© 2017 James Andrew Babcock
To my Mom
ACKNOWLEDGMENTS

I would first and foremost like to thank my chair Paul Koonce for his dedication and investment in my growth as a composer, scholar, teacher and overall person. Dr. Koonce’s guidance has significantly boosted my creativity and scholarship in unexpected ways, giving me many tools to draw from in the future. Thank you to Paul Richards for instilling confidence in my work and giving me the freedom to find my own compositional voice. Thanks to James Paul Sain, whose teaching, friendliness, humor, and overall guidance made my doctoral degree easier to navigate and helped me learn many things about myself. Thank you to Jennifer Thomas, whose passion for meticulous research and writing has helped guide my dissertation project every step of the way. Thanks to Jack Stenner for opening me up to many new ways of making and conceptualizing art. Also, I would like to thank my previous composition teachers—Cort Lippe, Sam Pellman, Jeffery Stadelman—your wisdom continues to crop up unexpectedly and help guide my decisions both in and out of music.

I would also like to thank all of my UF peers who accompanied me along the way, particularly Travis Garrison, Adam Scott Neal, Benjamin O’Brien, Sean Peuquet, Michael Polo, Thomas Royal, Rob Seaback, Michael Smith, and Chet Udell. Your friendship and unique perspectives have made my time at UF most enjoyable and rewarding. In addition, I would like to thank the Stewart, Brown, Williams, and Horner families for all their support through the years and encouraging me to do what I love. Specifically, thank you to Miski Horner for sticking with me through this entire process of strict routine, bodily injury, and philosophical babble—your patience and warm spirit I am forever grateful for. Finally, I would like to thank my mother, whose unconditional love, thoughtfulness, kindness, and ability to listen are qualities I will always aspire to.
# TABLE OF CONTENTS

**ACKNOWLEDGMENTS**......................................................................................................................... 4

**LIST OF FIGURES**........................................................................................................................... 7

**LIST OF OBJECTS** ............................................................................................................................ 8

**ABSTRACT** ......................................................................................................................................... 9

**CHAPTER**

1 **INTRODUCTION** ........................................................................................................................... 11

2 **THE PARADOX OF VOICE** ........................................................................................................... 13

   - Defining Voice .......................................................................................................................... 13
   - Myths of Voice .......................................................................................................................... 15
   - Technological Ventriloquisms and Virtual Effects .................................................................. 19
   - Voice and Musical Persona ....................................................................................................... 21
   - Linguistics of the Voice ............................................................................................................. 23
   - Nonsense and Beyond Singing .................................................................................................... 27
   - Vocality ..................................................................................................................................... 32
   - Artificial Vocality ..................................................................................................................... 37

3 **THE LISTENING VOICE** ............................................................................................................... 41

   - Mimesis, Mimicry, Imitation ................................................................................................. 41
   - Ecological Perception and Embodied Music Cognition ......................................................... 43
   - Metaphor and Vocal Embodiment .......................................................................................... 47
   - Motor-Mimetic Participation ..................................................................................................... 49
   - Subvocalization and Timbre Appraisal ..................................................................................... 53
   - Expression, Exertion, Emotion ................................................................................................. 59
   - Mishearing and Vocal Reflexivity ............................................................................................. 71
   - Cross-Modal Vocality .............................................................................................................. 74
   - Hyperproximity and Oral Entanglements .............................................................................. 77

4 **TOWARD A THEORY OF HYPERVOCALITY** .............................................................................. 86

   - Dissolution of Voice and Re-enchantment ............................................................................. 86
   - Attention, Distraction, Counterpoint ...................................................................................... 90
   - Spatial Counterpoint in *Walkabout* ...................................................................................... 94
   - Holistic and Biomorphic Space ............................................................................................. 98
   - Biomorphic Voice-Bodies in *Red Snow* ............................................................................ 101
   - A Theory of Hypervocality ..................................................................................................... 105
5 REVEALING VOCAL ENTANGLEMENTS: CONSTRUCTING THE
HYPERVOCAL ........................................................................................................107

*Reconstruction: Voices Through Mimetic Reciprocity* ..............................................109
*Vocal Silhouettes: Short of Touch and Chatterbox* ..................................................114
*Ventriloquy: Composer as Ventriloquist and Listener as Dummy, or Listener as
Ventriloquist and Composer as Dummy? .................................................................118
  Unveiling the Voices ..............................................................................................121
  The Ventriloquial Reply .........................................................................................125
Final Remarks ...........................................................................................................127

