can't dance. I'm really awful. And I'm not supple. I can't touch my toes or anything like that. But I can imagine myself, you know, doing the most beautiful arabesque and so on. I can visualise that. And I can feel it. I just can't do it."

In a performance, one's voice might crack, we might run out of breath too early in a passage because we are tired, or we might simply not have the flexibility to sing higher notes on that day. Technology and devices can also fail, independently of our intention and our bodies (Bin, 2018). These are all normal parts of human life; things go wrong, and we often just have to live in the body we wake up with on any particular day. However, there are sensations of guilt, remorse, or regret that can come with this. We may also believe that we should conform our bodies to different kinds of interactions (Mice and McPherson, 2022). However, with the voice or any other kind of interaction, we can only do so much to extend the tools we are given; rather than attempting to blame ourselves, we must be able to accept that bodies are messy and often do not do what we want them to do (Klemmer et al., 2006; Spiel, 2021). Rather, it is important to embrace the parts we cannot control and work toward improving interaction and doing what we can with the things we cannot change.

8.5 Discussion

The system was developed to collect information about the singing voice, yet the experience of singing with it was completely changed to work with the system in this musical exercise. In this context, it is a perhaps a poor tool for collecting data on musical practice, which was not done at all in the end, However, we see that it is an excellent tool in allowing the user to practice manipulating their body or how they move in different contexts, to unpick parts of their practice, and to examine different physical and personal facets of their interaction. This resonates with ideas of postphenomenology or entanglement and using design practices in HCI (Ihde, 1975; Zimmerman et al., 2007), where we as users are inseparable from our tools and, through their use, they have the potential to change our goals (Frauenberger, 2019; Wakkary et al., 2018).

Rather than demonstrating any particular technical skill or performing in a way shaped by many years of experience or instruction, the intentions through performance became focused on this bodily relationship in the moment and on how the ingrained musical imagery and internalised movements could be manipulated. When applying the system in this way, the user and their embodied techniques and experience are inseparable from the system itself and they cannot exist alone; the sEMG system was designed around the vocalist's body, and the vocalist's practice then revolves around how the system reacts (Homewood et al., 2021). The system becomes a collaborator in this performance, assisting the improvisation by providing feedback about the body and sonifying performer intentions which have become background processes over time. As well, the expression of internal sensory experience can exemplify parts of our typical interaction in new ways, forcing us to examine them. The give and take between the user and the system as two agents creates a space to examine the perception of our actions.

Additionally, when designing tools and interactions, we must be aware of the constraints enforced by our bodies and the constraints we force through design. In choosing to sonify different parts of movement and interaction, we can bring attention to these elements and allow performers to explore their technique and even unconscious movement. This provides an outlet to become closer and understand our bodies better, as partners. However, it is necessary to note where some elements of movement or the body are out of our control; in bringing attention to them or providing them with a voice of their own, we can also run the risk of reinforcing disconnect we feel. When we design for individual bodies and work with these interactions, we should also work to embrace the mess and the mistakes.

8.6 Conclusion

Through my long-term relationship with this tool, I demonstrate how sEMG and biofeedback can be used to detect musical mental imagery use and sense vocalists' intention in performance. The development of musical imagery through experience allows performers to focus on their intention and high-level action, rather than on technique. With sEMG, musical systems can be based around this performer intention and use both conscious and unconscious body movement to form interaction. sEMG can be useful especially in bringing attention to embodied techniques which have become highly internalised and allow performers to play with their knowledge and experience in performance, and I demonstrate this potential through the changes in my own behaviour when confronted with the sonification. Highlighting a particular body-based technique for interaction in digital synthesis can shape the content of a performance and strengthen the performer's relationship with the body and use it for creative control. As well, such an interaction can highlight communication and perception of the body and the constraints of interaction from both our physical forms and the design of the interaction.

Chapter 9

Exploring Vocal Movement & Perception

Fundamental Techniques Perceived through Sonification

The final study of this thesis aimed to explore the perceptions of other vocalists when working with sonified laryngeal sEMG for an extended period of time. The study aimed to explore other aspects of the vocalist-voice relationship which became present to vocalists after interacting with novel feedback. Having already lived with the system for an extended period, I wanted to see how others felt about this feedback and whether the same understanding of technique arose for them as for me, or if different and new perspectives would be found in different individuals. Rather than seeking any kind of ground truth about vocal interaction with biosignal feedback, I aimed to explore how this technology would be received through individual perspectives.

Vocalists experience intensely personal embodiment, as vocalisation has few outwardly visible effects and kinaesthetic sensations occur largely within the body, rather than through external touch. We explored this embodiment using a probe which sonified laryngeal muscular movements and provided novel auditory feedback to two vocalists over a month-long period. Somatic and micro-phenomenological approaches revealed that the vocalists understand their physiology *through* its sound, rather than awareness of the muscular actions themselves. The feedback shaped the vocalists' perceptions of their practice and revealed a desire for reassurance about exploration of one's body when the body-as-sound understanding was disrupted. Vocalists experienced uncertainty and doubt without affirmation of perceived correctness. This research also suggests that technology is viewed as infallible and highlights expectations that exist about its ability to dictate success, even when we desire or intend to explore.

Portions of this chapter have been published in:

Courtney N. Reed and Andrew P. McPherson. 2023. The Body as Sound: Unpacking Vocal Embodiment through Auditory Biofeedback. In TEI '23: Proceedings of the Seventeenth International Conference on Tangible, Embedded, and Embodied Interaction (TEI '23), February 26-March 1, 2023, Warsaw, Poland. ACM, New York, NY, USA, 15 pages. DOI: 10.1145/3569009.3572738

9.1 Method

Lived experience is comprised of aspects that are internal and others that are external and also measurable. The affordance of examining these modes with the voice is in the interlinking of these experiences; both internal kinaesthetic and external auditory feedback provide the basis for the vocalist's understanding. The perspective taken in this paper is unique in that it shifts the perspective: the boundary between internal and external feedback is manipulated to provide vocalists with an external representation of something they would normally perceive internally. This re-introduces a familiar sensory perception in an unfamiliar way. This study aimed to investigate how the vocalists' awareness and understanding of their movement changed with the introduction of this novel feedback.

We therefore adopted a somatic approach in creating a design probe which would allow vocalists to interact with their embodied practice through novel auditory feedback. Given that the existing relationship relies on coordination of sensorimotor control and auditory feedback, we use sEMG as a way to externalise the internal kinaesthetic feedback. We worked with two vocalists as they explored their embodied vocal practices while engaging with this novel feedback about their movement. This study focused on co-exploration through a long-term interaction with this feedback, similar to the structure of the autoethnographic study mentioned previously in Chapter 8.

The singers incorporated the sEMG into their individual practice, having it become part of their routine, and documented their experience through an introspective review of recordings and journalling after each use. Additionally, I worked with the singers at different stages of their exploration in focused interviews (Howell et al., 2021; Mice and McPherson, 2022) and micro-phenomenological explorations. The goal of this study was ultimately to observe the new aspects of interaction which arose while working with the sEMG, how this impacted the vocalists' perception of their movement over time, revealed insights about their understanding, and enabled them to communicate about their experience. I present here the two vocalists' perspectives and interactions and then discuss in general how this long-term interaction revealed insights into their perception and relationship with their voices.

9.1.1 Participants

A call for participants was made through personal channels, including academic networks of voice scientists and music educators and musical social networks on Facebook and Twitter. I worked with two singers during the course of the study. Both were female, aged 29 and 31. The participants were from Brazil and Egypt and are now working as vocalists in major cities: Berlin and Barcelona, respectively. The vocalists were chosen based on their current musical activity: Vocalist 1 is a singer-songwriter and vocal teacher who also works from time-to-time in audio production. She performs Hindustani classical music and is studying Indian classical vocal techniques with another teacher on a weekly basis. Vocalist 2 works in computational music research and is pursuing a PhD in music information retrieval. She performs regularly and rehearses weekly with a small jazz ensemble. She has recently branched into generative electronic composition with an all-female computer music group in her city.

9.1.2 Materials

The VoxBox

After agreeing to participate, Participants received a kit for sEMG interaction at home — the VoxBox (Figure 9.1). We created the VoxBox as a probe to externalise an internal sensory experience: the

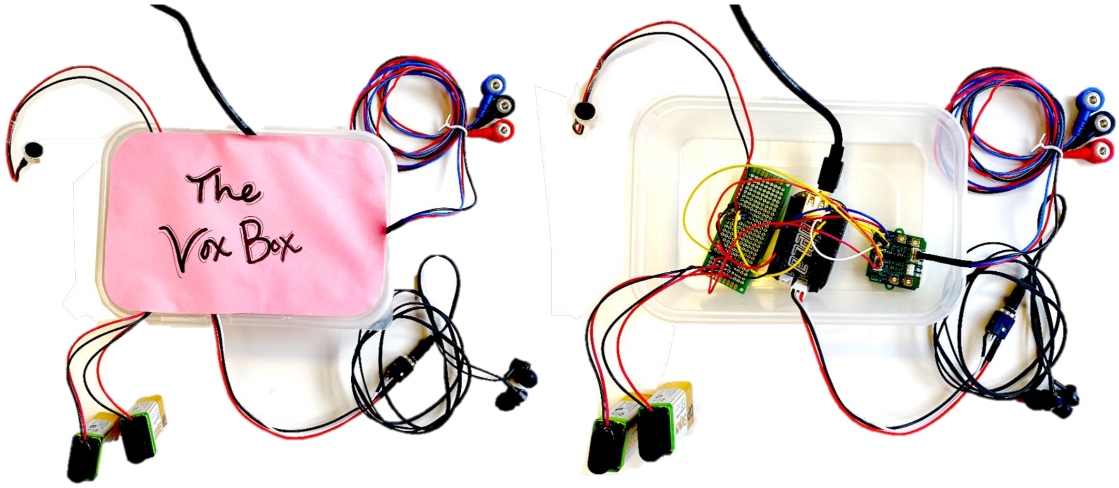


Figure 9.1: The VoxBox: external cables for electrodes, headphones, power supply (USB and 9 V batteries), and microphone (left), and internal VoxEMG, Bela Mini, and power routing (right).

control over the laryngeal muscles (Figure 9.2). Through providing a sonification of the laryngeal movements while singing and externalising this proprioception, we aimed to afford vocalists a novel context to experience their living bodies. The VoxBox collects the analog electrical signals of the muscle activations during singing and sonifies them, using the muscle contractions to generate sounds which can be interacted with in real-time by the vocalist. The VoxBox uses the VoxEMG board (Reed and McPherson, 2020; Reed et al., 2022b) and a Bela Mini (McPherson, 2017; McPherson and Zappi, 2015) for processing the sEMG data and rendering the sonification (Figure 9.1). The VoxBox allowed the vocalists to easily set up an sEMG feedback system using their personal computers so that they could sing in their usual rehearsal spaces, disrupting the habitual in the vocal action but not in the practice environment itself. The study components were enclosed within a plastic box so that the vocalists could focus on the interaction without needing to do any setup other than plugging in the power supply and their headphones, but were able to open the box if curious or if needing to troubleshoot. As discussed previously in Chapter 7, Section 7.6.2, we return to a traditional rigid electrode setup because of the time and skill needed to create, try on, and test wearable Singing Knit collars for the vocalists.

Included in the VoxBox kit are pre-gelled adhesive disposable electrodes (Kendall H124SG ECG electrodes, Cardinal Health), cabled electrode clips (CAB-12970 sensor cables, Sparkfun Electronics) for gathering sEMG signals, as well as kinesio tape (Kinesiologie-Tape, Altapharma) for securing the cables, if needed. The kit also included a pair of basic, wired in-ear headphones (Aurora, iFrogz) to ensure that the listening environment was the same for each participant; these are intentionally non-noise cancelling so that the participants would be able to hear themselves at the same time as the sonification played back from the Bela, effectively blending the external stimuli of the vocal audio with the sonification audio.

Participants also received a digital guide, *Working with the VoxBox*, to reference at any time (Figure 9.3). This guide details the components, how the box works, a tutorial for using the Bela browser IDE to run the sound design, and other setup/troubleshooting steps.[1]

[1]The full *Working with the VoxBox* guide can be found here: https://bit.ly/vox_box

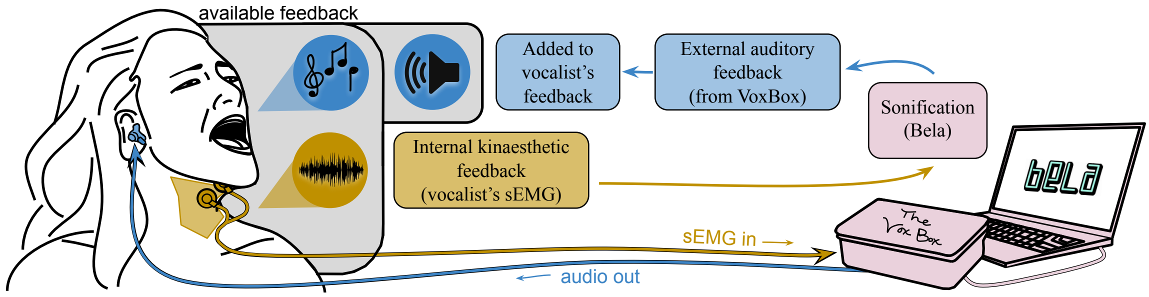


Figure 9.2: Using the VoxBox to externalise internal sensory experiences: the vocalist's internal laryngeal movements (internal kinaesthetic feedback, yellow) are captured with surface electromyography. The sEMG data is used by the VoxBox to generate a sonification (external auditory feedback, blue), which is added to the vocalist's existing auditory feedback of their singing (grey box).

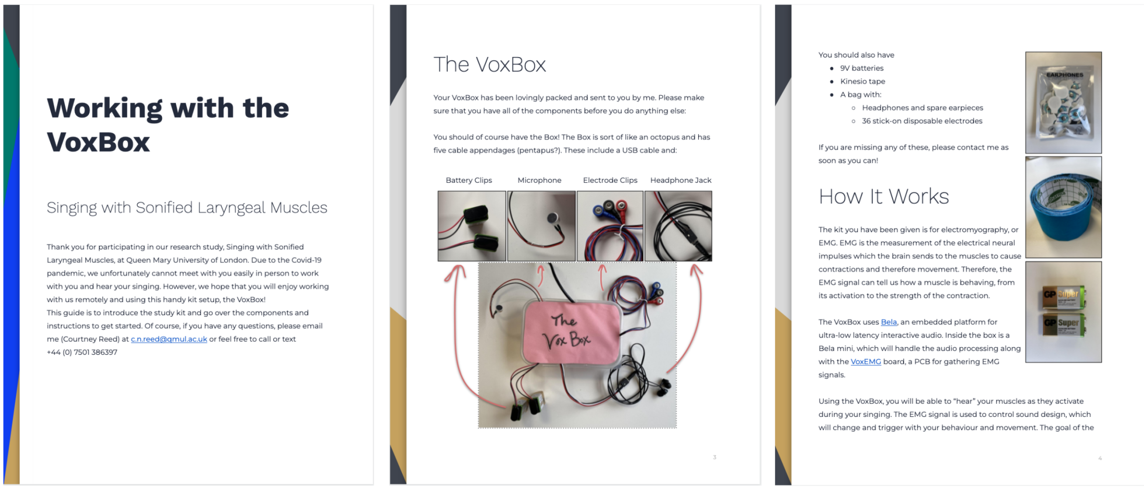


Figure 9.3: Pages 1, 3, and 4 from the digital *Working with the VoxBox* guide given to the vocalists, showing the study information, kit and its components, and setup.

Sonification

The sound design used for the sonification of the vocal sEMG was the same as the one I had tested when working and performing with the Singing Knit, as described previously in Section 8.3: the differential of the vocal sEMG signal is calculated and used to control a high-pass filter on a white noise generator, resulting in a *whooshing* wind or air sound (Figure 9.4). Again, the goal was to provide a non-vocal soundscape in which the body might have a sort of crackly yet airy quality about it, like a breath of its own. Through the sEMG sonification, I hoped to pull the body's movement out of the existing action paths and make it distinct, so that a vocalist might be able to interact specifically with this movement, where normally it might be unconscious or understood at a higher level in their action. The sound design was intentionally non-vocal and non-tonal to ensure that it did not interfere with whatever the singer wanted to do, and also to provide a degree of separation between the muscles and the vocalist's high-level understanding of their practice. This ideally positions the body, expressed through the sEMG sonification, as a separate entity and a

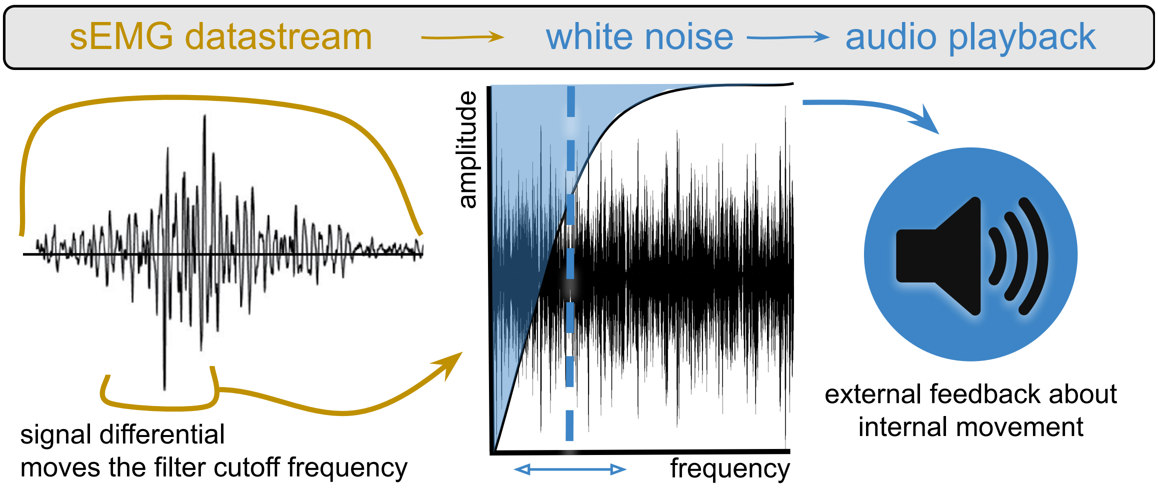


Figure 9.4: Translating the vocalist's sEMG signals to audio feedback: the differential of the sEMG signal is used to move the cutoff frequency of a filtered white noise. This audio is played back to the vocalist as an external feedback about their internal movement.

collaborator to highlight the control aspects the vocalist has with it. Additionally, I had extensively explored it myself, providing a frame of comparison to my own self-study.

Interactivity Questionnaire

Throughout the study, an adaptation of the Homebrew'd Musical Instruments interactivity and control questionnaire (Pigrem, 2021; Pigrem et al., 2022) was used to assess how participants' connection to and controllability over the sound design changed through the course of its use. This Interactivity Questionnaire used a ranking scale (Strongly Disagree to Strongly Agree) to examine the vocalists' feelings of control and connection, and how natural or comfortable the VoxBox was to work with. The full questionnaire can be found in Section C.1.

9.1.3 Procedure

After agreeing to participate, the VoxBox kits were mailed to the participants at home. All communication was done over remote audio-video calling on Zoom. After receiving the kit, I met individually with each vocalist for a Briefing session. This involved making sure all components had been received and that everything in the box worked correctly and going through the setup instructions as outlined in the guide provided. Participants were instructed on how to set up the VoxEMG and Bela with their computer and check that they could hear sound and that the electrodes were properly working by listening to the sound design changes while tapping them.

Briefing

After ensuring the kit functioned correctly after its trip in the post, I provided each vocalist with a short tutorial on identifying the position of one of the suprahyoid muscles, the geniohyoid, which helps to position the floor of the mouth and widens the throat for deep breathing and swallowing and the mouth for large opening positions. This consisted of the self-palpation exercise while opening the mouth and moving the tongue, to identify the placement of the hyoid bone and the muscle bodies. This was introduced to me by a speech-language pathologist. We continued this exercise together over video chat until the participants were able to recognise the muscle location with a little practice. The end- and mid-electrode sites, just above the hyoid and in the centre-middle of the flesh under the chin, respectively, were located with my guidance and the participants practiced placing the electrodes, feeling the movement while holding them in place and opening and closing their mouth, and listening to the sound design with this action. The vocalists were informed that it might take time to get acquainted with their body in this way and to spend time doing self-palpation as both a way to understand the muscles better independent of the sonification and to relax the muscles through pressure. The vocalists also practiced changing the position of the electrodes slightly in applications to see where they felt they had better response; as this was done remotely, this was done at the discretion of the vocalists as a way to connect better with their individual anatomies and to provide flexibility for exploration of the body.

Study Phases

The study then consisted of two parts: 1) an Exploratory phase, and 2) a Targeted Technique phase. Each phase lasted two weeks, for a total of one month of extended use with the VoxBox. For the Exploratory phase, the vocalists were asked to practice as they would normally, incorporating the VoxBox into their usual routine and practicing the repertoire and genres they would normally. The goal of this phase was to establish a connection with the sonification and provide a free-form exploration of the feedback. The second phase added an additional task through Targeted Technique. This involved beginning with a set of targeted vocalises (exercises for vocal warmup) which focused on the four vocal fundamentals outlined in Chapter 6: supported breathing, posture, sound production, and sound shaping. The vocalists were asked to complete these tasks at the start of each use of the VoxBox. The specific exercises chosen (Section C.2) were intended to cause noticeable movement with the suprahyoid muscles for the vocalists to focus on. After, the vocalists were free to continue exploring the sonification as they pleased.

With each use, the sonification script running on Bela would record the input signal voltages from the VoxBox, as well as the input from a small capsule microphone in order to synchronise the data with the singing. The vocalists were asked to record themselves with audio-video, for instance on their computer or on a phone, and then to save the data from Bela with their recording into a folder for each session. The vocalists would review their singing for reflection after the use. They were tasked to keep a basic journal with their use, to write down anything they had found notable during the interaction, their perception or feeling about the interaction, and anything else they wished to share.

Debriefing

In between each phase (end of Week 2 and Week 4), the vocalists conducted a debrief which lasted approximately 45 minutes. For Week 2, this consisted of a short semi-structured interview about how the exercises were going, initial impressions, and feelings about working with the sound. Vocalists chose a moment of interest, either a connection or a disconnection with the sonification, that they wished to explore in detail and participated in co-investigation of the experience through a micro-phenomenological interview. Finally, the vocalists completed the Interactivity Questionnaire while verbally discussing why they indicated each response.

The same was done for the Week 4 debrief, with the addition of the vocal fundamentals. The vocalists were asked to describe the vocalises as they noticed them with the sonification and answer a few questions on their expectations of the auditory feedback they received. This included discussion of how they might use the sound, as a metaphor, to discuss the technique in each exercise with their students. Finally, the vocalists were allowed to share anything else they wanted to conclude the study. In a follow-up micro-phenomenological interview, the vocalists were asked to recall a moment where they experienced a similar moment to the one previously explored (e.g., if the vocalist felt a sense of connection while doing a particular behaviour, another instance of this connection from the most recent two-week period was explored).

The full sets of interview questions for the two debriefing sessions can be found in Section C.3.

9.2 Analysis

Debriefing interviews with the participants were transcribed at the level of utterances. A bottom-up, inductive reflexive thematic analysis was used to organise the vocalists' communication of their interaction into an initial set of codes, which was then iterated through and organised into themes (Braun and Clarke, 2012, 2020a). For the micro-phenomenological interviews, the details co-investigated with the vocalists were first transcribed. Satellite dimensions — that is, moments where the vocalist slipped away from their evocation of the experience and spoke more generally or about other, similar experiences (Valenzuela-Moguillansky and Vásquez-Rosati, 2019) — were noted. The remaining details of the experience examined were structured to reveal the unfolding of the event in both a diachronic (happening chronologically in time) and synchronic (detailing a singular moment in time, expanding the depth of the experience) details of the interaction. From this structure, the aspects of different sensory modalities perceived during the experience are noted.

The Interactivity Questionnaire was analysed further, jointly between the change in the vocalists' ratings for each of the items and a thematic analysis of the responses to explore *why* these responses might or might not have changed.

9.3 Results

I will first discuss the vocalists' interactions separately, with respect to their indications and open-ended feedback on the Interactivity Questionnaire and micro-phenomenological interviews, and then, using thematic analysis, will explore 3 key themes which emerged in the vocalists' perception and interaction with the sonified laryngeal movement during the study period.

Overall, the vocalists reported spending about 6 hours (V1 = 5.25, V2 = 6.5 hours) working with the kit during the course of the month. I will narrate the vocalists' experiences in present tense, as they would have described the evocation during the micro-phenomenological inspired interview. Each specific experience is noted in **bold**. The structure of each experience is outlined and presented in a figure, where the x-axis depicts the diachronic succession of the experience. The y-axis depicts the synchronic depth of the sensory perception in a singular moment. I have used arrows to show how these small perception details form the larger overarching moment in the experience. As well, I have depicted these details by their modality: tangible sensations in red, auditory in yellow, and emotional characteristics in blue. There were no visual details uncovered when the vocalists and I inquired further about what they noted during the experience. If the vocalist was able to identify the location of these sensations somewhere in the body, this is also noted in a bubble placed above their description. It is important to note that, if we had explored a different moment or other aspects of the chosen experiences, the details revealed would likely have been different. However, we expect to see that the overall structure of repeated experiences might be similar or reveal consistent stages or aspects of interaction through the experiences (Petitmengin, 2006; Petitmengin et al., 2018).

I would like to reiterate that the goal of this analysis is not to uncover some sort of ground truth about work with biosignal feedback in this way, but rather to better understand the change in perception and relationships with the body as a result of interacting with such novel information about it. The interactions revealed similarities in the perception of the voice and ways in which such interaction can shape that understanding. It is important to note the relationship between this change and their individual backgrounds, genres of focus, and other related experiences which might have been brought into this interaction. Within the long-term interaction with the VoxBox, the goal was to build a relationship with the singers as they explored their own relationships with their voice. I will discuss each of the vocalists' experiences in detail before moving on to a more general discussion of the study.

9.3.1 Vocalist 1

Vocalist 1 (V1), while finding small connection in uncovering some of her unconscious movements, in general felt frustrated and struggled to incorporate the VoxBox into her practice routine. In her Interactivity Questionnaire, she answered that she felt strongly disconnected from the sound design and that she was not in control of it; however, she was able to connect her movement and the resulting sound and indicated that it influenced her movements during her practice. This relationship with the VoxBox and the biosignal feedback remained constant through the study and she indicated the same responses both times the questionnaire was given (Table 9.1). Based on V1's responses, her interaction with the VoxBox stagnated over the study period. This appears to be due to a mismatch between V1's perception of her body and her singing and the system's behaviour, which meant she struggled to negotiate the system's reaction and her expectations for it.

	Week 2 Debrief	**Week 4 Debrief**	**Change**
I felt connected to the sound	Strongly Disagree	Strongly Disagree	=
I was able to communicate musically through the sound	Strongly Disagree	Strongly Disagree	=
I felt in control of the sound	Strongly Disagree	Strongly Disagree	=
The sound felt unnatural to work with	Strongly Agree	Strongly Agree	=
The sound felt like an part of me, an extension of my body	Strongly Disagree	Strongly Disagree	=
I found the sound unresponsive and hard to control	Strongly Agree	Strongly Agree	=
I was able to make connections between my movement and the sound	Agree	Agree	=
The sound influenced the movements I was making	Agree	Agree	=

Table 9.1: Vocalist 1's responses on the Interactivity Questionnaire, demonstrating the lack of change in her interaction perception.

V1: Week 2 Debriefing

In the first debrief, V1 commented that she had been able to find a few connections with her movement. This was mostly related to movement of her neck and head while not singing; she often sits at her piano while practicing and noted that she would hear responses more clearly in her body sway while playing, with the sound design making a more pronounced "whoosh" as she moved around her environment. This made her more aware of her position while getting ready to practice and in accompanist gestures — the postures and movements performed to facilitate or support her singing, for instance moving her head or changing posture to help resonance in her chest — while warming up. She noted that, although she usually sits to play, she had not thought too much about how this was happening, focusing primarily on being relaxed above all else. Over time, she was able to connect these movements, which were otherwise unconscious, to the response of the sound design. Other non-sung vocalisations also produced a noticeable reaction in the sound — something was different when she spoke in a lower register while wearing the electrodes, but she elaborated that she could not pin down exactly what was happening.

While explicitly focusing on singing, the interaction was less clear. V1 remarked that she would practice her usual pieces but was unable to make any connections when focused on the repertoire; in response, she began to focus on some more specific techniques, such as vowel formation, placement in the resonant space of the chest, and tension and release in the throat (note that this was before the prompting in the second half of the study). She was able to make a few connections to the sound design, notably that she would hear when she was lower in her register. However, she struggled to make connections to the actions she was taking and what she wanted to focus on and found herself very frustrated. She commented that, at first, while she felt herself highly attuned to the sound design, the response was "mild." However, she also remarked that she expected this to be the case as she did not know much about the system yet:

> "I don't expect to fully, deeply understand it, because obviously, there's so much information I don't know... I was very, very sensitive about anything in particular, because it's new."

As time progressed, this feeling of uncertainty about what she heard culminated in uncertainty of the movements and her technique. V1 initially indicated that she perceived her singing as the combination of the muscles and tissues of her larynx working together. She describes this as the muscles making the sound of her voice, which she hears and is able to understand.

> "Muscle is very complex, because it's tissue... the combination of all of them creates this kind of sound that I hear. I try to picture it like this: so that you have three textures. And then they are all constantly mixing up with each other while I'm using my muscles, but because they are my muscles and it's the technique I have, I'm always getting the same sound, because... the main point is to always be relaxed within the technique."

This indicates that there may have been a fundamental difference between V1's understanding of her body and the feedback she received from the system: while attentive to her movement and the activation of her muscles, she was striving to be relaxed. Further, when talking about her singing, she described her voice not through her body and its movements but rather through her sound and resonance, noting that she feels resonance in her body more than her muscles moving. She remarked that she expected and wanted the sound design to match what she was doing in terms of pitch, mirroring her sound instead of producing this noise. This was because she wanted the sound to indicate to her how the muscle was moving and, ideally, to be able to hear different muscles making different sounds. The critical aspect was that she felt the sound should be pitched and resemble some other aspect of her practice; otherwise the sound was too loud and got in the way of understanding her voice:

> "Obviously, in the practice, the most important thing is that you really hear your voice and what you're doing in order for like, like understanding what you're doing with your muscles... If those three muscles had their own particular pitch, then I would know exactly which muscle is working. The sound that I get is like an analogue sound sort of like another static, which is like, constantly very abstract. When you get like a very determined pitch on the piano, you know, very clear."

The system does not respond to the pitch of the voice in this way, so it is easy to understand why V1 would have become frustrated as she tried to focus on this in her practice. Despite her understanding the reaction to changing her posture and small ancillary gestures at the piano, these were largely connections to movements while playing and speaking, but not singing. When singing, the examination of pitch response and audio cues of performance became critical and was not able to be met.

V1: Week 4 Debriefing

At the end of the study, the same issues remain: "I would notice that if I would speak, then something different would happen. Or if I would move, when I would hear that something happened, if I move if I bend it to the front, or the back or something, but if I was still, it was like a constant sound." After speaking during the previous debrief about the VoxBox's response to movement, V1 expressed frustration at both the system's lack of responsiveness to pitch and also at her own struggle to separate these concepts in her mind. Sound is V1's medium in her art and understanding of her technique, and the VoxBox's sonification of her muscles interfered with this practice:

> "[I would tell myself] okay, this is not responding to my sound, but it's responding actually just to my muscular movement, like physical muscular movement... if you would hear something that is imitating the same pitch that you're doing, following your rhythm, or doing a harmony instead, if you would have found the fundamental of the pitch. That also gives a feeling that it's relating to the sound that I'm projecting, and sound is the thing that I'm working with. I didn't see any relationship to it. That's why it felt unnatural."

Regardless of the specific qualities of the sound, V1 comments that it is the non-pitched aspect and its contrast to her voice which caused her to "have to divide my attention." She felt it was "really tricky" and that she had to "remove one side of the headphones and "uncover my other ear to actually notice and connect deeply with my sound and my resonance." She felt that, perhaps with more training or guidance in a traditional teaching setting that she could have better understood the sound design, as she did her own voice:

> "Perhaps I needed help recognising different pitches that the device could produce, so I could have paid more attention to all the sounds. I think I would have needed you to show me. This is how this it should sound when it's interacting with changes on your muscles, you know, and then I would have known what to pay attention to."

This mirrors back to her own teaching; although she speaks with her students about relaxation and fluidity in the body while singing, V1's students are also learning to focus heavily on their sound. She elaborated that, like in her own practice, she has her students use a tuner app while singing, using the app's XY plot GUI to get her students to "reach the line." At the same time, she remarks that "what makes a great singer more than the actual fact of being extremely accurate on the pitch. I don't think any I don't I don't think the being in pitch makes you a great singer. I think it's how you create the moment, and that has to do with the ability to connect with your body with yourself in the moment and just explore it." This suggests that, although there feels to be some dichotomy between the physical and audio presence of the voice, V1 explores her body and movement through her sounds. The control over and connection to the body is felt physically through this pitch-based audio representation.

In response, V1 suggested that, if not pitched, the audio feedback should be very different from the voice and have a percussive feedback, acting as an indicator or more of a cue when the muscular activation reached a certain tension or other point. This further adds to the desire for a guidance; such a percussive or cued feedback would provide a point of reassurance of the system, much like the tuner app. In this sense, if the feedback where "less of a [constant] pad, more of like a percussive frequency, then perhaps it would indicate, for example, if there was some more tension in the muscle or not."

V1: Micro-phenomenological Perspectives

When I present the micro-phenomenological accounts, I will narrate the vocalists' experiences in present tense, as they would have described the evocation to me. I discuss some of the details revealed through the evocation of these experiences. Each specific experience explored is noted in **bold**. The structure of each experience is then presented in a figure, where the x-axis depicts the diachronic (successive) structure of the experience. The y-axis depicts the synchronic depth of the sensory perception in a singular moment. I have used arrows to show how these small perception details form the larger overarching moment in the experience. As well, I have depicted these details by their modality: tangible sensations in red, auditory in yellow, and emotional characteristics in blue. There were no visual details uncovered when the vocalists and I inquired further about what they noted during the experience. If the vocalist was able to identify the location of these sensations somewhere in the body, this is also noted in a bubble placed above their description.

It is important to note that, if we had explored a different moment or other aspects of the chosen experiences, the details revealed would likely have been different. However, we expect to see that the overall structure of repeated experiences might be similar or reveal consistent stages or aspects of interaction through the experiences (Petitmengin, 2006; Petitmengin et al., 2018).

Experience 1: In a micro-phenomenological inspired interview, we explore some of V1's feelings of frustration and disconnect (Figure 9.5). In the first moment, V1 explored a moment where she was not able to find a notable response from the VoxBox when she was changing her pitch and moving between different registers: "**I didn't feel like I was interacting with the sound while I was doing something, which made it a little bit like useless to me. I felt useless for the device,** let's put it like that way."

V1 begins an exercise where she moves between her chest and head registers to capture the greatest changes in her range and explore the VoxBox's reaction. After a few alternations, she feels a sense of mismatch and notices no clear reactions in the pitch of the sonification, although she notes a "wobbling" in the audio feedback. She feels a sense of frustration and tries to consider why she is not hearing anything. When I ask her *When you are feeling this mismatch, what do you feel?*. She imagines her muscles as different pitches, where the muscles have different textures and layer over each other. They combine to make the tone of her voice; she hears them as separate tones which work together to form a whole sound. She wants the sound design to behave similarly, where there would be an indication of each muscle's movement (she imagines multiple muscles, even though only one is being measured) and similar layering to create a harmony. In the end, she hears only small changes in the noise and feels frustration that the sound is too abstract. This is difficult for her to explain, but she remarks that she feels nothing happens when she expects it to. In the end, she moves on to try another exercise.

Experience 2: In the Week 4 interview, V1 and I again explore a moment of doubt and feeling useless for the device. V1 questioned her technique while practicing the sound production vocalise and **wondered if really that she was not moving her muscles, because of her own fault in her technique.** We uncovered some tangible sensations and as well explored her emotional experience while this moment unfolded (Figure 9.6).

The moment begins with V1 singing the first phrase of the articulation given for the sound production vocalise (singing descended from sol to do on *ta*). As she repeats the phrase, she becomes more and more frustrated that she cannot hear anything notable from the VoxBox in response. She hears only the "static" of the noise generated by the sonification. In this moment, she thinks that maybe there is something about her technique which does not register or cannot be picked up by the voice. She worries that, based on the tension needed to belt and use her chest register most of the time in Hindustani classical music, that her muscles are not moving properly: "if they [the sounds] are not moving, that means that my muscle is not moving... I'm either hesitant to think that perhaps my technique is not great. That's why nothing's happening... Or like, it's just that the muscle is moving always in the same way. So therefore, also everything sounds the same." When we explore *What does it feel like when you feel the sound is not working with your technique specifically?* she tells me that she feels a physical hesitance or that her movement is smaller or less active than it should be. Emotionally, this is paired with a sense of doubt and of personal misunderstanding of how the box is meant to work. In the end, the moment finishes with her conclusion that the VoxBox must not work with her specific vocal practices.

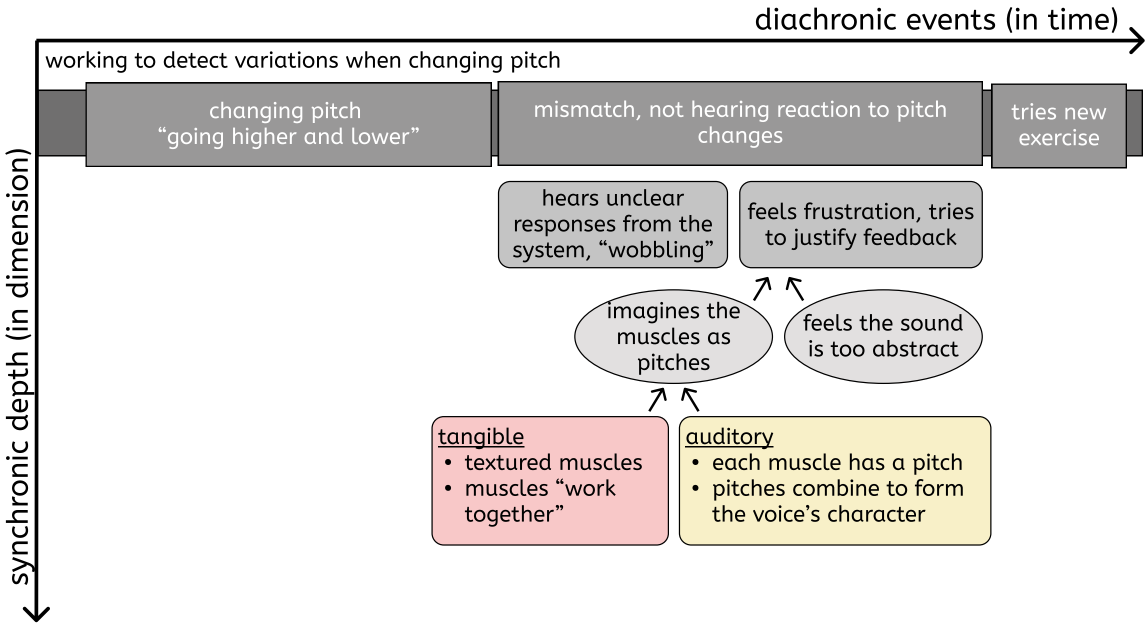


Figure 9.5: Experience 1 (Vocalist 1): V1's sensory perception during an experience of mismatch and feeling "useless" about her interaction while exploring feedback for her register switches.