LIST OF REFERENCES .................................................................................................130

BIOGRAPHICAL SKETCH ..........................................................................................147
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3-1</td>
<td>Sonogram and waveform of Skrillex's <em>Scary Monsters and Nice Sprites</em>, 0:41–0:51.</td>
<td>66</td>
</tr>
<tr>
<td>3-2</td>
<td>Sonogram and waveform of Merzbow's <em>Uzumushi</em>, 0:00–2:40.</td>
<td>68</td>
</tr>
<tr>
<td>3-3</td>
<td>Roughness in Merzbow's <em>Uzumushi</em>, 1:51–1:53.</td>
<td>69</td>
</tr>
<tr>
<td>3-4</td>
<td>Kiki and Bouba (adapted from Ramachandran and Hubbard, 2001).</td>
<td>77</td>
</tr>
<tr>
<td>3-5</td>
<td>Sonogram of glissandi and roughness in Autechre's <em>M39 Diffain</em>, 4:10–4:13.</td>
<td>81</td>
</tr>
<tr>
<td>4-1</td>
<td>Sonogram of high-friction, hyperproximate voice in <em>Walkabout</em>, 2:08–2:13.</td>
<td>96</td>
</tr>
<tr>
<td>4-3</td>
<td>Sonogram of wide voice-body containing low formants in <em>Red Snow</em>, 2:29–2:45.</td>
<td>103</td>
</tr>
<tr>
<td>5-1</td>
<td>Sonogram of cartoon-like voice glissando in <em>Reconstruction</em>, 3:16–3:19.</td>
<td>114</td>
</tr>
<tr>
<td>5-2</td>
<td>Image of Rubin vase (image taken from <a href="https://commons.wikimedia.org/wiki/File:Rubin2.jpg">https://commons.wikimedia.org/wiki/File:Rubin2.jpg</a>).</td>
<td>115</td>
</tr>
<tr>
<td>5-3</td>
<td>Sonogram of barking piano interjection in <em>Chatterbox</em>, 2:04–2:05.</td>
<td>117</td>
</tr>
<tr>
<td>5-4</td>
<td>Sonogram of rapid prosody in <em>Short of Touch</em>, 1:29–1:33.</td>
<td>118</td>
</tr>
<tr>
<td>Object</td>
<td>page</td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>---------------</td>
<td></td>
</tr>
<tr>
<td>5-1 Reconstruction: electroacoustic composition soundfile (.wav file 117MB)</td>
<td>110</td>
<td></td>
</tr>
<tr>
<td>5-2 Short of Touch: electroacoustic composition soundfile (.wav file 97MB)</td>
<td>114</td>
<td></td>
</tr>
<tr>
<td>5-3 Chatterbox: electroacoustic composition soundfile (.wav file 36MB)</td>
<td>115</td>
<td></td>
</tr>
<tr>
<td>5-4 Ventriloquy: electroacoustic composition soundfile (.wav file 156MB)</td>
<td>119</td>
<td></td>
</tr>
</tbody>
</table>
The voice plays a crucial, performative role in understanding timbre and navigating musical experience. As a mediator of desire, empathy, affect, exertion, and cross-modal perception, our voice allows us to imagine what it is like to be something else, opening an intersubjective dialog between our body and the objects we entangle ourselves with. As a consequence of our vocality’s multisensorial nature, the electroacoustic medium affords the composer a platform to reveal the vocal continuum in sound, which disrupts the listener’s subjective grounding through immersive experiences that conjoin body and timbre, perception and action, and self and other.

Proposing a theory of hypervocality, this investigation first expands the concept of vocality to all sound and its cross-modal implications, followed by its application to electroacoustic music, where it exists as an amalgamation of interdependent components realized through spatial counterpoint. The theory develops out of a transdisciplinary approach, using philosophy, linguistics, psychoacoustics, psychoanalysis, psychology, neuroscience, musicology, critical theory, media studies, art history, and sound studies; its goal is to enrich the composing and listening process, while working to better understand how and why we make meaning out of music and
sound. In illustration of these new forms of musical expression, which move beyond song, language, and the human figure, the document includes analyses of various electroacoustic works, the last of which focus on four of the author’s works: 

Reconstruction, Chatterbox, Short of Touch, and Ventriloquy.
CHAPTER 1
INTRODUCTION

Studies in vocality have steadily grown from a trickle in 1987 until the present, with numerous, current investigations examining the relationship between audio technology and human vocal utterance. Although scholars, including musicologists, music theorists, and digital media specialists, probe such issues as vocal performance and voice-based audio reproduction, most studies fail to address vocal embodiment from a listener’s perspective. This dissertation examines the voice’s role in mediating musical experience, using electroacoustic music to understand the workings of our imagination and our needs and desires, both in relation to music and to other humans, animals, and objects. The document proposes a theory of hypervocality, first expanding the concept of vocality to all sound and its cross-modal implications, followed by its application to electroacoustic music, where it exists as an amalgamation of interdependent components realized through spatial counterpoint. The theory develops out of a transdisciplinary approach, using philosophy, linguistics, psychoacoustics, psychoanalysis, psychology, neuroscience, musicology, critical theory, media studies, art history, and sound studies; its goal is to enrich the composing and listening process, while working to better understand how and why we make meaning out of music and sound.