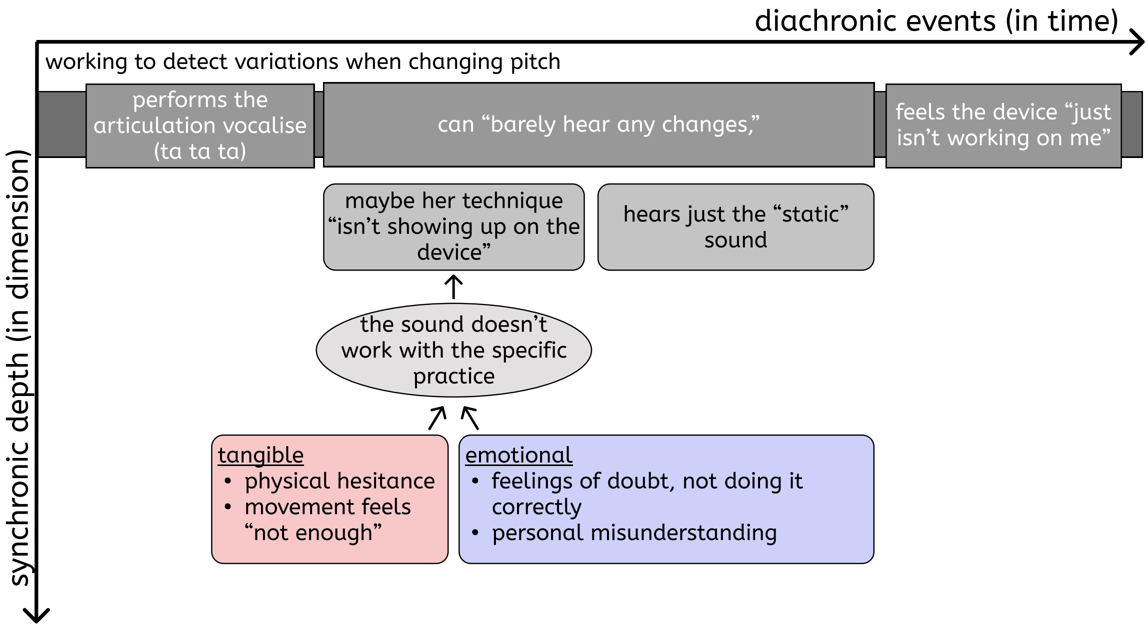


Figure 9.6: Experience 2 (Vocalist 1): V1's sensory perception during an experience discerning the sEMG activation during her breath before a long phrase.

9.3.2 Vocalist 2

Vocalist 2 (V2) was able to connect better to her body and the sound while using the VoxBox. In her Interactivity Questionnaire, we see that her connection improved over the course of the study. (Table 9.2). She felt more in control in some moments than others, but overall more in control; additionally, the sound design felt more natural and she was able to make more connections between her movement and the sound over time. She remarked that she felt she was not able to communicate musically through the sound in the end; this was related to V1's experience, wherein V2 struggled to multitask with the auditory feedback from the VoxBox and her own voice. As well, V2 felt the sound influenced her movements less over time, due to the nature of her connection with the system and also physical constraints of the design, which inform future versions of the VoxBox.

	Week 2 Debrief	Week 4 Debrief	Change
I felt connected to the sound	Disagree	Agree	+
I was able to communicate musically through the sound	Neither Agree nor Disagree	Disagree	−
I felt in control of the sound	Disagree	Neither Agree nor Disagree	+
The sound felt unnatural to work with	Agree	Disagree	+
The sound felt like an part of me, an extension of my body	Disagree	Disagree	=
I found the sound unresponsive and hard to control	Neither Agree nor Disagree	Neither Agree nor Disagree	=
I was able to make connections between my movement and the sound	Disagree	Agree	+
The sound influenced the movements I was making	Agree	Disagree	−

Table 9.2: Vocalist 2's responses on the Interactivity Questionnaire, depicting some positive (e.g., being more connected to the sound, feeling more in control) and negative changes (e.g., feeling more certain about being unable to communicate musically).

V2: Week 2 Debriefing

V2's initial experience was of intrigue and exploration around the spontaneous connections she was able to make between her movement and the audio feedback. She was able to make connections between activity in her body and an increase in the activations she heard in the sound, that when she was engaged and started to sing. Specifically, she attributed what she could hear as a tension, for instance in the resistance of holding long notes for varying time lengths. This was also associated with a relaxation or movement in timekeeping or acting as a "metronome" with the body, wherein the looseness and tension "would help in also these muscles moving."

However, she worried a bit about her other non-vocal gestures interfering with what she felt were more technical muscular activations expressed through the sonification, that "they would like confound, I don't want to stop the movement." This might suggest that, although the muscular activations are still laryngeal, V2 perceives a separation between accompanist gestures such as postural shifts, figurative gestures like ancillary body sway with the groove, and effective gestures to achieve vocalisation specifically (e.g., those "technical" muscular activations she wanted to focus on). This separation was related to a struggle V2 normally feels in her practice, where she finds it hard to remember that her physical tension and reducing strain on her body are still, if not more, important when she focuses on the execution of more difficult technical passages. Being aware of this relaxation helped V2 to become aware of the unintended tension in her body:

> "When I'm singing a hard passage, and I move a bit, like I [try to] liberate. I always felt that the harder and the stiffer I should be, I can focus on getting it correct. And I sometimes realise it's actually it's the other way around. If I move a bit, then actually, I don't know, something in my body makes me able to sing that passage."

However, V2 described her relationship with the VoxBox at this time as only a "basic communication." She wasn't sure if she would ultimately be able to musically communicate through the sound because of the spontaneity of the relationship. Often, she would be able to hear a response in the auditory feedback but unable to determine quickly enough in the moment what she had done to influence the system. When trying to go back and find that activation again, V2 noticed her consciousness and efforts to recreate the behaviour removed the natural approach she originally had, making it hard to receive the same interaction from the system — the awareness undoes the instinctive behaviour.

> "I heard a change and I tried to seize it later. So it was like I was, I can't remember what I thought at the beginning. Because at the beginning, I wasn't thinking. I reacted to it."

> "When I'd be singing, like the chorus of a song twice, I'd feel that in some specific part, there could be something [an element of the audio feedback] that keeps happening. But of course, because I repeat that, it doesn't happen again. When I'm spontaneously singing it... I guess your body will change. I mean, you will not be singing it as spontaneously as you were... Every song and every day has its mood."

As with V1, we see V2 also attributes difficulty with the system to a personal failure: she wonders if her technique is lacking in complexity or is too comfortable and not engaged enough, resulting in missed connections with the system. In spots where she things she should hear a response from the VoxBox, for instance in her breathing or in her register changes, she feels as though she has limited control over her voice. In some of her breathing exercises, she is able to match the expansion of the breath to the sound feedback; at other times, the response is more ambiguous and "harder to interpret."

> "It's hard to control my voice. I was starting to wonder if maybe the way I sing doesn't, because like I'm an amateur singer at the end, like I'm not training my muscles. I'm not breathing the right way. I'm just singing [standard] works."

V2, although a very active vocalist, took her lack of professional training as the reason for the system's varying response to her movement. She remarked that she worried she was singing in a non-engaged way and was being too "lazy" about her body and movements during her practice.

> "I hardly ever feel the need to stand up in order to sing in my jam sessions. There is no microphone, and I'm just sitting crossed [legged], singing like this. So I mean, this probably says something. Also, it's maybe I'm too comfortable in that singing style."

V2 also struggled at times to work with the feedback as a sonification. For her, she felt distracted when trying to listen to both her voice and the sonification, adding to the feeling that she would miss the spontaneity of the interaction. This distraction caused a "sense of anxiety" and she wondered if a visual representation might be helpful although "even though it would be harder" in terms of concentration while singing. As well, she remarked that the feedback being noise felt unnatural, but that she understood why this was the case:

> "If it's very pitched, obviously, it would intervene with your own singing and you don't want that. So it has to be something that would never come from you naturally. It has to be something sort of external that. It can't be musical.

V2 was unsure of what qualities the sonification should have, but the suggestion that the sound should be something separate and unnatural from the voice suggests there is some difficulty in interacting with the muscular movement at the same time as the usual audio of the voice. The "external" quality should not interfere with the voice but perhaps is viewed as parallel to the existing auditory feedback.

V2: Week 4 Debriefing

By the end of the study, V2 was able to form several concrete connections with her body during her practice. She felt that the system was easier to work with after having spent more time with it and felt that "it's definitely responsive. It's a bit hard to control, but it's not unresponsive. I could hear lots of variations in the sound." She responded that her movements became more pronounced during her practice as she sought to trigger more response from the system, although this was more tricky to do while singing difficult passages as she worried it would be too much movement. Especially in warming up and working with her posturing, she noted that she moved more and her "movements are exaggerated... it's very much easier to hear sound" when playing with her body and positioning in this way.

As follow-up from the previous debrief, she also uncovered aspects of her auditory relationship with her voice. For V2, listening to herself is very critical during her practice and she felt it difficult at times to focus on the sonification because "I'm just so accustomed to always being listening to my pitch, that I cannot listen to anything else. I just can't." This was a revelation for her, in comparison to the previous two weeks of practice, as to why she did not notice some sound changes easily or quickly, compared to others:

> "I think remember what last time, I was telling you that when I take a big breath, this is when I hear some change. It's because, when I'm taking a breath, that's when I'm not singing. And that's when I can hear... In the beginning, I was saying, 'I'm not hearing any changes,' no, there are subtle things happening [while singing]... I think I have a good ability to drown out sounds, which is something I do even like, if I'm concentrating."

V2 was able to make more connections in the moments where she was not actively listening to her voice (e.g., while breathing and working with posture). This did help her to become more aware of her breath and the regulation of her air through the breathing vocalises, where she was not singing:

> "I'm going to do eight counts now because I want to run out of breath... It helped me understand what's happening. How so? Just split in the sense that okay, it's not, it's not the breathlessness. It's the silence."

Rather than focusing on her feelings of breathlessness, she found she was able to focus on the sound versus silence aspects of the sonification. This helped her to be able to tell where her focus was as she increased the duration of her extended phrases.

However, she had some other feelings about using the VoxBox in practice and, in terms of feeling as though the sounds were an extension of her body, she noted that "it's not quite where I want it to be." This connection was beginning to form and she felt that, with more time and work to refocus her attentiveness, the connection could be improved to this stage. The ideal for this extension, as in my interactions, would be to be aware of the relation between movement and sound and begin to associate the sonification as an externalisation of herself which could be interacted with in a novel way. She felt that, in order to really qualify as an extension, she would needed to understand the feedback she was hearing in a more in-depth way and be able to know more specifically what was influencing the sound design:

> "I could hear [the changes] but like 'extension of my body' is a level of knowledge... it has to be like a higher level of knowledge. That's why I didn't feel that yet. Although with more time, I'm not denying that it could be felt."

Again, the idea that the addition of the sonification was tricky to reconcile with the existing vocal audio remained. V2 felt again that it would be better to do something in a different "channel" that where she was already focused:

> "What's the best form of feedback? I think if it's the same channel as the other thing that you're actually doing it becomes really hard. Yeah, it's like if someone is doing something visual, don't give them visual feedback, give them other feedback."

Similar to V1, V2 also discussed a bit of her desire to have confirmation that she was using the VoxBox "correctly." Although she was able to make new connections between her action and sound and felt she could understand her behaviour better, V2 would have liked to have an external reassurance. She suggested having some kind of cue, similar to V1, that she was singing in an appropriate or ideal way:

> "I don't think that our bodies are alike so I don't think it's possible because of different references and other things, but if it were possible to get a beep when I'm in like, 'Okay, you're good,' like a calibration?... It will be nice."

She also mentioned that having others to work with or a social aspect to learning and understanding the device would be beneficial, particularly when she was feeling lost or uncertain:

> "To sort of see and understand what happened with other people, what other people are saying, and then sort of to build my expectation a bit would be helpful. Because even though that might bias the way I think, and I might be reinforcing things that I've read, at the same time, when a when you're feeling a bit lost, it's nice to grab onto something."

> "Generally I would have liked to do this study with another participant like, putting us in pairs a bit. Yeah, maybe with another participant, just to get their feedback or to what potentially is not an option. But yeah, in a more collaborative fashion. Because sometimes it felt a bit lonely. Like, sometimes I didn't feel motivated."

This highlighted some of the need for reassurance with one's own body, to be able to gauge expectations and results according to what others were feeling. Although V2 understood that different bodies are not alike and acknowledged that it could bias her own perception to hear from others, she felt the benefits of communication about the experience in moments of feeling lost or lonely in her experience would outweigh this potential risk.

Finally, it is worth mentioning that V2 felt at this point that her movements were less influenced by the sound she heard, although she connected them better. There was a sense of safety in exploring her breath and posture, and moving to other interactions was limited at times by the equipment itself being a bit precarious. She understood that "obviously, it's a prototype" but that the setup process made her a bit apprehensive to move "as much as I would have liked to" or use it in contexts outside of her home, for instance with her ensemble. This affirms some of the original concerns presented in Chapter 7 during the development of the Singing Knit. When mentioning this to V2 after the conclusion of the study, she remarked "that's a dream like to wear."

V2: Micro-phenomenological Perspectives

Together with V2, we used a micro-phenomenological interview in each of the debrief sections to explore her connection to her breath, which was made more obvious through use of the VoxBox.

Experience 1: In the initial Week 2 interview, V2 was exploring the sonification and still making this connection to her breath. She remarked that, when she sang longer notes, **she noticed a change in the sound but she couldn't quite pin down what was happening.**. We further explored the sensory interactions that made up the moment of her noticing this interaction — what was going on at the time of her realisation (Figure 9.7).

First, V2 prepares her breath to sing a longer phrase. She notices that there is a change in the sonification, and this dies away as she begins vocalising. When I asked her *How do you feel that has something has changed?*, she says that she hears something "out of the norm." She is unsure initially of what that is, but notices the departure from the constant noise and then "tries to seize" the cause of the sound, but is unable. She determines that there is a change and hears an audible "rise and fall" in the sonification. Interestingly, she remarks that "her brain decides, not me," again reiterating that this is a more physical response than a cognisant or even auditory sensation. When I ask her *What do you feel when your brain makes this decision?*, she replies that her body responds — the decision of her brain is felt in her body. we further explored this sensation in the body as having a physical reaction, albeit not an easy-to-describe one, and that there is a "knowing" in her body as it "senses a change." As well, this change is felt with a notable emotion of happiness and V2 thinks to herself "I got it to work!" She does something that the VoxBox has picked up on and feels a satisfaction as she receives the response from the sonification. The experience ends as she notices the response dies away after she begins to sing.

Experience 2: In the second interview, another experience with the sEMG sonification around the breath is explored: V2 has begun to uncover her connection to her breath being a result of her changing focus while singing. In moments where she sings, it becomes harder to focus on the sonification as she "tunes out" other feedback besides her voice. We explore a moment together where **she realises her shifting focus when she is breathing before beginning to sing, and "that's when I can hear"** (Figure 9.8).

V2 knows that her breath is something which the VoxBox responds to. She is working on one of the breathing vocalises to test this interaction in repetition. She begins and takes an intentionally "big breath." She hears the sonification and again connects this to her movement. She then feels a sense of contrast, going from a relatively silent and continuous sonification to hearing the sound design. She realises this contrast is obvious without the sound from her voice. She notices this silence and focuses her attention to the sound design. In this moment, she notices more subtle changes in the sound design. When asked *What do you feel when you notice these subtle changes?*, she is aware of her posture and small movements in her body. The awareness centers in her neck and shoulders. She comments as well that she knows these small adjustments to her posture provide easy-to-hear responses from the VoxBox, although it appears this comment was a more general satellite dimension — rather than being specific to this experience, it is a justification she makes based on her previous encounters. Her focus feels different than normal. She concentrates on the details of the sonification becoming clear. When I ask her to explore *What do you feel when this sonification becomes clear?* she feels that her attention shifts as she breathes and then just again before she begins to sing. The experience ends when she begins to sing and the sound of the VoxBox is removed or "tuned out."

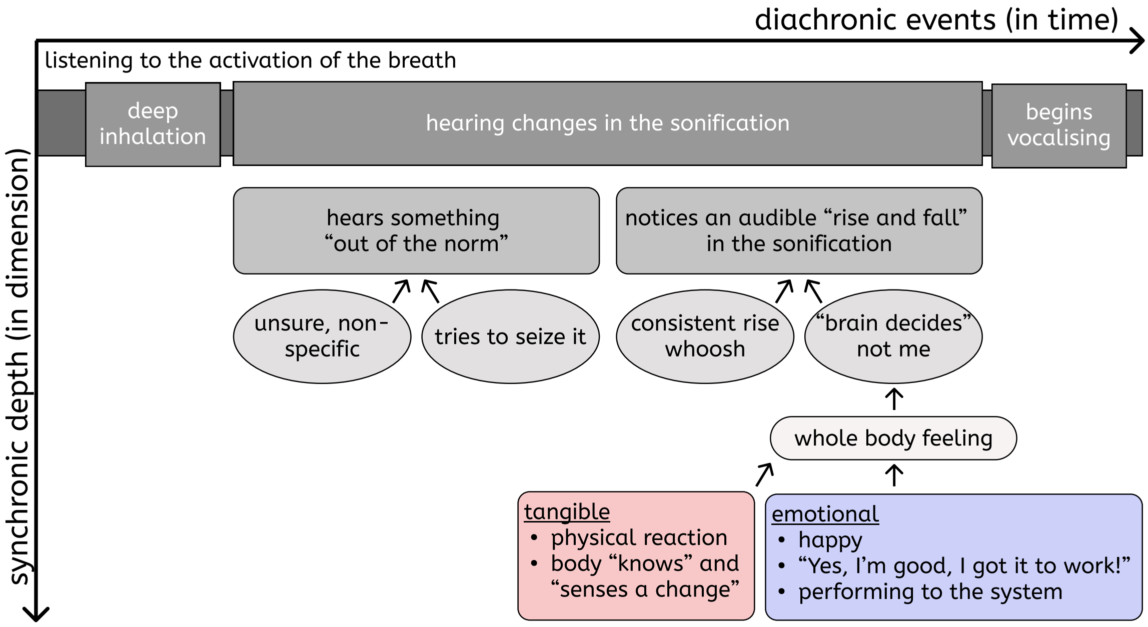


Figure 9.7: Experience 1 (Vocalist 2): V2's sensory perception during an experience discerning the sEMG activation during her breath before a long phrase.

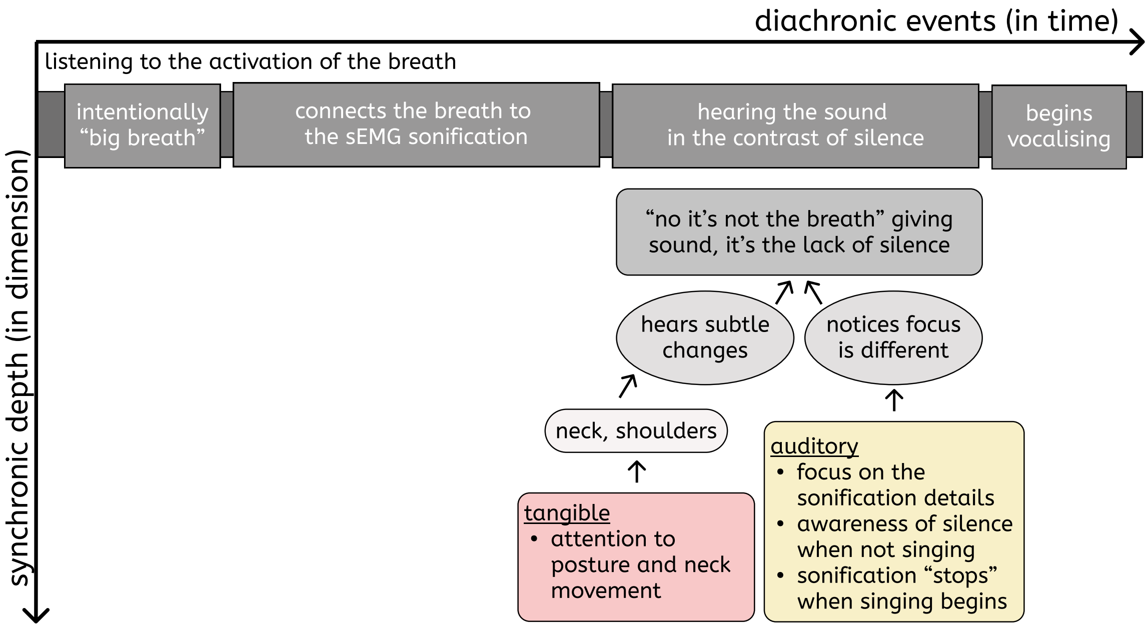


Figure 9.8: Experience 2 (Vocalist 2): V2's sensory perception during an experience where she feels her changing focus while breathing.

9.4 Discussion

Thematic analysis of the two vocalists' interviews revealed three main themes, which I have named 1) *The voice is its audio*, 2) *The necessity of community, encouragement, and "correctness"*, and 3) *The infallible technology and the body and self to blame.* These themes capture many of the individual points in the analysis highlighted above, and reveal further detail about similarities between the two experiences for discussion. In a reflexive manner, I again use my experience in the vocal context to organise the results; as well, I detail how the participants' specific musical backgrounds (Hindustani classical and jazz/contemporary improvisation, respectively).

9.4.1 *The voice is its audio.*

This theme captures some of the major difficulties both vocalists found in working with the VoxBox. Feedback about the body was delivered through sound; however, this created a barrier to understanding at times because, as indicated by both of the vocalists, it occupied the same sensory channel they were already focusing on — the auditory feedback of their voice. The VoxBox was designed for auditory feedback intentionally to operate in the same channel already being used, to avoid forcing attention to another sensory modality. Again, the sound was intended to be as non-overlapping with the vocal audio as possible, using filtered noise as its base and having no pitch or percussive aspects which might interfere with what the vocalists were singing.

V1 expected the feedback to respond similarly to her voice in terms of pitch; for her, this would have provided more valuable feedback about what she was doing. For V1, sound is her media for creation: she describes it as "the most important" part of her practice and needed the sonification to relate more directly to the sound she was projecting in order to feel natural. V1's primary practice in Hindustani classical music revolves around her pitching; she uses a tuner for constant focus on the demands of the specific raga she is practicing. As we see from the micro-phenomenological analysis, V1 also had an existing image of her voice and how her physiology worked. In her mind, the most important components of her interaction were based around her sound. In some sense, as we see within the described images of her muscles combining to create the timbre and characteristic of her voice, she understands her body *as a sound.* To her, her body occupies a very important in both her own practice and in teaching her students; yet, we see there is a massive importance also placed on the pitch of the voice, rather than the kinaesthetic experience itself.

V2 expresses a similar feeling of distraction and inevitably tuning out the sonification at times when she was singing. We see that she was able to connect to the sonification well when she was not actively vocalising, for instance in her breathing exercises and when working on posture and alignment. She remarked that the feedback should be given in another channel, for instance as a visual representation; even though it would have been "harder to learn," it would have been separate from the audio feedback she was already listening to. Her focus was also very divided at times.

This further affirms that the understanding of the voice is entangled with its auditory feedback, as is the basis for the sEMG as a form of direct control. Both vocalists were very conscious of their bodies and felt that this connection was more important than anything else, yet the audio feedback was how they understood what they were doing. They both felt additional feedback about the body *should not* be communicated as audio. If they were relating to their practice and understanding their voices and bodies through aspects of her timbre and pitch, then it makes sense that the vocalisation would be disruptive to their existing imagery. The sonification of muscle movement is "too abstract," as a sound, compared to the existing feedback they are normally focusing on.

This suggests that the vocalists use audio feedback as a metaphor already; they have found a mapping to what they hear and know how to react physically, even if they cannot describe verbally what that reaction is. The audio feedback is an explanation for the physical action; in a sense, when hearing X the vocalist knows to take one course of action, and when hearing Y, they know to do another. The existing vocal audio feedback is, in these cases, being used as an explanation for and is entangled with the physical action. This seems to be a sort of sensory translation process (Ekman and Rinott, 2010; Wirfs-Brock et al., 2022), which mirrors the different perceptions and descriptions of vocal interaction discussed so far in this thesis: Aspects of the sound utilise a tacit understanding of what is going on internally. This is then described or communicated to others through abstract metaphor. Competition for attention appears to disrupt this sensory translation for understanding. In moments where there was no active vocalisation for V2, using audio feedback might not have been so disruptive because there was no pre-existing attention to sound. Rather, V2 explained that she was focusing on her movement and feelings of relaxation in her posture. Perhaps, with attention to the physical feelings of tension and relaxation, the sonification was able to provide a parallel source of reference, making it easier to understand and react to.

Overall, this theme perhaps leaves us with more questions than answers; although we see in Chapter 5 that audio is not the only sufficient source of feedback for vocalists in their performance and in Chapter 6 how vocalists rely significantly on tactile and visual references to understand and teach the voice, it is clear vocalists critically rely on hearing their own voices. In this sense, perhaps less of audio feedback's importance is noticed because it is extremely difficult to talk about normally. Vocalists hear their movements, as it were, but explain them through visualisations or movement. V2 was able to connect some specific sound characteristics of the sEMG feedback to her movements, which demonstrates that sound feedback can provide auditory images to accompany kinetic experiences; similarly, although V1 was not able to make a specific connection, she did mention that it might be useful for the VoxBox to convey muscular tension in the body as consonant and dissonant chords instead. Both vocalists also mentioned briefly that having a percussive sound might also work, as it was extremely different from the vocal sounds and might be better to act as more of a "cue" to muscular activation, rather than as something musical to interact with.

This suggests that the vocalists are actively making connections and building imagery between their sensory experiences in different modalities, which can provide a reference for this more abstract(Wirfs-Brock et al., 2022); however, providing an additional sound to an action pathway already so reliant on sound was difficult for the vocalists to understand, even though the sound design was intended to be as "non-vocal" as possible in using a non-pitched noise as its source. This may also explain why V1 wanted the feedback to behave as her voice did, perhaps to then match the existing way she understood her practice. For V2, she did not give attention to the feedback, similar to what she does with other sound sources while singing; her musical background means many of her performances are in noisy environments. Perhaps moving towards tangible or visual interaction might be the most beneficial for some singers who rely more on this existing audio imagery. Even in this case, where the auditory interaction was designed to occupy a space around the vocal audio feedback, the addition of new auditory stimuli forced the vocalists to divide their attention.

The voice seems to then be understood inseparably from its audio. Although the singers were very cognisant of their bodies when describing their practice, they struggled to work with the separation of the laryngeal movement and the disruption from the sonification. The audio is heavily responsible for the innate understanding of the voice, almost to the point of exclusivity. When describing the voice in her micro-phenomenological interaction, V1 relates the movement of her muscles to tones and pitches; V2 describes her awareness as a change she can hear based on something her "brain decides." The link between movement and fine-tuned motor control is determined on sound, linking the awareness and sensory experience of the body to its sound is a very fuzzy, overlapping way. This reliance on audio feedback for motor control and understanding of physical interaction is a sort of translation between the senses; this is not a sensory experience found in other places in the body and is unique to the voice.

9.4.2 *The necessity of community, encouragement, and correctness.*

In the vocalists' feedback, we see also the need for reassurance and correctness in their movement. I say "correctness," because this is a difficult aspect of vocal pedagogy to assess: different bodies move in different ways and the experience is unique to each vocalist; in Chapter 6, we see how vocalists use metaphor to approximate this experience, leaving a wide range of "correct" behaviours. As well, singing practice is usually meted on the resulting sound, rather than the movement of the body (except in cases where it is overtly visual that a behaviour is less than ideal, such as with poor posture).

Despite the study being explained as a chance to explore their relationships with their body and being reassured that there was no expected behaviour or outcome, both V1 and V2 expected and wanted to have some kind of affirmation that what they were doing was correct. This might suggest some kind of participant bias, where the openness of the study directive left the vocalists wanting to make sure they were hitting the mark with their participation (Howell et al., 2021), but might also suggest that, especially when learning a new interaction method or practice, the reassurance or confirmation that what they were doing was more important than how connected they felt with their own body. V1 usually measures her performance with a tuner or against the piano to attend her focus on pitch. V2 relies often on her ensemble to gauge her practice and is typically operating in a semi-structured environment where the musicians each supply improvisation ideas and direction. For both musicians, the task might have been generally daunting considering the specific demands of their respective genres and limited experience with completely free-form exploration. This is similar to my own experiences breaking out of my classical Western background; it is difficult to create outside of the boundaries of where we have been trained, as this creation is inherently conflated with what we view to be correct and "good."

Both vocalists commented that they would have wanted me to show them where their movements were and what to pay attention to; V2 also discussed her desire to work with others while using the VoxBox, despite knowing and acknowledging that it would bias her. Again, reflecting on the difficulty of vocal pedagogy, a teacher can only guide while working with the body; although there are certain things to be aware of, the teacher would not be able to tell the student a yes or no that they were moving and acting correctly because the correctness is too varied. A vocalist must be able to learn what is healthy or comfortable and what is not; some singers have different physiological abilities than others, making the mark of "correct" a constantly moving goal post. In the study, I did act as a teacher when introducing the feedback and when leading the vocalists through the exercises they did in the latter half of the study. However, I would not have been able to tell them explicitly where they would feel a connection (doing so would have likely biased them to find or force connections), and in a realistic sense it was uncertain, from a researcher standpoint, where these connections would happen anyway. It was as much an exploration from the research side as it was on the participant side to see what connection resulted from the interaction. My own experience with the VoxBox is unique to my body, imagery, and other perceptions while singing. As with other teachers, I was only able to guide, not confirm or score the participants.

However, the need for this reassurance and indeed the community aspects are very intertwined with learning and technology use in unfamiliar contexts: we learn from watching and mirroring others' behaviour forms a good deal of our own practice. In this sense, we are biased towards what we are taught or what others are doing; perhaps by providing the singers with a specific reference or example, I might have been able to teach them to listen for particular sounds and learn the sonification as a data source (Wirfs-Brock et al., 2021), but there is no real ground truth to this interaction and it would have likely influenced what they heard in their own interaction. This highlights how the feedback and information provided in the learning process influences perception of individual, personal parts of our lives, and indeed our bodies themselves (Homewood et al., 2021). If it is important and natural to seek confirmation from others about individual experiences, particularly those involving the body, it is important to make sure information is shared in a way where interpretation and internalisation in one's own body is still the chief goal. The use of the VoxBox and similar technology then shares a similar risk with vocal pedagogy and other physical practice (e.g., sports education); it is easy to judge our experience against someone else working with a different tool, and this might potentially lead to pushing too hard or physical harm. When engaging with technology, we must make sure to actively direct focus to the body and work against instincts to compare, as we would with other instances of learning.

This may also be a reflection of the learning process being conducted online instead of in-person; the difficulty in understanding and measuring abilities is something which voice teachers are generally struggling with in a post-pandemic era. When working with voice teachers throughout this PhD, several told me that they were struggling to keep their lessons filled because students saw "no point" in doing virtual voice lessons. There is a value (whether actual or perceived) for vocalists in being able to reference another person and to experience "ideal" use during the learning process. This may also draw again on the importance of the audio to understanding what is right or wrong — vocalists assess their sound and learn over time and with practice to recreate behaviours which reproduce that desired sound goal. Without a reference and consistently working independently, both V1 and V2 felt it was hard to tell what should be expected of them and their behaviour. Both commented that they "didn't understand what it is doing," at times without this guidance and affirmation of the "expected behaviour," as V1 notes.

9.4.3 *The infallible technology and the body-self to blame.*

Another interesting facet of the interactions observed with the VoxBox lies in the moments of disconnect and failure. There was a personal association to the interaction seen with the both vocalists: getting a clear reaction and connection from the VoxBox was reassuring and encouraging for V2 in her ability. For the moments where she was disconnected, and a majority of V1's experience, the disconnect was generally viewed as the fault of the self and body, rather than the technology. Neither vocalist commented that they thought the VoxBox was broken, worked badly, or was poorly designed. If there were such feelings, perhaps they were not shared with me, as the designer, to avoid sharing negative feedback. On the other hand, the vocalists shared many worries that, somehow, their actions or techniques were to blame:

V1 worried her technique was somehow incompatible or that she just did not understand the device as a fault of her own. In not understanding the relevance specifically of pitch-related feedback in her Hindustani classical practice, my design within an assumption of Western classical practice potentially influenced this feeling of "poor" technique, going so far as to reinforce colonial ideals of Western classical as the pinnacle of vocal technique. V2 doubted her practice routine and wondered if her practice was too "lazy," resulting in underdeveloped muscular movements. She remarked that she had not been practicing technique recently, using her limited time to rehearse with her ensemble. Although this encouraged her a bit to spend more time "challenging herself" in the future, we see the negativity that can be placed on one's own perspective of themselves by technology. This may also link to some of the elitist narratives around classical Western vocal technique as being the established "best" practice; knowing my own vocal work, V2 remarked at one point how her work is "not like what you do," and diminished her practice in comparison to my own. However, other than cultural preference toward particular vocal genres as being more "skilled," there was no explicit reason to think that her doing jazz improvisation, as compared to my Western art song and choral work, would result in less success with the VoxBox.

Beyond perceptions about vocal culture, there is a feeling that bodies must adapt to technology, rather than the other way around (Mice and McPherson, 2022), or that technology is somehow "infallible" and knows best. When something goes wrong, the vocalists jumped to blame themselves, rather than considering that maybe the device was at fault. We see how then technology can shape perception and the body itself (Homewood et al., 2021; Spiel, 2021). Reiterating again the previous theme, using technology, whether intentionally or not, as the source of ground truth or "typical" qualities about the body can neglect the individuality in experience. With the entanglement of our bodies with technology, this can influence our behaviour and perception of our movement, for good or for bad. There are any number of reasons that the sEMG might have been difficult to use in this study. Perhaps, in a way interesting to design research, it was because the feedback disrupted an existing audio-motor pathway, as discussed above. It might have been as well difficulties in employing the technology at times, for instance, in getting a consistent electrode placement as discussed in Chapter 7 or just getting used to a new interaction method. However, these vocalists placed some kind of trust in the technology. When it did not work the way they expected it, this was interpreted as a personal fault.

When designing technology, it is important to acknowledge the role that this interaction has on our perception of self and ability. With a restrictive view, we may fall into the trap of the "quantified self" and influence thinking about our bodies by conveying "ideal" or "normal" response in biodata feedback (Lee, 2014; Prpa, 2020; Spiel, 2021). In this study, it is clear that this is also dependent on the design of the technology itself and our entanglement with it (Mice and McPherson, 2022): by providing a context which was too open-ended or exploratory in nature, participants did not have enough confirmation of their actions or the ability to gauge whether their expectations were appropriate. This can also create feelings of being lost or uncertain. Individual interpretation and perception should be acknowledged to avoid over generalising experience; yet, guidance with reflection on that could be used to create encouraging environments for difficult tasks such as exploring movement or learning new skills. This mirrors how vocal teachers instruct on technique as well.

There is however a mismatch between the design goal, to create a probe which allowed for exploration of the vocal technique through novel feedback, and the vocalists' expectations of the technology. Both vocalists had an expectation of the VoxBox *to tell them* something about their body and practice, rather than for it to be used as a channel for them *to explore their action*. Most of the technology we interact with in a daily basis tells us something about the world; it is very rare for technology to be oriented towards exploration, leaving participants of somatic studies looking for an answer (Howell et al., 2021). Although I had specified when briefing the participants initially that each individual interaction would be different, this expectation remained. I included the musical connection aspect on the Interaction Questionnaire to observe how the sonification quality and interaction played a role in their practice. In fact, the sonification had been designed to be as "non-musical," (I would argue that such textured sounds are indeed "music," but this is another entire debate) being without pitch or rhythm and acting as a texture to support the voice. This fact was given to participants during the briefing and in the end neither V1 nor V2 felt musically connected to the sound (as I expected); *however*, the vocalists were disappointed in this, with V1 experiencing a feeling of it being "useless" in her practice and V2 being unable to multitask with her musical performance and an external sound. In the design, I may have addressed my own interests or needs for my practices, which did not align with the needs of the vocalists.

It is also important here to then discuss the burden of the participants in working with technology designed for exploration of the body. Because there were no set goals for its use, we see that the vocalists were unsure of their interaction and wanted to have more guidance. V2 mentioned that doing this study on her own was a bit "lonely" at times. Without an expectation, or with an expectation that did not match the actuality, as in the case of V1, the vocalists might have placed blame on themselves (Howell et al., 2021). Additionally, the pressure of the situation of participating at all might have caused these negative associations; both vocalists are very passionate about teaching and holistic connection with the body. They were excited to work with the technology as a contrast to some of the more traditional methods of teaching, which ignore individual perception and bodies; even after facing some difficulty with the VoxBox, both vocalists still contributed helpful feedback on the sound design and expressed interest in the continuation of the work. With such personal investment, it is natural that a participant bias might have emerged, creating pressure on the vocalists and resulting in blame when things went wrong in their understanding (Howell et al., 2021). For the researcher or designer, it is still valuable to acknowledge failure and to build on feedback. When working with participants, it is important to acknowledge how technology shapes perception, be proactive about these feelings of doubt, and foster a communicative and caring environment, especially when working with movement and body perception.