Chapter 2 looks at the contradictions in defining voice according to its relationship to myth, language and subjectivity, bodily excess, the history of Western music, and technology. The voice’s paradoxes open electroacoustic music to a more symbiotic vocality: one that disseminates the heterogeneity of our identities. Chapter 3 examines the voice’s role in music listening perception. Using embodied cognition to
understand how our voice mediates empathy, affect, exertion, expressivity, and imagination, the chapter proposes that our vocality emanates from something we do while listening to music; we understand timbre not as an objective acoustic fact, but through a performative activity involving vocal participation, providing us with a multisensorial experience. Chapter 4 looks at detrimental effects of our technology, as well as our evolving attention spans, expanding the notion of vocality to an orchestration or mesh of interdependent parts. Paul Koonce's *Walkabout* and Natasha Barrett's *Red Snow* illustrate ways composers can employ distraction and suggest sonic ambiguity, creating more immersive, information-rich, and elusive experiences for the listening subject. The chapter concludes with a proposed theory of *hypervocality*. Chapter 5 explores the self-reflexive dialogue of the electroacoustic composer, offering insights to four of the author’s works: *Reconstruction*, *Chatterbox*, *Short of Touch*, and *Ventriloquy*. The first three works act as building blocks to *Ventriloquy*, which serves as the centerpiece composition for this dissertation and the various topics discussed throughout.
“Voice” is a loaded term. Its multiple definitions and connotations make it difficult to pin down. Our voice is one of our most familiar companions, yet in some ways is completely foreign to us. I have a voice, and it comes from my body, but my body is also a place where other voices come in and intermingle, obfuscating boundaries between inside and outside. Unlike an instrument, “which can be locked up and put away after use,” we carry our voice wherever we go. Of course, we know voices as the discrete, contrapuntal lines in a polyphonic composition. But voice also serves as a connotation of individual autonomy and uniqueness. ‘Use your voice!,’ we are often told when another finds us being timid or jumbling our speech. Such a metaphor, which stresses the importance of articulating clear and authoritative values through verbal speech, often undercuts the resonant material and multisensorial aspects of a voice.

Voice is perhaps the most important sound we encounter on a daily basis, because it points toward meaning: “it is as if there is an arrow in it which raises the expectation of meaning.” The presence of a voice “instantly sets up a hierarchy of

---


3 Having ‘a voice’ spreads beyond individual autonomy and into various groups of people: in our democratic society, there is “the people’s voice.” Likewise, for many oppressed groups, voice serves as indicator of empowerment. Last, voice is a place where sexual delineation regularly takes place.

perception,” structuring the sonic space it lies in.\(^5\) In our attempts to assess our surroundings, we go searching for it:

Present somebody with a complex or unintelligible arrangement of dots or strips, and they are likely to begin by trying to resolve it into a face. Presented with a similarly confused set of sounds, human beings seem equivalently impelled to wonder at the outset if there is a voice to be made out in them.\(^6\)

Voices in our environment can offer indications of safety and well-being or aggression and danger. When I hear a sudden shout behind me, my body turns instantly to locate and identify the source in my visual field. Voices are also “necessarily erotic,” for “there is no human voice which is not an object of desire—or of repulsion.”\(^7\)

Acoustically speaking, voice can be defined narrowly in physiological terms as “sound produced by vibration of the vocal folds.”\(^8\) Another commonly accepted definition, which highlights a more detailed act of articulation, is “the acoustic results of the coordinated action of the respiratory system, tongue, jaw, lips, and soft palate.”\(^9\)

However, most aspects of the body, including posture and musculoskeletal movement, affect the voice’s output. On top of this, a voice never resounds in isolation—it always exists in an entanglement with other objects and their interiors, such as other bodies,


\(^6\) Steven Connor, Beyond Words: Sobs, Hums, Stutters and Other Vocalizations (Chicago: The University of Chicago Press, 2014), 7.


\(^9\) ibid., 6.
smaller objects, and objects that enclose us (which we typically call “space”). A voice, or any sound for that matter, is always multiplicitous, always a resonant convocation of indefinite objects. Regardless, both scientific definitions focus on the acoustic signal, excluding “the perceptual impression that occurs as a result.” This is problematic because “the voice that is produced cannot be separated from the act of listening that provides the context for production.” Moreover, what we can measure in the acoustic world, using various digital tools, does not always correspond to our subjective perception; “[a]uditory perception more often than not depends on factors external to—sometimes in contradiction to—the acoustic stimulus that provokes it.” In this sense, a voice “exists in a very real sense only in the mind of the listener, not in the objective world.”

We can see that trying to define voice, with its various manifestations, is a difficult task, as only contradictions emerge the deeper one looks. There are several topics and terms, however, that illuminate how voice might function in the context of electroacoustic listening and composition, as well as the sonic environments we inhabit.

**Myths of Voice**

Prescientific definitions of voice demonstrate its elusive nature and the power of unlocated sound. “Western mythologies ascribe a beguiling or terrifying role to any
awesome, incorporeal voice.” Barry Blesser and Ruth Salter note that “[f]or our ancestors, voices included those of powerful spirits, the sounds from large acoustic structures and objects like caves and mountains, from thunder, and from wind itself.”

Without science, “preliterate cultures used religion, along with its associated art and myths, to explain a wide range of otherwise inexplicable events, to invest the unknowable with some kind of causal meaning.” Blesser and Salter comment on this further:

The voice of a resonant cave is more than a literary metaphor. You would have felt the cave was alive when it acknowledged your presence by responding to your footsteps with a voice of its own. From an experiential perspective, a cave is something that has a voice and sounds alive. Only from a modern, scientific perspective is it simply a natural hollow with sonic reflections and resonances.

Although vocalizing in a large reverberant cavern or church today may sound eerie and mysterious, we typically do not ascribe the space as having some form of external agency: we hear it as a multiplication and diffusion of one voice. This highlights the voice as a “built-in biological flashlight” that can “illuminate the environment” and attend to “how the environment changes the sound,” producing “an aural image of an acoustic space.” This process is known as echolocation.

---


17 ibid., 72.

18 ibid., 71.

19 ibid., 343.
Primitive cultures’ tendency towards animism fueled the power of the unlocated, omnipotent voice. Such voices show up in “the cacophonous frenzies of Dionysus,” “the joyous noisiness of Hindu rituals,” and the means to “bring about conditions of trance.”

Michel Chion further adds to the list:

This interdiction against looking, which transforms the Master, God, or Spirit into an *acousmatic* voice, permeates a great number of religious traditions, most notably Islam and Judaism. We find it also in the physical setup of Freudian analysis: the patient on the couch should not see the analyst, who does not *look* at him. And finally we find it in the cinema, where the voice of the acousmatic master who hides behind a door, a curtain or offscreen, is at play in some key films.

In the 1950s, French composer Pierre Schaeffer dug up the obscure term *acousmatic* to describe “sound heard without seeing its source.” The term originated from the ancient group the Akousmatikoi: a gathering of students who listened to Pythagoras teach behind a veil:

> [T]he cults of disciples surrounding the great mathematician listened to him teach from behind a curtained *arras*, the better to focus their thoughts on the content of his speech, not be distracted by his body or his gestures. Power accrues to the utterance and not the person; words are also freer, something more than the speech of a human being; they point not merely to Pythagoras and his earthly form, but become symbols that detach entirely from an agent of utterance to take on other meanings.

---


For several psychoanalysts, myths of vocal omnipotence begin with the very first acousmatic voice we experience: the mother’s. Didier Anzieu’s “sound envelope” and “audio-phonic skin” posit an imaginary cavern enclosed by the protective touch of the mother’s voice. Such a voice buffers against the undifferentiated “rumblings, echoes and resonances” that shock and disturb the infant and instigates the infant’s first buccal, audio-tactile explorations through imitation. In addition, the mother’s voice and face play a constant game of hide and seek with the child, creating a dialectic between presence and absence, appearance and disappearance, that the infant, driven by the demand for other, learns to master through the use of his own voice. Although considered a retrospective fantasy, the infant’s relationship to the unseen voice provides a primary negotiation between self and other and opens the idea of sound as not something that we merely hear with our ears, but something that touches us and that we touch back through the “acoustic mirror” of another’s voice.

---


Technological Ventriloquisms and Virtual Effects

Today, acousmatic voices constantly permeate our surroundings—whether riding a bus, traversing an airport, listening to a phone’s operating system answer questions, or working out at the gym—“[v]ocal sound . . . is no longer tied to a hole in the head but is free to issue from anywhere in the landscape.”28 We live in a world full of voices without bodies, where loudspeakers talk to us constantly. Such “technological ventriloquisms”—the telephone, radio, phonograph, and various digital devices today—actualize this forgotten fantasy of the omnipresent voice.29 Yet, despite our technologies’ abilities to remove the voice from its source, send it over long distances, and replay it indefinitely, “there is no disembodied voice—no voice that does not have somebody, something of somebody’s body, in it.”30 Connor’s claim seems to contradict the ubiquitous presence of disembodied voices, but it stems from the fact that we do not allow the sonically incongruous to occur: the listener’s “perceptual mechanism has been biologically conditioned to explain causation and find meaning.”31 Our need to create coherence and visually locate sound, means that we project bodies onto voices that we cannot see. For every voice, there is a second body, a body double, that competes with, contradicts, supersedes, or even reshapes the body that actually emits the sound. Gilles Deleuze echoes this interdependence between actual and virtual perception:


30 Connor, Beyond Words, 17.

An actual perception has its own memory as a sort of immediate, consecutive or even simultaneous double. . . . Memory is not an actual image which forms after the object has been perceived, but a virtual image coexisting with the actual perception of the object. . . . Purely actual objects do not exist. Every actual surrounds itself with a cloud of virtual images.32

This doubling effect leads Connor to his simple concept, the “vocalic-body:” bodies produce voices, and voices themselves produce bodies.33 While this appears to make the voice subservient to the body, it is phenomenologically the opposite: a voice “seems to colour and model its container.”34 Moreover, we do not need the mediation of voice technologies to witness these uncanny, virtual effects. For example, upon first meeting someone, we often find something out of sync between that person’s voice and her appearance before we become accustomed to it. We may think to ourselves, ‘that voice cannot possibly be coming from her’:

The voice acquires a spectral autonomy, it never quite belongs to the body we see, so that even when we see a living person talking, there is always some degree of ventriloquism at work: it is as if the speaker’s own voice hollows him out and in a sense speaks "by itself," through him.35

---

32 Gilles Deleuze and Claire Parnet, "The Actual and the Virtual," trans. Eliot Ross Albert, in Dialogues II (New York: Columbia University Press, 2007), 148-150. Deleuze’s ideas stem, in part, from Immanuel Kant’s theory idea that imagination is a synthetic lie between its productive and reproductive aspects: “The power of imagination (facultas imaginandi), as a faculty of intuition without the presence of the object, is either productive, that is, a faculty of the original presentation of the object (exhibitio originaria), which thus precedes experience; or reproductive, a faculty of the derivative presentation of the object (exhibitio derivativa), which brings back to mind an empirical intuition that it had previously.—Pure intuitions of space and time belong to the productive faculty; all others presuppose empirical intuition, which, when is connected with the concept of the object and thus becomes empirical cognition, is called experience.—The power of imagination, in so far as it also produces images involuntary, is called fantasy.” See Immanuel Kant, Anthropology from a Pragmatic Point of View, trans. Robert B. Louden, ed. Robert B. Louden (Cambridge: Cambridge University Press, 2006), 60 (emphasis in original).