9.4.4 Motivations in Externalising the Body

The major affordance of sEMG is that we are able to capture aspects of internal movement, which is normally perceived through proprioceptive sensing, as external feedback. The technology could function within the quantified self paradigm, providing a marker against which vocalists measure how much laryngeal tension they should have or judging the control over their muscles as being sufficient or not. Whether such a device is possible is unclear, but probably unlikely given differences in physiology. Philosophically, the VoxBox comes from the opposite direction, intending to provide a backdrop and context for exploring lived experience by externalising sensory experiences which are not normally conscious and providing new insights into individual interaction. This suggests that other sensing methods can also be used in the design of interactions for exploration with the internal, particularly through biofeedback.

However, considering the above theme, it is important to acknowledge the role that this interaction has on our perception of self and ability from an ethical consideration. In this study, it is clear that connection to the data through embodied understanding is also dependent on the design of the technology itself and our entanglement with it (Mice and McPherson, 2022): by providing a context which was too open-ended or exploratory in nature, participants did not have enough confirmation of their actions or the ability to gauge whether their expectations were appropriate. This can also create feelings of being lost or uncertain. Individual interpretation and perception should be acknowledged to avoid over generalising experience; yet, guidance with reflection on that could be used to create encouraging environments for difficult tasks such as exploring movement or learning new skills. Although there was no direct quantification of the self, as one might see on a fitness tracker, and the interaction was designed specifically as a probe to explore embodiment, the inability to connect with the feedback provided resulted in misinterpretation of action and ability (Fogg, 1998; Kitson et al., 2018; Prpa, 2020). This can be seen in related work, wherein participants attempt to fit themselves and their bodies to an interface (Mice and McPherson, 2022), rather than believing the interaction should be adjusted to their needs. This highlights that "quantification" of the self is not just a numbers game, and these expectations of performing to a system, rather than the system performing to you, are entrenched in the way we view and interact with technology (Homewood et al., 2021; Mice and McPherson, 2022).

9.4.5 Links to Personal Experience

This experience with the VoxBox was very different from my own work with it, which I believe is derived from the expectations and goals of the interaction. As discussed above, the vocalists who worked with the VoxBox experienced moments of disconnect and interruption with the sonification. They struggled to balance the reliance on the audio of their voice with the audio produced by the VoxBox. Looking back on my experience in Chapter 8, it is a stark contrast to my feelings of connection with the same sonification.

The largest disparity between the experiences is the use of auditory feedback as an effective method for engaging with the vocal embodiment. In a sense, the VoxBox did achieve its goal in awareness of physical movement, largely centered around awareness of the breath as an unconscious activity, and reflection on the engagement in one's own practices. However, V1 and V2 did not feel this was an appropriate channel; rather, they felt it interrupted their practice at times and were only able to really engage with it when they weren't focused on singing. Although initially I was surprised by the overwhelming difficulty in working with it, the feedback suggests that the attention to the vocalists' existing perceptions and expectations was perhaps missing in the design:

My position as designer allowed a different perspective to the work with the sonification and perhaps shaped my focus when using it. The vocalists were asked to incorporate the VoxBox into their existing practice. On the other hand, while I did use the VoxBox in my existing practice, I went into it with the flexibility of a designer: I was able to change things which did not work and I evolved with it through its development. In a sense, the state of the VoxBox perhaps reflects my own interaction and perception of my body — the setup and choice of the muscles examined was based on my own connection with it and the sonification itself, as a breathy, noisy entity, reflects the more compelling relationship I experienced with the system in my breathing. In this way, I have shaped the system more as an artifact of my relationship with my own body. This opens up avenues for interaction and expectation: I am experimenting with what works best and tweaking parameters as I find relationships, allowing me to shape it to my understanding of my body. When something does not align with my expectations, I instead feel fault as a designer or engineer, rather than as a singer, because I know the inner workings of the design.

For the other vocalists, I have simply provided them with a tool for exploring their vocal practice and they have an expectation of how it works. The VoxBox was designed as a probe for exploration, not as a device for providing feedback on correct singing practice; however, the participants of this study attempted to fit it into a success narrative. This authority on the part of the technology is something I did not experience in the same way: after working with it for so long and adapting with it, I was more apt to think something was wrong with the sonification or sensing (although I did typically blame my skills). On the other hand, the vocalists trusted more the authority of the VoxBox, without the experience of its inner working, and were more prone to thinking there was something wrong with their own behaviour.

In this way, the role of the designer provides an agency over the technology in some forms. Although it may be an annoying, pesky thing to work with, the designer generally would place frustration on the design and their influence over it. For another user, the technology has an authority and expected role, leaving the blame to be passed to the body and the self. In this way, I had a very different experience than the other vocalists did. I worked with the sonification as with learning any other skill, gaining better connection the longer I used it. For the vocalists, the expectation of "success" in using an exploration probe might have prevented them from reaching this level of dialogue between the body and the feedback.

However, the VoxBox did achieve for each vocalists what it was intended to do — to disrupt and destabilise the existing vocal practices by highlighting and bringing attention to a very small, often unconscious movement in the body. Namely, there was an experience of divided attention when expectations were not being met which caused the vocalists to evaluate their movement and response: in the micro-phenomenological interview I participated in, we uncovered the feelings of being torn and loss of control as a result of unexpected response from the sonification. Similarly, V1 and V2 both experience this sensation of divided attention which clashes with their images of what they are doing with their voice. V1 expects to hear a pitch as she visualises and hears an image of her muscles as tones, layering together. V2's awareness of the sonification fades away when she begins singing, contrary to her expectation, because she cannot bring her focus away from her voice. I remain silent and expect the sonification to follow my lead, but it activates at my small movements and pulls me away from my focus in the duet. In all cases, there is a feeling of self-fault at this divided attention. For the other vocalists, the blame was placed on imagined poor technique or lazy practice. In my case, as the designer, I blamed myself and my design practices for the unexpected response. The "bespoke" design centred around my practice perhaps reveals more about my evolving relationship with it. Were the other vocalists able to engage in this way with the inner working of the technology, perhaps this role in the relationship with it would have been more of a dialogue, as I experienced, than with a sense of authority.

9.4.6 Future Work

This work with the VoxBox suggests future study in conveying biofeedback for interacting and exploring tacit knowledge. Although the VoxBox accomplished its goal as a probe for exploring embodied relationships between the vocalist and voice, it is possible to inform further "success" in designs where it would be beneficial to provide sensations of correctness and reassurance to the user. We see in this specific case that the auditory channel of interaction revealed an understanding about vocal embodiment *because* it disrupted the existing interaction. Namely, the need for interaction in different and combined modalities seems chief: while I was able to interact well with auditory feedback, V1 and V2 had varying levels of success. V2's encounter with auditory feedback in her non-vocalised movements suggests that more attention needs to be given to the different, individual imagery pathways being used. This could suggest interaction modules developed for quick switching of sensory modalities; for instance, using a voltage output device which allows a user to test audio, haptic, and visual feedback methods in practice, to compare or adapt as they find is most helpful for them. Varying imagery abilities might require more flexibility or options in biofeedback. Again, this would benefit from lessons learned from studying voice teachers: providing the ability for an individual to internalise information in their own understanding. Perhaps using tangible or visual externalisations might be more useful for some singers who rely more on this existing audio connection; future work might incorporate flexible rendering of this externalisation through different modalities to allow for more reactive exploration of lived experience through sensory domains which better match individual embodiment. In these cases, frustration and connection may be linked to different modalities for different people. It will likewise be worthwhile to work on bespoke designs and mappings for individual users, for instance through a co-design strategy or workshop environment where vocalists are able to explore different modalities of biofeedback, mappings between their biodata and the feedback, or even other sensors beyond sEMG.

Continuing with the attention to bodies and diversity of experience, it is worth noting that working with other participants would have likely yielded different results and interactions. Future work would benefit from the exploration of similar internal-to-external sensory translation, either within the vocal context outlined here or through the development of further probes for internal sensory experiences. For instance, sports sciences and other movement-based artistic practices would be key areas for further iterations of this type of study. In this vein, it would likewise be beneficial to further explore vocalists' perceptions of their bodies as instruments and their interaction with their physical experiences in singing, aside from technological mediated activities; adding to previous research on singer identity (O'Bryan, 2015), this type of introspection will provide a "baseline" of how vocalists view this relationship with their body-instrument and highlight individual perspectives going into further studies with biofeedback. Additionally, the use case presented in this paper deals with a month-long exploration in an isolated, remote study. Conducting the study in person or with a longer time frame might have shaped the experiences differently, providing better support contexts to work through frustration during the interaction and ample time for learning the internal-to-external translation as it evolved over different lengths of use. As suggested by V2, it might be worthwhile to conduct this type of study by comparing explorations of such lived experiences in a group setting to a solo activity.

As well, it would be worthwhile to study this connection between imagery and which feedback methods work better for different people. Imagery ability was not measured for this study, and it would be interesting to conduct a study focused more specifically on feedback modalities compared to individual imagery aptitudes. In allowing participants to switch between different modalities and work with methods they find they are more comfortable with, we may also see further connection between multi-modal imagery and understanding of the body. Letting vocalists find their own modalities would potentially allow more control over the entanglement with the technology, creating environments which work to suit individuality. This is further reflected in my own experience with the sEMG sonification; through adapting and evolving the technology to suit my needs and explore paradigms I found interesting, I was able to explore my singing more enjoyably and thoroughly with the sound design. By providing bespoke elements to the interaction design or allowing users to adapt it over time to their needs, we would likely find similar aspects of understanding as I experienced, and the creation of more artifacts which express an individual relationship. If the VoxBox is a culmination of my own understanding of my voice, perhaps further artifacts and probes using sEMG for vocal interaction would emerge in different ways when designed by other individuals.

9.5 Conclusion

In conclusion, extensive work with other singers using VoxBox toolkits for vocal sEMG interaction revealed how sonified biofeedback played a role in the perception and understanding of movement. The vocalists had varying experiences with the feedback, including feelings of connection to breath and posture alignment, but also distraction and disconnect with vocalised action. This suggests that, for some users, sonification interrupts with existing auditory-motor associations and that vocalists understand, albeit in a non-verbal way, their voice through its audio. There is a sensory translation mechanism between the feedback a vocalist receives: physical sensations within the body are mapped to auditory information, which is further coded through abstract metaphor to be easier to understand and articulate. These working parts are blurred together and can be difficult to parse out individually. Technology can influence movement and also personal, emotional perception of the body or self. Users can find exploration difficult, seeking confirmation or "correctness" from technology. When this is lacking or unclear, it is easy to believe the disconnect to be because of personal fault or poor interaction technique. Through this study, we see the importance of feedback and how technology shapes the body itself. Attention to individual understanding and existing perception is important; rather than creating wholly new interaction paradigms, it is important to acknowledge and work with existing imagery to augment experiences, rather than replace or recreate them. This suggests the use of multi-modal feedback when providing information about the body, and working with users to create understanding for their individual perception.

Chapter 10

Discussion

Vocalisation for singing is a complex physiological process. Control over these actions and understanding of action-result mappings is rooted in often wordless tacit knowledge of the body. Because the connection to the voice is obscured from others and known tacitly from the point of view of the vocalist, it can be difficult to interact with the physical voice directly. More often, technology is developed using indirect methods of audio processing and translation. In this thesis, I have worked with my own vocal practice and other vocalists and voice teachers to explore the physical side of singing.

From the studies in this thesis, we can see the entanglement between technology and user, as well as between the user and their body. The experiences and interactions with the world shape how we perceive our movement and bodies. Through these interactions, we form ideas and understanding, and this experience is continually formed with more information in a cyclical way. Focusing on the body and interacting directly with movement, I have been able to uncover more understanding of this embodied use of the voice from the perspective of the vocalist.

From the initial study, we see how vocalists use multi-modal imagery to adapt to performance conditions with AAF. Actions are regulated with both auditory and tactile feedback; when auditory feedback is disrupted, tactile connections can be used to maintain timing and continue singing. Depending on the goals of the performance (e.g., whether to stay explicitly in time or to have more expressive play with the beat), vocalists are also able to change their focus and execution. From working with voice teachers, we see how this connection between imagery and understanding of the voice is expressed. Teachers use abstract references and metaphors to explain and teach fundamental vocal practice. The abstract experiences are explained in a way which is non-domain specific and appears to be pre-linguistic; that is, the understanding does not depend on the linguistic metaphor itself, but rather the abstract understanding is articulated through language.

In order to understand the underlying images and focus on the body, this thesis work developed a novel method of vocal interaction through sEMG. Focusing on the physiological movement of the body, vocalists were able to interact with directly movements they use to control their singing, but which they are not normally conscious of. This externalisation of the movements, in breaking away from normal practice, revealed several facets of the perception of the body active in singing and insights into vocal behaviour, including feelings of control and disconnect from the body and awareness of unconscious movement. As well, the use of the sEMG interaction impacted the movement and perception, encouraging certain behaviours such as breathing and examination of posture, as well as causing doubts in other parts of the practice.

10.1 Addressing Research Questions

There were several key takeaways from this thesis work which address the research questions outlined in Chapter 1. I will start with the sub-questions and build back up to address the main research question as a summary of the findings of this thesis.

SQ1: How do vocalists use abstract mental representations of their actions, through musical imagery and metaphor, to perform, understand, and speak about their vocal practice?

This research demonstrates that the pathway from vocalist intention to vocal execution is composed of many working parts. The underlying understanding of the voice and using it is very blurred between different interaction modalities, so much to the point where the vocalists have difficulty separating them. This reveals some of the true intricacy of tacit knowledge and perhaps why it can be very difficult to explain or teach such practices. For this particular research question, the compilation of the work with vocalists in adapting to AAF, teaching vocal fundamentals, and interacting with the body through biosignals reveal a a cyclical understanding of this action that is adapted depending on the environment. From this work, I have extended Dunbar-Wells's model of teacher metaphor translation (see Figure 3.2 in Chapter 3) to include the cyclical adaptation of this understanding (Figure 10.1):

To recap Dunbar-Wells's understanding, the teacher's metaphor is encoded into mental, sensory-based imagery, which then informs neural responses (Dunbar-Wells, 1997). As in the Functional Equivalence theory of imagery, this neural response is active whether the action is actually executed or only imagined. This neural response triggers movement, then vocal execution. In this revised model, I instead begin with the other side of this process, from the understanding of the vocalist/teacher. Combined feedback from auditory and kinaesthetic sources, for instance the timbre of the voice and the tension in the abdomen, respectively, is received and processed. From here, the physical adjustments are noted, mapping behaviours to what they experience. This mapping is stored in imagery based on the composite sensory information. This process happens unconsciously; the body receives feedback and the mapping is internalised. The way the vocalist verbalises or expresses this is through metaphor. This can take the form of different modalities, depending on personal meaning. As discovered through interviews with the vocal teachers, this metaphor is often non-domain specific, meaning vocalists pull references from other life and outside, non-vocal contexts to explain this understanding.

On the bottom half of Figure 10.1, we see Dunbar-Wells's original model of metaphorical translation to execution, which is nearly the same as the process of mapping reversed. This decoding happens when metaphors are used as references for actions; the process then begins again as new feedback is received and the encoding is adjusted. It is likely this happens when vocalists rehearse and are able to run through this cycle repetitively with practice. It appears then that voice teachers have co-opted a portion of this feedback cycle, supplementing the source of the mental image to help stimulate the formation of imagery, again with the vocalist applying a personal meaning to the reference and creating that mental imagery.

This model is likely generalisable to other similar body-based practices (e.g., dance, sports, meditative practices) when executing and learning other activities wherein abstract action-reaction is mapped to imagery and expressed through metaphor. In this case, I have listed the sensory modalities which were uncovered through working with vocalists during this PhD; it is of course reasonable to believe olfactory and gustatory experiences are also involved in forming imagery, but were not prominent in these instances.

From this research and the resulting model, we can gather a couple key takeaways which are extendable beyond vocal practice:

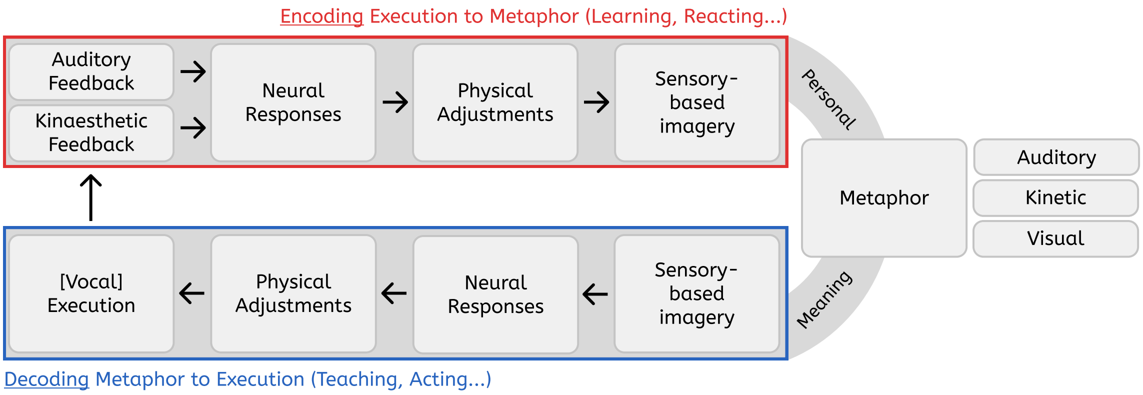


Figure 10.1: A revised model of translation of vocal metaphor, understanding, and execution, suggested by this research.

Metaphors and abstract language work in communication by getting around tacit knowledge's linguistic, contextual barrier. As seen in the work done with voice teachers in this thesis, metaphors work because they directly address what makes tacit knowledge so difficult to express or communicate to others: they do not directly address the knowledge. Metaphors rather approximate understanding and are intentionally vague in a way where they can be interpreted by an individual within their own understanding. As with imagery itself, they play with personal meaning to strengthen the connection: by referencing other life outside of the target context, they establish a relevance which makes sense. As well, they subvert the difficulty of describing linguistically one's tacit knowledge by using unrelated easy-to-understand language or move completely away from language, instead using gesture or visual references to describe. In this way, metaphor facilitates sensory translation between modalities and helps us to express one sensory experience in terms of another.

Vocalists view the body and voice in a congruent, hard-to-separate way, which reflects the multi-modality of mental imagery. As seen from work with the VoxBox, the understanding of the voice is very tangled between auditory and kinaesthetic feedback. This is so much so that the modalities are linked together, for instance V1 understanding her body and the movement of her laryngeal muscles through an auditory reference. Because of the way that both auditory and kinaesthetic feedback are mapped in the vocalist's imagery, the movement and the control of the body can be physically understood through sound. Even in practices which are very body-based, with vocalists detailing the practices they carry out to become aware of their body, the modality of understanding the physicality is not always physical or tactile. As seen also with the voice teachers, visual references were heavily used to understand the alignment and movement of the body and the muscles. Vocalists understand their bodies through a mix of modalities, which are hard to separate from one another, as shown by the areas of frustration when working with the VoxBox. It is possible that the audio representations of the voice are so important in this understanding because of the emphasis placed on sound over body; although vocalists turn their attention to more holistic pedagogy, the imagery they rely on has been constructed with an emphasis on auditory feedback, interlinking even further the understanding of the body through sound.

This entanglement between imagery modalities may also explain why auditory imagery alone was not linked to better adaptation to delayed auditory feedback in the study outlined in Chapter 5. At the beginning of this thesis, I had assumed auditory references were straightforward and the primary driving factor for vocalisation; however, we see that multi-modality entangled individual understanding is more likely. Thinking of the voice as a vibrational practice, as discussed with reference to Nina Sun Eidsheim's work (Eidsheim, 2015) in Chapter 3, Section 3.2.4, might contextualise these observations. The experiences of the living body in a vocal context "transcend audition," experiencing sound not only through auditory but also tactile, spatial, physical, and material means (Eidsheim, 2015; Stadnicki, 2016).

SQ2: How can we use biosignals capture the internal sensory experience of singing by conveying aspects of low-level muscular movement during singing?

This thesis applied surface electromyography for biosignal feedback to convey information about internal movements while singing back to the vocalist. This thesis explored ways in which sEMG can be applied in a vocal context to detect laryngeal movement. As well, the development of the Singing Knit demonstrated ways in which technology developed in one context can be co-opted into other areas. By using familiar materials and creating bespoke garments, existing technologies can be adapted to fit into the cultural context of another domain: in this case, sEMG sensing methods with rigid electrodes were employed in a performance context through a knit wool collar, designed specifically for the wearer's anatomy. The medical perception of the electrodes was adapted for a performance context, allowing both contexts to have presence; the knit collar provided a flexible, comfortable application, but the reference to the placement highlighted aspects of physiology present with the original use.

We see, both from an autoethnographic perspective and from working with others, that this provided a sense of the body's movement, but rather than capturing the internal sensory experience of the vocalists allowed for an examination of the typical understanding and perception of movement:

Biosignals such as sEMG can be used as a probe for understanding experiences and to help with sensory translation. The ambiguous nature of sEMG, combined with the fact that most of the muscular movement examined was very unconscious, provided rather a "playground" through which vocalists could explore and experiment with their bodies through sonification. The sEMG does not always trigger when expected and this instead causes a sort of dialogue with the system. To quote Nicolls (2010), it seems beneficial to approach such sensor-based biosignal feedback as "an entirely collaborative approach and therefore one that involves at least directed improvisation and, more likely, fairly extensive improvised exploration." The improvisatory nature of similar work with sEMG (Erdem and Jensenius, 2020; Tanaka, 2015) and other biofeedback sensors relating to internal sensory experiences (Cotton et al., 2021a; Tsaknaki et al., 2021), the nature of the behaviours as unconscious allows performers to explore the physicality of their movement, note effort and resistance in their body, and unpick elements of interaction which they were not previously aware of. At other times, there are still unexplained interactions; sometimes nothing works at all and sometimes the feedback is visceral and provides a sense of connection and understanding with the body. In a sense, biosignal feedback presented in this way mirrors the experience we have with our bodies on a day-to-day basis; the ever-changing nature of our bodies is not completely understood by us and there are many aspects of physiology which are still a mystery to science. Biofeedback in this way presents a way to translate some of these movements and aspects of experience which we perceive but are not consciously aware of, into another modality through which we can learn and explore.

Biosignal feedback can provide insight into existing imagery and help form new images. As seen in my own interactions and the interactions with other vocalists, biosignal feedback can provide insight into imagery use by subverting our typical actions. Much like the exploration with AAF in Chapter 5, biosignal feedback disrupts the existing action paths in singing; in our somaesthetic context, we break away from the habitual, and in doing so realise qualities of our actions and behaviours that might have otherwise been subconscious or forgotten about. For instance, my autoethnographic exploration revealed interesting residuals from vocal upbringing; my posture and attention to "proper" classical behaviours was revealed when I started to move my neck and jaw to receive feedback from the vocal sEMG. I had also forgotten so much about my own breathing. These intact images of posture are pervasive and deeply ingrained in my performance behaviour. Likewise, we see in Vocalist 1's difficulty with the audio-centred biosignal feedback that her existing imagery is already audio and pitch dependent. With the disruption of her normal practice, we were together able to uncover how she views her voice — visually aligning to produce the timbre of her voice — and understand more of the imagery she uses to understand her pitch and timbre. In the interaction with this new feedback, our imagery and the feedback also become entangled and evolve together (Frauenberger, 2019; Tuuri et al., 2017); knowing more about our interaction through this biosignal feedback causes us to change our actions, furthering this cyclical relationship. This requires that shortcomings of systems like the VoxBox and its single-modality feedback be addressed to ensure that feedback is most useful for individuals. In addressing this, I see two facets where biosignal interaction would be most interesting to research within vocal pedagogy: first, exploring how vocal educators and students can use biosignal feedback *as a metaphor* for mapping understanding of the feedback to vocal experiences. Similarly, this could involve iterating through and adjusting the feedback to provide references to and encourage new imagery formation. Secondly, designing around individual interpretation, as a necessity when using biosignal feedback, as a way to help students to understand themselves and their bodies through an iterative co-design.

SQ3: How does the real-time sonification of sEMG signals influence perception of movement and create new connections between the vocalist and their body?

Through using sEMG as a probe to explore the connection between vocalist and voice, we have uncovered a number of ways in which interaction with novel feedback can bring awareness to an existing practice and perceptions of the body. The voice is not only part of the body, but also a source of identity and character, as are other parts of our physicality. The entanglement with technology, especially in contexts which are very personal and deal with a sense of identity such as singing is for vocalists, shapes the practice itself. With interaction, this feedback can dictate movement and change perception of the body, whether for good or bad. The expectations and actuality for design do not always match, particularly in somaesthetic-based research such as this, where the technology was placed into an existing practice to observe the change in action and perception. From the work with other vocalists, we see the positives and negatives that this kind of probe can bring to an interaction — as researchers, we must strive to work with others in a way that, while challenging practice and learning about interaction, does not subvert the existing relationships in a negative way. Some general takeaways from this research question include:

Bringing internal experience outside the body allows us to perceive the interaction differently and influences our behaviour. Interaction, with and without technology, shapes our understanding of the world in a cyclical way. In this research, we focused specifically on an internal sensory experience which was embedded in larger action paths and largely unconscious. With biosignal feedback through sEMG, we can make someone aware of their movement in novel ways. This awareness changes perception of movement and indeed the movement itself. From my own interaction with sonified sEMG, I found myself more aware of aspects of my breathing and felt more inclined to move in particular ways, at times in ways that I had been specifically discouraged from doing as a vocal student. The addition of this feedback led to its pursuit and actions which produced more detectable movements, rather than "good" vocal behaviour. For other singers, while being able to make connections between movement and sound, this experience changed the existing imagery pathways and made their typical actions more difficult at times. At times, this led to play, for instance experimentation with posture and attention to the breath through longer vocal phrases. At other times, the feedback made the vocalists aware of how crucial auditory feedback was in their normal activity and helped them to explore how they understood their movement as a result of their vocal sound. Through changing the sensory feedback in an activity as well-known and refined as singing for these vocalists, we were able to not only see how perception changed, but also details about the relationship as it exists normally.

Control over the body is a variable sensation which must be negotiated in interaction. As mentioned previously, the variability of the body is an aspect of daily life, and one which is very well-known to singers. The body is not the same every day; engaging in practices which rely on refined action execution requires a dialogue with the body and taking care of the relationship with it. Knowing when we are tired and going easy on ourselves is just one example of this negotiation of control. Working with sonified sEMG augmented these feelings of control and being controlled. Where vocalists were able to successfully feel in control, there was a sense of understanding and empowerment over the interaction; at other times, the sonification did not respond as expected, leaving feelings of frustration and disconnect, as if the body were a separate entity. In this way, the sEMG is representative of existing control and communication with the body. For myself, after having lived with the sEMG for such a long period of time, I began to feel more flexible with the communication, allowing a conversation rather than forcing a behaviour from the sonification. As with the body itself, I can only force a behaviour so much; however, I can learn to listen and react, to trust and collaborate with my voice and acknowledge some of the constraints on my movement as part of my own physiology. This is dynamic with an ever-changing physiology; work with sEMG feedback highlights this relationship and partnership with the body.

Even when explicitly exploring, users still desire reassurance and confirmation of "correctness" in their interaction. As seen in work with other vocalists using the VoxBox, exploration of movement and negotiating this relationship can be difficult. Even when aware that they were meant to be exploring and that there was not an expectation for the interaction, the vocalists wanted reassurance that they were behaving correctly. But, what really is correct behaviour in this case? The ideal scenario would be that the vocalists were able to uncover something interesting about their movement; this is not something which can be deemed correct or incorrect with a research probe such as the VoxBox, or even a person other than the individual using it. This might indicate a sense of uncertainty when working with this kind of technology — working with an exploration probe is an uncommon experience, as we are used to technology providing some kind of ground truth or performing a specific task.

Users can view technology as being infallible: we must be aware of the influence of the technology on self-perception. To add to the previous point, there is often expectation that technology is always right. As the designer, I had a different, more intricate understanding of the system than the other vocalists. From their view, the VoxBox was expected to behave a particular way; for instance, for V1, she expected the sonification would reflect her vocal audio. This difference in expectation is often attributed to a failure on the part of the user; if the system is not responding in the expected way, then it must be that the user is not doing something correct or well. This did result in an awareness of one's actions, but the idea that technology is infallible leaves difficulty for such exploratory probes. Self-blame or questioning of one's well-refined practice is obviously not the goal. However, it is interesting that this expectation mismatch fell on the fault of the user.

We can then return to the main research question addressed by this thesis:

> **How are vocalists able to control parts of and interact with their bodies through internal sensory feedback, when such feedback and action is hard to articulate or even conceptualise except in abstract representation.**

The most notable finding of this thesis is perhaps that vocalists can *not* control their voice — at least, not every part of it. Rather, this relationship is more of a collaborative one, much like other partnerships with the body. Rather than controlling the vocal physiology, vocalists have an inherent understanding of their voice and work with it. In fact, although I focus on control dynamics when interfacing with the vocal physiology, I would, in future research, not refer to body-based interaction with an idea of control at all. Instead, ideas of partnership and cooperation might be more appropriate for this kind of interaction with the body; this is extendable to other kinds of work in NIME and HCI research.

Through multi-modal feedback and mental imagery, vocalists understand their action-result paths in individual ways. The body and voice are one and the same, and collaboration with it, while difficult to explain outright, can be captured through the flexibility and ambiguity of metaphorical representations. This allows another person to understand an individual's connection with their body and provoke new understanding in their own connection. Similarly, we use sonification of internal laryngeal feedback —a metaphor in its own right — to encompass knowledge about the body in an alternate modality and express the vocal presence as an external sound. This can help vocalists to create new connections with their bodies.

On the other hand, without proper situation of technological interventions and their role in body-based interaction, representations like this data sonification can lead to confusion and self-blame. Much like metaphor mistranslation in vocal pedagogy, modality flexibility and iterative communication are needed to form novel connections to the body. Designers of such musical instruments and body-based interactive devices, such as myself, must be aware of our influence in the relationships between individuals and their bodies. We might think about bodies in a more material-conscious way, as having their own qualities and influence in the interaction. We would like to have complete control over our bodies, but we do not and cannot possibly achieve this. This is even more true (rightly so) for the bodies of others. We are made up of our environments, locations, age, backgrounds, genetics, cultures, contexts, and physiology. Many of those things cannot be controlled or are variable on a day-to-day basis; there are many background process in the body we do not understand in current science, and *many* which are autonomous and operate even in our conscious absence. Recognising we cannot control everything but instead embracing uniqueness and variability and, indeed, messiness can in fact offer insight into new designs and interactions and lead to deeper connection and appreciation of the body itself.

10.2 Implications for HCI

From these takeaways, there are a number of implications for HCI which could inform future work.

The understanding of metaphors and how they operate with ambiguity can inform many further design choices. In working with technology that deals with skill transfer or other learning environments, it is worthwhile to acknowledge that metaphors go beyond the simple UI symbols representing particular tasks (thinking specifically of Shneiderman and Maes's (1997b) discussion of Desktop Metaphors and direct manipulation) and are capable of representing more complex tasks (Blackwell, 2006). Here, we see how multi-modal feedback can function as metaphor; incorporating abstract references in the same ways as teachers do, with enough ambiguity and non-domain specific content to guide but not direct, would likely help users to create stronger associations of workflows and personal connection to their knowledge. In the same way, the incorporation of multi-sensory feedback would likely allow users to adopt relationships which suit their individual experience and understanding. Through the study of metaphor as used by vocal teachers, I have proposed a model of sensory-based knowledge transfer which uses the key principles of vocal metaphor — requiring no pre-existing domain-specific knowledge, working independently of language, providing ambiguity, and intentionally limiting what is communicated (Chapter 6, Figure 6.2). Extrapolated beyond the vocal context, we can see the balance between the two agents (human or technological, or some combination thereof) and how mutual understanding is negotiated by centring the metaphor in and translating knowledge through lived experiences.

This work also encourages the use of biosignals such as sEMG for the design of systems which can be used to probe experiences and perceptions of action. During the course of this PhD I have created valuable hardware which will provide the basis for future designs and explorations with sEMG feedback, namely the VoxEMG platform (Chapter 7, Figure 7.17) and the Singing Knit (Chapter 7, Figure 7.22). By providing feedback in different formats or allowing users to define the interactions as they learn more about it, as I did during the design of the VoxEMG, it is likely that we could create bespoke instruments and other interactive devices which work with the user's understanding of their body. Rather than producing a one-size-fits all device, we could allow users to create artifacts which reflect their own perception and experience.

The biggest implications from this thesis are in the understanding of human perception of action and the self and how this plays a role in interaction with technology; and, with entanglement theory, how this interaction then plays a role in perception of the self. The pathways which humans understand and communicate about sensory experience are indeed complex and often rely on the blending of different sensory modalities together; however, this understanding can be disseminated through the use of metaphor. When we insert technology into these existing relationships, we must be conscious and careful of how it will impact the existing imagery and expectations. As well, we would likely benefit from being flexible in how we convey information, as a good voice teacher would.

In understanding more about sensory translation and how information about the body can be understood in other modalities, such as the voice being understood through its audio, we can direct future research in sensory sketching and sharing of experience through more than just verbal channels. Practices such as body mapping (Boydell et al., 2020; Cochrane et al., 2022) and material speculation (Friske et al., 2020; Wakkary et al., 2015; Wirfs-Brock et al., 2022) are garnering interest in their ability to describe wordless experiences in new ways and study human perception. Through such multi-modal "sketching," future studies may be able to uncover further detail of the blending of sensory modalities. In this way, we maybe be able to determine the more intricate ways that we understand tacit knowledge, especially in regards to internal sensory experiences. As well, we may be able to provide wider sensory varieties in communicating and sharing our experiences with others. I have also demonstrated through this work, as well as through papers published around this work, how the micro-phenomenology discipline can be used within an HCI context to investigate the unfolding of experiences and general structures in interaction. This work has provided the basis of micro-phenomenology to the NIME and CHI communities, outlining methodology and demonstrating strategies for examining tacit knowledge for future work in HCI.

Finally, interaction with the body relies on a variable level of control; we are used to situations where we are unable to do things we would like, at the restriction of our physical bodies. However, when expectations are not met and the known or desired action-to-result pathways are disrupted, this can lead to self-doubt, feelings of disconnect from the body, and questioning of one's behaviour (potentially in a negative light). In situations where technology is designed for exploration or challenging behaviour, it should be noted that this relationship can be played with successfully along a spectrum. If the control or expectation is completely violated, the result will likely be a feeling of total disconnect or even "uselessness." This research suggests that users view technology as producing some kind of ground truth; if it does not interact the way they expect it to when relaying information about their movement or body, it can happen that they view themselves at fault. However, if a user is able to maintain a dialogue and understand the changing control, the interaction becomes more natural, as if improvising; some moments will work well, and others will not. In this balance lies creativity and challenging dynamics for exploration, playing with the movements of the body, and understanding more of how we move and interact with the world around us.

10.3 Personal Reflections

Before concluding this thesis, I want to briefly discuss some of my own personal reflections on having done this research. The goal of this research was to understand more about the vocalist-voice relationship. From my own interest, this research was valuable to understanding a piece of my own tacit knowledge, having practiced this craft over many years but knowing just how little I understood about it. Singing feels like having a super power at times; I can just change key or transpose something without thinking at all about it, which is something other musicians must consciously do. However, the general consensus which vocalists sometimes receive from other musicians is that singing is something everyone can do, and therefore it requires less skill than other instruments.

Reflecting on this research, I see just how much more complex and multi-faceted this relationship with the voice, of control over the body as an instrument, and adaptation on a day-to-day basis is. When I began this thesis, I had really envisaged the vocal experience as many of us do: purely as its audio. This is perhaps a factor of my own bias and musical culture creeping in; I hope that, since the beginning of this thesis I have broken away from the neglect of the physicality and "fetisisation" of aurality in Western classical vocal practice (Stadnicki, 2016). Although I knew the effort of my body while singing, I had never made many of the connections between the feedback and experience I was having. Through the awareness of my body and the role of even a very small muscle in my practice, I am better poised to give attention to the breadth of sensory experiences that shape our experiences in the world. Especially following my autoethnographic work, I also appreciate that the way I understand my body — we do not have a direct, spoken communication, but I know and trust that we understand each other. And through acknowledgement of the things I cannot change, I can learn to listen the materiality of myself and grow, adapt, and create with it.

Having gone through this research, I can safely conclude that anything humans do — from the most mundane, daily rituals, to precise artistic craft — requires refined and delicate skill, control, restraint, and understanding. We are able to fully realise this through careful and practiced attention to our bodies and the knowledge they hold. Contrary to some of the modern ideologies of efficiency and rapid learning, taking time to slow down, explore, and unpick our habits is where the magic really happens. With this knowledge and practice, I hope this thesis is beneficial to research, art, and appreciation of ourselves and bodily understanding of the world, in all that we do.

10.4 Conclusion

This research explored the relationship between vocalists and their voices through sensory feedback and perception of movement. Though a part of the body, attention when examining or interacting with the voice through technology is often focused solely on its sound. The vocalist however, has a multi-sensory experience, relying on interlinked musical mental imagery to perform their craft. This understanding and connection to the body is tacit but the control over the body and the vocal physiology is innate and highly refined.

Through different perspectives of inquiry, this research uncovered aspects of this relationship which further inform understanding of human perception and sensory experience. I have used mixed methods from a variety of contexts, uniting computer science, engineering, cognitive science, psychology, design, and musical interaction. Objective evaluations of vocal performance when relying extensively on musical imagery, as well as biosensing through surface electromyography were united with the subjective experiences of vocalists engaging in and teaching their vocal practice, as well as their interactions with their body through sonified laryngeal movement. I also approach this exploration as both a researcher and as a practitioner; applying my understanding of vocal pedagogy and personal experience to this interaction design has allowed me not only to inquire into the perception of other vocalists but also about my own practices.

This work uncovered how vocalists rely on multi-modal imagery to understand their voices. To share and communicate these experiences, abstract metaphor is used to make the sensory experience relatable to others. These metaphors work by relating to pre-linguistic understanding in a non-domain-specific way; they are informative enough to be understood widely and ambiguous enough to provide ample room for others to infer their own understanding.

From this research, I have developed a novel method for direct interaction with the vocal physiology through surface electromyography. As well, I have explored the adaptation of this technology in wearable contexts. Using this sensing method, I have uncovered a number of details about the embodied relationship between vocalist and voice. Namely, that vocalists formed new perceptions about their movement with the addition of sonified feedback. The interaction with this technology induced feelings of control and being controlled, mirroring the existing relationship with the body. As well, the interaction encouraged vocalists to play with their movement and to question their behaviour.

From this thesis, we can also see the influence of technology on these well-refined actions. When interrupting existing imagery pathways, the feedback became distracting or was tuned out. This caused feelings of self-doubt, blame, and disconnect from the body. However, in moments of control, the technology allowed the vocalists to explore their actions and make new connections between their movement and what they heard, creating awareness of previously unconscious behaviour.