33 Connor, Dumbstruck, 35.

34 ibid., 35.

Every emission of voice comes from an undisclosed interior, making it impossible to see where the voice stems from. In this sense, the voice becomes an effect liberated from its source. This leads Dolar to reach a paradoxical conclusion about the voice: “there is no such thing as disacousmatization. The source of the voice . . . cannot possibly match what we cannot see.”  

Add Deleuze and Guattari, “[t]he voice is far ahead of the face, very far ahead.”

These insights demonstrate that a voice’s effect is much more plastic than we initially might consider. In the electroacoustic realm, where unseen voices can run amok, our virtual engagement with sound brings forth many questions: what types of voice-bodies do we imagine? Is a voice clothed in everyday language just a ‘talking head’? How do we cope with voices that arise from contortions and movements that exceed the capabilities of the human body? What if we think it is a voice, but in fact it is not?

**Voice and Musical Persona**

In art, literature, and music, the ‘authorial voice’ often suggests a certain persona(e) that emerges in the reception of a creative work. Edward Cone’s *The Composer’s Voice* proposes that a composition has “the constant presence of a single musical persona.” The composer’s voice is like a “virtual author:” “a creative mind whom we assume to have made the work as a whole” in which “all its utterances are

---


heard as emanating from a single composing subject.” Although Cone’s focus is on acoustic works, his theory does touch on the electroacoustic realm:

The electronic persona is unitary, and it is uniquely embedded and embodied by a single performance (or execution) of its music. . . . The presence of a single commanding figure who sums up and represents the complete musical persona is especially helpful to the average listener in his effort to make contact with that persona.  

Cone’s stance is analogous to literary theory’s concept of “omniscient narration:” the idea that a novel is controlled by a single, all-knowing, all-seeing storyteller.  

Mikhail Bakhtin’s concept of “heteroglossia,” one in which novels contain many different voices, which may compete or overlap one another, offers an antidote to the unitary, authoritative voice. Bakhtin shows us that the reader, or in our case, the listener, finds authority where it is most appropriate for herself, often in defiance of the creator’s intentions.  

Drawing off Bakhtin’s theory, Abbate offers a counter to Cone’s outlook:

To Cone’s monologic and controlling “composer’s voice,” I prefer an aural vision of music animated by multiple, decentered voices localized in several invisible bodies. This vision proposes an interpretation of music shaped by prosopopoeia, the rhetorical figure that grants human presence to nonhuman objects or phenomena, and one that traditionally entails a strongly visual fantasy in which we imagine faces and eyes upon nonhuman forms. By speaking of music’s “voices,” we reconstrue the trope in an auralized form, in which imagining human faces or bodies means figuring forth sounds from those faces’ lips and throats.

---


40 Cone, The Composer’s Voice, 111-139.

41 Paul Cobley, Narrative (London: Routledge, 2001), 100-114.


43 Abbate, Unsung Voices, 13.
Abbate offers a phantasmagorical account of the listening experience: one in which extrinsic references reach well beyond the 19th century operatic score and stage that she analyzes. How would Abbate’s aural vision sound in the electroacoustic realm, where the lack of human performers can facilitate vivid, visual imagery in the listener’s imagination? Moreover, what types of human faces, eyes, and bodies might we envisage, and is this vision an anthropomorphic trap? How would one create a music that rejuvenates preliterate cultures’ “tendencies to animism,” where sounds appear to speak?44

Linguistics of the Voice

For many, voice is synonymous with verbal speech. Language manifests itself everywhere: from the silent voice of consciousness that narrates our subjectivity, to the voice of the superego that tells us to behave and obey the written law; from the internet voices we devour, to the voices of face to face conversations.45 Whereas John Cage teaches us that music is everywhere in the sounds around us, William Burroughs suggests the same is true of language:

The word is now a virus. . . . It is now a parasitic organism that invades and damages the central nervous system. Modern man has lost the option of silence. Try halting your sub vocal speech. Try to achieve even ten seconds of inner silence. You will encounter a resisting organism that forces you to talk. That organism is the word.46


45 Following Freud, Dolar describes the superego as “an internalization of the law” realized through the subject’s inner voice. See Dolar, A Voice and Nothing More, 40.

The infant’s acquisition of verbal speech is met with a simultaneous “phonic amnesia” in which the finite, sonic palette of language replaces the child’s “infinite capacity for undifferentiated articulation.”\(^{47}\) Roman Jakobson details this process further:

A child, during his babbling period, can accumulate articulations which are never found within a single language or even a group of languages—consonants of any place of articulation, palatalized and rounded consonants, sibilants, affricates, clicks, complex vowels, diphthongs, etc. . . . As all observers acknowledge with great surprise, the child then loses nearly all of his ability to produce sounds in passing over from the pre-language stage to the first acquisition of words, i.e., to the first genuine stage of language. . . . Babbling, on the one hand, and the so-called “Hörstummheit” (comprehension without speech) of the child, on the other, prove that he is deficient neither in motor nor acoustic impressions, but nevertheless most of his sounds are suddenly lost.\(^{48}\)

Once acquired, language becomes a voice that is inescapable; one that, even when we do not speak aloud, is our constant companion with which we form an inner dialogue. The voice of consciousness, or *inner voice*, fuels our subjectivity through the production of outwardly silent, yet internally audible language—a form of “pure auto-affection,” with “the unique experience of the signified producing itself spontaneously, from within the self.”\(^{49}\) It is the voice that tells us what we are doing—‘I’m walking to the car’ or ‘I really need some sleep.’ It is also the sound you hear while quietly reading this text. As a whole, inner voice “serves to ground the illusion of the transparent self-presentation of the speaking subject” and is summed up by Jacques Derrida’s phrase “s’entendre parler”—


to hear oneself speak. Such a voice “lies on the side of the good, because it works against your risks of slipping into dissociation or incoherence, inclining you instead to live in critical harmony with yourself.”