In all, we see the importance of balance in control when working with exploratory tools and design probes. With a little ambiguity, these interactions can encourage creativity and expand on interaction, even going so far as to subvert existing behaviour in favor of pursuing the interaction. On the other hand, technology has the ability to create rifts in understanding. In acknowledging the variety of individual understanding and adapting to different sensory experiences, as done in vocal pedagogy, this balance can be further refined.

In future work in both HCI and in vocal pedagogy, this attentiveness to the multi-faceted sensory experiences of vocalists will be beneficial to encouraging dialogue with the body and understanding of fine-detailed perception. Through the findings of this thesis in how this embodied relationship is perceived, we see better how humans understand the connections with their bodies. In communicating tacit knowledge with respect to these varied experiences, we can better share our skills and understanding of the world and, hopefully, of each other.

Appendix A

Vocalists' Use of Auditory Imagery

Additional figures and tables from Chapter 5.

A.1 Participant Demographics & Experience

ID	Principal Instrument	Other Instruments	Nationality	BAIS-V	BAIS-C	Performance Experience (years)	Music Theory Study (years)	Formal Training (Y/N)
1	Piano	-	Taiwan	4.64	5	8	10	Yes
2	Guitar	-	Ireland, USA	5	4.86	10	10	Yes
3	Flute	-	UK	6.14	6.14	16	10	Yes
4	Voice	-	UK	4.79	5.5	18	11	Yes
5	Voice	Violin, Guitar	UK	5.21	4.64	3	3	Yes
6	Voice	-	UK	4.86	4.71	2	0.5	No
7	Voice	Recorder, Piano, Clarinet	Portugal	5.79	5.71	20	4	Yes
8	Piano	-	China	5.07	3.79	12	2	Yes
9	Sampler/EDI*	-	UK, Australia	5.93	6	24	1	No
10	Voice	-	UK, USA	4.36	3.79	1	0.5	No
11	Voice	Guitar, Ukulele, Piano	UK	4.86	5.36	11	7	Yes
12	Voice	-	Costa Rica	5.57	5.79	11	7	No
13	Voice	-	India	5.64	6.21	9	20	Yes
14	Dhol	-	India	6.57	6.43	12	6	Yes
15	Voice	Violin, Piano	UK	4.5	4.86	7	3	Yes
16	Guitar	-	Italy	4.14	4.86	10	10	Yes

Table A.1: Participant information including principal instrument, demographics, performance experience, and musical training provided alongside respective scores on the BAIS-V and BAIS-C subscales (*Electronic Digitial Instrument).

A.2 Complexity Measure Correlations

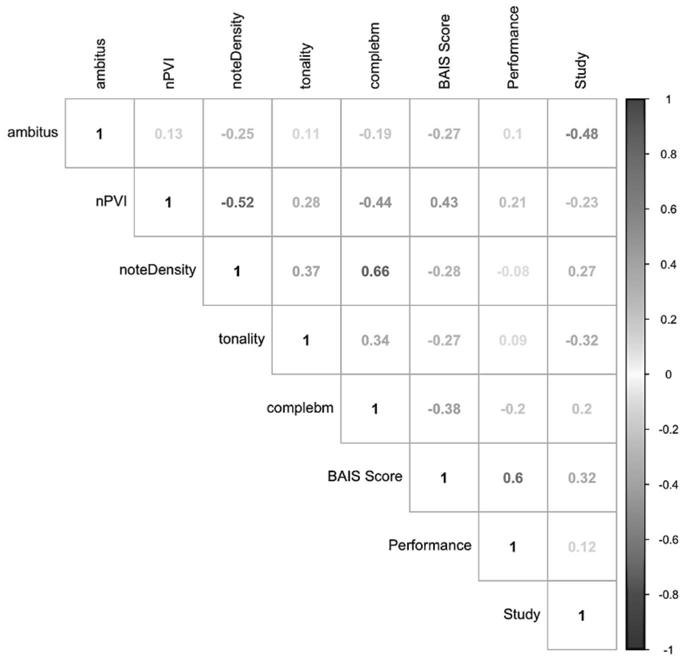


Figure A.1: Correlation matrix between participant experience measures (BAIS score, years of performance experience, and years of theory study) and respective complexity measures for each piece calculated with MIDI toolbox functions (NB: *ambitus* = melodic range (semitones), *complebm* = melodic complexity, *nPVI* = durational variability of note events).

A.3 Participant-Chosen Pieces

Table A.2 follows on the subsequent page.

ID	Song (Artist/Soundtrack)	Key	Tempo (bpm)	Meter	Beat Length (ms)	Bar Length (ms)	MIDI toolbox Complexity Measures				
							ambitus	nPVI	notedensity	tonalstability	complebm
1	Bizarre Love Triangle (Frente!)	Db maj	250	8/8	240	1920	12	39.117	1.1938	4.392	4.7504
2	New Slang (The Shins)	C maj	130	4/4	462.54	1846.15	16	25.113	0.96104	3.1655	3.7666
3	Where No One Stands Alone (The Peasall Sisters)	Bb maj	70	3/4	857.14	2571.43	17	67.21	0.80288	3.9503	2.7102
4	There's a Fine Fine Line (from the musical *Avenue Q*)	G maj	60	4/4	1000	4000	19	38.486	0.95789	3.8457	3.8514
5	Mr. Snow (from the musical *Carousel*)	G maj	65	2/2	923.08	1846.15	17	36.571	0.92545	3.7231	3.6752
6	Tears in the Typing Pool (Broadcast)	A min	80	9/8	750	6750	15	37.887	0.92364	3.8757	3.6609
7	Fantoches (*Fêtes Galantes*, Debussy)	C maj	110	2/4	545.45	1090.91	19	40.137	1.272	4.0661	4.1736
8	City of Stars (Mia's Solo) (from the film *La La Land*)	F maj	120	4/4	500	2000	20	50.939	0.72189	4.7393	3.7685
9	We Only Come Out At Night (Smashing Pumpkins)	C maj	60	4/4	1000	4000	19	43.81	0.84679	4.1044	3.5586
10	American Pie (Don McLean)	E min	140	4/4	428.57	1714.29	24	42.181	1.1435	4.2247	3.9507
11	Voi Che Sapete (*Le nozze di Figaro*, Mozart)	Bb maj	65	2/4	800	1600	17	40.844	1.5809	4.6399	4.9772
12	Proud Mary (Creedence Clearwater Revival)	D maj	60	4/4	1000	4000	14	35.367	1.0896	3.7356	3.5572
13	Bharat (from the film *Manikarnika*)	D maj	60	4/4	1000	4000	15	33.401	1.1255	3.4381	4.3185
14	My Heart Will Go On (Celene Dion)	E maj	60	4/4	1000	4000	14	68.401	0.75	4.1741	3.6365
15	Agnus Dei (*Krönungsmesse*, Mozart)	F maj	60	3/4	1000	4000	17	51.99	0.85535	3.7956	4.2806
16	Back Pocket (Vulfpeck)	D maj	90	4/4	666.67	2666.69	13	29.955	1.8895	4.5944	4.3794

Table A.2: Participant-chosen songs as performed with the reference tempo and key centre agreed at the start of the trials. Reference tempo and the first two bars for tonal reference were provided at the start of each trial. Timing factors and complexity measures are presented for each piece (NB: *ambitus* = melodic range (semitones), *complebm* = melodic complexity, *nPVI* = durational variability of note events).

A.3.1 Performance Combinations

Task	Normal Feedback	Headphone Feedback	200 ms Delay	600 ms Delay	+ 1/4 Tone Pitch Shift	+ Whole Tone Pitch Shift
Normal	16	16	16	16	11	16
Toggled	16	16	16	16	11	16
Toggled & Voice Distraction	14	14	14	14	9	13

Table A.3: Performances included for each task-condition combination: The 1/4 Tone Pitch Shift condition was introduced after the initial five participants. Participants 6 and 13 did not complete the Toggled & Voice Distraction tasks and Participant 5 was not able to complete the performance in the Whole Tone Pitch Shift condition in the Toggled & Voice Distraction task due to time constraints.

A.3.2 Participant BAIS Grouping

ID	Aggregate BAIS	BAIS Group	Performance Experience (years)	Music Theory Study (years)
10	4.075	Low	1	0.5
8	4.43	Low	12	2
16	4.5	Low	10	10
15	4.68	Low	7	3
6	4.785	Low	2	0.5
1	4.82	Low	8	10
5	4.925	Low	3	3
2	4.93	Low	10	10
11	5.11	High	11	7
4	5.145	High	18	11
12	5.68	High	11	7
7	5.75	High	20	4
13	5.925	High	9	20
9	5.965	High	24	1
3	6.14	High	16	10
14	6.5	High	12	6

Table A.4: Participant Demographics: Participants are ordered by aggregate BAIS score, demonstrating the median split and the relation of other demographic information.

A.4 Individual-Adjusted TRD Analyses

Effect	DFn	F	p
Task	2	0.672	0.512
Condition	3	1.154	0.330
BAIS Group	1	3.799	0.053
Task : Condition	6	0.796	0.575
Task : BAIS Group	2	3.304	**0.040** *
Condition : BAIS Group	3	3.628	**0.015** *
Task : Condition : BAIS Group	6	1.481	0.189

(a) 2x3x4 Analysis of Variance (ANOVA Type II Tests)

BAIS Group	Effect	DFn	F	p
High	Condition	3	1.28	0.283
High	Task	2	2.9	0.058
High	Condition : Task	6	1.27	0.277
Low	Condition	3	3.70	**0.013** *
Low	Task	2	1.03	0.359
Low	Condition : Task	6	1.01	0.421

(b) Two-Way Interaction by Group (ANOVA Type II Tests)

Condition	Statistic	p
Delay 200 ms	-0.646	0.519
Delay 600 ms	0.721	0.472
1/4 Pitch Shift	-0.145	0.885
Whole Tone Pitch Shift	-3.69	**3.05e-4** *

(c) Group Pairwise Comparisons (Bonferroni-adjusted), Condition

Task	Statistic	p
Toggle	-0.395	0.694
Toggled & Voice Dist.	-2.97	**0.00345** *

(d) Group Pairwise Comparisons (Bonferroni-adjusted), Task

Condition A	Condition B	Estimate	p
Delay 200 ms	Delay 600 ms	-0.127	0.984
Delay 200 ms	1/4 Pitch Shift	-0.276	0.872
Delay 200 ms	Whole Tone Pitch Shift	0.660	0.261
Delay 600 ms	Whole Tone Pitch Shift	-0.149	0.976
Delay 600 ms	1/4 Pitch Shift	0.786	0.139
1/4 Pitch Shift	Whole Tone Pitch Shift	0.935	0.0712

(e) Tukey's HSD, Low BAIS Group, Condition

Task A	Task B	Estimate	p
Toggled	Toggled & Voice Dist.	-0.229	0.902

(f) Tukey's HSD, Low BAIS, Task

Table A.5: Full-factorial results from analysis of the effect on individual-adjusted TRD by interaction between BAIS Group, Condition, and Task.

A.5 Individual-Adjusted CV Analyses

Effect	DFn	F	p
Task	2	0.233	0.792
Condition	3	7.321	**1.33e-4 ***
BAIS Group	1	7.323	**0.008 ***
Task : Condition	6	0.683	0.664
Task : BAIS Group	2	0.179	0.836
Condition : BAIS Group	3	0.657	0.179
Task : Condition : BAIS Group	6	0.457	0.839

(a) 2x3x4 Analysis of Variance (ANOVA Type II Tests)

BAIS Group	Effect	DFn	F	p
High	Condition	3	0.884	0.451
High	Task	2	0.521	0.959
High	Condition : Task	6	0.451	0.842
Low	Condition	3	8.22	**4.38e-5 ***
Low	Task	2	0.002	0.998
Low	Condition : Task	6	0.688	0.659

(b) Two-Way Interaction by Group (ANOVA Type II Tests)

Condition	Statistic	p
Delay 200 ms	2.85	**4.95e-3 ***
Delay 600 ms	2.34	**0.0205 ***
1/4 Pitch Shift	0.243	0.808
Whole Tone Pitch Shift	-0.0108	0.991

(c) Group Pairwise Comparisons (Bonferroni-adjusted), Condition

Condition A	Condition B	Estimate	p
Delay 200 ms	Delay 600 ms	0.512	0.962
Delay 200 ms	1/4 Pitch Shift	3.21	**0.0325 ***
Delay 200 ms	Whole Tone Pitch Shift	3.29	**0.021 ***
Delay 600 ms	Whole Tone Pitch Shift	2.69	0.0894
Delay 600 ms	1/4 Pitch Shift	2.78	0.0625
1/4 Pitch Shift	Whole Tone Pitch Shift	0.0899	0.99

(d) Tukey's HSD, Low BAIS Group, Condition

Table A.6: Full-factorial results from analysis of the effect on individual-adjusted CV by interaction between BAIS Group, Condition, and Task.

A.6 Individual-Adjusted MBs Analyses

Effect	DFn	F	p
Task	1	2.646	0.107
Condition	3	2.615	0.056
BAIS Group	1	0.319	0.574
Task : Condition	3	0.977	0.407
Task : BAIS Group	1	0.051	0.822
Condition : BAIS Group	3	0.363	0.780
Task : Condition : BAIS Group	3	0.261	0.853

(a) 2x3x4 Analysis of Variance (ANOVA Type II Tests)

BAIS Group	Effect	DFn	F	p
High	Condition	3	1.75	0.162
High	Task	1	1.57	0.214
High	Condition : Task	3	0.902	0.443
Low	Condition	3	1.26	0.293
Low	Task	1	1.10	0.298
Low	Condition : Task	3	0.336	0.8

(b) Two-Way Interaction by Group (ANOVA Type II Tests)

Table A.7: Full-factorial results from analysis of the effect on individual-adjusted MBs by interaction between BAIS Group, Condition, and Task.

A.7 Group-Adjusted TRD Analyses

Effect	DFn	F	p
Task	2	0.355	0.702
Condition	3	1.223	0.304
BAIS Group	1	0.010	0.921
Task : Condition	6	0.470	0.830
Task : BAIS Group	2	0.549	0.579
Condition : BAIS Group	3	1.815	0.147
Task : Condition : BAIS Group	6	1.081	0.377

(a) 2x3x4 Analysis of Variance (ANOVA Type II Tests)

BAIS Group	Effect	DFn	F	p
High	Condition	3	0.367	0.777
High	Task	2	0.115	0.891
High	Condition : Task	6	0.817	0.558
Low	Condition	3	2.78	**0.043 ***
Low	Task	2	0.757	0.471
Low	Condition : Task	6	0.734	0.623

(b) Two-Way Interaction by Group (ANOVA Type II Tests)

Condition	Statistic	p
Delay 200 ms	0.00421	0.997
Delay 600 ms	1.125	0.213
1/4 Pitch Shift	0.613	0.541
Whole Tone Pitch Shift	-2.00	**0.047 ***

(c) Group Pairwise Comparisons (Bonferroni-adjusted), Condition

Condition A	Condition B	Estimate	p
Delay 200 ms	Delay 600 ms	-0.118	0.996
Delay 200 ms	1/4 Pitch Shift	-0.261	0.961
Delay 200 ms	Whole Tone Pitch Shift	0.578	0.679
Delay 600 ms	Whole Tone Pitch Shift	-0.143	0.993
Delay 600 ms	1/4 Pitch Shift	0.696	0.54
1/4 Pitch Shift	Whole Tone Pitch Shift	0.839	0.41

(d) Tukey's HSD, Low BAIS Group, Condition

Table A.8: Full-factorial results from analysis of the effect on group-adjusted TRD by interaction between BAIS Group, Condition, and Task.

A.8 Group-Adjusted CV Analyses

Effect	DFn	F	p
Task	2	0.262	0.77
Condition	3	12.721	**2.01e-07 ***
BAIS Group	1	0.017	0.91
Task : Condition	6	0.513	0.79
Task : BAIS Group	2	0.225	0.79
Condition : BAIS Group	3	1.220	0.31
Task : Condition : BAIS Group	6	0.394	0.88

(a) 2x3x4 Analysis of Variance (ANOVA Type II Tests)

BAIS Group	Effect	DFn	F	p
High	Condition	3	3.36	**0.02 ***
High	Task	2	0.574	0.565
High	Condition : Task	6	0.22	0.97
Low	Condition	3	10.6	**2.37e-6 ***
Low	Task	2	0.0007	0.993
Low	Condition : Task	6	0.687	0.66

(b) Two-Way Interaction by Group (ANOVA Type II Tests)

Condition	Statistic	p
Delay 200 ms	0.995	0.321
Delay 600 ms	0.633	0.527
$1/4$ Pitch Shift	-0.208	0.835
Whole Tone Pitch Shift	-1.64	0.104

(c) Group Pairwise Comparisons (Bonferroni-adjusted), Condition

Table A.9: Full-factorial results from analysis of the effect on group-adjusted CV by interaction between BAIS Group, Condition, and Task.

A.9 Group-Adjusted MBs Analyses

Effect	DFn	F	p
Task	1	3.442	0.067
Condition	3	2.421	0.071
BAIS Group	1	0.106	0.745
Task : Condition	3	0.883	0.453
Task : BAIS Group	1	2.000	0.161
Condition : BAIS Group	3	0.567	0.638
Task : Condition : BAIS Group	3	0.155	0.926

(a) 2x3x4 Analysis of Variance (ANOVA Type II Tests)

BAIS Group	Effect	DFn	F	p
High	Condition	3	2.09	0.107
High	Task	1	0.081	0.777
High	Condition : Task	3	0.733	0.535
Low	Condition	3	0.94	0.425
Low	Task	1	5.57	**0.02 ***
Low	Condition : Task	3	0.305	0.821

(b) Two-Way Interaction by Group (ANOVA Type II Tests)

Task	Statistic	p
Toggle	-0.629	0.531
Toggled & Voice Dist.	1.46	0.146

(c) Group Pairwise Comparisons (Bonferroni-adjusted), Task

Table A.10: Full-factorial results from analysis of the effect on group-adjusted MBs by interaction between BAIS Group, Condition, and Task.

Appendix B

Understanding Vocal Perception

Interview prompts and imagery assessments from Chapter 6.

B.1 Interview Script & Prompts

The interview lasts approximately 30 minutes, but is open to last as long as the participant wants to discuss their teaching methods or background in vocal pedagogy. The interview was conducted in a semi-structured way, following these questions as conversational prompts and discussing in detail the aspects the teacher conveyed.

B.1.1 Intro & Background Questions:

This portion of the interview is just to get a general background of your teaching and singing career.

1. How long have you been singing, approximately? E.g., is it something you have always done, or did you develop an interest in it later?
2. What is your own most predominant style/what musical style do you most enjoy singing?
3. Are you regularly performing?
4. How long have you been teaching, approximately?
5. What style(s) do you predominantly teach?
6. Have you studied teaching practices (voice or otherwise) or have any certifications?
7. Do you speak any other languages (other than English)?
    - If YES: Did you undertake voice instruction in another language or do you teach mainly in a language other than English?
    - *In this case, take note of the exact wording of any metaphors based in a language other than English from the teachers going forward.*

B.1.2 How You Learned to Sing:

I want to know more about how you learned when you were a student. When you were a beginning voice student...

8. At what age did you have your first one-to-one voice instruction?

9. Do you remember your teacher using any metaphorical language in the voice lesson, especially any that stuck with you?
    - *This is to encourage teachers to think about some of the more evocative metaphors they have heard and to get them into thinking of and remembering the abstract references they had heard as students.*
10. Did your teacher include voice physiology in the lesson? If yes, to what extent did your teacher discuss/show/mention physiological technique?
11. How might your teacher have described these fundamental techniques to you?:
    - Supported breathing?
    - Body posture and avoiding tension or strain (particularly in the neck or the larynx)?
    - Sound production in the larynx?
    - Resonance spaces and shaping vowels/mouth space?

B.1.3 How You Teach Others to Sing:

Let's now go through metaphors as if you were teaching a beginner student. You can instruct to me, as if I'm a beginner voice student, on the following metaphors:

12. How would you instruct on:
    - Supported breathing; for instance, if I'm using too much air too quickly and don't have good control over my airflow?
    - Body posture; for instance, if I'm slouching and maybe have my shoulders rolled forward?
    - Avoiding tension; for instance, if I've locked my muscles or if I'm reaching with my chin and getting pinched sound when I sing high notes?
    - Resonance; for instance, I'm struggling to project my sound or I'm not shaping my vowels well and I've started to sing through a nasal sound?
13. In general, what is your feeling on the use of metaphor in voice education compared to physiology?
14. Would you prefer physiology or metaphor used more often/less often, or in some combination?
15. How do you feel physiological and metaphorical instruction might benefit or detract from the lesson?

B.2 Imagery Self-Assessments

Teachers self-assessed their imagery abilities using the Bucknell Auditory Imagery Scale (BAIS) for auditory imagery ability (Halpern, 2015) and the Movement Imagery Questionnaire 3 (MIQ-3) for visual and kinetic imagery ability (Williams et al., 2012). The scores for each participant are included here. The scoring ranges from 1 to 7 on all scales.

	MIQ-3			BAIS	
Pt	**Internal Visual**	**External Visual**	**Kinaesthetic**	**Auditory Vividness**	**Auditory Control**
1	6.00	6.25	5.75	4.00	5.14
4	4.75	6.00	3.75	5.42	4.86
5	6.00	6.00	6.25	5.14	4.36
8	5.00	6.00	3.25	4.36	4.64
10	7.00	6.00	6.50	6.07	5.79
11	5.50	6.75	6.75	6.43	6.71
12	6.75	6.00	5.00	5.57	4.93
17	5.00	5.00	5.75	4.07	4.14
18	6.50	6.75	6.00	6.71	6.86
19	5.75	6.50	5.50	6.00	6.57
22	5.75	5.75	5.75	5.93	5.64
24	6.50	5.75	5.75	6.50	6.79
M	5.88	6.06	5.50	5.52	5.54
SD	0.73	0.478	1.04	0.94	0.99

Table B.1: Participant self-assessed scores on the MIQ-3 and BAIS for visual, kinaesthetic, and auditory imagery.

Appendix C

Exploring Vocal Movement & Perception

Materials and interview prompts from Chapter 9.

C.1 Interactivity Questionnaire

The Interactivity Questionnaire was completed during both debriefings:

	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
I felt connected to the sound					
I was able to communicate musically through the sound					
I felt in control of the sound					
The sound felt unnatural to work with					
The sound felt like a part of me, an extension of my body					
I found the sound unresponsive and hard to control					
I was able to make connections between my movement and the sound					
The sound influenced the movements I was making					

Table C.1: The Interactivity Questionnaire rating scale.

C.2 Vocalises

These four vocalises were tasked to the vocalists for the Targeted Technique phase. The exercises were sung to the vocalists during the Week 2 debriefing, which was recorded and offered back to the vocalists for reference. It was also offered to provide a transcription but all three vocalists declined this, working with the auditory reference.

1. Comfortable and flexible posture during singing.
At the start of the practice, stand or sit comfortably and align your posture to eliminate tension and create flexibility in the neck. Focus on release of tension. Keep the shoulders back, neck long and relaxed, chin tucked slightly down. Take as long as you need to feel comfortable.

2. Sustained and controlled breathing.
First work on a hissing vocalise to warm up and get the breath going. This should use a sustained *Sss* hiss sound, to move the air without pitch. Exhale on the hiss starting with four counts, then move to eight, and further if you wish. You can slow or speed up the tempo. Try to focus on sustaining the breath for longer and longer each time. Feel the tension in your abdomen but not in your neck and back as you control the breath.

3. Sound production with articulation.
Use a descending pattern on *Ta ta ta ta ta* (sol to do) to begin to create sound. Use the articulation to get the breath going and focus the sound. You can start as high or as low as you like and end where you choose. Focus on each pulse and the feeling of the articulation as you sing.

4. Vowel formation.
Use a sustained pitch to go through different vowel sounds *Aa Eh Ee Oh Oo*, ascending after each group. Again, start as high or as low as you like and end where you choose. Focus on the quality of the sound and creating clear, distinct sounds.

After completing these vocalises, sing again as you please - whatever you want to try or focus on. This can be from your normal repertoire if you are working on something in particular, or explore new vocalises or exercises as you please.

C.3 Interview Script & Prompts

Vocalists answered a series of semi-structured interview questions during each debriefing session:

C.3.1 Exploratory (Week 2)

- General

- How are things going with the interaction experience?
- Are there any [initial] impressions you want to share?

- Interactivity Questionnaire

[Complete the Interactivity Questionnaire] Please briefly explain your thinking or elaborate on why you select the answer.

- Micro-phenomenological Interview

Vocalists are asked to choose a specific moment of connect between movement and sound they noticed to explore in a micro-phenomenological interview.

- Controllability and Working with the Sound

- What are your overall impressions of the quality of the sound?
- Do you find the sound pleasing to work with?
- What would you wish to change about the sound?
- Can you describe the connections between your movement and the resulting sound?

C.3.2 Targeted Technique (Week 4)

The second debriefing used the same components as above:

- General

- Interactivity Questionnaire

- Micro-phenomenological Interview

- Vocal Fundamentals

This was the only section added to the interview script, to address the vocalises added:
For each vocalise (*go through each one-by-one*):

- Did you notice anything in particular about the sound while you performed the exercise?
- Did this change over time (noticing anything more or less, different impressions)?
- What was surprising? What was not?
- If you were to teach someone else (a beginner student), how might you explain this technique?
- What would you say about the sound while performing this? What would you tell that student to listen for?

- Controllability and Working with the Sound

Appendix D

Equipment & Materials

Here you can find a full list of materials used in each study, in the case that you wish to order any of them or reproduce this work exactly. For the other electrical components listed throughout, any version meeting the specifications listed is suitable.

D.1 Vocalists' Use of Auditory Imagery

Equipment:

- **ART 412-A speakers (RCF)** - listening in-room to directions and audio feedback
- **DT100 over-ear studio headset (Beyerdynamic)** - listening to audio feedback
- **C414B-XLII condenser microphone (AKG)** - recording participants' singing
- **MG16XU mixing console (Yamaha)** - routing microphone input and audio outputs
- **4-710d Tone-Blending mic preamp (Universal Audio)** - amplifying recorded audio, boosting vocals in recordings
- **24" LCD monitor (BenQ)** - viewing visual stimuli
- **Mac Pro (Mac OS 10.14.1)** - running other software

Software:

- **Logic Pro X (Apple)** - recording audio tracks
- **MAX/MSP, Max 7 (Cycling'74)** - producing and timing visual stimuli
- **Tony (Queen Mary University of London** - annotating f_0 of sung pitches
- **Sonic Visualiser (Queen Mary University of London)** - annotating f_0 of sung pitches

Questionnaires:

- **Goldsmiths Musical Sophistication Index (Gold-MSI)** - assessing participant musical aptitude and ability
- **Bucknell Auditory Imagery Scale (BAIS)** - assessing participant auditory imagery

D.2 Understanding Vocal Perception

Other Materials:

- **Goldsmiths Musical Sophistication Index (Gold-MSI)** - assessing participant musical aptitude and ability
- **Bucknell Auditory Imagery Scale (BAIS)** - assessing participant auditory imagery
- **Movement Imagery Questionnaire-3 (MIQ-3)** - assessing participant kinetic and visual imagery

D.3 Surface Electromyography for Vocal interaction

Equipment:

- **Gold-plated silver cup electrodes (MediMaxTech, New Malden, UK)** - sEMG sensing
- **Hypafix non-woven fabric tape (BSN Medical GmbH, Hamburg, Germany)** - securing electrodes to skin
- **Ten20 Conductive Paste (Weaver and Company, Aurora, CO, USA)** - securing electrodes to skin
- **Bela board (Bela)** - signal processing
- **OPA1612 (Texas Instruments)** - op-amp for differential amplification of sEMG signals
- **Hand-flat double bedded knitting machine (Dubied)** - constructing the body of the Singing Knit
- **Conductive zebra jersey (Hitek)** - constructing soft fabric electrode pads
- **MSO 2024B oscilloscope (Tektronix)** - comparing signals in prototyping

D.4 Autoethnographic Interaction & Evaluation

Software:

- **Supercollider v. 3.12.2** - sonification of sEMG data
- **Pure Data, pd vanilla v. 0.52-2** - sonification of sEMG data

D.5 Vocalists' Use of Auditory Imagery

Equipment:

- **Kendall H124SG disposable ECG electrodes (Cardinal Health)** - sEMG sensing
- **CAB-12970 sensor cables (Sparkfun Electronics)** - connecting electrodes to the Vox-EMG boards
- **Kinesiologie-Tape non-woven fabric tape (Altapharma)** - securing electrodes to skin
- **Aurora in-ear headphones (iFrogz)** - listening to sonified feedback

Software:

- **Pure Data, pd vanilla v. 0.52-2** - sonification of sEMG data

Bibliography

Gizem Acar, Ozberk Ozturk, Ata Jedari Golparvar, Tamador Alkhidir Elboshra, Karl Böhringer, and Murat Kaya Yapici. Wearable and flexible textile electrodes for biopotential signal monitoring: A review. *Electronics*, 8(5):479, 2019. DOI 10.3390/electronics8050479. (Cited on page 120).

Meredith A. Achey, Mike Z. He, and Lee M. Akst. Vocal Hygiene Habits and Vocal Handicap Among Conservatory Students of Classical Singing. *Journal of Voice*, 30(2):192–197, 2016. DOI 10.1016/j.jvoice.2015.02.003. (Cited on pages 13 and 42).

Ozgun Kilic Afsar, Ali Shtarbanov, Hila Mor, Ken Nakagaki, Jack Forman, Karen Modrei, Seung Hee Jeong, Klas Hjort, Kristina Höök, and Hiroshi Ishii. OmniFiber: Integrated Fluidic Fiber Actuators for Weaving Movement Based Interactions into the 'Fabric of Everyday Life'. In *The 34th Annual ACM Symposium on User Interface Software and Technology*, UIST '21, page 1010–1026, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450386357. DOI 10.1145/3472749.3474802. (Cited on page 22).

Anand Agarawala and Ravin Balakrishnan. Keepin' It Real: Pushing the Desktop Metaphor with Physics, Piles and the Pen. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '06, page 1283–1292, New York, NY, USA, 2006. Association for Computing Machinery. ISBN 1595933727. DOI 10.1145/1124772.1124965. (Cited on pages xvi, 35, 106, and 108).

André Aleman and Mascha van't Wout. Subvocalization in auditory-verbal imagery: just a form of motor imagery? *Cognitive Processing*, 5(4):228–231, 2004. DOI 10.1007/s10339-004-0034-y. (Cited on page 57).

Vinoo Alluri and Petri Toiviainen. Exploring Perceptual and Acoustical Correlates of Polyphonic Timbre. *Music Perception*, 27(3):223–242, February 2010. DOI 10.1525/mp.2010.27.3.223. (Cited on page 48).

Ron Amadeo. The (updated) history of Android, Oct 2016. URL https://arstechnica.com/gadgets/2016/10/building-android-a-40000-word-history-of-googles-mobile-os/. (Cited on page 108).

Robert St. Amant, Clayton T. Morrison, Yu-Han Chang, Wei Mu, Paul R. Cohen, and Carole Beal. An Image Schema Language. In *7th International Conference on Cognitive Modelling (ICCM), April 5-8, Trieste, Italy*, pages 2–6, 2006. (Cited on page 35).

Kristina Andersen, Ron Wakkary, Laura Devendorf, and Alex McLean. Digital Crafts-Machine-Ship: Creative Collaborations with Machines. *Interactions*, 27(1):30–35, Dec 2019. ISSN 1072-5520. DOI 10.1145/3373644. (Cited on page 32).

John Annett. Motor Skills. In N. J. Mackintosh, N. J., and A. M. Colman, editors, *Learning and Skills*, pages 56–75. Longman, London/New York, 1994. (Cited on page 42).

Alissa N. Antle, Greg Corness, and Milena Droumeva. What the body knows: Exploring the benefits of embodied metaphors in hybrid physical digital environments. *Interacting with Computers*, 21 (1-2):66–75, 2009. DOI 10.1016/j.intcom.2008.10.005. (Cited on pages 35, 37, and 42).

Mihailo Antović. Schemas, grounds, meaning: On the emergence of musical concepts through conceptual blending. *Musicae Scientiae*, 22(1):57–71, 2018. DOI 10.1177/1029864917711218. (Cited on page 39).

Newton B. Armstrong. *An enactive approach to digital musical instrument design–Theory, Models, Techniques*. VDM Verlag, 2007. ISBN 978-3836419260. (Cited on pages 32 and 38).

Ozgur Atalay, Asli Tuncay, Muhammad D. Husain, and William R. Kennon. Comparative study of the weft-knitted strain sensors. *Journal of Industrial Textiles*, 46(5):1212–1240, 2017. DOI 10.1177/1528083715619948. (Cited on page 120).

Chester J. Atkinson. Adaptation to delayed side-tone. *Journal of Speech and Hearing Disorders*, 18 (4):386–391, 1953. (Cited on page 66).

Jean-Julien Aucouturier and Emmanuel Bigand. Mel Cepstrum & Ann Ova: The Difficult Dialog Between MIR and Music Cognition. In *13th International Society for Music Information Retrieval Conference (ISMIR), Porto, Portugal, Oct 8-12*, pages 397–402, 2012. (Cited on page 48).

Krotos Audio. Dehumaniser II User Manual. 2022. URL https://s3-us-west-2.amazonaws.com/dehumaniser/Manuals/Dehumaniser+2+Manual.pdf. (Cited on page 22).

Waves Audio. OVox Vocal ReSynthesis: Waves OVox User Guide. 2020. URL https://www.waves.com/1lib/pdf/plugins/ovox-vocal-resynthesis-v2.pdf. (Cited on page 21).

Alan Baddeley, Marge Eldridge, and Vivien Lewis. The role of subvocalisation in reading. *The Quarterly Journal of Experimental Psychology Section A*, 33(4):439–454, 1981. DOI 10.1080/14640748108400802. (Cited on page 57).

Freya Bailes. The use of experience-sampling methods to monitor musical imagery in everyday life. *Musicae Scientiae*, 10(2):173–190, 2006. DOI 10.1177/102986490601000202. (Cited on pages 27 and 59).

Karen Barad. Posthumanist performativity: Toward an understanding of how matter comes to matter. *Signs: Journal of women in culture and society*, 28(3):801–831, 2003. (Cited on page 32).

Karen Barad. *Meeting the Universe Halfway: Quantum Physics and the Entanglement of Matter and Meaning*. Duke University Press, 2007. (Cited on pages 32, 46, and 47).

Christopher Bartlette, Dave Headlam, Mark Bocko, and Gordana Velikic. Effect of Network Latency on Interactive Musical Performance. *Music Perception*, 24(1):49–62, September 2006. DOI 10.1525/mp.2006.24.1.49. (Cited on page 67).

Eliot Bates. The Social Life of Musical Instruments. *Ethnomusicology*, 56(3):363–395, 2012. DOI 10.5406/ethnomusicology.56.3.0363. (Cited on page 38).

Franziska Baumann. Interview by Franziska Baumann with Pamela Z, 2021. URL http://www.franziskabaumann.ch/en/press/interview1-pamela.php. (Cited on page 23).

Franziska Baumann. *Embodied Human–Computer Interaction in Vocal Music Performance*. Springer, 2023. (Cited on page 23).

Janet Beavin Bavelas, Alex Black, Charles R. Lemery, and Jennifer Mullett. I *Show* How You Feel: Motor Mimicry as a Communicative Act. *Journal of Personality and Social Psychology*, 50(2): 322–329, 1986. DOI 10.1037/0022-3514.50.2.322. (Cited on page 12).

Michel Beaudouin-Lafon, Wendy E. Mackay, Peter Andersen, Paul Janecek, Mads Jensen, Michael Lassen, Kasper Lund, Kjeld Mortensen, Stephanie Munck, Katrine Ravn, Anne Ratzer, Søren Christensen, and Kurt Jensen. CPN/Tools: Revisiting the Desktop Metaphor with Post-WIMP Interaction Techniques. In *CHI '01 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '01, page 11–12, New York, NY, USA, 2001a. Association for Computing Machinery. ISBN 1581133405. DOI 10.1145/634067.634076. (Cited on page 36).

Michel Beaudouin-Lafon, Wendy E. Mackay, Mads Jensen, Peter Andersen, Paul Janecek, Henry Michael Lassen, Kasper Lund, Kjeld Høyer Mortensen, Stephanie Munck, Anne V. Ratzer, Katrine Ravn, Søren Christensen, and Kurt Jensen. CPN/Tools: A Tool for Editing and Simulating Coloured Petri Nets ETAPS Tool Demonstration Related to TACAS. In *Proceedings of the 7th International Conference on Tools and Algorithms for the Construction and Analysis of Systems*, TACAS 2001, page 574–577, Berlin, Heidelberg, 2001b. Springer-Verlag. ISBN 3540418652. (Cited on pages 36 and 102).

Vincent Becker, Pietro Oldrati, Liliana Barrios, and Gábor Sörös. TouchSense: Classifying and Measuring the Force of Finger Touches with an Electromyography Armband. In *Proceedings of the 9th Augmented Human International Conference*, AH '18, New York, NY, USA, 2018. Association for Computing Machinery. ISBN 9781450354158. DOI 10.1145/3174910.3174947. (Cited on pages 55, 56, and 119).

Guillermo Bernal, Dishaan Ahuja, and Federico Casalegno. EMG-Based Biofeedback Tool for Augmenting Manual Fabrication and Improved Exchange of Empirical Knowledge. In *Proceedings of the XVI International Conference on Human Computer Interaction*, Interacción '15, New York, NY, USA, 2015. Association for Computing Machinery. ISBN 9781450334631. DOI 10.1145/2829875.2829932. (Cited on page 58).

Michael Bernays and Caroline Traube. Verbal expression of piano timbre: Multidimensional semantic space of adjectival descriptors. In *Proceedings of the International Symposium on Performance Science (ISPS2011)*, pages 43–53, Augustt, 2011. ISBN 978-94-90306-02-1. (Cited on page 12).

Roberta Bianco, Giacomo Novembre, Peter E. Keller, Arno Villringer, and Daniela Sammler. Musical genre-dependent behavioural and EEG signatures of action planning. A comparison between classical and jazz pianists. *NeuroImage*, 169:383–394, 2018. DOI 10.1016/j.neuroimage.2017.12.058. (Cited on page 27).

Emmanuel Bigand, Richard Parncutt, and Fred Lerdahl. Perception of musical tension in short chord sequences: The influences of harmonic function, sensory dissonance, horizontal motion, and musical training. *Perception & Psychophysics*, 58(1):125–141, 1996. DOI 10.3758/bf03205482. (Cited on page 31).