Despite the centripetal pull towards verbal coherence, our prelinguistic tendencies, epitomized by the infant at “the apex of babble,” lurk in the backdrop, waiting to rupture the very stability of linguistic meaning we seek:

But there is another dimension, that of depth, the depths of the body, where another language emerges, raucous, violent, full of consonants, and unpronounceable sounds, of screams and hoarse whispers. This is always threatening to emerge from the orifices of the body, to overcome and destroy the fragile language of the surface, to plunge the subject who is made to utter the sounds into the deepest madness for its effect is to disrupt language as the vehicle of meaning. It carries with it the much deeper nonsense of the body: language is no longer effect (a system of differential values promoting the exchange of meaning) but affect, where the overpowering passions of the body find a direct outlet.

Julia Kristeva calls the extra-verbal means by which instinctual, bodily drives make their way into language “the semiotic.” The semiotic, unconfined by the rules of grammar

---


52 Denise Riley, “‘A Voice Without a Mouth’: Inner Speech,” *Qui Parle* 14, no. 2 (Spring/Summer, 2004): 89.


55 Julia Kristeva “Revolution in Poetic Language,” trans. Margaret Waller, in *The Kristeva Reader*, ed. Toril Moi (New York: Columbia University Press, 1986), 93 (my emphasis). The “semiotic” should not be confused with “semiotics” (the study of signs), despite the fact that, according to Kristeva, the “semiotic” plays an integral role in the signification process.
and syntax, manifests itself in the rhythm, intonations, glossolalias, and other sonic peculiarities of discourse.\textsuperscript{56-57} Moreover, the semiotic is “definitely heterogeneous to meaning but always in sight of it or in either a negative or surplus relationship to it.”\textsuperscript{58} To Kristeva, the subject that articulates is not a stable subject, but rather a “subject-in-process” that forms out of the repression of the voice’s materiality via linguistic negation, or the “symbolic.”\textsuperscript{59} Although primarily suppressed, these “psychical inscriptions” creep into the everyday \textit{paralanguage} of the speaker\textsuperscript{60}

Paralanguage is some kind of articulation of the vocal apparatus, or significant lack of it, i.e. hesitation, between segments of vocal articulation. This includes all noises and sounds which are extra-speech sounds, such as hissing, shushing, whistling, and imitation sounds, as well as a large variety of speech modifications, such as quality of voice (sepulchral, whiney, giggling), extra high-pitched utterances, or hesitation and speed in talking.\textsuperscript{61}

We know paralanguage as ‘tone of voice’ or ‘the way she said it’ and use it to emphasize important aspects of verbal information, demarcate rhythm, and broadcast emotional intent. Paralanguage serves as “the main carrier of affective signals” and is present in every speech act, whether we swear at a politician, cheer on a favorite

\footnotesize
\begin{itemize}
\item \textsuperscript{57} In certain religions, such as Pentecostal Christianity, a human subject might ‘speak in tongues’ by entering a trance-like state and enunciating in an unknown language. This act, otherwise known as glossolalia, emphasizes that the subject is taken over by some form of spiritual being, that speaks through him/her, using the body as a conduit for vocal expression. For an overview of glossolalia, see Mark J. Cartlidge ed., \textit{Speaking in Tongues: Multi-disciplinary Perspectives}, ed. (Milton Keynes: Paternoster Press, 2006).
\item \textsuperscript{58} Kristeva, \textit{Desire in Language}, 133.
\item \textsuperscript{59} Kristeva, \textit{The Kristeva Reader}, 91.
\item \textsuperscript{60} Edith Kurzweil, “An Interview with Julia Kristeva,” \textit{Partisan Review} 53, no. 2 (1986): 222.
\item \textsuperscript{61} Mary Ritchie Key, \textit{Paralanguage and Kinesics: Nonverbal Communication} (Metuchen: The Scarecrow Press, Inc., 1975), 10.
\end{itemize}
athlete, growl for a cup of coffee, whisper in the cinema, or impersonate a beloved pet.\footnote{Kreiman and Sidtis, \textit{Foundations of Voice Studies}, 303.}

Within paralanguage, phonologists often use the term \textit{prosody} to encompass “[t]he pitch, timing, and volume of speech.”\footnote{Ann Wennerstrom, \textit{The Music of Everyday Speech: Prosody and Discourse Analysis} (New York: Oxford University Press, 2001), 274.} Paralanguage’s lack of a stable descriptive system, met by “an ocean of possible prosodic patterns,” make it “difficult (if not impossible) to identify consistent correlations between the cues, these functions, the meanings they convey, and linguistic structure.”\footnote{Kreiman and Sidtis, \textit{Foundations of Voice Studies}, 261-278.} Vocabulary can “distinguish nuances of meaning, but words fail us when we are faced with the infinite shades of the voice, which infinitely exceed meaning.”\footnote{Dolar, \textit{A Voice and Nothing More}, 13.}

\textbf{Nonsense and Beyond Singing}

The voice’s sonorous excesses—its bodily overflows and waste products—pose a semantic asymmetry that has been historically met with an opposing force of “dogmatism and fear.”\footnote{David Appelbaum, \textit{Voice} (Albany: SUNY Press, 1990), 4.} David Appelbaum notes that the entire history of philosophy shows “the erasure of voiced interruptions (squeals, rasps, gurgles, clicks, clacks, and wheezes) from the acknowledged interval of speaking,” demonstrating “an increasing anxiety to retain control over the speaking voice and to defend against lapses in articulated sound.” This is not simply philosophy’s problem either: we see it crop up in the battle between music and text in various places of Western music history. As Lawrence Kramer notes, “vocal music always seems to be struggling against a latent
impulse to dissolve its language away.”  

In ancient Greece, Plato declared that “the music and the rhythm must follow the speech.” Moreover, Plato had a particular problem with the flute and banned its widespread usage. Is not a flute simply an “acoustic prosthesis of the mouth,” an object placed where the voice leaves the body, modifying the breath into a sound devoid of language? With the advent of polyphony, “music scattered the sense of the words, and made them unintelligible to the unpracticed public.” As a consequence, in the 16th Century, text declamation became paramount in Josquin’s ‘musica reservata’ (reserved music), as well in the music of Palestrina, who was ordered by Pope Paul III and the Council of Trent to not obscure the text, thus avoiding “music intended merely to give pleasure (‘inanem aurium oblectationem’) to the listeners.”

---


In more recent years, recording technology has given composers the green light to use nonsensical voices that reach beyond the confines of singing. Utterances can be as perverse, mutated, grotesque, or downright weird as the composer pleases:73

Unless under exceptional circumstances, to have a singer mouthing unintelligible syllables would be to invite the danger of absurdity. However, when the voice appears on tape this restriction is lifted, the unseen singer no longer needing to behave with rational decorum.74

The proclivity to extend the use of the voice beyond singing harkens back to Arnold Schoenberg’s use of ‘Sprechstimme’ (speech-voice, also called ‘Sprechgesang,’ or speech-song) in Peirot Lunnaire (1912) to express “the beauty of the of the speaking, declaiming voice,” creating “a more subtle interpretation than was thought possible through a bel canto delivery.”75 The Italian Futurists also played a significant role. In one of his several manifestos (1913), poet and playwright Filippo Tommaso Marinetti calls for a poetic aesthetic that destroys “the canals of syntax” through the use of aggressive onomatopoeia to mimic the industrial machines of the early 20th century.76 Also noteworthy is the work of Dadaist sound poet Kurt Schwitters. In Ursonate (1922-32), Schwitters splices and stutters German phonemes into incomprehensible words that resemble the prosody of everyday speech, as well as elongating many of these sounds

---

73 This was not always the case, as Berio’s Visage was banned from Italian Radio in 1962 for sounding pornographic. See David Osmond-Smith and Cathy Berberian “The Tenth Oscillator: The Work of Cathy Berberian 1958-1976,” Tempo 58, no. 227 (January 2004): 2-13.


to create song-like gestures.\(^77\) Moreover, French Surrealist Antonin Artaud left an important mark with his “Theatre of Cruelty” manifestos:

>What the theater can still take over from speech are its possibilities for extension beyond words, for development in space, for dissociative and vibratory action upon the sensibility. This is the hour of intonations, of a world’s particular pronunciation.\(^78\)

Artaud’s 1947 radiophonic work *Pour en finir avec le jugement de dieu* (To Have Done With The Judgment of God), utilizes glossolalia and tortured screams to mirror the speech of infants, poets, and the mentally ill, producing a semiotic outpouring that “suspends the representative value of language.”\(^79\) Artaud’s “terror was dark, filthy, emanating from the deepest recesses of his body, a body which his discourse tried, always unsuccessfully, to rejoin.”\(^80\)

Karlheinz Stockhausen’s *Gesang der Jünglinge* (1955/56) is an innovative landmark in electroacoustic composition. Combining the recordings of a single boy and sounds made from oscillators, filters, and noise generators, the German composer demonstrates similarities between the acoustic structure of speech sounds and electronic timbres through its serial-based organization. Italian composer Luciano Berio extended this relationship further in his work *Visage* (1961). Berio uses non-verbal,

\(^77\) A phoneme is “[a]n abstract minimal unit of sound in a language that can distinguish meaning, as in /p/ vs. /b/ in *pig* and *big*.” Phoneme and timbre are not interchangeable. For example, /e/ represents all the pronunciations that differ from other linguistically meaningful units; /e/ can be screamed or whispered, producing vastly different timbres, but they still maintain the same phonemic identity. See Ann Wennerstrom, *The Music of Everyday Speech: Prosody and Discourse Analysis* (New York: Oxford University Press, 2001), 62 and 274.


vocal materials—including sobs, laughter, and cries—and an imitative electronics portion, to reveal a paralinguistic resemblance between human voice and synthetic sound, “attaining a secret neuter language without constants and entirely in indirect discourse where the synthesizer and the instrument speak no less than the voice, and the voice plays no less than the instrument.”