S. M. Astrid Bin. *The Show Must Go Wrong: Towards an Understanding of Audience Perception of Error in Digital Musical Instrument Performance.* Doctoral dissertation, Queen Mary University of London, London, UK, 2018. (Cited on page 149).

Laura Bishop, Freya Bailes, and Roger T. Dean. Musical Imagery and the Planning of Dynamics and Articulation During Performance. *Music Perception: An Interdisciplinary Journal*, 31(2): 97–117, 2013. DOI 10.1525/mp.2013.31.2.97. (Cited on page 28).

John W. Black. The effect of delayed side-tone upon vocal rate and intensity. *Journal of Speech and Hearing Disorders*, 16(1):56–60, 1951. (Cited on page 66).

Alan F. Blackwell. The Reification of Metaphor as a Design Tool. *ACM Trans. Comput.-Hum. Interact.*, 13(4):490–530, 2006. ISSN 1073-0516. DOI 10.1145/1188816.1188820. (Cited on pages 36 and 188).

Pasquale Bottalico, Simone Graetzer, and Eric J. Hunter. Effect of Training and Level of External Auditory Feedback on the Singing Voice: Volume and Quality. *Journal of Voice*, 30(4):434–442, July 2016. DOI 10.1016/j.jvoice.2015.05.010. (Cited on page 83).

Pasquale Bottalico, Simone Graetzer, and Eric J. Hunter. Effect of Training and Level of External Auditory Feedback on the Singing Voice: Pitch Inaccuracy. *Journal of Voice*, 31(1):122.e9–122.e16, January 2017. DOI 10.1016/j.jvoice.2016.01.012. (Cited on page 83).

Arend Bouhuys, Donald F. Proctor, and Jere Mead. Kinetic aspects of singing. *Journal of Applied Physiology*, 21(2):483–496, 1966. DOI 10.1152/jappl.1966.21.2.483. (Cited on pages 13, 15, 16, and 17).

Katherine M. Boydell, Angela Dew, Susan Collings, Kate Senior, and Louisa Smith, editors. *Applying Body Mapping In Research*. Routledge, 2020. DOI 10.4324/9780429340260. (Cited on pages 37, 49, and 188).

Virginia Braun and Victoria Clarke. Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2):77–101, 2006. DOI 10.1191/1478088706qp063oa. (Cited on pages 52 and 92).

Virginia Braun and Victoria Clarke. Thematic Analysis. In H. Cooper, P. M Camic, D. L. Long, A. T. Panter, D. Rindskopf, and K. J. Sher, editors, *PA Handbook of Research Methods in Psychology*, volume 2: Research Designs: Quantitative, Qualitative, Neuropsychological, and Biological. American Psychological Association, Washington, 2012. ISBN 978-1-4338-1005-3. (Cited on pages 52 and 157).

Virginia Braun and Victoria Clarke. One size fits all? What counts as quality practice in (reflexive) thematic analysis? *Qualitative Research in Psychology*, 18(3):328–352, 2020a. DOI 10.1080/14780887.2020.1769238. (Cited on pages 52 and 157).

Virginia Braun and Victoria Clarke. One size fits all? What counts as quality practice in (reflexive) thematic analysis? *Qualitative Research in Psychology*, 18(3):328–352, August 2020b. DOI 10.1080/14780887.2020.1769238. (Cited on page 92).

Warren Brodsky, Yoav Kessler, Bat-Sheva Rubinstein, Jane Ginsborg, and Avishai Henik. The mental representation of music notation: Notational audiation. *Journal of Experimental Psychology: Human Perception and Performance*, 34(2):427–445, 2008a. DOI 10.1037/0096-1523.34.2.427. (Cited on page 27).

Warren Brodsky, Yoav Kessler, Bat-Sheva Rubinstein, Jane Ginsborg, and Avishai Henik. The mental representation of music notation: Notational audiation. *Journal of Experimental Psychology: Human Perception and Performance*, 34(2):427–445, 2008b. DOI 10.1037/0096-1523.34.2.427. (Cited on page 57).

Rachel M. Brown and Caroline Palmer. Auditory–motor learning influences auditory memory for music. *Memory & Cognition*, 40(4):567–578, 2012. DOI 10.3758/s13421-011-0177-x. (Cited on pages 28 and 83).

Bryony Buck, Jennifer MacRitchie, and Nicholas J. Bailey. The Interpretive Shaping of Embodied Musical Structure in Piano Performance. *Empirical Musicology Review*, 8(2):92–119, 2013. DOI 10.18061/emr.v8i2.3929. (Cited on pages 11, 12, and 19).

Theresa A. Burnett, Marcia B. Freedland, Charles R. Larson, and Timothy C. Hain. Voice F0 responses to manipulations in pitch feedback. *The Journal of the Acoustical Society of America*, 103(6):3153–3161, June 1998. DOI 10.1121/1.423073. (Cited on page 81).

Jean Callaghan. Singing Teachers and Voice Science - An Evaluation of Voice Teaching in Australian Tertiary Institutions. *Research Studies in Music Education*, 10(1):25–41, 1998. DOI 10.1177/1321103x9801000103. (Cited on pages 39, 41, 42, and 104).

Yves Candau, Jules Françoise, Sarah Fdili Alaoui, and Thecla Schiphorst. Cultivating kinaesthetic awareness through interaction. In *Proceedings of MOCO'17, , United Kingdom, June 2017, 8 pages*, pages 28–30. 2017. DOI 10.1145/3077981.3078042. (Cited on page 34).

David Canfield Smith. SIGCHI Lifetime Research Award Talk: Icons, Metaphor, and End-User Programming. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI EA '20, page 1–9, New York, NY, USA, 2020. Association for Computing Machinery. ISBN 9781450368193. DOI 10.1145/3334480.3386148. (Cited on pages 35 and 36).

Chris Cannam, Christian Landone, and Mark Sandler. Sonic Visualiser: An Open Source Application for Viewing, Analysing, and Annotating Music Audio Files. In *ACM Multimedia 2010 International Conference (MM'10), Firenze, Italy, 25-29 Oct* , 2010. (Cited on page 68).

Jonathan J. Cannon and Anirudhh D. Patel. How Beat Perception Co-opts Motor Neurophysiology. *Trends in Cognitive Sciences*, 25(2):137–150, 2021. DOI 10.1016/j.tics.2020.11.002. (Cited on page 81).

Dario Cazzani. Posture Identification of Musicians Using Non-Intrusive Low-Cost Resistive Pressure Sensors. In Edgar Berdahl and Jesse Allison, editors, *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 54–57, Baton Rouge, Louisiana, USA, 2015. Louisiana State University. DOI 10.5281/zenodo.1179042. (Cited on page 54).

Lisa P. Chan, Steven R. Livingstone, and Frank A. Russo. Facial Mimicry in Response to Song. *Music Perception: An Interdisciplinary Journal*, 30(4):361–367, 2013. DOI 10.1525/mp.2013.30.4.361. (Cited on page 13).

Olivier Chapuis and Nicolas Roussel. Metisse is Not a 3D Desktop! In *Proceedings of the 18th Annual ACM Symposium on User Interface Software and Technology*, UIST '05, page 13–22, New York, NY, USA, 2005. Association for Computing Machinery. ISBN 1595932712. DOI 10.1145/1095034.1095038. (Cited on page 108).

Hillel J. Chiel and Randall D. Beer. The brain has a body: adaptive behavior emerges from interactions of nervous system, body and environment. *In Trends in Neurosciences*, 20(12):553–557, 1997. DOI 10.1016/s0166-2236(97)01149-1. (Cited on pages 34 and 36).

Rubana H. Chowdhury, Mamun B. I. Reaz, Mohd Alauddin Bin Mohd Ali, Ashrif A. A. Bakar, Kalaivani Chellappan, and Tae G. Chang. Surface Electromyography Signal Processing and Classification Techniques. *Sensors*, 13(9):12431–12466, 2013. DOI 10.3390/s130912431. (Cited on pages 55 and 56).

Miha Ciglar. An Ultrasound Based Instrument Generating Audible and Tactile Sound. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 19–22, Sydney, Australia, 2010. DOI 10.5281/zenodo.1177745. (Cited on page 54).

Terry Clark, Aaron Williamon, and Aleksandar Aksentijevic. Musical imagery and imagination: The function, measurement, and application of imagery skills for performance. In D. Hargreaves, D. Miell, and R. MacDonald, editors, *Musical Imaginations: Multidisciplinary perspectives on creativity, performance and perception*, pages 351–366. Oxford University Press, December 2011. DOI 10.1093/acprof:oso/9780199568086.003.0022. (Cited on page 30).

Thomas F. Cleveland. Vocal pedagogy in the twenty first century: Mental imaging and the teaching of voice. *The National Association of Teachers of Singing Journal*, 45(3):54, 1989. (Cited on page 42).

Martin Coath, Susan L. Denham, Leigh M. Smith, Henkjan Honing, Amaury Hazan, Piotr Holonowicz, and Hendrik Purwins. Model cortical responses for the detection of perceptual onsets and beat tracking in singing. *Connection Science*, 21(2-3):193–205, 2009. DOI 10.1080/09540090902733905. (Cited on page 82).

Karen Anne Cochrane, Kristina Mah, Anna Ståhl, Claudia NúNúñezez-Pacheco, Madeline Balaam, Naseem Ahmadpour, and Lian Loke. Body Maps: A Generative Tool for Soma-Based Design. In *Sixteenth International Conference on Tangible, Embedded, and Embodied Interaction*, TEI '22, New York, NY, USA, 2022. Association for Computing Machinery. ISBN 9781450391474. DOI 10.1145/3490149.3502262. (Cited on pages 37, 39, 49, 101, 102, and 188).

Berton Coffin. *Historical Vocal Pedagogy Classics.* Scarecrow Press, Inc., Lanham, Maryland, 1989. ISBN 0-8108-4412-5. (Cited on page 41).

Ian D. Colley, Peter E. Keller, and Andrea R. Halpern. Working memory and auditory imagery predict sensorimotor synchronisation with expressively timed music. *Quarterly Journal of Experimental Psychology*, 71(8):1781–1796, January 2018a. DOI 10.1080/17470218.2017.1366531. (Cited on page 81).

Ian D. Colley, Manuel Varlet, Jennifer MacRitchie, and Peter E. Keller. The influence of visual cues on temporal anticipation and movement synchronization with musical sequences. *Acta Psychologica*, 191:190–200, 2018b. DOI 10.1016/j.actpsy.2018.09.014. (Cited on pages 12 and 64).

Geoffrey L. Collier and R. Todd Ogden. Adding Drift to the Decomposition of Simple Isochronous Tapping: An Extension of the Wing-Kristofferson Model. *Journal of Experimental Psychology: Human Perception and Performance*, 30(5):853–872, 2004. DOI 10.1037/0096-1523.30.5.853. (Cited on page 70).

Perry R. Cook. Real-Time Performance Controllers for Synthesized Singing. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 236–237, Vancouver, BC, Canada, 2005. DOI 10.5281/zenodo.1176846. (Cited on page 22).

Enrico Costanza, Samuel A. Inverso, and Rebecca Allen. Toward Subtle Intimate Interfaces for Mobile Devices Using an EMG Controller. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '05, page 481–489, New York, NY, USA, 2005. Association for Computing Machinery. ISBN 1581139985. DOI 10.1145/1054972.1055039. (Cited on pages 56, 119, and 140).

Kelsey Cotton, Ozgun Kilic Afsar, Yoav Luft, Priyanka Syal, and Fehmi Ben Abdesslem. SymbioSinging: Robotically Transposing Singing Experience across Singing and Non-Singing Bodies. In *Creativity and Cognition*, C&C '21, New York, NY, USA, 2021a. Association for Computing Machinery. ISBN 9781450383769. DOI 10.1145/3450741.3466718. (Cited on pages 22, 54, 136, and 184).

Kelsey Cotton, Pedro Sanches, Vasiliki Tsaknaki, and Pavel Karpashevich. The Body Electric: A NIME designed through and with the somatic experience of singing. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Shanghai, China, 2021b. DOI 10.21428/92fbeb44.ec9f8fdd. (Cited on pages 2, 21, 22, 49, 54, and 136).

Roddy Cowie, Gary Mckeown, and Ellen Douglas-Cowie. Tracing Emotion: An Overview. *International Journal of Synthetic Emotions*, 3:1–17, 01 2012. DOI 10.4018/jse.2012010101. (Cited on page 48).

Jennifer Cumming and Sarah E. Williams. *The Role of Imagery in Performance*. Oxford University Press, Oxford, 2012. DOI 10.1093/oxfordhb/9780199731763.013.0011. (Cited on pages xv, 25, 26, 27, 29, 30, 40, and 60).

Luke Dahl. Triggering Sounds from Discrete Air Gestures: What Movement Feature Has the Best Timing? In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 201–206, London, United Kingdom, 2014. Goldsmiths, University of London. DOI 10.5281/zenodo.1178738. (Cited on page 54).

Jiajie Dai, Matthias Mauch, and Simon Dixon. Analysis of Intonation Trajectories in Solo Singing. In *16th International Society for Music Information Retrieval Conference (ISMIR), Malaga, Spain, Oct 26-30*, pages 420–426, 2015. (Cited on pages 68, 69, and 86).

Christophe d'Alessandro, Lionel Feugère, Sylvain Le Beux, Olivier Perrotin, and Albert Rilliard. Drawing melodies: Evaluation of chironomic singing synthesis. *The Journal of the Acoustical Society of America*, 135(6):3601–3612, 2014. DOI 10.1121/1.4875718. (Cited on page 22).

Nicolas d'Alessandro, Christophe d'Alessandro, Sylvain Le Beux, and Boris Doval. Real-time CALM Synthesizer: New Approaches in Hands-Controlled Voice Synthesis. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 266–271, Paris, France, 2006. DOI 10.5281/zenodo.1176863. (Cited on page 22).

Ayoub Daliri, Sara-Ching Chao, and Lacee C. Fitzgerald. Compensatory responses to formant perturbations proportionally decrease as perturbations increase. *Journal of Speech, Language, and Hearing Research*, 63(10):3392–3407, 2020. (Cited on page 79).

S. Dalla Bella and M. Berkowska. Singing proficiency in the majority. *Annals of the New York Academy of Sciences*, 1169:99–10, 2009. (Cited on page 86).

J. L. Daniels. Pedagogical Opinion: Finding square one: A message to teachers. *The National Association of Teachers of Singing Bulletin*, 40(2):36–37, 1983. (Cited on page 41).

Claudia Daudén Roquet and Corina Sas. Body Matters: Exploration of the Human Body as a Resource for the Design of Technologies for Meditation. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference*, DIS '20, page 533–546, New York, NY, USA, 2020. Association for Computing Machinery. ISBN 9781450369749. DOI 10.1145/3357236.3395499. (Cited on pages 37 and 49).

Claudia Daudén Roquet and Corina Sas. Interoceptive Interaction: An Embodied Metaphor Inspired Approach to Designing for Meditation. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450380966. DOI 10.1145/3411764.3445137. (Cited on pages 35, 37, 42, and 101).

Jane W. Davidson. Bodily movement and facial actions in expressive musical performance by solo and duo instrumentalists: Two distinctive case studies. *Psychology of Music*, 40(5):595–633, 2012. DOI 10.1177/0305735612449896. (Cited on pages 11, 12, and 28).

Carlo J. De Luca. *Surface Electromyography: Detection and Recording.* DelSys Inc., 2002. (Cited on page 55).

Gamhewage C. de Silva, Tamara Smyth, and Michael J. Lyons. A Novel Face-tracking Mouth Controller and its Application to Interacting with Bioacoustic Models. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 169–172, Hamamatsu, Japan, 2004. DOI 10.5281/zenodo.1176667. (Cited on page 21).

Franfois Delalande. Human movement and the interpretation of music. In *Second International Colloquium on the Psychology of Music, Ravello, Italy*, 1990. (Cited on page 11).

Samuel Delalez and Christophe d'Alessandro. Vokinesis: Syllabic Control Points for Performative Singing Synthesis. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 198–203, Copenhagen, Denmark, 2017. Aalborg University Copenhagen. DOI 10.5281/zenodo.1176220. (Cited on page 22).

Natalie Depraz, Francisco J. Varela, and Pierre Vermersch. *On becoming aware: A pragmatics of experiencing.* John Benjamins Publishing, 2003. (Cited on pages 34, 36, 38, 50, and 101).

Laura Devendorf and Daniela K. Rosner. Beyond Hybrids: Metaphors and Margins in Design. In *Proceedings of the 2017 Conference on Designing Interactive Systems*, DIS '17, page 995–1000, New York, NY, USA, 2017. Association for Computing Machinery. ISBN 9781450349222. DOI 10.1145/3064663.3064705. (Cited on page 32).

Laura Devendorf and Kimiko Ryokai. Being the Machine: Reconfiguring Agency and Control in Hybrid Fabrication. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, page 2477–2486, New York, NY, USA, 2015. Association for Computing Machinery. ISBN 9781450331456. DOI 10.1145/2702123.2702547. (Cited on page 33).

Kristin N. Dew and Daniela K. Rosner. Lessons from the Woodshop: Cultivating Design with Living Materials. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, page 1–12, New York, NY, USA, 2018. Association for Computing Machinery. ISBN 9781450356206. DOI 10.1145/3173574.3174159. (Cited on page 33).

Robert B. Dewell. Over again: Image-schema transformations in semantic analysis. *Cognitive Linguistics*, 5(4):351–380, 1994. DOI 10.1515/cogl.1994.5.4.351. (Cited on page 35).

Balandino Di Donato, Atau Tanaka, Michael Zbyszynski, and Martin Klang. EAVI EMG board. In Marcelo Queiroz and Anna Xambó Sedó, editors, *Proceedings of the International Conference on New Interfaces for Musical Expression*, Porto Alegre, Brazil, 2019. UFRGS. (Cited on page 56).

Carl DiSalvo and Jonathan Lukens. Nonanthropocentrism and the Nonhuman in Design: Possibilities for Designing New Forms of Engagement with and Through Technology. *From Social Butterfy to Engaged Citizen: Urban Informatics, Social Media, Ubiquitous Computing, and Mobile Technology to Support Citizen Engagemen*, 421:421–436, 2011. DOI https://doi.org/10.7551/mitpress/8744.003.0034. (Cited on page 32).

Marco Donnarumma, Baptiste Caramiaux, and Atau Tanaka. Muscular Interactions. Combining EMG and MMG sensing for musical practice. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 128–131, Daejeon, Republic of Korea, 2013. Graduate School of Culture Technology, KAIST. DOI 10.5281/zenodo.1178504. (Cited on page 56).

Maurin Donneaud, Cedric Honnet, and Paul Strohmeier. Designing a Multi-Touch eTextile for Music Performances. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 7–12, Copenhagen, Denmark, 2017. Aalborg University Copenhagen. DOI 10.5281/zenodo.1176151. (Cited on page 121).

Mine Doğantan-Dack. In the Beginning Was Gesture. In A. Gritten and E. King, editors, *New Perspectives on Music and Gesture*. Ashgate, Surrey, England, 2011. ISBN 9780754664628. (Cited on pages 12 and 19).

Paul Dourish. *Where the action is: the foundations of embodied interaction*. MIT Press, Cambridge, 1980. (Cited on pages 100 and 103).

Graham Dove and Sara Jones. Narrative Visualization: Sharing Insights into Complex Data. In *Interfaces and Human Computer Interaction Conference (IHCI 2012), 21 - 23 July 2012, Lisbon, Portugal*, 07 2012. (Cited on page 107).

Marco Dozza, Fay B. Horak, and Lorenzo Chiari. Auditory biofeedback substitutes for loss of sensory information in maintaining stance. *Experimental Brain Research*, 178(1):37–48, 2007. DOI 10.1007/s00221-006-0709-y. (Cited on page 144).

Tim Duente, Max Pfeiffer, and Michael Rohs. Zap++ a 20-channel electrical muscle stimulation system for fine-grained wearable force feedback. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services*, pages 1–13, 2017. (Cited on pages 119 and 120).

Susan T. Dumais and William P. Jones. A Comparison of Symbolic and Spatial Filing. *SIGCHI Bull.*, 16(4):127–130, Apr 1985. ISSN 0736-6906. DOI 10.1145/1165385.317479. (Cited on page 36).

Roslyn Dunbar-Wells. *The relevance of metaphor in voice teaching: A comparative study of Sinus Tone Production and Vocal Cord Theories*. Doctoral dissertation, University of Reading, Reading, UK, 1997. (Cited on pages xv, 40, 41, 108, and 182).

Roslyn Dunbar-Wells. The Relevance of Metaphor to Effective Voice Teaching Strategies. *Australian Voice*, 5:50–59, 1999. (Cited on pages 39 and 40).

Roslyn Dunbar-Wells. Using appropriate language modes and explicit teaching aids. *Australian Voice*, 9:63–68, 2003. (Cited on page 40).

Brian D. Ebie. The effects of verbal, vocally modeled, kinesthetic, and audio-visual treatment conditions on male and female middle-school vocal music students' abilities to expressively sing melodies. *Psychology of Music*, 32(4):405–417, 2004. DOI 10.1177/0305735604046098. (Cited on pages 31 and 39).

Matthias Echternach, Michael Markl, and Bernhard Richter. Dynamic real-time magnetic resonance imaging for the analysis of voice physiology. *Current Opinion in Otolaryngology & Head and Neck Surgery*, 20(6):450–457, 2012. DOI 10.1097/MOO.0b013e3283585f87. (Cited on page 54).

Frank A. Edmonson. Effect of Interval Direction on Pitch Acuity in Solo Vocal Performance. *Journal of Research in Music Education*, 20(2):246–254, 1972. DOI 10.2307/3344090. (Cited on pages 28 and 83).

Tuomas Eerola and Adrian C. North. Expectancy-Based Model of Melodic Complexity. In C. Woods, G.B. Luck, R. Brochard, S. A. O'Neill, and J. A. Sloboda, editors, *Proceedings of the Sixth International Conference on Music Perception and Cognition, Keele, Staffordshire, UK: Department of Psychology. CD-ROM*, 2000. (Cited on page 72).

Tuomas Eerola and Petri Toiviainen. MIDI toolbox: MATLAB tools for music research. 2004. (Cited on page 72).

Tuomas Eerola, Kelly Jakubowski, Nikki Moran, Peter E. Keller, and Martin Clayton. Shared periodic performer movements coordinate interactions in duo improvisations. *Royal Society Open Science*, 5(2):1–24, 2018. DOI 10.1098/rsos.171520. (Cited on pages 12 and 19).

Alfred O. Effenberg. Using Sonification to Enhance Perception and Reproduction Accuracy of Human Movement Patterns. In *Proc. International Workshop on Interactive Sonification, Bielefeld, Germany*, pages 1–5, 2004. DOI 10.1109/mmul.2005.31. (Cited on page 58).

Nina Sun Eidsheim. Sensing Voice: Materiality and the Lived Body in Singing and Listening. *The Senses and Society*, 6(2):133–155, July 2011. DOI 10.2752/174589311x12961584845729. (Cited on pages 33 and 34).

Nina Sun Eidsheim. *Sensing Sound: Singing and Listening as Vibrational Practice*. Duke University Press, 2015. (Cited on pages 33, 34, 42, 111, and 184).

Nina Sun Eidsheim. Maria Callas's Waistline and the Organology of Voice. *The Opera Quarterly*, 33(3-4):249–268, 2017. DOI 10.1093/oq/kbx008. (Cited on pages 33, 34, and 111).

Inger Ekman and Michal Rinott. Using Vocal Sketching for Designing Sonic Interactions. In *Proceedings of the 8th ACM Conference on Designing Interactive Systems*, DIS '10, page 123–131, New York, NY, USA, 2010. Association for Computing Machinery. ISBN 9781450301039. DOI 10.1145/1858171.1858195. (Cited on page 172).

Shirlee Emmons and Alma Thomas. *Power performance for singers: Transcending the barriers*. Oxford University Press, Oxford, 1998. (Cited on page 40).

Cagri Erdem and Alexander Refsum Jensenius. RAW: Exploring Control Structures for Muscle-based Interaction in Collective Improvisation. In *Proceedings of the International Conference on New Interfaces for Musical Expression, Birmingham City University, Birmingham, UK*, pages 477–482, 2020. DOI 10.5281/zenodo.4813485. (Cited on pages 33, 56, 58, and 184).

Donna M. Erickson, Thomas Baer, and Katherine S. Harris. The role of strap muscles in pitch lowering. In D. M. Bless, editor, *Vocal fold physiology: Contemporary research and clinical issues*, pages 279–285. College-Hill Press, San Diego, CA, 1983. (Cited on page 15).

Thomas Erickson. The Design and Long-Term Use of a Personal Electronic Notebook: A Reflective Analysis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '96, page 11–18, New York, NY, USA, 1996. Association for Computing Machinery. ISBN 0897917774. DOI 10.1145/238386.238392. (Cited on page 53).

Georg Essl and Sile O'Modhrain. An enactive approach to the design of new tangible musical instruments; The design of new tangible musical instruments. *Organised Sound*, 11(3):285–296, 2006. DOI 10.1017/s135577180600152x. (Cited on pages 32 and 141).

Georg Essl, Michael Rohs, and Sven Kratz. Use the Force (or something) — Pressure and Pressure — Like Input for Mobile Music Performance. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 182–185, Sydney, Australia, 2010. DOI 10.5281/zenodo.1177761. (Cited on page 54).

Evan Balster and Richard Hogg. imitone, 2022. URL http://imitone.com. (Cited on page 21).

Grant Fairbanks. Selective vocal effects of delayed auditory feedback. *Journal of speech and hearing disorders*, 20(4):333–346, 1955. (Cited on pages 66 and 83).

Morwaread M. Farbood. A Parametric, Temporal Model of Musical Tension. *Music Perception: An Interdisciplinary Journal*, 29(4):387–428, 2012. DOI 10.1525/mp.2012.29.4.387. (Cited on page 31).

Morwaread M. Farbood. A Global Model of Musical Tension. In *Music Perception and Cognition – 10th International Conference (ICMPC 10), Sapporo, Japan*, Aug 25-29, 2008. (Cited on page 31).

Morwaread M. Farbood and Khen C. Price. Timbral Features Contributing to Perceived Auditory and Musical Tension. In *Music Perception and Cognition – 13th International Conference, (ICMPC 13), Aug 4-8, Seoul, South Korea*, 2014. (Cited on page 31).

Morwaread M. Farbood and Khen C. Price. The contribution of timbre attributes to musical tension. *Journal of the Acoustical Society of America*, 141(1):419–427, 2017. DOI 10.1121/1.4973568. (Cited on page 31).

Morwaread M. Farbood and Finn Upham. Interpreting expressive performance through listener judgments of musical tension. *Frontiers in Psychology*, 4(998):1–15, 2013. DOI 10.3389/fpsyg.2013.00998. (Cited on page 31).

Megan Farokhmanesh. *Why is this floppy disk joke still haunting the internet?* 2017. URL https://web.archive.org/web/20220424120448/https://www.theverge.com/2017/10/24/16505912/floppy-disk-3d-print-save-joke-meme. (Cited on page 103).

Emanuel Felipe Duarte, Luiz Ernesto Merkle, and M Cecília C. Baranauskas. The interface between Interactive Art and human-Computer Interaction: Exploring dialogue genres and evaluative practices. *Journal of Interactive Systems*, 10:20, 2019. (Cited on page 37).

Tara J. Fenwick. Expanding Conceptions of Experiential Learning: A Review of the Five Contemporary Perspectives on Cognition. *Adult Education Quarterly*, 50(4):243–272, 2000. DOI 10.1177/07417130022087035. (Cited on page 32).

José Miguel Fernandez, Thomas Köppel, Nina Verstraete, Grégoire Lorieux, Alexander Vert, and Philippe Spiesser. GeKiPe, a gesture-based interface for audiovisual performance. In *Proceedings* *of the International Conference on New Interfaces for Musical Expression*, pages 450–455, Copenhagen, Denmark, 2017. Aalborg University Copenhagen. DOI 10.5281/zenodo.1176312. (Cited on page 54).

Lionel Feugère and Christophe d'Alessandro. Digitartic: bi-manual gestural control of articulation in performative singing synthesis. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 331–336, Daejeon, Republic of Korea, 2013. Graduate School of Culture Technology, KAIST. DOI 10.5281/zenodo.1178520. (Cited on page 22).

Lionel Feugère, Christophe d'Allesandro, and Adrien Bartoli. Contrôle gestuel de la synthèse vocale. Les instruments Cantor Digitalis et Digitartic. *Traitement du signal*, 32(4):417–442, 2015. DOI 10.3166/ts.32.417-442. (Cited on page 22).

Steven A. Finney and Caroline Palmer. Auditory feedback and memory for music performance: Sound evidence for an encoding effect. *Memory & Cognition*, 31(1):51–64, 2003. DOI 10.3758/bf03196082. (Cited on pages 28 and 83).

Steven A. Finney and William H. Warren. Delayed auditory feedback and rhythmic tapping: Evidence for a critical interval shift. *Perception & Psychophysics*, 64(6):896–908, August 2002. DOI 10.3758/bf03196794. (Cited on pages 67 and 83).

Georgia A. Floridou, Victoria J. Williamson, Lauren Stewart, and Daniel Müllensiefen. The Involuntary Musical Imagery Scale (IMIS). *Psychomusicology: Music, Mind, and Brain*, 25(1):28–36, 2014. DOI 10.1037/pmu0000067. (Cited on page 60).

BJ Fogg. Persuasive Computers: Perspectives and Research Directions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, page 225–232, USA, 1998. ACM Press/Addison-Wesley Publishing Co. ISBN 0201309874. DOI 10.1145/274644.274677. (Cited on page 176).

Federico Fontana, Hanna Järveläinen, Stefano Papetti, Federico Avanzini, Giorgio Klauer, and Lorenzo Malavolta. Rendering and Subjective Evaluation of Real Vs. Synthetic Vibrotactile Cues on a Digital Piano Keyboard. In *Proceedings of the 12th International Conference on Sound and Music Computing (SMC-15), July 30-31 & Aug 1, Maynooth, Ireland*, pages 161–167, 2015. (Cited on page 27).

Laura Forlano. Decentering the Human in the Design of Collaborative Cities. *Design Issues*, 32(3): 42–54, July 2016. DOI 10.1162/desi_a_00398. (Cited on page 32).

Nicholas E. V. Foster, Andrea R. Halpern, and Robert J. Zatorre. Common parietal activation in musical mental transformations across pitch and time. *NeuroImage*, 75:27–35, 2013. DOI 10.1016/j.neuroimage.2013.02.044. (Cited on pages 60 and 64).

Matthias K. Franken, Daniel J. Acheson, James M. McQueen, Peter Hagoort, and Frank Eisner. Opposing and following responses in sensorimotor speech control: Why responses go both ways. *Psychonomic Bulletin & Review*, 25(4):1458–1467, June 2018. DOI 10.3758/s13423-018-1494-x. (Cited on page 80).

Matthias K. Franken, Robert J. Hartsuiker, Petter Johansson, Lars Hall, and Andreas Lind. Don't blame yourself: Conscious source monitoring modulates feedback control during speech production. *Quarterly Journal of Experimental Psychology*, pages 1–13, February 2022. DOI 10.1177/17470218221075632. (Cited on page 80).

Christopher Frauenberger. Entanglement HCI The Next Wave? *ACM Transactions in Computer-Human Interaction*, 27(1), 2019. ISSN 1073-0516. DOI 10.1145/3364998. (Cited on pages 32, 38, 149, and 185).

Christopher Frauenberger, Judith Good, and Wendy Keay-Bright. Phenomenology, a Framework for Participatory Design. In *Proceedings of the 11th Biennial Participatory Design Conference*, PDC '10, page 187–190, New York, NY, USA, 2010. Association for Computing Machinery. ISBN 9781450301312. DOI 10.1145/1900441.1900474. (Cited on page 48).

Adrian Freed. Application of new Fiber and Malleable Materials for Agile Development of Augmented Instruments and Controllers. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 107–112, Genoa, Italy, 2008. DOI 10.5281/zenodo.1179539. (Cited on page 121).

Limor Fried. Lady Ada, 2012. URL https://www.ladyada.net/. (Cited on page 120).

Mikhaila Friske, Jordan Wirfs-Brock, and Laura Devendorf. Entangling the Roles of Maker and Interpreter in Interpersonal Data Narratives: Explorations in Yarn and Sound. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference*, DIS '20, page 297–310, New York, NY, USA, 2020. Association for Computing Machinery. ISBN 9781450369749. DOI 10.1145/3357236.3395442. (Cited on pages 107 and 188).

Alf Gabrielsson and Patrik N. Juslin. Emotional Expression in Music Performance: Between the Performer's Intention and the Listener's Experience. *Psychology of Music*, 24(1):68–91, 1996. DOI 10.1177/0305735696241007. (Cited on pages 31 and 39).

Shaun Gallagher. *How the Body Shapes the Mind*. Oxford University Press, Oxford, 2005. ISBN 978-0-19-920416-8. (Cited on page 38).

Shaun Gallagher and Dan Zahavi. *The Phenomenological Mind: An Introduction to Philosophy of Mind and Cognitive Science, 2nd edition*. Routledge, 2012. (Cited on page 38).

Isabel García-López and Javier Gavilán Bouzas. The singing voice. *Acta Otorrinolaringologica (English Edition)*, 21(2):483–496, 2010. DOI 10.1016/s2173-5735(10)70082-x. (Cited on pages 13, 15, 16, 17, and 31).

Linda I. Garrity. Electromyography: A review of the current status of subvocal speech research. *Memory & Cognition*, 5(6):615–622, 1977. DOI 10.3758/bf03197407. (Cited on page 57).

Denise Gastaldo, Natalia Rivas-Quarneti, and Lilian Magalhaes. Body-Map Storytelling as a Health Research Methodology: Blurred Lines Creating Clear Pictures. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, Vol 19:No 2 (2018), 2018. DOI 10.17169/fqs-19.2.2858. (Cited on pages 37 and 49).

William W. Gaver. The video window: my life with a ludic system. *Personal and Ubiquitous Computing*, 10(2-3):60–65, 2006. DOI 10.1007/s00779-005-0002-2. (Cited on page 53).

William W. Gaver, Jacob Beaver, and Steve Benford. Ambiguity as a Resource for Design. CHI '03, page 233–240. Association for Computing Machinery, New York, NY, USA, 2003. ISBN 1581136307. DOI 10.1145/642611.642653. (Cited on pages 58, 103, and 104).

Rebecca W. Gelding, William Forde Thompson, and Blake W. Johnson. The pitch imagery arrow task: Effects of musical training, vividness, and mental control. *PloS One*, 10:e0121809, 2015. DOI 10.1371/journal.pone.0121809. (Cited on page 60).

Marylou Pausewang Gelfer and Diane K. Bultemeyer. Evaluation of vocal fold vibratory patterns in normal voices. *Journal of Voice*, 4(4):335–345, 1990. DOI 10.1016/s0892-1997(05)80051-7. (Cited on page 13).

Raymond W. Gibbs, Paula Lenz Costa Lima, and Edson Francozo. Metaphor is grounded in embodied experience. *Journal of Pragmatics*, 36(7):1189–1210, 2004. ISSN 0378-2166. DOI 10.1016/j.pragma.2003.10.009. (Cited on pages 35, 37, and 102).

Rolf Inge Godøy and Harald Jørgensen. *Musical Imagery*. Number 5 in Studies on New Music Research. Swets & Zeitlinger, Lisse, Netherlands, 2001. ISBN 978-90-265-1831-7. (Cited on pages 11, 26, 29, 34, 36, and 40).

Rolf Inge Godøy and Marc Leman. *Musical Gestures: Sound, Movement, and Meaning*. Routledge, New York, 2010. ISBN 978-0-415-99886-4. (Cited on pages 11, 19, and 38).

Werner Goebl and Caroline Palmer. Tactile feedback and timing accuracy in piano performance. *Experimental Brain Research*, 186(3):471–479, 2008. DOI 10.1007/s00221-007-1252-1. (Cited on pages 12, 28, 58, and 83).

Patrick Gomez and Brigitta Danuser. Relationships Between Musical Structure and Psychological Measures of Emotion. *Emotion*, 7(2):377–387, 2007. DOI 10.1037/1528-3542.7.2.377. (Cited on page 31).

Elizabeth Goodman, Erik Stolterman, and Ron Wakkary. Understanding Interaction Design Practices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, page 1061–1070, New York, NY, USA, 2011. Association for Computing FgaMachinery. ISBN 9781450302289. DOI 10.1145/1978942.1979100. (Cited on page 53).

Chelsea L. Gordon, Patrice R. Cobb, and Ramesh Balasubramaniam. Recruitment of the motor system during music listening: an ALE meta-analysis of fMRI data. *PLos ONE*, 13:e0207213, 2018. DOI 10.1371/journal.pone.0207213. (Cited on page 81).

Edwin E. Gordon. All about audiation and music aptitudes: Edwin E. Gordon discusses using audiation and music aptitudes as teaching tools to allow students to reach their full music potential. *Music Educators Journal*, 86(2):41–44, 1999. (Cited on page 26).

Edwin E. Gordon. Roots of music learning theory and audiation. 2011. (Cited on page 27).

Esther Grabe and Ee Ling Low. Durational variability in speech and the rhythm class hypothesis. In C. Gussen-hoven and N. Warner, editors, *Laboratory Phonology*, volume 7, pages 515–546. Berlin: Mouton de Gruyter, 1995. (Cited on page 72).

Niccolò Granieri and James Dooley. Reach: a keyboard-based gesture recognition system for live piano sound modulation. In Marcelo Queiroz and Anna Xambó Sedó, editors, *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 375–376, Porto Alegre, Brazil, 2019. Ufrgs. DOI 10.5281/zenodo.3673000. (Cited on page 54).

Irene Gratton, Maria A. Brandimonte, and Nicola Bruno. Absolute Memory for Tempo in Musicians and Non-Musicians. *PLoS ONE*, 11(10):e0163558, 2016. DOI 10.1371/journal.pone.0163558. (Cited on page 81).

Barry Green and W. Timothy Gallwey. *The inner game of music*. Anchor Press/Doubleday, New York, 1986. (Cited on page 40).

Emma B. Greenspon, Peter Q. Pfordresher, and Andrea R. Halpern. Pitch Imitation Ability in Mental Transformations of Melodies. *Music Perception: An Interdisciplinary Journal*, 34(5):585–604, 2018. DOI 10.1525/mp.2017.34.5.585. (Cited on pages 60, 64, and 79).

Berit Greinke, Giorgia Petri, Pauline Vierne, Paul Biessmann, Alexandra Börner, Kaspar Schleiser, Emmanuel Baccelli, Claas Krause, Christopher Verworner, and Felix Biessmann. An Interactive Garment for Orchestra Conducting: IoT-Enabled Textile & Machine Learning to Direct Musical Performance. In *Proceedings of the Fifteenth International Conference on Tangible, Embedded, and Embodied Interaction*, TEI '21, New York, NY, USA, 2021a. Association for Computing Machinery. ISBN 9781450382137. DOI 10.1145/3430524.3442451. (Cited on page 121).

Berit Greinke, Emma Wood, Sophie Skach, Arantza Vilas, and Pauline Vierne. Folded Electronic Textiles: Weaving, Knitting, Pleating and Coating Three-dimensional Sensor Structures. *Leonardo*, pages 1–9, 2021b. DOI 10.1162/leon_a_02183. (Cited on page 120).

Barbara Griffin, Peak Woo, Raymond Colton, Janina Casper, and David Brewer. Physiological characteristics of the supported singing voice: A preliminary study. *Journal of Voice*, 9(1):45–56, 1995. DOI 10.1016/s0892-1997(05)80222-x. (Cited on pages 16, 17, 54, and 55).

George Grouios. Mental practice: A review. *Journal of Sport Behavior*, 15(1):42–59, 1992. (Cited on pages 26 and 101).

Horst Günter. Mental concepts in singing: A psychological approach, Part 1. *The National Association of Teachers of Singing Journal*, 48(5):46, 1992a. (Cited on page 40).

Horst Günter. Mental concepts in singing: A psychological approach, Part 2. *The National Association of Teachers of Singing Journal*, 49(1):4–6, 1992b. (Cited on page 40).

Bruce D. Hale. The effects of internal and external imagery on muscular and ocular concomitants. *Journal of Sport Psychology*, 4:379–387, 1982. (Cited on page 101).

Bruce D. Hale. Imagery Perspectives and Learning in Sports Performance. In A. A. Sheikh and E. R. Korn, editors, *Imagery in Sports and Physical Performance*, pages 75–96. Baywood, New York, 1994. (Cited on pages 26 and 101).

Craig R. Hall and Kathleen A. Martin. Measuring movement imagery abilities: a revision of the Movement Imagery Questionnaire. *Journal of Mental Imagery*, 1997. (Cited on page 60).

Craig R. Hall and John Pongrac. *Movement imagery questionnaire*. University of Western Ontario, London, Ontario, 1983. (Cited on page 60).

Andrea R. Halpern. Perceived and Imagined Tempos of Familiar Songs. *Music Perception: An Interdisciplinary Journal*, 6(2):193–202, 1988. DOI 10.2307/40285425. (Cited on page 81).

Andrea R. Halpern. Differences in auditory imagery self-report predict neural and behavioral outcomes. *Psychomusicology: Music, Mind, and Brain*, 25(1):37–47, 2015. DOI 10.1037/pmu0000081. (Cited on pages 59, 60, 64, 71, 91, and 203).

Andrea R. Halpern and Katie Overy. Voluntary auditory imagery and music pedagogy. *The Oxford handbook of sound and imagination*, 2:391–407, 2019. (Cited on page 27).

Andrea R. Halpern and Robert J. Zatorre. When that tune runs through your head: A PET investigation of auditory imagery for familiar melodies. *Cerebral Cortex*, 9:697–704, 1999. DOI 10.1093/cercor/9.7.697. (Cited on pages 25, 26, 60, and 117).

Andrea R. Halpern, Robert J. Zatorre, Marc Bouffard, and Jennifer A Johnson. Behavioral and neural correlates of perceived and imagined musical timbre. *Neuropsychologia*, 42(9):1281–1292, 2004. DOI 10.1016/j.neuropsychologia.2003.12.017. (Cited on page 25).

Sandra L. Hamlet and John M. Reid. Transmission of Ultrasound Through the Larynx as a Means of Determining Vocal-Fold Activity. *IEEE Transactions on Biomedical Engineering*, Bme-19(1): 34–37, 1972. DOI 10.1109/tbme.1972.324156. (Cited on page 54).

Beate Hampe and Joseph E. Grady. *From Perception to Meaning: Image Schemas in Cognitive Linguistics*. De Gruyter, Inc., Berlin, 2005. ISBN 978-3-11-019753-2. (Cited on page 35).

Jihyun Han and Nicolas Gold. Lessons Learned in Exploring the Leap Motion(TM) Sensor for Gesture-based Instrument Design. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 371–374, London, United Kingdom, 2014. Goldsmiths, University of London. DOI 10.5281/zenodo.1178784. (Cited on page 54).

Thomas Hanna. *What is somatics?* North Atlantic Books, Berkeley, CA, 1995. (Cited on page 49).

William J. Hardcastle. *Physiology of Speech Production: An Introduction for Speech Scientists*. Academic Press Inc., London, 1976. ISBN 0-12-324950-3. (Cited on pages xix, 13, 57, 117, and 123).

Sarah-Indriyati Hardjowirogo. Instrumentality. On the Construction of Instrumental Identity. In *Musical Instruments in the 21st Century*, pages 9–24. Springer Singapore, 2016. DOI 10.1007/978-981-10-2951-6_2. (Cited on page 33).

Jacob Harrison, Robert H Jack, Fabio Morreale, and Andrew P. McPherson. When is a Guitar not a Guitar? Cultural Form, Input Modality and Expertise. In Thomas Martin Luke Dahl, Douglas Bowman, editor, *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 299–304, Blacksburg, Virginia, USA, 2018. Virginia Tech. ISBN 978-1-949373-99-8. DOI 10.5281/zenodo.1302589. (Cited on page 19).

Steve Harrison, Deborah Tatar, and Phoebe Sengers. The Three Paradigms of HCI. page 24, 2007. (Cited on page 37).

Kate Hartman, Jackson McConnell, Boris Kourtoukov, Hillary Predko, and Izzie Colpitts-Campbell. Monarch: Self-Expression Through Wearable Kinetic Textiles. In *Proceedings of the Ninth International Conference on Tangible, Embedded, and Embodied Interaction*, TEI '15, page 413–414, New York, NY, USA, 2015. Association for Computing Machinery. ISBN 9781450333054. DOI 10.1145/2677199.2690875. (Cited on page 120).

Kate Hartman, Boris Kourtoukov, and Erin Lewis. Kinetic Body Extensions for Social Interactions. In *Proceedings of the Twelfth International Conference on Tangible, Embedded, and Embodied Interaction*, TEI '18, page 736–739, New York, NY, USA, 2018. Association for Computing Machinery. ISBN 9781450355681. DOI 10.1145/3173225.3173333. (Cited on pages 56, 120, and 140).

Mariam Hassib, Max Pfeiffer, Stefan Schneegass, Michael Rohs, and Florian Alt. Emotion Actuator: Embodied Emotional Feedback through Electroencephalography and Electrical Muscle Stimulation. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, page 6133–6146, New York, NY, USA, 2017. Association for Computing Machinery. ISBN 9781450346559. DOI 10.1145/3025453.3025953. (Cited on pages 119 and 120).

Thomas Hemsley. *Singing and imagination: A human approach to a great musical tradition.* Oxford University Press, Oxford/New York, 1998. (Cited on pages 30 and 40).

Christian T. Herbst, Markus Hess, Frank Müller, Jan G. Švec, and Johan Sundberg. Glottal Adduction and Subglottal Pressure in Singing. *Journal of Voice*, 29(4):391–402, 2015. DOI 10.1016/j.jvoice.2014.08.009. (Cited on page 54).

Sibylle C. Herholz, Andrea R. Halpern, and Robert J. Zatorre. Neuronal correlates of perception, imagery, and memory for familiar tunes. *Journal of Cognitive Neuroscience*, 24:1382–1397, 2012. DOI 10.1162/jocn_a_00216. (Cited on pages 60 and 64).

Dorien Herremans and Elaine Chew. Tension ribbons: Quantifying and visualising tonal tension. In *Proceedings of the 2nd International Conference on Technologies for Music Notation and Representation, Cambridge, UK*, pages 27–29, 2016. (Cited on page 31).

Stellan Hertegård. What have we learned about laryngeal physiology from high-speed digital videoendoscopy? *Current Opinion in Otolaryngology & Head and Neck Surgery*, 13(3):152–156, 2005. DOI 10.1097/01.moo.0000163451.98079.ba. (Cited on page 54).

Kate Hevner. Experimental Studies of the Elements of Expression in Music. *The American Journal of Psychology*, 48(2):246–268, 1936. DOI 10.2307/1415746. (Cited on page 31).

Zebulon Highben and Caroline Palmer. Effects of Auditory and Motor Mental Practice in Memorized Piano Performance. *Bulletin of the Council for Research in Music Education*, pages 1–8, 2004. (Cited on pages 28 and 83).

Jerome Hines. *Great singers on great singing.* Victor Gollancz, London, 1983. (Cited on pages 1, 30, 40, and 42).

Atsushi Hiyama, Yusuke Doyama, Kento Kakurai, Hidetoshi Namiki, Masaaki Miyasako, and Michitaka Hirose. Archiving and transferring of traditional artisanship focused on interaction between artisan and tools. In *Proc. 16th International Conference on Virtual Systems and Multimedia, Seoul, South Korea*, pages 171–176, 2010. DOI 10.1109/vsmm.2010.5665987. (Cited on page 144).

Megan Hofmann, Lea Albaugh, Ticha Sethapakadi, Jessica Hodgins, Scott E. Hudson, James McCann, and Jennifer Mankoff. KnitPicking Textures: Programming and Modifying Complex Knitted Textures for Machine and Hand Knitting. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*, UIST '19, page 5–16, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 9781450368162. DOI 10.1145/3332165.3347886. (Cited on page 120).

Megan Hofmann, Jennifer Mankoff, and Scott E. Hudson. KnitGIST: A Programming Synthesis Toolkit for Generating Functional Machine-Knitting Textures. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*, UIST '20, page 1234–1247, New York, NY, USA, 2020. Association for Computing Machinery. ISBN 9781450375146. DOI 10.1145/3379337.3415590. (Cited on page 121).

Charles Holbrow, Elena Jessop, and Rebecca Kleinberger. Vocal Vibrations: A Multisensory Experience of the Voice. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 431–434, London, United Kingdom, 2014. Goldsmiths, University of London. DOI 10.5281/zenodo.1178800. (Cited on pages 21 and 22).

Patricia Holmes. Imagination in practice: a study of the integrated roles of interpretation, imagery and technique in the learning and memorisation processes of two experienced solo performers. *British Journal of Music Education*, 32(3):217–235, 2005. DOI 10.1017/s0265051705006613. (Cited on pages 27 and 29).

Sarah Homewood, Amanda Karlsson, and Anna Vallgårda. Removal as a Method: A Fourth Wave HCI Approach to Understanding the Experience of Self-Tracking. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference*, DIS '20, page 1779–1791, New York, NY, USA, 2020. Association for Computing Machinery. ISBN 9781450369749. DOI 10.1145/3357236.3395425. (Cited on page 49).

Sarah Homewood, Marika Hedemyr, Maja Fagerberg Ranten, and Susan Kozel. Tracing Conceptions of the Body in HCI: From User to More-Than-Human. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450380966. DOI 10.1145/3411764.3445656. (Cited on pages 32, 48, 149, 174, 175, and 176).

Cedric Honnet, Hannah Perner-Wilson, Marc Teyssier, Bruno Fruchard, Jürgen Steimle, Ana C. Baptista, and Paul Strohmeier. PolySense: Augmenting Textiles with Electrical Functionality Using In-Situ Polymerization. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, page 1–13, New York, NY, USA, 2020. Association for Computing Machinery. ISBN 9781450367080. DOI 10.1145/3313831.3376841. (Cited on page 120).

Kristina Höök. Transferring Qualities from Horseback Riding to Design. In *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries*, NordiCHI '10, page 226–235, New York, NY, USA, 2010. Association for Computing Machinery. ISBN 9781605589343. DOI 10.1145/1868914.1868943. (Cited on pages 1, 32, 49, 101, 103, and 110).

Kristina Höök. *Designing with the Body: Somaesthetic Interaction Design*. MIT Press, 2018. (Cited on pages 37, 49, and 50).

Kristina Höök and Jonas Löwgren. Strong Concepts: Intermediate-Level Knowledge in Interaction Design Research. *ACM Trans. Comput.-Hum. Interact.*, 19(3), oct 2012. ISSN 1073-0516. DOI 10.1145/2362364.2362371. (Cited on page 109).

Kristina Höök, Martin P. Jonsson, Anna Ståhl, and Johanna Mercurio. Somaesthetic Appreciation Design. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, page 3131–3142, New York, NY, USA, 2016. Association for Computing Machinery. ISBN 9781450333627. DOI 10.1145/2858036.2858583. (Cited on page 49).

Kristina Höök, Baptiste Caramiaux, Cumhur Erkut, Jodi Forlizzi, Nassrin Hajinejad, Michael Haller, Caroline C. M. Hummels, Katherine Isbister, Martin Jonsson, George Khut, Lian Loke, Danielle Lottridge, Patrizia Marti, Edward Melcer, Florian F. Müller, Marianne G. Petersen, Thecla Schiphorst, Elena M. Segura, Anna Ståhl, Dag Svanæs, Jakob Tholander, and Helena Tobiasson. Embracing First-Person Perspectives in Soma-Based Design. *Informatics*, 5(1):8, 2018. DOI 10.3390/informatics5010008. (Cited on page 103).

Kristina Höök, Steve Benford, Paul Tennent, Vasiliki Tsaknaki, Miquel Alfaras, Juan Pablo Martinez Avila, Christine Li, Joseph Marshall, Claudia Daudén Roquet, Pedro Sanches, Anna Ståhl, Muhammad Umair, Charles Windlin, and Feng Zhou. Unpacking Non-Dualistic Design: The Soma Design Case. *ACM Transactions on Computer-Human Interaction*, 28(6):1–36, 2021. ISSN 1073-0516, 1557-7325. DOI 10.1145/3462448. (Cited on pages 37, 49, and 101).

Eva Hornecker and Jacob Buur. Getting a Grip on Tangible Interaction: A Framework on Physical Space and Social Interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '06, page 437–446, New York, NY, USA, 2006. Association for Computing Machinery. ISBN 1595933727. DOI 10.1145/1124772.1124838. (Cited on page 36).

Fred Hosken. The subjective, human experience of groove: a phenomenological investigation. *Psychology of Music*, 46:1–17, 2018. DOI 10.1177/0305735618792440. (Cited on page 82).

David M. Howard. Intonation Drift in A Capella Soprano, Alto, Tenor, Bass Quartet Singing With Key Modulation. *Journal of Voice*, 21(3):300–315, May 2007. DOI 10.1016/j.jvoice.2005.12.005. (Cited on pages 67, 79, and 80).

David M. Howard. Acoustics of the trained versus untrained singing voice. *Current Opinion in Otolaryngology & Head and Neck Surgery*, 17(3):155–159, 2009. DOI 10.1097/MOO.0b013e32832af11b. (Cited on pages 13 and 15).

Noura Howell, Greg Niemeyer, and Kimiko Ryokai. Life-Affirming Biosensing in Public: Sounding Heartbeats on a Red Bench. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, page 1–16, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 9781450359702. DOI 10.1145/3290605.3300910. (Cited on pages xvi, 108, and 109).

Noura Howell, Audrey Desjardins, and Sarah Fox. Cracks in the Success Narrative: Rethinking Failure in Design Research through a Retrospective Trioethnography. *ACM Trans. Comput.-Hum. Interact.*, 28(6), 2021. ISSN 1073-0516. DOI 10.1145/3462447. (Cited on pages 152, 173, 175, and 176).

Peter Howell. Assessment of some contemporary theories of stuttering that apply to spontaneous speech. *Contemporary issues in communication science and disorders*, 31(Spring):123–140, 2004. (Cited on pages 67 and 83).

Peter Howell and Alexandra Archer. Susceptibility to the effects of delayed auditory feedback. *Perception and Psychophysics*, 3(36):296–302, 1984. DOI 10.3758/bf03206371. (Cited on page 35).

Peter Howell, David J. Powell, and Ian Khan. Amplitude contour of the delayed signal and interference in delayed auditory feedback tasks. *Journal of Experimental Psychology: Human Perception and Performance*, 9(5):772, 1983. (Cited on page 67).

Ching-Tang Huang, Chien-Fa Tang, Ming-Chen Lee, and Shuo-Hung Chang. Parametric design of yarn-based piezoresistive sensors for smart textiles. *Sensors and Actuators A: Physical*, 148(1): 10–15, 2008. DOI 10.1016/j.sna.2008.06.029. (Cited on page 120).

Kunpeng Huang, Md. Tahmidul Islam Molla, Kat Roberts, Pin-Sung Ku, Aditi Galada, and Cindy Hsin-Liu Kao. Delocalizing Strain in Interconnected Joints of On-Skin Interfaces. In *2021 International Symposium on Wearable Computers*, ISWC '21, page 91–96, New York, NY, USA, 2021a. Association for Computing Machinery. ISBN 9781450384629. DOI 10.1145/3460421.3478812. (Cited on page 121).

Kunpeng Huang, Ruojia Sun, Ximeng Zhang, Md. Tahmidul Islam Molla, Margaret Dunne, Francois Guimbretiere, and Cindy Hsin-Liu Kao. WovenProbe: Probing Possibilities for Weaving Fully-Integrated On-Skin Systems Deployable in the Field. In *Designing Interactive Systems Conference 2021*, DIS '21, page 1143–1158, New York, NY, USA, 2021b. Association for Computing Machinery. ISBN 9781450384766. DOI 10.1145/3461778.3462105. (Cited on page 121).

Madeline Huberth and Takako Fujioka. Performers' Motions Reflect the Intention to Express Short or Long Melodic Groupings. *Music Perception: An Interdisciplinary Journal*, 35(4):437–453, 2018. DOI 10.1525/mp.2018.35.4.437. (Cited on pages 11 and 12).

Jörn Hurtienne. How Cognitive Linguistics Inspires HCI: Image Schemas and Image-Schematic Metaphors. *International Journal of Human-Computer Interaction*, 33(1):1–20, 2016. DOI 10.1080/10447318.2016.1232227. (Cited on page 42).

Jörn Hurtienne, Diana Löffler, Clara Hüsch, Daniel Reinhardt, Robert Tscharn, and Stephan Hube. *Happy Is Up, Sad Is Down: 65 Metaphors for Design.* Bis B.V., Uitgeverij (BIS Publishers), 2020. ISBN 9789063695934. (Cited on pages 35 and 37).

Edmund Husserl. *Ideas: General Introduction to Pure Phenomenology.* Taylor & Francis, 2014. ISBN 9781317832270. (Cited on pages 37 and 46).

Kristina Höök, Baptiste Caramiaux, Cumhur Erku, Jodi Forlizzi, Nassrin Hajinejad, Michael Haller, Caroline C. M. Hummels, Katherine Isbister, Martin Jonsson, George Khut, Lian Loke, Danielle Lottridge, Patrizia Marti, Edward Melcer, Florian Floyd Müller, Marianne Graves Petersen, Thecla Schiphorst, Elena Márquez Segura, Anna Ståhl, Dag Svanæs, Jakob Tholander, and Helena Tobiasson. Embracing First-Person Perspectives in Soma-Based Design. *Informatics*, 5(8):1–26, 2015. DOI 10.3390/informatics5010008. (Cited on page 53).

Dawn Iacobucci, Steven S. Posavac, Frank R. Kardes, Matthew J. Schneider, and Deidre L. Popovich. Toward a more nuanced understanding of the statistical properties of a median split. *Journal of Consumer Psychology*, 25(4):652–665, 2015. DOI 10.1016/j.jcps.2014.12.002. (Cited on page 85).

Naoto Igarashi, Kenji Suzuki, Hiroaki Kawamoto, and Yoshiyuki Sankai. bioLights: Light emitting wear for visualizing lower-limb muscle activity. In *Proc. International Conference of the IEEE Engineering in Medicine and Biology, Buenos Aires, Argentina*, pages 6393–6396, 2010. DOI 10.1109/iembs.2010.5627306. (Cited on pages 58 and 119).

Don Ihde. The Experience of Technology: Human-Machine Relations. *Cultural Hermeneutics*, 2(3): 267–279, 1975. DOI 10.1177/019145377500200304. (Cited on pages 1, 32, 141, and 149).

Don Ihde. Chapter 4. Human Beginnings and Music: Technology and Embodiment Roles. In *Sound and Affect: Voice*, pages 99–107. University of Chicago Press, Music, World, edited by Judith Lochhead, Eduardo Mendieta and Stephen Decatur Smith, Chicago, 2021. DOI 10.7208/9780226758152-006. (Cited on page 32).

Jordi Janer. Voice-controlled plucked bass guitar through two synthesis techniques. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 132–135, Vancouver, BC, Canada, 2005. DOI 10.5281/zenodo.1176758. (Cited on page 21).

Marc Jeannerod. The representing brain: Neural correlates of motor intention and imagery. *Behavioral and Brain Sciences*, 17(2):187–245, 1994. (Cited on pages 26 and 34).

Marc Jeannerod. Mental imagery in the motor context. *Neuropsychologia*, 33(11):1419–1432, 1995. (Cited on pages 26, 34, and 36).

Marc Jeannerod. The 25th Bartlett Lecture - To act or not to act: Perspectives on the representation of actions. *Quarterly Journal of Experimental Psychology*, 52:1–29, 1999. (Cited on pages 26, 34, and 36).

Alexander Refsum Jensenius, Victor Gonzalez Sanchez, Agata Zelechowska, and Kari Anne Vadstensvik Bjerkestrand. Exploring the Myo controller for sonic microinteraction. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 442–445, Copenhagen, Denmark, 2017. Aalborg University Copenhagen. DOI 10.5281/zenodo.1176308. (Cited on page 56).

Jennifer A. Jestley. *Metaphorical and Non-Metaphorical Imagery Use in Vocal Pedagogy: An Investigation of Underlying Cognitive Organisational Constructs.* Doctoral dissertation, University of British Columbia, Vancouver, Canada, 2011. (Cited on pages xix, 1, 25, 26, 38, 40, 41, and 104).

Jeff A. Johnson, Teresa L. Roberts, William Verplank, David C. Smith, Charles H. Irby, Marian Beard, and Kevin Mackey. The Xerox Star: a retrospective. *Computer (Long Beach Calif.)*, 22 (9):11–26, September 1989. (Cited on page 35).

Mark Johnson. The Body in the Mind: The Bodily Basis of Meaning, Imagination, and Reason. *The Journal of Aesthetics and Art Criticism*, 47(4):400, 1989. DOI 10.2307/431155. (Cited on pages 34 and 36).

Gerald Jonas. Aglow (An Interview with Diana Dew). *The New Yorker*, 1967. URL https://www.newyorker.com/magazine/1967/01/28/aglow-2. (Cited on page 121).

Benjamin Jones, Yuxuan Mei, Haisen Zhao, Taylor Gotfrid, Jennifer Mankoff, and Adriana Schulz. Computational Design of Knit Templates. *ACM Trans. Graph.*, 41(2), 2021. ISSN 0730-0301. DOI 10.1145/3488006. (Cited on page 121).

Jeffery A. Jones and Dwayne Keough. Auditory-motor mapping for pitch control in singers and nonsingers. *Experimental brain research*, 190(3):279–287, 2008. (Cited on page 80).

Patrik N. Juslin and Petri Laukka. Communication of Emotions in Vocal Expression and Music Performance: Different Channels, Same Code? *Psychological Bulletin*, 129(5):770–814, 2003. DOI 10.1037/0033-2909.129.5.770. (Cited on pages 18 and 31).

Patrik N. Juslin and Guy Madison. The Role of Timing Patterns in Recognition of Emotional Expression from Musical Performance. *Music Perception*, 17(2):197–221, 1999. DOI 10.2307/40285891. (Cited on page 31).

Patrik N. Juslin and Daniel Västfjäll. Emotional responses to music: The need to consider underlying mechanisms. *Behavioral and Brain Sciences*, 31(5):559—621, 2008. DOI 10.1017/s0140525x08005293. (Cited on pages 18 and 31).

Stuart B. Kamenetsky, David S. Hill, and Sandra E. Trehub. Effect of Tempo and Dynamics on the Perception of Emotion in Music. *Psychology of Music*, 25(2):149–160, 1997. DOI 10.1177/0305735697252005. (Cited on page 31).

Prithvi Kantan and Sofia Dahl. Communicating Gait Performance Through Musical Energy: Towards an Intuitive Biodfeedback System for Neurorehabilitation. In *Proc. International Workshop on Interactive Sonification, Stockholm, Sweden*, pages 108–115, 2019. DOI 10.5281/zenodo.3756783. (Cited on page 58).

Ajay Kapur, Ariel J. Lazier, Philip L. Davidson, Scott Wilson, and Perry R. Cook. The Electronic Sitar Controller. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 7–12, Hamamatsu, Japan, 2004. DOI 10.5281/zenodo.1176623. (Cited on pages 20 and 118).

Arnav Kapur, Shreyas Kapur, and Pattie Maes. AlterEgo: A Personalized Wearable Silent Speech Interface. In *23rd International Conference on Intelligent User Interfaces*, IUI '18, page 43–53, New York, NY, USA, 2018. Association for Computing Machinery. ISBN 9781450349451. DOI 10.1145/3172944.3172977. (Cited on pages 55, 57, 116, 117, 121, 140, and 141).

Jakob Karolus, Hendrik Schuff, Thomas Kosch, Paweł W. Wozniak, and Albrecht Schmidt. EMGuitar: Assisting Guitar Playing with Electromyography. In *Proceedings of the 2018 Designing Interactive Systems Conference*, DIS '18, page 651–655, New York, NY, USA, 2018. Association for Computing Machinery. ISBN 9781450351980. DOI 10.1145/3196709.3196803. (Cited on page 121).

Jakob Karolus, Annika Kilian, Thomas Kosch, Albrecht Schmidt, and Paweł W. Wozniak. Hit the Thumb Jack! Using Electromyography to Augment the Piano Keyboard. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference*, DIS '20, page 429–440, New York, NY, USA, 2020. Association for Computing Machinery. ISBN 9781450369749. DOI 10.1145/3357236.3395500. (Cited on pages 56 and 121).

Jakob Karolus, Felix Bachmann, Thomas Kosch, Albrecht Schmidt, and Paweł W. Woźniak. Facilitating Bodily Insights Using Electromyography-Based Biofeedback during Physical Activity. In *Proceedings of the 23rd International Conference on Mobile Human-Computer Interaction*, MobileHCI '21, New York, NY, USA, 2021a. Association for Computing Machinery. ISBN 9781450383288. DOI 10.1145/3447526.3472027. (Cited on page 119).

Jakob Karolus, Francisco Kiss, Caroline Eckerth, Nicolas Viot, Felix Bachmann, Albrecht Schmidt, and Pawel W. Wozniak. EMBody: A Data-Centric Toolkit for EMG-Based Interface Prototyping and Experimentation. *Proc. ACM Hum.-Comput. Interact.*, 5(EICS), May 2021b. DOI 10.1145/3457142. (Cited on page 119).

Alexandre Kaspar, Liane Makatura, and Wojciech Matusik. Knitting Skeletons: A Computer-Aided Design Tool for Shaping and Patterning of Knitted Garments. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*, UIST '19, page 53–65, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 9781450368162. DOI 10.1145/3332165.3347879. (Cited on pages 121 and 135).

Peter E. Keller. Mental imagery in music performance: underlying mechanisms and potential benefits. *Annals of the New York Academy of Sciences*, 1252(1):206–213, 2012. DOI 10.1111/j.1749-6632.2011.06439.x. (Cited on pages 26 and 27).

Peter E. Keller and Mirjam Appel. Individual Differences, Auditory Imagery, and the Coordination of Body Movements and Sounds in Musical Ensembles. *Music Perception: An Interdisciplinary Journal*, 28(1):27–46, 2010. DOI 10.1525/mp.2010.28.1.27. (Cited on page 12).

Roger A. Kendall and Edward C. Carterette. The Communication of Musical Expression. *Music Perception: An Interdisciplinary Journal*, 8(2):129–64, 1990. DOI 10.2307/40285493. (Cited on page 31).

Michael Kennedy and Joyce Bourne Kennedy. *The Concise Oxford Dictionary of Music*. Oxford University Press, Oxford, 1980. DOI 10.1093/acref/9780199203833.001.0001. (Cited on page 69).

Zeeshan O. Khokhar, Zhen G. Xiao, and Carlo Menon. Surface EMG pattern recognition for real-time control of a wrist exoskeleton. *BioMedical Engineering OnLine*, 9(41):2010, 2010. DOI 10.1186/1475-925x-9-41. (Cited on pages 119 and 140).

Annika Kilian, Jakob Karolus, Thomas Kosch, Albrecht Schmidt, and Paweł W. Woźniak. EMPiano: Electromyographic Pitch Control on the Piano Keyboard. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI EA '21, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450380959. DOI 10.1145/3411763.3451556. (Cited on pages 119 and 121).

Ozgun Kilic Afsar, Yoav Luft, Kelsey Cotton, Ekaterina R. Stepanova, Claudia Núñez Pacheco, Rebecca Kleinberger, Fehmi Ben Abdesslem, Hiroshi Ishii, and Kristina Höök. Corsetto: A Kinesthetic Garment for Designing, Composing for, and Experiencing an Intersubjective Haptic Voice. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23, New York, NY, USA, 2023. Association for Computing Machinery. ISBN 9781450394215. DOI 10.1145/3544548.3581294. (Cited on pages 20 and 22).

Su Jeong Kim, So Yeon Jeong, and Tae Lim Yoon. The Effect of Visual Feedback of Head Angles With Using a Mobile Posture-Aware System on Craniocervical Angle and Neck and Shoulder Muscles Fatigue During Watching the Smartphone. *The Journal of Korean Physical Therapy*, 30, 2018. ISSN 1229-0475. DOI 10.18857/jkpt.2018.30.2.47. (Cited on page 119).

Naoki Kimura, Michinari Kono, and Jun Rekimoto. SottoVoce: An Ultrasound Imaging-Based Silent Speech Interaction Using Deep Neural Networks. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, page 1–11, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 9781450359702. DOI 10.1145/3290605.3300376. (Cited on pages 21 and 54).

Mathias S Kirkegaard, Mathias Bredholt, Christian Frisson, and Marcelo Wanderley. TorqueTuner: A self contained module for designing rotary haptic force feedback for digital musical instruments. In Romain Michon and Franziska Schroeder, editors, *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 273–278, Birmingham, UK, 2020. Birmingham City University. DOI 10.5281/zenodo.4813359. (Cited on page 54).

Alexandra Kitson, Mirjana Prpa, and Bernhard E Riecke. Immersive Interactive Technologies for Positive Change: A Scoping Review and Design Considerations. *Frontiers in Psychology*, 9:1354, 2018. (Cited on page 176).

Boris Kleber, Niels Birbaumer, Ralf Veit, Tracy Trevorrow, and Martin Lotze. Overt and imagined singing of an Italian aria. *NeuroImage*, 36(3):889–900, 2007. DOI 10.1016/j.neuroimage.2007.02.053. (Cited on pages 25, 26, and 117).

Rébecca Kleinberger, Nikhil Singh, Xiao Xiao, and Akito van Troyer. Voice at NIME: a Taxonomy of New Interfaces for Vocal Musical Expression. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, The University of Auckland, New Zealand, June 2022. DOI 10.21428/92fbeb44.4308fb94. (Cited on page 20).

Scott R. Klemmer, Björn Hartmann, and Leila Takayama. How Bodies Matter: Five Themes for Interaction Design. In *Proceedings of the 6th Conference on Designing Interactive Systems*, DIS '06, page 140–149, New York, NY, USA, 2006. Association for Computing Machinery. ISBN 1595933670. DOI 10.1145/1142405.1142429. (Cited on pages 37, 46, and 149).

Jarrod Knibbe, Paul Strohmeier, Sebastian Boring, and Kasper Hornbæk. Automatic Calibration of High Density Electric Muscle Stimulation. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 1(3), Sep 2017. DOI 10.1145/3130933. (Cited on page 120).

Jarrod Knibbe, Rachel Freire, Marion Koelle, and Paul Strohmeier. Skill-Sleeves: Designing Electrode Garments for Wearability. In *Proceedings of the Fifteenth International Conference on Tangible, Embedded, and Embodied Interaction*, TEI '21, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450382137. DOI 10.1145/3430524.3440652. (Cited on pages 119, 120, 121, 124, and 137).

Marion Koelle, Thomas Olsson, Robb Mitchell, Julie Williamson, and Susanne Boll. What is (Un)Acceptable? Thoughts on Social Acceptability in HCI Research. *Interactions*, 26(3):36–40, 2019a. ISSN 1072-5520. DOI 10.1145/3319073. (Cited on page 119).

Marion Koelle, Torben Wallbaum, Wilko Heuten, and Susanne Boll. Evaluating a Wearable Camera's Social Acceptability In-the-Wild. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI EA '19, page 1–6, New York, NY, USA, 2019b. Association for Computing Machinery. ISBN 9781450359719. DOI 10.1145/3290607.3312837. (Cited on page 119).

Yasuharu Koike, Kumiyo Nakakoji, and Yasunori Yamamoto. Tele-Kinesthetic Interaction: Using Hand Muscles to Interact with a Tangible 3D Object. In *ACM SIGGRAPH 2006 Emerging Technologies*, SIGGRAPH '06, page 33–es, New York, NY, USA, 2006. Association for Computing Machinery. ISBN 1595933646. DOI 10.1145/1179133.1179167. (Cited on pages 56 and 119).

Stephen M. Kosslyn. *Image and mind*. Harvard University Press, Cambridge, MA, 1980. (Cited on pages 26, 34, 36, and 37).

Stephen M. Kosslyn, Marlene Behrmann, and Marc Jeannerod. The cognitive neuroscience of mental imagery. *Neuropsychologia*, 33(11):1335–1344, 1995. (Cited on page 34).

Stephen M. Kosslyn, Giorgio Ganis, and William L. Thompson. Neural foundations of imagery. *Nature Reviews Neuroscience*, 2(9):635–642, 2001. DOI 10.1038/35090055. (Cited on pages 25, 34, 36, 57, and 101).

Carol L. Krumhansl. *Cognitive Foundations of Musical Pitch*. New York: Oxford University Press, 1990. (Cited on page 72).

Carol L. Krumhansl. An Exploratory Study of Musical Emotions and Psychophysiology. *Canadian Journal of Experimental Psychology*, 51(4):336–352, 1997. DOI 10.1037/1196-1961.51.4.336. (Cited on page 31).

Carol L. Krumhansl. Rhythm and pitch in music cognition. *Psychological Bulletin*, 126(1):159–179, 2000. DOI 10.1037/0033-2909.126.1.159. (Cited on page 31).

Carol L. Krumhansl. Music: A Link Between Cognition and Emotion. *Current Directions in Psychological Science*, 11(2):45–50, 2002. DOI 10.1111/1467-8721.00165. (Cited on page 31).

Robert Kubey, Reed Larson, and Mihaly Csikszentmihalyi. Experience Sampling Method Applications to Communication Research Questions. *Journal of Communication*, 46(2):99–120, 1996. DOI 10.1111/j.1460-2466.1996.tb01476.x. (Cited on page 59).

George Lakoff. The contemporary theory of metaphor. In Andrew Ortony, editor, *Metaphor and Thought*, pages 202–251. Cambridge University Press, Cambridge, November 1993. (Cited on pages 36, 37, and 102).

George Lakoff and Mark Johnson. *Metaphors We Live By*. University of Chicago Press, Chicago, 1980. ISBN 0-226-46801-1. (Cited on page 35).

George Lakoff and Mark Johnson. The embodied mind. In *Philosophy in the flesh: Embodied mind and its challenges to Western thought*, pages 16–41. New York, Basic Books, 1999. (Cited on pages 34, 36, and 103).

Daniel R. Lametti, Sazzad M. Nasir, and David J. Ostry. Sensory preference in speech production revealed by simultaneous alteration of auditory and somatosensory feedback. *Journal of Neuroscience*, 32(27):9351–9358, 2012. (Cited on page 81).

Ronald W. Langacker and George Lakoff. Women, Fire, and Dangerous Things: What Categories Reveal about the Mind. *Language*, 64(2):384, 1988. DOI 10.2307/415440. (Cited on pages 34 and 36).

William Langston. Violating Orientational Metaphors Slows Reading. *Discourse Processes*, 34(3): 281–310, 2002. DOI 10.1207/s15326950dp3403_3. (Cited on pages 35 and 37).

Charles R. Larson, Kenneth W. Altman, Hanjun Liu, and Timothy C. Hain. Interactions between auditory and somatosensory feedback for voice F 0 control. *Experimental Brain Research*, 187(4): 613–621, 2008. (Cited on page 30).

Astrid Twenebowa Larssen, Toni Robertson, and Jenny Edwards. The Feel Dimension of Technology Interaction: Exploring Tangibles through Movement and Touch. In *Proceedings of the 1st International Conference on Tangible and Embedded Interaction*, TEI '07, page 271–278, New York, NY, USA, 2007. Association for Computing Machinery. ISBN 9781595936196. DOI 10.1145/1226969.1227024. (Cited on page 103).

John Law and Marianne Elisabeth Lien. Slippery: Field notes in empirical ontology. *Social Studies of Science*, 43(3):363–378, September 2012. DOI 10.1177/0306312712456947. (Cited on page 33).

Amber M. Leaver, Jennifer Van Lare, Brandon Zielinski, Andrea R. Halpern, and Josef P. Rauschecker. Brain Activation during Anticipation of Sound Sequences. *Journal of Neuroscience*, 29(8):2477–2485, 2009. DOI 10.1523/jneurosci.4921-08.2009. (Cited on pages 27 and 28).

Bernard S. Lee. Some Effects of Side-Tone Delay. *The Journal of the Acoustical Society of America*, 22(5):639–640, September 1950. DOI 10.1121/1.1906665. (Cited on page 67).

Myungin Lee. Entangled: A Multi-Modal, Multi-User Interactive Instrument in Virtual 3D Space Using the Smartphone for Gesture Control. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Shanghai, China, 2021. DOI 10.21428/92fbeb44.eae7c23f. (Cited on page 54).

Victor R. Lee. What's happening in the" Quantified Self" movement? *Proceedings of the 2014 International Conference of the Learning Sciences (ICLS)*, page 1032, 2014. (Cited on page 175).

Andreas C. Lehmann, John A. Sloboda, and Robert H. Woody. *Psychology for Musicians: Understanding and Acquiring the Skills*. Oxford University Press, Oxford, 2007. ISBN 978-0-19-514610-3. (Cited on page 12).

Marc Leman. *Embodied Music Cognition and Mediation Technology*. MIT Press, Cambridge, MA, 2008. ISBN 978-0-262-12293-1. (Cited on pages 12 and 25).

Marc Leman and Pieter-Jan Maes. The Role of Embodiment in the Perception of Music. *Empirical Musicology Review*, 9(3-4):236–246, 2015. DOI 10.18061/emr.v9i3-4.4498. (Cited on pages 11, 12, and 27).

Marc Leman, Luc Nijs, Pieter-Jan Maes, and Edith Van Dyck. What Is Embodied Music Cognition? In R. Bader, editor, *Springer Handbook of Systematic Musicology*. Springer-Verlag, Berlin, Germany, 2017. DOI 10.1007/978-3-662-55004-5_34. (Cited on page 11).

Giacomo Lepri and Andrew McPherson. Making Up Instruments: Design Fiction for Value Discovery in Communities of Musical Practice. In *Proceedings of the 2019 on Designing Interactive Systems Conference*, DIS '19, page 113–126, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 9781450358507. DOI 10.1145/3322276.3322353. (Cited on page 33).

Fred Lerdahl. Calculating Tonal Tension: Analysis of the First Movement of Mozart's Piano Sonata K. 282. *Music Perception: An Interdisciplinary Journal*, 13(3):319–363, 1996. DOI 10.2307/40286174. (Cited on page 31).

Fred Lerdahl and Carol L. Krumhansl. Modeling Tonal Tension. *Music Perception: An Interdisciplinary Journal*, 24(4):329–366, 2006. DOI 10.1525/mp.2007.24.4.329. (Cited on page 31).

LessEMF. Shielding and Conductive Fabrics, 2020. URL https://www.lessemf.com/fabric.html. (Cited on page 127).

Dan J. Levitin and Perry R. Cook. Memory for musical tempo: Additional evidence that auditory memory is absolute. *Perception & Psychophysics*, 58(6):927–935, 1996. DOI 10.3758/bf03205494. (Cited on page 81).

Qing Li, Bruce Clark, and Ian Winchester. Instructional design and technology grounded in enactivism: A paradigm shift? *British Journal of Educational Technology*, 41(3):403–419, 2010. DOI 10.1111/j.1467-8535.2009.00954.x. (Cited on page 32).

An Liang, Rebecca Stewart, Rachel Freire, and Nick Bryan-Kinns. Knit Stretch Sensor Placement for Body Movement Sensing. In *Proceedings of the Fifteenth International Conference on Tangible, Embedded, and Embodied Interaction*, TEI '21, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450382137. DOI 10.1145/3430524.3440629. (Cited on page 120).

Chin Guan Lim, Chin Yi Tsai, and Mike Y. Chen. MuscleSense: Exploring Weight Sensing Using Wearable Surface Electromyography (SEMG). In *Proceedings of the Fourteenth International Conference on Tangible, Embedded, and Embodied Interaction*, TEI '20, page 255–263, New York, NY, USA, 2020. Association for Computing Machinery. ISBN 9781450361071. DOI 10.1145/3374920.3374943. (Cited on pages 56, 119, and 140).

César F. Lima, Nadine Lavan, Samuel Evans, Zarinah Agnew, Andrea R. Halpern, Pradheep Shanmugalingam, Sophie Meekings, Dana Boebinger, Markus Ostarek, Carolyn McGettigan, Jane E. Warren, and Sophie K. Scott. Feel the Noise: Relating Individual Differences in Auditory Imagery to the Structure and Function of Sensorimotor Systems. *Cerebral Cortex*, 25(11):4638–4650, 2015. DOI 10.1093/cercor/bhv134. (Cited on page 60).

Sebastian Löbbers and George Fazekas. Sketching Sounds: an exploratory study on sound-shape associations. In *Proceedings of the International Computer Music Conference (ICMC), Santiago, Chile*, pages 1–6, 2021. (Cited on page 31).

Sebastian Löbbers, Mathieu Barthet, and George Fazekas. Sketching Sounds: Using sound-shape associations to build a sketchbased sound synthesiser. In *Proceedings of the DMRN+16 Digital Music Research Network Workshop, London, UK*, page 1, 2021. (Cited on page 31).

Nina Loimusalo and Erkki Huovinen. Silent Reading and Aural Models in Pianists' Mental Practice. In *Proceedings of the 14th International Conference on Music Perception and Cognition (ICMPC), San Francisco, California, USA*, pages 609–614, July 5-9, 2016. URL http://icmpc.org/icmpc14/files/ICMPC14_Proceedings.pdf. (Cited on page 27).

Pedro Lopes, Alexandra Ion, Willi Mueller, Daniel Hoffmann, Patrik Jonell, and Patrick Baudisch. Proprioceptive Interaction. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, page 939–948, New York, NY, USA, 2015. Association for Computing Machinery. ISBN 9781450331456. DOI 10.1145/2702123.2702461. (Cited on pages 119 and 120).

Pedro Lopes, Sijing You, Lung-Pan Cheng, Sebastian Marwecki, and Patrick Baudisch. Providing Haptics to Walls & Heavy Objects in Virtual Reality by Means of Electrical Muscle Stimulation. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, page 1471–1482, New York, NY, USA, 2017. Association for Computing Machinery. ISBN 9781450346559. DOI 10.1145/3025453.3025600. (Cited on page 120).

Alex Loscos and Thomas Aussenac. The Wahwactor: A Voice Controlled Wah-Wah Pedal. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 172–175, Vancouver, BC, Canada, 2005. DOI 10.5281/zenodo.1176776. (Cited on page 21).

Yiyue Luo, Kui Wu, Tomás Palacios, and Wojciech Matusik. KnitUI: Fabricating Interactive and Sensing Textiles with Machine Knitting. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450380966. DOI 10.1145/3411764.3445780. (Cited on page 120).

Hugh S. Lusted and R. Benjamin Knapp. Biomuse: Musical performance generated by human bioelectric signals. *The Journal of the Acoustical Society of America*, 84(S1):S179–s179, 1988. DOI 10.1121/1.2025994. (Cited on page 56).

Michael J. Lyons, Michael Haehnel, and Nobuji Tetsutani. Designing, Playing, and Performing with a Vision-based Mouth Interface. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 116–121, Montreal, Canada, 2003. DOI 10.5281/zenodo.1176529. (Cited on pages 18 and 21).

Anna Macaranas, Alissa N. Antle, and Bernhard E. Riecke. What is Intuitive Interaction? Balancing Users' Performance and Satisfaction with Natural User Interfaces. *Interacting with Computers*, 27(3):357–370, 2015. DOI 10.1093/iwc/iwv003. (Cited on page 42).

Ewen N. MacDonald, Robyn Goldberg, and Kevin G. Munhall. Compensations in response to real-time formant perturbations of different magnitudes. *The Journal of the Acoustical Society of America*, 127(2):1059–1068, 2010. (Cited on page 79).

Jennifer MacRitchie and Hubert Eiholzer. Exploring the perceptual effects of performers' interpretations. *Journal of Interdisciplinary Music Studies*, 6(2):177–200, 2012. DOI 10.4407/jims.2014.02.004. (Cited on page 39).

Jennifer MacRitchie and Andrew J. Milne. Exploring the Effects of Pitch Layout on Learning a New Musical Instrument. *Applied Sciences*, 7(1218):1–19, 2017. DOI 10.3390/app7121218. (Cited on pages 28, 58, and 83).

Jennifer MacRitchie and Massimo Zicari. The Intentions of Piano Touch. In *Joint 12th International Conference on Music Perception and Cognition and 8th Triennial Conference of the European Society for the Cognitive Sciences of Music, Thessaloniki, Greece*, July 23–28, 2012. (Cited on page 12).

Jennifer MacRitchie, Stuart Pullinger, Nicholas J. Bailey, and Graham Hair. Communicating Phrasing Structure with Multi-Modal Expressive Techniques in Piano Performance. In *The 2nd International Conference on Music Communication Science, Sydney, Australia, Dec 3-4*, 2009. (Cited on page 12).

Guy Madison. Detection of linear temporal drift in sound sequences: empirical data and modelling principles. *Acta Psychologica*, 117(1):95–118, September 2004. DOI 10.1016/j.actpsy.2004.05.004. (Cited on pages 70 and 82).

Guy Madison and Björn Merker. On the limits of anisochrony in pulse attribution. *Psychological Research*, 66(3):201–207, August 2002. DOI 10.1007/s00426-001-0085-y. (Cited on page 82).

Kristina Mah, Lian Loke, and Luke Hespanhol. Towards a Contemplative Research Framework for Training Self-Observation in HCI: A Study of Compassion Cultivation. *ACM Trans. Comput.-Hum. Interact.*, 28(6), 2021. ISSN 1073-0516. DOI 10.1145/3471932. (Cited on page 50).

Jessica R. Malloy, Dominic Nistal, Matthias Heyne, Monique C. Tardif, and Jason W. Bohland. Delayed Auditory Feedback Elicits Specific Patterns of Serial Order Errors in a Paced Syllable Sequence Production Task. *Journal of Speech, Language, and Hearing Research*, 65(5):1800–1821, May 2022. DOI 10.1044/2022_jslhr-21-00427. (Cited on pages 67, 80, 81, 82, and 83).

Paul Marshall, Rowanne Fleck, Amanda Harris, Jochen Rick, Eva Hornecker, Yvonne Rogers, Nicola Yuill, and Nick Sheep Dalton. Fighting for Control: Children's Embodied Interactions When Using Physical and Digital Representations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, page 2149–2152, New York, NY, USA, 2009. Association for Computing Machinery. ISBN 9781605582467. DOI 10.1145/1518701.1519027. (Cited on page 36).

Andrea Martelloni and Courtney N. Reed. Music From the Augmented Instruments Lab: 23 November 2021, 2021. URL https://www.youtube.com/watch?v=axn_wQM_I_c&t=3426s. (Cited on page 122).

Charles Patrick Martin. Myo-to-OSC: A pure-python cross-platform solution for simply connecting Myo armbands to OSC-connected software., 2018. (Cited on page 56).

Charles Patrick Martin, Alexander Refsum Jensenius, Kari Anne Vadstensvik Bjerkestrand, and Victoria Johnson. Stillness Under Tension: Performance for Myo armbands and Bela embedded computers. In *MusicLab Vol.1: Biophysical Music*, 2017. (Cited on page 56).

Charles Patrick Martin, Alexander Refsum Jensenius, and Jim Torresen. Composing an Ensemble Standstill Work for Myo and Bela. In Thomas Martin Luke Dahl, Douglas Bowman, editor, *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 196–197, Blacksburg, Virginia, USA, 2018. Virginia Tech. ISBN 978-1-949373-99-8. DOI 10.5281/zenodo.1302543. (Cited on page 56).

Juan Pablo Martinez Avila, Vasiliki Tsaknaki, Pavel Karpashevich, Charles Windlin, Niklas Valenti, Kristina Höök, Andrew P. McPherson, and Steve Benford. Soma Design for NIME. In Romain Michon and Franziska Schroeder, editors, *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 489–494, Birmingham, UK, 2020. Birmingham City University. DOI 10.5281/zenodo.4813491. (Cited on page 49).

HITEK Electronic Materials. Technical Textiles, 2021. URL https://www.hitek-ltd.co.uk/technical-textiles. (Cited on page 127).

Max Mathews. Electronic Violin: A Research Tool. *Journal of the Violin Society of America*, 8(1), 1984. (Cited on page 19).

Masaki Matsubara, Hideki Kadone, Masaki Iguchi, Hiroko Terasawa, and Kenji Suzuki. The Effectiveness of Auditory Biofeedback on a Tracking Task for Ankle Joint Movements in Rehabilitation. In *Proc. International Workshop on Interactive Sonification, Erlangen, Germany*, pages 81–86, 2013. (Cited on page 58).

Matthias Mauch and Simon Dixon. pYIN: A Fundamental Frequency Estimator Using Probabilistic Threshold Distributions. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2014)*, 2014. (Cited on page 68).

Matthias Mauch, Chris Cannam, Rache Bittner, George Fazekas, Justin Salamon, Jiajie Dai, Juan Bello, and Simon Dixon. Computer-aided Melody Note Transcription Using the Tony Software: Accuracy and Efficiency. In *Proceedings of the First International Conference on Technologies for Music Notation and Representation (TENOR)*, 2015. (Cited on page 68).

Mattias Mauch, Klaus Frieler, and Simon Dixon. Intonation in unaccompanied singing: Accuracy, drift, and a model of reference pitch memory. *The Journal of the Acoustical Society of America*, 136(1):401–411, 2014. DOI 10.1121/1.4881915. (Cited on pages 67, 68, and 69).

Guerino Mazzola. Expressive Semantics. In *The Topos of Music: Geometric Logic of Concepts, Theory, and Performance*, volume 2. Springer, Basel, Switzerland, 2002. ISBN 978-3-0348-9454-8. (Cited on page 12).

Stephen McAdams and Bruno L. Giordano. The Perception of Musical Timbre. *The Oxford Handbook of Music Psychology*, pages 72–80, 2009. (Cited on page 31).

Gary H. McClelland, John G. Jr. Lynch, Julie R. Irwin, Stephen A. Spiller, and Gavan J. Fitzsimons. Median splits, Type II errors, and false–positive consumer psychology: Don't fight the power. *Journal of Consumer Psychology*, 25(4):679–689, 2015. DOI 10.1016/j.jcps.2015.05.006. (Cited on page 85).

Denisa Qori McDonald, Richard Vallett, Erin Solovey, Geneviève Dion, and Ali Shokoufandeh. Knitted Sensors: Designs and Novel Approaches for Real-Time, Real-World Sensing. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 4(4), Dec 2020. DOI 10.1145/3432201. (Cited on page 120).

Jess McIntosh, Charlie McNeill, Mike Fraser, Frederic Kerber, Markus Löchtefeld, and Antonio Krüger. EMPress: Practical Hand Gesture Classification with Wrist-Mounted EMG and Pressure Sensing. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, page 2332–2342, New York, NY, USA, 2016. Association for Computing Machinery. ISBN 9781450333627. DOI 10.1145/2858036.2858093. (Cited on page 119).

Keith A. McMillen. Stage-Worthy Sensor Bows for Stringed Instruments. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 347–348, Genoa, Italy, 2008. DOI 10.5281/zenodo.1179597. (Cited on page 19).

Andrew P. McPherson. Portable Measurement and Mapping of Continuous Piano Gesture. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 152–157, Daejeon, Republic of Korea, 2013. Graduate School of Culture Technology, KAIST. DOI 10.5281/zenodo.1178610. (Cited on page 54).

Andrew P. McPherson. Bela: An embedded platform for low-latency feedback control of sound. *Journal of the Acoustical Society of America*, 141(3618), 2017. DOI 10.1121/1.4987761. (Cited on pages 114 and 153).

Andrew P. McPherson and Victor Zappi. An Environment for Submillisecond-Latency Audio and Sensor Processing on BeagleBone Black. In *Proceedings of the Audio Engineering Society (AES) 138th Convention, Warsaw, Poland, 2015 May 7–10*, 2015. (Cited on pages 114 and 153).

Geoffrey S. Meltzner, Jason J. Sroka, James T. Heaton, L. Donald Gilmore, Glen Colby, Serge H. Roy, Nancy F. Chen, and Carlo J. De Luca. Speech Recognition for Vocalized and Subvocal Modes of Production using Surface EMG Signals from the Neck and Face. In *9th Annual Conference of the International Speech Communication Association (INTERSPEECH 2008), Brisbane, Australia, September 22-26, 2008*, pages 2667–2670, 2008. DOI 10.21437/Interspeech.2008-661. (Cited on pages 57, 119, and 140).

Maurice Merleau-Ponty. *Phenomenology of perception.* Routledge, 2014. (Cited on pages 1, 37, and 46).

Lia Mice and Andrew P. McPherson. Super Size Me: Interface Size, Identity and Embodiment in Digital Musical Instrument Design. In *CHI Conference on Human Factors in Computing Systems*, CHI '22, New York, NY, USA, 2022. Association for Computing Machinery. ISBN 9781450391573. DOI 10.1145/3491102.3517626. (Cited on pages 33, 38, 48, 54, 149, 152, 175, and 176).

Richard Miller. The invisible instrument? *The National Association of Teachers of Singing Bulletin*, 37(2):1, 1980. (Cited on page 41).

Richard Miller. The law of contingency and vocal pedagogies. *The National Association of Teachers of Singing Journal*, 51(1):31, 1995. (Cited on page 41).

Richard Miller. Imagery and the Teaching of Singing. In *On the Art of Singing*, chapter 1, pages 3–5. Oxford University Press, Oxford, 1996. DOI 10.1093/acprof:osobl/9780195098259.001.0001. (Cited on pages 39, 42, and 110).

Richard Miller. The reluctant student. *The National Association of Teachers of Singing Journal*, 54(3):41–43, 1998a. (Cited on pages 41 and 110).

Richard Miller. The singing teacher in the age of voice science. In R. T. Sataloff, editor, *Vocal health and pedagogy*, pages 297–300. Singular, San Diego, CA, 1998b. (Cited on page 41).

Richard Miller. Historical overview of vocal pedagogy. In R. T. Sataloff, editor, *Vocal health and pedagogy*, pages 301–313. Singular, San Diego, CA, 1998c. (Cited on page 41).

Richard Miller. The syntax of voice technique. *The National Association of Teachers of Singing Journal*, 58(1):49–50, 2001. (Cited on page 41).

Ali Momeni. Caress: An Electro-acoustic Percussive Instrument for Caressing Sounds. In Edgar Berdahl and Jesse Allison, editors, *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 245–250, Baton Rouge, Louisiana, USA, 2015. Louisiana State University. DOI 10.5281/zenodo.1179142. (Cited on page 19).

Fabio Morreale, Andrea Guidi, and Andrew P. McPherson. Magpick: an Augmented Guitar Pick for Nuanced Control. In Marcelo Queiroz and Anna Xambó Sedó, editors, *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 65–70, Porto Alegre, Brazil, 2019. UFRGS. DOI 10.5281/zenodo.3672868. (Cited on page 54).

Daniel Müllensiefen, Bruno Gingras, Lauren Stewart, and Jason Jií Musil. Goldsmiths Musical Sophistication Index (Gold-MSI) v1. 0: Technical Report and Documentation Revision 0.3. *London: Goldsmiths, University of London*, 2013. (Cited on pages 59, 65, and 91).

Daniel Müllensiefen, Bruno Gingras, and Lauren Stewart. The Musicality of Non-Musicians: An Index for Assessing Musical Sophistication in the General Population. *PLoS ONE*, 9(2):e89642, 2014. DOI 10.1371/journal.pone.0089642. (Cited on pages 59, 65, and 91).

Meinard Müller, Peter Grosche, and Frans Wiering. Automated Analysis of Performance Variations in Folk Song Recordings. In *Proceedings ACM MIR'10, March 29–31, 2010, Philadelphia, Pennsylvania, USA*, 2010. ISBN 78-1-60558-815-5/10/03. DOI 10.1145/1743384.1743429. (Cited on page 82).

Sara Nabil, Lee Jones, and Audrey Girouard. Soft Speakers: Digital Embroidering of DIY Customizable Fabric Actuators. In *Proceedings of the Fifteenth International Conference on Tangible, Embedded, and Embodied Interaction*, TEI '21, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450382137. DOI 10.1145/3430524.3440630. (Cited on page 121).

Andrea Nacci, Giovanna Baracca, Salvatore Osvaldo Romeo, Maria Denise Cavaliere, Maria Rosaria Barillari, Stefano Berrettini, Francesco Ursino, and Bruno Fattori. Endoscopic and Phoniatric Evaluation in Singing Students. *Journal of Voice*, 33(2):135–142, 2017. DOI 10.1016/j.jvoice.2017.10.006. (Cited on page 54).

Keith V. Nesbitt. Modelling Human Perception to Leverage the Reuse of Concepts across the Multi-Sensory Design Space. In *Proceedings of the 3rd Asia-Pacific Conference on Conceptual Modelling - Volume 53*, APCCM '06, page 65–74, AUS, 2006. Australian Computer Society, Inc. ISBN 192068235X. (Cited on page 108).

Carman Neustaedter and Phoebe Sengers. Autobiographical Design in HCI Research: Designing and Learning through Use-It-Yourself. In *Proceedings of the Designing Interactive Systems Conference*, DIS '12, page 514–523, New York, NY, USA, 2012. Association for Computing Machinery. ISBN 9781450312103. DOI 10.1145/2317956.2318034. (Cited on pages 53 and 103).

Sarah Nicolls. Seeking Out the Spaces Between: Using Improvisation in Collaborative Composition with Interactive Technology. *Leonardo Music Journal*, 20:47–55, 2010. DOI 10.1162/LMJ_a_00012. (Cited on page 184).

Luc Nijs, Micheline Lesaffre, and Marc Leman. The Musical Instrument as a Natural Extension of the Musician. In H. Castellengo, M. Genevois and J.-M. Bardez, editors, *Music and Its Instruments*, pages 467–484. Editions Delatour France, 2013. (Cited on pages 38 and 47).

Jun Nishida, Kanako Takahashi, and Kenji Suzuki. A Wearable Stimulation Device for Sharing and Augmenting Kinesthetic Feedback. In *Proceedings of the 6th Augmented Human International Conference*, AH '15, page 211–212, New York, NY, USA, 2015. Association for Computing Machinery. ISBN 9781450333498. DOI 10.1145/2735711.2735775. (Cited on page 120).

Aditya Shekhar Nittala, Andreas Karrenbauer, Arshad Khan, Tobias Kraus, and Jürgen Steimle. Computational design and optimization of electro-physiological sensors. *Nature Communications*, 12(1), 2021. DOI 10.1038/s41467-021-26442-1. (Cited on page 120).

Kristin Norderval. Electrifying Opera: Amplifying agency for opera singers improvising with interactive audio technology (Colloquial Paper). In *Proceedings of the 5th International Conference on Live Interfaces*, pages 9–11, 2020, Trondheim, Norway, 2020. March. (Cited on pages 22 and 123).

Charlotte Nordmoen and Andrew P. McPherson. Making Space for Material Entanglements: A Diffractive Analysis of Woodwork and the Practice of Making an Interactive System. In *Designing Interactive Systems Conference*, DIS '22, page 415–423, New York, NY, USA, 2022. Association for Computing Machinery. ISBN 9781450393584. DOI 10.1145/3532106.3533572. (Cited on pages 32, 33, and 46).

Charlotte Nordmoen, Jack Armitage, Fabio Morreale, Rebecca Stewart, and Andrew McPherson. Making Sense of Sensors: Discovery Through Craft Practice With an Open-Ended Sensor Material. In *Proceedings of the 2019 on Designing Interactive Systems Conference*, DIS '19, page 135–146, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 9781450358507. DOI 10.1145/3322276.3322368. (Cited on page 32).

Donald A. Norman and Stephen W. Draper (Eds.). *User Centered System Design: New Perspectives on Human-computer Interaction (1st ed.).* CRC Press, 1986. DOI 10.1201/9780367807320. (Cited on page 37).

Alex Nowitz. Moving Tongues : Playing Space: solo performance featuring voice, strophonion with four loudspeakers, and video, 2018. URL https://www.diva-portal.org/smash/record.jsf?pid=diva2%3A1269577&dswid=7129. (Cited on page 23).

Alex Nowitz. *Monsters I Love: On Multivocal Arts.* PhD thesis, Stockholms Konstnärliga Högskola, 2019. (Cited on page 23).

Alex Nowitz. Assemblages of Multivocal and Schizophonic Practices. *Machinic Assemblages of Desire: Deleuze and Artistic Research 3*, page 129, 2021. (Cited on page 23).

Claudia Núñez Pacheco and Lian Loke. Felt-Sensing Archetypes: Analysing Patterns of Accessing Tacit Meaning in Design. In *Proceedings of the 28th Australian Conference on Computer-Human Interaction*, OzCHI '16, page 462–471, New York, NY, USA, 2016. Association for Computing Machinery. ISBN 9781450346184. DOI 10.1145/3010915.3010932. (Cited on pages 1, 34, and 38).

Claudia Núñez-Pacheco. Tangible Body Maps of Felt-Sensing Experience. In *Proceedings of the Fifteenth International Conference on Tangible, Embedded, and Embodied Interaction*, TEI '21, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450382137. DOI 10.1145/3430524.3442700. (Cited on pages 37, 49, and 101).

Claudia Núñez-Pacheco and Lian Loke. Getting Into Someone Else's Soul: Communicating Embodied Experience. *Journal of Digital Creativity*, 31, 2020. (Cited on pages 39 and 104).

Kristian Nymoen, Ståle A. Skogstad, and Alexander Refsum Jensenius. SoundSaber – A Motion Capture Instrument. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 312–315, Oslo, Norway, 2011. DOI 10.5281/zenodo.1178125. (Cited on page 54).

Kristian Nymoen, Mari Romarheim Haugen, and Alexander Refsum Jensenius. MuMYO — Evaluating and Exploring the MYO Armband for Musical Interaction. In Edgar Berdahl and Jesse Allison, editors, *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 215–218, Baton Rouge, Louisiana, USA, 2015. Louisiana State University. DOI 10.5281/zenodo.1179150. (Cited on page 56).

Rory O' Keeffe, Seyed Yahya Shirazi, Sarmad Mehrdad, Tyler Crosby, Aaron M. Johnson, and S. Farokh Atashzar. Perilaryngeal-Cranial Functional Muscle Network Differentiates Vocal Tasks: A Multi-Channel sEMG Approach. *IEEE Transactions on Biomedical Engineering*, pages 1–1, 2022. DOI 10.1109/tbme.2022.3175948. (Cited on pages 54 and 55).

Jessica O'Bryan. "We ARE our instrument!": Forming a singer identity. *Research Studies in Music Education*, 37(1):123–137, 2015. DOI 10.1177/1321103x15592831. (Cited on pages 13, 38, 42, 111, and 179).

Dora Ohrenstein. Insights into training aural and kinaesthetic awareness. *The National Association of Teachers of Singing Journal*, 60(1):29–35, 2003. (Cited on pages 34, 37, and 40).

Kotaro Okada and Shigeyuki Hirai. Interactive Sonification for Correction of Poor Sitting Posture While Working. In *Proc. International Workshop on Interactive Sonification, Stockholm, Sweden*, pages 101–107, 2019. DOI 10.5281/zenodo.3743339. (Cited on page 144).

Nicola Orio. A gesture interface controlled by the oral cavity. In *Proceedings of the International Computer Music Conference (ICMC)*, 1997. (Cited on page 21).

Maggie Orth. Maggie Orth, 2009. URL http://www.maggieorth.com/art_Instruments.html. (Cited on page 121).

Maggie Orth, J. R. Smith, E. R. Post, J. A. Strickon, and E. B. Cooper. Musical Jacket. In *ACM SIGGRAPH 98 Electronic Art and Animation Catalog*, SIGGRAPH '98, page 38, New York, NY, USA, 1998. Association for Computing Machinery. ISBN 1581130457. DOI 10.1145/281388.281456. (Cited on page 121).

Dan Overholt. The Overtone Fiddle: an Actuated Acoustic Instrument. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 30–33, Oslo, Norway, 2011. DOI 10.5281/zenodo.1178127. (Cited on page 20).

Friedemann Pabst and Johan Sundberg. Tracking multi-channel electroglottograph measurement of larynx height in singers. *Scandinavian Journal of Logopedics and Phoniatrics*, 18(4):143–152, 1993. DOI 10.3109/14015439309101360. (Cited on page 15).

Allan Paivio. Imagery and synchronic thinking. *Canadian Psychological Review/Psychologie canadienne*, 16(3):147–163, 1975. DOI 10.1037/h0081801. (Cited on page 26).

Allan Paivio. The Relationship Between Verbal And Perceptual Codes. In *Perceptual Coding*, pages 375–397. Elsevier, 1978. DOI 10.1016/b978-0-12-161908-4.50017-6. (Cited on page 26).

Pamela Z. Pamela Z, 2020. URL http://pamelaz.com/. (Cited on page 22).

Laurel S. Pardue and Andrew P. McPherson. Near-Field Optical Reflective Sensing for Bow Tracking. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 363–368, Daejeon, Republic of Korea, 2013. Graduate School of Culture Technology, KAIST. DOI 10.5281/zenodo.1178628. (Cited on page 20).

Laurel S. Pardue, Dongjuan Nian, Christopher Harte, and Andrew P. McPherson. Low-Latency Audio Pitch Tracking: A Multi-Modal Sensor-Assisted Approach. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 54–59, London, United Kingdom, 2014. Goldsmiths, University of London. DOI 10.5281/zenodo.1178899. (Cited on page 20).

Laurel S. Pardue, Christopher Hart, and Andrew P. McPherson. A Low-Cost Real-Time Tracking System for Violin. *Journal of New Music Research*, 44(4), 2015. DOI 10.1080/09298215.2015.1087575. (Cited on page 20).

Laurel S. Pardue, Kurijn Buys, Dan Overholt, Andrew P. McPherson, and Michael Edinger. Separating sound from source: sonic transformation of the violin through electrodynamic pickups and acoustic actuation. In Marcelo Queiroz and Anna Xambó Sedó, editors, *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 278–283, Porto Alegre, Brazil, 2019. UFRGS. DOI 10.5281/zenodo.3672958. (Cited on page 20).

Benjamin Parrell and Caroline A. Niziolek. Increased speech contrast induced by sensorimotor adaptation to a nonuniform auditory perturbation. *Journal of Neurophysiology*, 125(2):638–647, February 2021. DOI 10.1152/jn.00466.2020. (Cited on page 86).

Aniruddh D. Patel and Joseph Daniele. An empirical comparison of rhythm in language and music. *Cognition*, 87(1):B35–B45, 2002. DOI 10.1016/S0010-0277(02)00187-7. (Cited on page 72).

Nadine Pecenka and Peter E. Keller. Auditory Pitch Imagery and Its Relationship to Musical Synchronization. *Annals of the New York Academy of Sciences*, 1169(1):282–286, July 2009. DOI 10.1111/j.1749-6632.2009.04785.x. (Cited on page 81).

M. Pehlivan and İ. Denizoğlu. Laryngoaltimeter: A New Ambulatory Device for Laryngeal Height Control, Preliminary Results. *Journal of Voice*, 23(5):529–538, 2009. DOI 10.1016/j.jvoice.2008.01.004. (Cited on page 55).

M. Pehlivan, N. Yüceyar, C. Ertekin, G. Çelebi, M. Ertaş, T. Kalayci, and I. Aydoğdu. An electronic device measuring the frequency of spontaneous swallowing: Digital Phagometer. *Dysphagia*, 11 (4):259–264, 1996. DOI 10.1007/bf00265212. (Cited on page 55).

Hannah Perner-Wilson and Mika Satomi. *The Crying Dress*. 2012. URL http://www.kobakant.at/?p=222. (Cited on page 121).

Hannah Perner-Wilson, Leah Buechley, and Mika Satomi. Handcrafting Textile Interfaces from a Kit-of-No-Parts. In *Proceedings of the Fifth International Conference on Tangible, Embedded, and Embodied Interaction*, TEI '11, page 61–68, New York, NY, USA, 2010. Association for Computing Machinery. ISBN 9781450304788. DOI 10.1145/1935701.1935715. (Cited on page 120).

Ryan A. Peterson and Joseph E. Cavanaugh. Ordered quantile normalization: a semiparametric transformation built for the cross-validation era. *Journal of Applied Statistics*, 47(13-15):2312–2327, 2019. DOI 10.1080/02664763.2019.1630372. (Cited on page 74).

Claire Petitmengin. Describing one's subjective experience in the second person: An interview method for the science of consciousness. *Phenomenology and the Cognitive sciences*, 5(3):229–269, 2006. (Cited on pages 34, 39, 50, 51, 157, and 161).

Claire Petitmengin. Anchoring in lived experience as an act of resistance. *Constructivist Foundations*, 16(2):172–181, 2021. (Cited on page 50).

Claire Petitmengin. Micro-phenomenology / La Micro-Phénoménologie, 2022. URL https://www.microphenomenology.com/. (Cited on page 50).

Claire Petitmengin and Michel Bitbol. The Validity of First-Person Descriptions as Authenticity and Coherence. *Journal of Consciousness Studies*, 16(11-12):363–404, 2009. (Cited on pages 50 and 51).

Claire Petitmengin, Anne Remillieux, Béatrice Cahour, and Shirley Carter-Thomas. A gap in Nisbett and Wilson's findings? A first-person access to our cognitive processes. *Consciousness and Cognition*, 22(2):654–669, 2013. (Cited on page 50).

Claire Petitmengin, Anne Remillieux, and Camila Valenzuela-Moguillansky. Discovering the structures of lived experience. *Phenomenology and the Cognitive Sciences*, 18(4):691–730, 2018. DOI 10.1007/s11097-018-9597-4. (Cited on pages 50, 51, 157, and 161).

Viggo Pettersen and Rolf H. Westgaard. The association between upper trapezius activity and thorax movement in classical singing. *Journal of Voice*, 18(4):500–512, 2004. DOI 10.1016/j.jvoice.2003.11.001. (Cited on pages 17 and 54).

Viggo Pettersen and Rolf H. Westgaard. The Activity Patterns of Neck Muscles in Professional Classical Singing. *Journal of Voice*, 19(2):238–251, 2005. DOI 10.1016/j.jvoice.2004.02.006. (Cited on pages 17 and 54).

Peter Q. Pfordresher. *Sound and Action in Music Performance*. Academic Press, London, 2019. ISBN 978-0-12-809196-8. (Cited on page 27).

Peter Q. Pfordresher and Andrea R. Halpern. Auditory imagery and the poor-pitch singer. *Psychonomic Bulletin & Review*, 20(4):747–753, 2013. DOI 10.3758/s13423-013-0401-8. (Cited on pages 27, 60, 64, 71, and 79).

Peter Q. Pfordresher and James T. Mantell. Effects of altered auditory feedback across effector systems: Production of melodies by keyboard and singing. *Acta Psychologica*, 139(1):166–177, January 2012. DOI 10.1016/j.actpsy.2011.10.009. (Cited on pages 30 and 86).

Peter Q. Pfordresher and Caroline Palmer. Effects of delayed auditory feedback on timing of music performance. *Psychological Research*, 66(1):71–79, 2002. DOI 10.1007/s004260100075. (Cited on pages 64, 67, 70, 73, 81, 85, and 86).

Peter Q. Pfordresher, S. Brown, K. M. Meier, M. Belyk, and M. Liotti. Imprecise singing is widespread. *Journal of the Acoustical Society of America*, 128:2182–2190, 2010. (Cited on page 86).

Peter Q. Pfordresher, Andrea R. Halpern, and Emma B. Greenspon. A Mechanism for Sensorimotor Translation in Singing: The Multi-Modal Imagery Association (MMIA) Model. *Music Perception: An Interdisciplinary Journal*, 32(3):242–253, 2015. DOI 10.1525/mp.2015.32.3.242. (Cited on pages 27, 30, 60, 79, and 81).

Jon Pigrem. HOMEBREW'D Digital Musical Instruments, 2021. URL https://jonpigrem.com/assets/files/STUDY-GUIDE.pdf. (Cited on page 155).

Jonathan Pigrem, Andrew McPherson, Nick Bryan-Kinns, and Robert Jack. Sound → Object → Gesture: Physical Affordances of Virtual Materials. In *Proceedings of the 17th International Audio Mostly Conference*, AM '22, page 59–66, New York, NY, USA, 2022. Association for Computing Machinery. ISBN 9781450397018. DOI 10.1145/3561212.3561230. (Cited on page 155).

Andreas Pointner, Thomas Preindl, Sara Mlakar, Roland Aigner, and Michael Haller. Knitted RESi: A Highly Flexible, Force-Sensitive Knitted Textile Based on Resistive Yarns. In *ACM SIGGRAPH 2020 Emerging Technologies*, SIGGRAPH '20, New York, NY, USA, 2020. Association for Computing Machinery. ISBN 9781450379670. DOI 10.1145/3388534.3407292. (Cited on page 120).

Cornelius Pöpel and Roger B. Dannenberg. Audio Signal Driven Sound Synthesis. In *Proceedings of the Annual International Computer Music Conference (ICMC'05)*, 2005. (Cited on page 19).

Cornelius Pöpel, Jochen Feitsch, Marco Strobel, and Christian Geiger. Design and Evaluation of a Gesture Controlled Singing Voice Installation. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 359–362, London, United Kingdom, 2014. Goldsmiths, University of London. DOI 10.5281/zenodo.1178905. (Cited on page 21).

Irene Posch. Crafting Tools for Textile Electronic Making. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '17, page 409–412, New York, NY, USA, 2017a. Association for Computing Machinery. ISBN 9781450346566. DOI 10.1145/3027063.3052972. (Cited on page 131).

Irene Posch. Crafting Tools. *Interactions*, 24(2):78–81, 2017b. ISSN 1072-5520. DOI 10.1145/3038227. (Cited on page 131).

Irene Posch. Tooling Textile Electronics, 2019. URL http://www.ireneposch.net/tooling/. (Cited on page 131).

E. Rehmi Post and Maggie Orth. Smart fabric, or "wearable clothing". In *Digest of Papers. First International Symposium on Wearable Computers*, pages 167–168, 1997. DOI 10.1109/iswc.1997.629937. (Cited on page 120).

Søren Bolvig Poulsen and Ulla Thøgersen. Embodied design thinking: a phenomenological perspective. *CoDesign*, 7(1):29–44, 2011. DOI http://dx.doi.org/10.1080/15710882.2011.563313. (Cited on page 37).

Thomas Preindl, Cedric Honnet, Andreas Pointner, Roland Aigner, Joseph A. Paradiso, and Michael Haller. Sonoflex: Embroidered Speakers Without Permanent Magnets. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*, UIST '20, page 675–685, New York, NY, USA, 2020. Association for Computing Machinery. ISBN 9781450375146. DOI 10.1145/3379337.3415888. (Cited on pages 120 and 121).

Daniela Prem and Richard Parncutt. Corporality in the timbre vocabulary of professional female jazz vocalists. In *M. M. Marin, M. Knoche, & R. Parncutt (Eds.), Proceedings of the First International Conference of Students of Systematic Musicology (SysMus08), Graz, Austria*, pages 69–71, 2008. (Cited on pages 13, 30, 31, and 40).

G. Kim Prisk, J. Hammer, and Christopher J. L. Newth. Techniques for measurement of thoracoabdominal asynchrony. *Pediatric Pulmonology*, 34(6):462–472, 2002. DOI 10.1002/ppul.10204. (Cited on page 54).

Shannon Proksch, Daniel C. Comstock, Butovens Médé, Alexandria Pabst, and Ramesh Balasubramaniam. Motor and Predictive Processes in Auditory Beat and Rhythm Perception. *Frontiers in Human Neuroscience*, 14(578546):1–13, 2020. DOI 10.3389/fnhum.2020.578546. (Cited on page 81).

Mirjana Prpa. *Attending to inner self: Designing and unfolding breath-based VR experiences through micro-phenomenology.* Doctoral dissertation, School of Interactive Arts and Technology, Simon Fraser University, British Columbia, Canada, 2020. (Cited on pages 175 and 176).

Mirjana Prpa, Sarah Fdili-Alaoui, Thecla Schiphorst, and Philippe Pasquier. Articulating Experience: Reflections from Experts Applying Micro-Phenomenology to Design Research in HCI. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, page 1–14, New York, NY, USA, 2020. Association for Computing Machinery. ISBN 9781450367080. DOI 10.1145/3313831.3376664. (Cited on pages 50 and 106).

Tim A. Pruitt, Andrea R.. Halpern, and Peter Q. Pfordresher. Covert singing in anticipatory auditory imagery: Psychophysiology. *Psychophysiology*, 56(3):e13297, 2019. DOI 10.1111/psyp.13297. (Cited on page 64).

David W. Purcell and Kevin G. Munhall. Compensation following real-time manipulation of formants in isolated vowels. *The Journal of the Acoustical Society of America*, 119(4):2288–2297, April 2006. DOI 10.1121/1.2173514. (Cited on page 81).

Zenon Pylyshyn. Return of the mental image: are there really pictures in the brain? *Trends in Cognitive Sciences*, 7(3):113–118, 2003. DOI 10.1016/s1364-6613(03)00003-2. (Cited on page 26).

Zenon W. Pylyshyn. The imagery debate: Analogue media versus tacit knowledge. *Psychological Review*, 88(1):16–45, 1981. DOI 10.1037/0033-295x.88.1.16. (Cited on page 26).

Zenon W. Pylyshyn. Mental imagery: In search of a theory. *Behavioral and Brain Sciences*, 25(2): 157–182, 2002. DOI 10.1017/s0140525x02000043. (Cited on page 26).

Courtney N. Reed. As the Luthiers Do: Designing with a Living, Growing, Changing Body-Material. In *In 2023 CHI Workshop on Body X Materials*, New York, NY, USA, 2023. (Cited on page 33).

Courtney N. Reed and Andrew P. McPherson. Surface Electromyography for Direct Vocal Control. In Romain Michon and Franziska Schroeder, editors, *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 458–463, Birmingham, UK, 2020. Birmingham City University. DOI 10.5281/zenodo.4813475. (Cited on pages 21, 54, 126, 132, 134, 140, and 153).

Courtney N. Reed and Andrew P. McPherson. Surface Electromyography for Sensing Performance Intention and Musical Imagery in Vocalists. In *Proceedings of the Fifteenth International Conference on Tangible, Embedded, and Embodied Interaction*, TEI '21, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450382137. DOI 10.1145/3430524.3440641. (Cited on pages 2, 119, 126, 133, and 136).

Courtney N. Reed, Charlotte Nordmoen, Andrea Martelloni, Giacomo Lepri, Nicole Robson, Eevee Zayas-Garin, Kelsey Cotton, Lia Mice, and Andrew P. McPherson. Exploring Experiences with

New Musical Instruments through Micro-phenomenology. *International Conference on New Interfaces for Musical Expression*, 2022a. DOI 10.21428/92fbeb44.b304e4b1. (Cited on pages 50, 51, and 147).

Courtney N. Reed, Sophie Skach, Paul Strohmeier, and Andrew P. McPherson. Singing Knit: Soft Knit Biosensing for Augmenting Vocal Performances. In *Proceedings of Augmented Humans 2022 (AHs 2022), March 13–15, 2022*, AHs '22, New York, NY, USA, 2022b. Association for Computing Machinery. DOI 10.1145/3519391.3519412. (Cited on page 153).

Nathan Renney, Benedict Gaster, Tom Mitchell, and Harri Renney. Studying How Digital Luthiers Choose Their Tools. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI '22, New York, NY, USA, 2022. Association for Computing Machinery. ISBN 9781450391573. DOI 10.1145/3491102.3517656. (Cited on page 33).

Bruno H. Repp. Automaticity and voluntary control of phase correction following event onset shifts in sensorimotor synchronization. *Journal of Experimental Psychology: Human Perception and Performance*, 28(2):410–430, 2002. DOI 10.1037/0096-1523.28.2.410. (Cited on page 70).

Matthew Rodger, Paul Stapleton, Maarten van Walstijn, Miguel Ortiz, and Laurel S. Pardue. What Makes a Good Musical Instrument? A Matter of Processes, Ecologies and Specificities . In Romain Michon and Franziska Schroeder, editors, *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 405–410, Birmingham, UK, 2020. Birmingham City University. DOI 10.5281/zenodo.4813438. (Cited on page 33).

Friederike Roers, Dirk Mürbe, and Johan Sundberg. Predicted Singers' Vocal Fold Lengths and Voice Classification–A Study of X-Ray Morphological Measures. *Journal of Voice*, 23(4):408–413, 2009. DOI 10.1016/j.jvoice.2007.12.003. (Cited on page 16).

Nelson Roy, Julie Barkmeier-Kraemer, Tanya Eadie, M. Preeti Sivasankar, Daryush Mehta, Diane Paul, and Robert Hillman. Evidence-Based Clinical Voice Assessment: A Systematic Review. *American Journal of Speech-Language Pathology*, 22(2):212–226, 2013. DOI 10.1044/1058-0360(2012/12-0014). (Cited on page 54).

Robert S. Sackett. The Influence of Symbolic Rehearsal upon the Retention of a Maze Habit. *The Journal of General Psychology*, 10(2):376–398, 1934. DOI 10.1080/00221309.1934.9917742. (Cited on page 26).

Esther Salaman. *Unlocking your voice*. V. Gollancz, London, 1989. (Cited on pages 30 and 40).

Sauro Salomoni, Wolbert van den Hoorn, and Paul Hodges. Breathing and Singing: Objective Characterization of Breathing Patterns in Classical Singers. *PLoS ONE*, 11(5):1–18, 2016. DOI 10.1371/journal.pone.0155084. (Cited on pages 17 and 54).

Geise Santos, Johnty Wang, Carolina Brum, Marcelo M. Wanderley, Tiago Tavares, and Anderson Rocha. Comparative Latency Analysis of Optical and Inertial Motion Capture Systems for Gestural Analysis and Musical Performance. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Shanghai, China, 2021. DOI 10.21428/92fbeb44.51b1c3a1. (Cited on page 54).

Jayanthi Sasisekaran. Effects of Delayed Auditory Feedback on Speech Kinematics in Fluent Speakers. *Perceptual and Motor Skills*, 115(3):845–864, December 2012. DOI 10.2466/15.22.pms.115.6.845-864. (Cited on page 66).

Mika Satomi and Hannah Perner-Wilson. How To Get What You Want, 2007. URL https://www.kobakant.at/DIY/?p=379. (Cited on page 126).

Helmuth Schaffrath. *The Essen folksong collection in kern format.* [computer database]. Menlo Park, CA: Center for Computer Assisted Research in the Humanities, 1995. (Cited on page 72).

E. Glenn Schellenberg, Ania M. Krysciak, and R. Jane Campbell. Perceiving Emotion in Melody: Interactive Effects of Pitch and Rhythm. *Music Perception*, 18(2):155–171, 2000. DOI 10.2307/40285907. (Cited on page 31).

Thecla Schiphorst. Self-Evidence: Applying Somatic Connoisseurship to Experience Design. In *CHI '11 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '11, page 145–160, New York, NY, USA, 2011. Association for Computing Machinery. ISBN 9781450302685. DOI 10.1145/1979742.1979640. (Cited on pages 37 and 49).

Donald Schön. *The Reflective Practitioner: How Professionals Think In Action.* Basic Books, 1984. (Cited on page 53).

Karen Sell. *The Disciplines of Vocal Pedagogy: Towards an Holistic Approach.* Ashgate, Burlington, VT, 2005. (Cited on page 41).

Phoebe Sengers and Bill Gaver. Staying Open to Interpretation: Engaging Multiple Meanings in Design and Evaluation. DIS '06, page 99–108, New York, NY, USA, 2006. Association for Computing Machinery. ISBN 1595933670. DOI 10.1145/1142405.1142422. (Cited on page 58).

Olivier Senn, Dawn Rose, Toni Bechtold, Lorenz Kilchenmann, Florian Hoesl, Rafael Jerjen, Antonio Baldassarre, and Elena Alessandri. Preliminaries to a Psychological Model of Musical Groove. *Frontiers in Psychology*, 10(1228), 2019. DOI 10.3389/fpsyg.2019.01228. (Cited on page 82).

Ali Shafti, Roger B. Ribas Manero, Amanda M. Borg, Kaspar Althoefer, and Matthew J. Howard. Embroidered Electromyography: A Systematic Design Guide. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 25(9):1472–1480, 2017. DOI 10.1109/tnsre.2016.2633506. (Cited on page 120).

Ben Shneiderman and Pattie Maes. Direct Manipulation vs. Interface Agents. *Interactions*, 4(6): 42–61, Nov 1997a. ISSN 1072-5520. DOI 10.1145/267505.267514. (Cited on page 35).

Ben Shneiderman and Pattie Maes. Direct Manipulation vs. Interface Agents. *Interactions*, 4(6): 42–61, 1997b. ISSN 1072-5520. DOI 10.1145/267505.267514. (Cited on pages 18 and 188).

R. Shusterman. *Thinking through the Body: Essays in Somaesthetics.* Cambridge University Press, Cambridge, UK, 2012. (Cited on page 49).

Richard Shusterman. *Body Consciousness: A Philosophy of Mindfulness and Somaesthetics.* Cambridge University Press, 2008. (Cited on pages 37 and 49).

Kai Siedenburg, Ichiro Fujinaga, and Stephen McAdams. A Comparison of Approaches to Timbre Descriptors in Music Information Retrieval and Music Psychology. *Journal of New Music Research*, 45(1):27–41, 2016. DOI 10.1080/09298215.2015.1132737. (Cited on page 31).

Sophie Skach, Rebecca Stewart, and Patrick G. T. Healey. Smarty Pants: Exploring Textile Pressure Sensors in Trousers for Posture and Behaviour Classification. *Proceedings*, 32(1):19, 2019. DOI 10.3390/proceedings2019032019. (Cited on page 120).

Matthew Skarha, Vincent Cusson, Christian Frisson, and Marcelo M. Wanderley. Le Bâton: A Digital Musical Instrument Based on the Chaotic Triple Pendulum. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Shanghai, China, 2021. DOI 10.21428/92fbeb44.09ecc54d. (Cited on page 54).

Ståle A. Skogstad, Alexander Refsum Jensenius, and Kristian Nymoen. Using IR Optical Marker Based Motion Capture for Exploring Musical Interaction. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 407–410, Sydney, Australia, 2010. DOI 10.5281/zenodo.1177895. (Cited on page 54).

John A. Sloboda. Music Structure and Emotional Response: Some Empirical Findings. *Psychology of Music*, 19(2):110–120, 1991. DOI 10.1177/0305735691192002. (Cited on page 31).

John A. Sloboda and Andreas C. Lehmann. Tracking Performance Correlates of Changes in Perceived Intensity of Emotion During Different Interpretations of a Chopin Piano Prelude. *Music Perception: An Interdisciplinary Journal*, 19(1):87–120, 2001. DOI 10.1525/mp.2001.19.1.87. (Cited on page 31).

John A. Sloboda, Susan A. O'Neill, and Antonia Ivaldi. Functions of Music in Everyday Life: An Exploratory Study Using the Experience Sampling Method. *Musicae Scientiae*, 5(1):9–32, 2001. DOI 10.1177/102986490100500102. (Cited on page 59).

David C. Smith, Frank E. Ludolph, Charles H. Irby, and Jeff A. Johnson. The Desktop Metaphor as an Approach to User Interface Design (Panel Discussion). In *Proceedings of the 1985 ACM Annual Conference on The Range of Computing: Mid-80's Perspective*, ACM '85, page 548–549, New York, NY, USA, 1985. Association for Computing Machinery. ISBN 0897911709. DOI 10.1145/320435.320594. (Cited on pages 36 and 106).

J. David Smith, Margaret Wilson, and Daniel Reisberg. The role of subvocalization in auditory imagery. *Neuropsychologia*, 33(11):1433–1454, 1995. DOI 10.1016/0028-3932(95)00074-d. (Cited on page 57).

Jonathan De Souza. *Music at Hand*. Oxford University Press, 2017. DOI 10.1093/acprof:oso/9780190271114.001.0001. (Cited on pages 38 and 47).

Katta Spiel. The Bodies of TEI – Investigating Norms and Assumptions in the Design of Embodied Interaction. In *Proceedings of the Fifteenth International Conference on Tangible, Embedded, and Embodied Interaction*, TEI '21, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450382137. DOI 10.1145/3430524.3440651. (Cited on pages 38, 104, 112, 136, 149, and 175).

Daniel Akira Stadnicki. Nina Sun Eidsheim. 2015. Sensing Sound: Singing and Listening as Vibrational Practice. Durham, NC: Duke University Press, 270 pp. ISBN 978-0-8223-6046-9 (hardcover), ISBN 978-0-8223-6061-2 (paperback), ISBN 978-0-8223-7469-5 (e-book). *Intersections: Canadian Journal of Music*, 36(2):135, 2016. DOI 10.7202/1051607ar. (Cited on pages 34, 42, 184, and 189).

Alex Stahl and Patricia Clemens. Auditory Masquing : Wearable Sound Systems for Diegetic Character Voices. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 427–430, Sydney, Australia, 2010. DOI 10.5281/zenodo.1177899. (Cited on page 20).

Nikolaus Steinbeis, Stefan Koelsch, and John A. Sloboda. The Role of Harmonic Expectancy Violations in Musical Emotions: Evidence from Subjective, Physiological, and Neural Responses. *Massachusetts Institute of Technology Journal of Cognitive Neuroscience*, 18(8):1380–1393, 2006. DOI 10.1162/jocn.2006.18.8.1380. (Cited on page 31).

Stelarc. Keynote: Contemporary Chimeras – Creepy, Uncanny, and Contestable Bodies. New York, NY, USA, 2020. Association for Computing Machinery. ISBN 9781450376037. (Cited on pages 119 and 121).

Becky Stern. Becky Stern, 2009. URL https://beckystern.com/. (Cited on page 120).

Rebecca Stewart. Cords and Chords: Exploring the Role of E-Textiles in Computational Audio. *Front. ICT*, 6:2, 2019. (Cited on page 121).

Rebecca Stewart. embelashed, 2020. URL http://embelashed.org/. (Cited on page 130).

Brad H. Story. An overview of the physiology, physics and modeling of the sound source for vowels. *Acoustical Science and Technology*, 23(4):195–206, 2002. DOI 10.1250/ast.23.195. (Cited on page 13).

Anna Ståhl, Vasiliki Tsaknaki, and Madeline Balaam. Validity and Rigour in Soma Design-Sketching with the Soma. *ACM Trans. Comput.-Hum. Interact.*, 28(6), 2021. ISSN 1073-0516. DOI 10.1145/3470132. (Cited on pages 37 and 49).

Johan Sundberg. Perceptual Aspects of Singing. *Journal of Voice*, 8(2):106–122, 1994. DOI 10.1016/s0892-1997(05)80303-0. (Cited on pages 13 and 31).

Ivan E. Sutherland. *Sketchpad: A Man-Machine Graphical Communication System*. PhD thesis, Massachusetts Institute of Technology, 1963. (Cited on page 35).

Dag Svanæs. Kinaesthetic thinking: The tacit dimension of interaction design. *Computers in Human Behavior*, 13(4):443–463, 1997. (Cited on pages 1, 34, and 38).

Dag Svanæs. Interaction design for and with the lived body: Some implications of Merleau-Ponty's phenomenology. ACM Trans. *Comput.-Hum. Interact*, 20:1.8, 2013. DOI http://dx.doi.org/10.1145/2442106.2442114. (Cited on pages 36 and 100).

Dag Svanæs. Phenomenology through Design: A Tale of a Human Tail. In *Proceedings of the 2019 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, volume 19 of *CHI EA '19*, pages 4–9, 2019. DOI 10.1145/3290607.x. (Cited on pages 36, 40, 47, and 101).

Dag Svanaes and Martin Solheim. Wag Your Tail and Flap Your Ears: The Kinesthetic User Experience of Extending Your Body. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '16, page 3778–3779, New York, NY, USA, 2016. Association for Computing Machinery. ISBN 9781450340823. DOI 10.1145/2851581.2890268. (Cited on pages 34, 40, 46, and 101).

Cantor Digitalis: Performative Singing Synthesis. Christophe d'Alessandro and Boris Doval and Lionel Feugère and Sylvain Le Beux and Olivier Perrotin, 2016. URL https://cantordigitalis.limsi.fr/. (Cited on page 22).

Emi Tamaki, Takashi Miyaki, and Jun Rekimoto. PossessedHand: A Hand Gesture Manipulation System Using Electrical Stimuli. In *Proceedings of the 1st Augmented Human International Conference*, AH '10, New York, NY, USA, 2010. Association for Computing Machinery. ISBN 9781605588254. DOI 10.1145/1785455.1785457. (Cited on pages 119 and 121).

Atau Tanaka. Intention, Effort, and Restraint: The EMG in Musical Performance. *Leonardo: Transactions in Live Interfaces*, 43(8):298–299, 2015. DOI 10.1162/LEON_a_01018. (Cited on pages 57, 58, 140, and 184).

Atau Tanaka. Embodied Musical Interaction. In *New Directions in Music and Human-Computer Interaction*, pages 135–154. Springer International Publishing, 2019. DOI 10.1007/978-3-319-92069-6_9. (Cited on page 56).

Atau Tanaka and R. Benjamin Knapp. Multimodal Interaction in Music Using the Electromyogram and Relative Position Sensing. In Alexander Refsum Jensenius and Michael J Lyons, editors, *A NIME Reader: Fifteen Years of New Interfaces for Musical Expression*, pages 45–58. Springer, 2017. ISBN 978-3-319-47213-3. (Cited on pages 55, 56, and 140).

Atau Tanaka and Miguel A. Ortiz. Gestural Musical Performance with Physiological Sensors, Focusing on the Electromyogram. In M. Lesaffre, P-J. Maes, and M. Leman, editors, *The Routledge Companion to Embodied Music Interaction*, pages 422–430. Oxon, Routledge, 2017. ISBN 9781138657403. (Cited on pages 54, 56, 58, 117, and 140).

Adán L. Benito Temprano and Andrew P. McPherson. A TMR Angle Sensor for Gesture Acquisition and Disambiguation on the Electric Guitar. In *Audio Mostly 2021*, AM '21, page 256–263, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450385695. DOI 10.1145/3478384.3478427. (Cited on page 54).

Adán L. Benito Temprano and Rebecca Stewart. Bela E-textile Capelet, 2019. URL https://oshpark.com/shared_projects/y0oSowUt. (Cited on page 130).

Paul Tennent, Kristina Höök, Steve Benford, Vasiliki Tsaknaki, Anna Ståhl, Claudia Dauden Roquet, Charles Windlin, Pedro Sanches, Joe Marshall, Christine Li, Juan Pablo Martinez Avila, Miquel Alfaras, Muhammad Umair, and Feng Zhou. Articulating Soma Experiences Using Trajectories. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450380966. DOI 10.1145/3411764.3445482. (Cited on pages 37 and 49).

Neil Thapen. Pink Trombone, https://dood.al/pinktrombone, 2017. URL https://dood.al/pinktrombone/. (Cited on page 22).

Marian Theiss, Philipp M. Scholl, and Kristof Van Laerhoven. Predicting Grasps with a Wearable Inertial and EMG Sensing Unit for Low-Power Detection of In-Hand Objects. In *Proceedings of the 7th Augmented Human International Conference 2016*, AH '16, New York, NY, USA, 2016. Association for Computing Machinery. ISBN 9781450336802. DOI 10.1145/2875194.2875207. (Cited on pages 56, 119, and 140).

Jakob Tholander, Klas Karlgren, Robert Ramberg, and Per Sökjer. Where All the Interaction is: Sketching in Interaction Design as an Embodied Practice. In *Proceedings of the 7th ACM Conference on Designing Interactive Systems*, DIS '08, page 445–454, New York, NY, USA, 2008. Association for Computing Machinery. ISBN 9781605580029. DOI 10.1145/1394445.1394493. (Cited on page 107).

Monica Thomasson and Johan Sundberg. Lung volume levels in professional classical singing. *Logopedics Phoniatrics Vocology*, 22(2):61–70, 1997. DOI 10.3109/14015439709075316. (Cited on page 16).

Evan Thompson and Francisco J. Varela. Radical embodiment: Neural dynamics and consciousness. *Trends in Cognitive Sciences*, 5(10):418–425, 2001. DOI 10.1016/s1364-6613(00)01750-2. (Cited on pages 38, 50, and 103).

C. William Thorpe, Stephen J. Cala, Janice Chapman, and Pamela J. Davis. Patterns of breath support in projection of the singing voice. *Journal of Voice*, 15(1):86–104, 2001. DOI 10.1016/s0892-1997(01)00009-1. (Cited on page 17).

Lisa R. Trainor and Andrea Bundon. Developing the craft: reflexive accounts of doing reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health*, 13(5):705–726, November 2020. DOI 10.1080/2159676x.2020.1840423. (Cited on page 92).

William H. Trusheim. Audiation and Mental Imagery: Implications for Artistic Performance. *The Quarterly*, 2(1-2):138–147, 1991. (Cited on pages 25, 26, 27, 29, 38, and 39).

Vasiliki Tsaknaki, Kelsey Cotton, Pavel Karpashevich, and Pedro Sanches. “Feeling the Sensor Feeling You”: A Soma Design Exploration on Sensing Non-Habitual Breathing. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI ’21, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450380966. DOI 10.1145/3411764.3445628. (Cited on pages 22, 54, 136, and 184).

Kyriakos Tsoukalas, Joseph Kubalak, and Ivica Ico Bukvic. L2OrkMote: Reimagining a Low-Cost Wearable Controller for a Live Gesture-Centric Music Performance. In Thomas Martin Luke Dahl, Douglas Bowman, editor, *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 275–280, Blacksburg, Virginia, USA, 2018. Virginia Tech. ISBN 978-1-949373-99-8. DOI 10.5281/zenodo.1302581. (Cited on page 54).

Yasunori Tsubouchi and Kenji Suzuki. BioTones: A wearable device for EMG auditory biofeedback. In *Proc. International Conference of the IEEE Engineering in Medicine and Biology, Buenos Aires, Argentina*, pages 6543–6546, 2010. DOI 10.1109/iembs.2010.5627097. (Cited on pages 55, 56, 58, and 119).

Kai Tuuri, Jaana Parvainen, and Antti Pirhonen. Who Controls Who? Embodied Control Within Human–Technology Choreographies. *Interacting with Computers*, pages 1–18, 2017. DOI 10.1093/iwc/iww040. (Cited on pages 1, 32, 37, 38, 47, 48, and 185).

Akria Utsumi. The Role of Feature Emergence in Metaphor Appreciation. *Metaphor and Symbol*, 20(3):151–172, 2005. DOI 10.1207/s15327868ms2003_1. (Cited on page 36).

Camila Valenzuela-Moguillansky and Alejandra Vásquez-Rosati. An Analysis Procedure for the Micro-Phenomenological Interview. *Constructivist Foundations*, 14(2):123–145, 2019. URL https://constructivist.info/14/2/123.valenzuela. (Cited on page 157).

Francisco Varela. Neurophenomenology: A Methodological Remedy for the Hard Problem. *AVANT. Pismo Awangardy Filozoficzno-Naukowej*, (1):31–73, 2010. (Cited on pages 37 and 50).

Francisco J. Varela. *Principles of biological autonomy*. Elsevier (North Holland), Amsterdam, 1979. (Cited on pages 38 and 50).

Francisco J. Varela and Jonathan Shear, editors. *The View from Within: First-person Approaches to the Study of Consciousness.* Imprint Academic, London, 1999. (Cited on page 50).

Francisco J. Varela, Evan Thompson, and Eleanor Rosch. *The embodied mind: Cognitive science and human experience (revised edition).* MIT Press, Cambridge, MA, 1991. (Cited on pages 38, 46, and 100).

Martin R. Vasilev, Julie A. Kirkby, and Bernhard Angele. Auditory Distraction During Reading: A Bayesian Meta-Analysis of a Continuing Controversy. *Perspectives on Psychological Science*, 13 (5):567–597, 2018. DOI 10.1177/1745691617747398. (Cited on page 66).

Niina Venetjoki, A. Kaarlela-Tuomaala, Esko Keskinen, and Valtteri Hongisto. The effect of speech and speech intelligibility on task performance. *Ergonomics*, 49:1068–91, 2006. DOI 10.1080/00140130600679142. (Cited on page 66).

William Vennard. *Singing: The mechanism and the technic.* Carl Fisher, New York, NY, 1967. (Cited on page 41).

Peter-Paul Verbeek. COVER STORY: Beyond interaction. *Interactions*, 22(3):26–31, 2015. DOI 10.1145/2751314. (Cited on page 38).

Dianne Verdonk. Visible Excitation Methods: Energy and Expressiveness in Electronic Music Performance. In Edgar Berdahl and Jesse Allison, editors, *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 42–43, Baton Rouge, Louisiana, USA, 2015. Louisiana State University. DOI 10.5281/zenodo.1179188. (Cited on page 20).

Dianne Verdonk. Bellyhorn, 2022. URL https://dianneverdonk.com/bellyhorn/. (Cited on page 20).

Pierre Vermersch. *L'entretien d'explicitation.* ESF Paris, 2006. (Cited on page 50).

Pierre Vermersch. Describing the practice of introspection. *Journal of Consciousness Studies*, 16 (10–12):20–57, 2009. (Cited on page 50).

J. Arif Verner. Midi Guitar Synthesis: Yesterday, Today and Tomorrow. *Recording Magazine*, 8(9), 1995. (Cited on page 19).

Florian Vogt, Graeme Mccaig, Mir A. Ali, and Sidney S. Fels. Tongue 'n' Groove: An Ultrasound based Music Controller. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 181–185, Dublin, Ireland, 2002. DOI 10.5281/zenodo.1176468. (Cited on pages 21 and 54).

Fred A. Voorhorst, Helmut Krueger, and Martin Bichsel. Menus beyond the Desktop Metaphor. In *CHI '00 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '00, page 271–272, New York, NY, USA, 2000. Association for Computing Machinery. ISBN 1581132484. DOI 10.1145/633292.633450. (Cited on pages 35 and 102).

A. Vurma and J. Ross. Production and perception of musical intervals. *Music Perception: An Interdisciplinary Journal*, 23(4):331–344, 2006. DOI 10.1525/mp.2006.23.4.331. (Cited on page 86).

Peter Vuust, Martin J. Dietz, Maria Witek, and Morten L. Kringelbach. Now you hear it: a predictive coding model for understanding rhythmic incongruity. *Annals of the New York Academy of Science*, 1423:19–29, 2018. DOI 10.1111/nyas.13622. (Cited on page 82).

Caroline Wakefield, Dave Smith, Aidan Patrick Moran, and Paul Holmes. Functional equivalence or behavioural matching? A critical reflection on 15 years of research using the PETTLEP model of motor imagery. *International Review of Sport and Exercise Psychology*, 6(1):105–121, 2013. DOI 10.1080/1750984x.2012.724437. (Cited on pages 29 and 84).

Ron Wakkary, William Odom, Sabrina Hauser, Garnet Hertz, and Henry Lin. Material Speculation: Actual Artifacts for Critical Inquiry. volume 1, 08 2015. DOI 10.7146/aahcc.v1i1.21299. (Cited on pages 37 and 188).

Ron Wakkary, Doenja Oogjes, Henry W. J. Lin, and Sabrina Hauser. Philosophers Living with the Tilting Bowl. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, page 1–12, New York, NY, USA, 2018. Association for Computing Machinery. ISBN 9781450356206. DOI 10.1145/3173574.3173668. (Cited on pages 1, 32, 53, 103, and 149).

Marcelo M. Wanderley, David Birnbaum, Joseph Malloch, Elliot Sinyor, and Julien Boissinot. SensorWiki.org: A Collaborative Resource for Researchers and Interface Designers. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 180–183, Paris, France, 2006. DOI 10.5281/zenodo.1177015. (Cited on page 54).

Jinfeng Wang, Hairu Long, Saeid Soltanian, Peyman Servati, and Frank Ko. Electromechanical properties of knitted wearable sensors: part 1 – theory. *Textile Research Journal*, 84(1):3–15, 2014. DOI 10.1177/0040517513487789. (Cited on page 120).

Simon Waters. The entanglements which make instruments musical: Rediscovering sociality. *Journal of New Music Research*, 50(2):133–146, 2021. DOI 10.1080/09298215.2021.1899247. (Cited on page 33).

Alan H. D. Watson. The Biology of Musical Performance and Performance-related Injury. In *The Voice: Management and Problems*, pages 139–192, Lanham, Maryland, USA, 2009. Scarecrow Press. (Cited on pages 48 and 144).

Peter J. Watson and Thomas J. Hixon. Respiratory Kinematics in Classical (Opera) Singers. *Journal of Speech, Language, and Hearing Research*, 28(1):104–122, 1985. DOI 10.1044/jshr.2801.104. (Cited on page 17).

Peter J. Watson, Jeannette D. Hoit, Robert W. Lansing, and Thomas J. Hixon. Abdominal muscle activity during classical singing. *Journal of Voice*, 3(1):24–31, 1989. DOI 10.1016/s0892-1997(89)80118-3. (Cited on page 17).

C. Watts, J. Murphy, and K. Barnes-Burroughs. Pitch Matching Accuracy of Trained Singers, Untrained Subjects with Talented Singing Voices, and Untrained Subjects with Nontalented Singing Voices in Conditions of Varying Feedback. *Journal of Voice*, 17(2):185–194, 1994. DOI 10.1016/s0892-1997(03)00023-7. (Cited on page 86).

Hasini R. Weerathunge, Tiffany Voon, Monique Tardif, Dante Cilento, and Cara E. Stepp. Auditory and somatosensory feedback mechanisms of laryngeal and articulatory speech motor control. *Experimental Brain Research*, 240(7-8):2155–2173, June 2022. DOI 10.1007/s00221-022-06395-7. (Cited on page 81).

Rudolf Weiss, W.S Brown Jr., and Jack Moris. Singer's Formant in Sopranos: Fact or Fiction? *Journal of Voice*, 15(4):457–468, 2001. DOI 10.1016/s0892-1997(01)00046-7. (Cited on pages 15 and 17).

David Whetten. What Constitutes A Theoretical Contribution? *Academy of Management Review*, 14:490–495, 10 1989. DOI 10.2307/258554. (Cited on page 108).

Ravindra Wijesiriwardana, Tilak Dias, and S. Mukhopadhyay. Resistive fibre-meshed transducers. In *Proceedings of the Seventh IEEE International Symposium on Wearable Computers (2003)*, pages 200–209, 2003. DOI 10.1109/iswc.2003.1241412. (Cited on page 120).

Sarah E. Williams, Jennifer Cumming, Nikos Ntoumanis, Sanna M. Nordin-Bates, Richard Ramsey, and Craig Hall. Further validation and development of the Movement Imagery Questionnaire. *Journal of Sport & Exercise Psychology*, 34:621–646, 2012. DOI 10.1123/jsep.34.5.621. (Cited on pages 25, 60, 91, and 203).

Pat H. Wilson. *Does real-time visual feedback improve pitch accuracy in singing?* Masters thesis, University of Sydney, Sydney, Australia, 2006. (Cited on page 29).

Robert A. Wilson and Frank Keil. The shadows and shallows of explanation. *Minds and Machines*, 8(1):137–159, 1998. (Cited on pages 100 and 104).

Kathleen Wilson Spillane. Breath support directives used by singing teachers: A Delphi study. *The National Association of Teachers of Singing Journal*, 45(3):9–15, 1989. (Cited on page 40).

Jordan Wirfs-Brock, Alli Fam, Laura Devendorf, and Brian Keegan. Examining Narrative Sonification: Using First-Person Retrospection Methods to Translate Radio Production to Interaction Design. *ACM Trans. Comput.-Hum. Interact.*, 28(6), Nov 2021. ISSN 1073-0516. DOI 10.1145/3461762. (Cited on pages 50, 107, and 173).

Jordan Wirfs-Brock, Maxene Graze, Laura Devendorf, Audrey Desjardins, Visda Goudarzi, Mikhaila Friske, and Brian C Keegan. Sketching Across the Senses: Exploring Sensory Translation as a Generative Practice for Designing Data Representations. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI EA '22, New York, NY, USA, 2022. Association for Computing Machinery. ISBN 9781450391566. DOI 10.1145/3491101.3503712. (Cited on pages 107, 172, and 188).

Kieran Woodward, Eiman Kanjo, Samuel Burton, and Andreas Oikonomou. EmoEcho: A Tangible Interface to Convey and Communicate Emotions. UbiComp '18, page 746–749, New York, NY, USA, 2018. Association for Computing Machinery. ISBN 9781450359665. DOI 10.1145/3267305.3267705. (Cited on page 56).

Robert H. Woody. Musicians' Cognitive Processing of Imagery-Based Instructions for Expressive Performance. *Journal of Research in Music Education*, 54(2):125–137, 2006. DOI 10.2307/4101435. (Cited on page 27).

Robert H. Woody. Emotion, Imagery and Metaphor in the Acquisition of Musical Performance Skill. *Music Education Research*, 4(2):213–224, 2010. DOI 10.1080/1461380022000011920. (Cited on pages 38 and 39).

David J. Wright, Caroline J. Wakefield, and Dave Smith. Using PETTLEP imagery to improve music performance: A review. *Musicae Scientiae*, 18(4):448–463, 2014. DOI 10.1177/1029864914537668. (Cited on pages 29 and 84).

Xiao Xiao, Grégoire Locqueville, Christophe d'Alessandro, and Boris Doval. T-Voks: the Singing and Speaking Theremin. In Marcelo Queiroz and Anna Xambó Sedó, editors, *Proceedings of the*

International Conference on New Interfaces for Musical Expression, pages 110–115, Porto Alegre, Brazil, 2019. UFRGS. DOI 10.5281/zenodo.3672886. (Cited on page 22).

Simin Yang, Courtney N. Reed, Elaine Chew, and Mathieu Barthet. Examining Emotion Perception Agreement in Live Music Performance. *IEEE Transactions on Affective Computing*, pages 1–17, 2021. DOI 10.1109/TAFFC.2021.3093787. (Cited on page 48).

Tomoko Yonezawa, Takahiko Suzuki, Kenji Mase, and Kiyoshi Kogure. HandySinger : Expressive Singing Voice Morphing using Personified Hand-puppet Interface. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 121–126, Vancouver, BC, Canada, 2005. DOI 10.5281/zenodo.1176844. (Cited on page 22).

Fou Yoshimura and Jo Kazuhiro. A "voice" instrument based on vocal tract models by using soft material for a 3D printer and an electrolarynx. In Marcelo Queiroz and Anna Xambó Sedó, editors, *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 411–412, Porto Alegre, Brazil, 2019. UFRGS. DOI 10.5281/zenodo.3673015. (Cited on page 22).

Kenneth S. Zagacki, Renee Edwards, and James M. Honeycutt. The role of mental imagery and emotion in imagined interaction. *Communication Quarterly*, 40(1):56–68, 1992. DOI 10.1080/01463379209369820. (Cited on pages 25, 64, and 65).

Jean Mary Zarate and Robert J. Zatorre. Experience-dependent neural substrates involved in vocal pitch regulation during singing. *NeuroImage*, 40(4):1871–1887, 2008. DOI 10.1016/j.neuroimage.2008.01.026. (Cited on pages 27, 30, 67, and 80).

Robert J. Zatorre and Andrea R. Halpern. Mental Concerts: Musical Imagery and Auditory Cortex. *Neuron*, 47(1):9–12, 2005. DOI 10.1016/j.neuron.2005.06.013. (Cited on page 60).

Robert J. Zatorre, Joyce L. Chen, and Virginia B. Penhune. When the brain plays music: auditory–motor interactions in music perception and production. *Nature Reviews Neuroscience*, 8(7): 547–558, 2007. DOI 10.1038/nrn2152. (Cited on pages 25, 26, and 28).

Robert J. Zatorre, Andrea R. Halpern, and Marc Bouffard. Mental reversal of imagined melodies: A role for the posterior parietal cortex. *Journal of Cognitive Neuroscience*, 22:775–789, 2010. DOI 10.1162/jocn.2009.21239. (Cited on page 60).

Lawrence M. Zbikowski. Chapter 15. Music, language, and multimodal metaphor. In C. J. Forceville and E. Urios-Aparisi, editors, *Multimodal Metaphor*, volume 11: Applications of Cognitive Linguistics. De Gruyter, Inc., Berlin, 2009. DOI 10.1515/9783110215366.6.359. (Cited on page 39).

Zhaoyan Zhang. Mechanics of human voice production and control. *The Journal of the Acoustical Society of America*, 140(4):2614–2635, 2016a. DOI 10.1121/1.4964509. (Cited on page 13).

Zhaoyan Zhang. Cause-effect relationship between vocal fold physiology and voice production in a three-dimensional phonation model. *The Journal of the Acoustical Society of America*, 139(4): 1493–1507, 2016b. DOI 10.1121/1.4944754. (Cited on page 16).

Jack Zhao and Andrew Vande Moere. Embodiment in Data Sculpture: A Model of the Physical Visualization of Information. In *Proceedings of the 3rd International Conference on Digital Interactive Media in Entertainment and Arts*, DIMEA '08, page 343–350, New York, NY, USA, 2008. Association for Computing Machinery. ISBN 9781605582481. DOI 10.1145/1413634.1413696. (Cited on page 107).

John Zimmerman, Jodi Forlizzi, and Shelley Evenson. Research through Design as a Method for Interaction Design Research in HCI. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '07, page 493–502, New York, NY, USA, 2007. Association for Computing Machinery. ISBN 9781595935939. DOI 10.1145/1240624.1240704. (Cited on pages 53 and 149).

G. Zimmermann, C. Brown, J.A.S. Kelso, Richard Hurtig, and K. Forrest. The association between acoustic and articulatory events in a delayed auditory feedback paradigm. *Journal of Phonetics*, 16(4):437–451, 1988. (Cited on pages 67 and 83).