While many of these efforts attempt to reduce the role of linguistic meaning, the very act of interpretation paradoxically serves to reinforce language’s presence. Philosopher Michel De Certeau comments on this further:

“[I]nterpretation supposes that somewhere there must be meaning. Interpretation searches for meaning, and finds it because it expects it to be there, because interpretation relies on the conviction that especially where meaning appears to be absent, it is hidden someplace, present “all the same.” Thus, the hermeneutic pursues its object most obstinately in those non-sense places where it postulates “secret languages.” It focuses upon that which it takes a challenge to meaning. And finally, because it believes in meaning, it is trapped into a semblance of language.”

According to Dolar, singing, laughing, coughing, crying, babbling, and screaming are all postlinguistic phenomena and “laden with signification precisely by virtue of being non-signifying.” Therefore, as “[p]hysiological and inarticulate” as a voice may be, “it cannot escape the structure” of language.

As Derrida reminds us, “[f]rom the moment that there is meaning there are nothing but signs. We think only in signs.” If this is the

---

81 Deleuze and Guattari, A Thousand Plateaus: Capitalism and Schizophrenia, 96.


83 Dolar, A Voice and Nothing More, 24

84 Jacques Derrida, Of Grammatology, 50. Charles S. Peirce defines a sign: “A sign, or representamen, is something which stands to somebody for something in some respect or capacity. It addresses somebody, that is, creates in the mind of that person an equivalent sign, or perhaps a more developed sign. That sign which it creates I call the interpretant of the first sign.” See Charles Peirce, Collected Papers of Charles Sanders Peirce, Volume 2: Elements of Logic. eds. Charles Hartshorne and Paul Weiss (Cambridge: Harvard University Press, 1960), 157.
case, then how does the inner voice function while listening to music? Furthermore, how does this voice contribute to our imagination’s production of virtual voice-bodies while listening? As Lawrence Kramer notes, music “incites us to intermittent, silently verbalized commentary on what we are hearing,” potentially opening the floodgates for extramusical signification.85 However, if we are silently humming along to an instrumental or non-verbal vocal part, are we not suspending, however momentarily, the production of language? Is this not something we aspire to while listening to music: to defer our linguistic output, placing ourselves into a territory of “intellectual uncertainty,” where “one does not know where one is, as it were,” and surrendering our inner voice to the resonance that envelops it?86

Vocality

Experiencing a voice purely as an object of pleasure, or ‘jouissance,’ to borrow Lacan’s term, is not something we have been comfortable with for very long.87 Throughout Western music, art, and literature, transgressive women became target for the projected unease of a senseless voice—one that is ripe with eros or hysteria; a body out of control, where the semiotic takes the upper hand and threatens to destroy the order encouraged by the text.88 Numerous myths tell of the seductive and

88 Although a comprehensive exploration of voice and gender in music is beyond the limits of this dissertation, I would like to point several sources that thoroughly deal with this topic: Embodied Voices: Representing Female Vocality in Western Culture, eds. by Leslie C. Dunn and Nancy A. Jones (Cambridge: Cambridge University Press, 1994); Suzanne G. Cusick, “Feminist Theory, Music Theory, and the Mind/Body Problem,” Perspectives of New Music 32, no. 1 (1994): 8-27; Suzanne G. Cusick, “Gender, Musicology, and Feminism,” in Rethinking Music, eds. Nicholas Cook and Mark Everist (Oxford:
dangerous powers of the female voice—a voice that becomes “an object once of desire and fear.”98 Such narratives, as seen in the Sirens of Homer’s Odyssey, “testify to the persistent desire of male artists to control through representation the anxieties aroused by the female voice, even while they license the display, and the enjoyment, of its powers.”99 In 19th century opera, the mastering of this threat is “enacted both thematically, through the defeat or death of the woman character, and discursively, through the containment of her utterance within a textuality identified as masculine.”100

In an effort to circumscribe the reinforcement of these narratives, which overall devalue the embodied voice, musicologists have gravitated towards the term vocality “to indicate a broader spectrum of utterance” that inhabits “an intersubjective acoustic space.”101 Paul Zumthor, the first scholar to devise the term, views it as “the whole of the activities and values that belong to the voice as such, independently of language.”102 Likewise, The Oxford English Dictionary defines vocality as “the quality of having voice or utterance; the possession or exercise of vocal powers” and “vocal quality or nature.”103

---


99 ibid., 3.

100 ibid., 7.

101 Dunn and Jones, Embodied Voices, 1-2.

102 as quoted in Cavarero, For More Than One Voice, 72.

These definitions raise some important questions: what are the boundaries of having a voice? Does this relate solely to the human voice or are other species included? Where do artificial and technologically-mediated voices fit into these definitions? Can inanimate objects speak?

Musicological studies in vocality rarely leave the confines of score-based, acoustic performance; they prioritize the human singer, as well as acoustic instruments’ potential to signify utterance. For Lawrence Kramer, previous theories of voice “center on either the musical expression of textual affect or meaning, the musical transformation (from assimilation to appropriation to deconstruction) of textual affect or meaning, or the relative independence of musical structure and expression from those of the text.”95 One of Kramer’s neologisms, “overvocalization,” situates itself in these generalizations: “the purposeful effacement of text by voice” using “melisma, tessitura, or sustained tones” for “the disintegration of language,” thus producing “emotional and metaphysical extremes, blurring of ego boundaries, and transits of identity.”96 Analyzing works from Beethoven, Schubert, and Brahms, Kramer limits the application of his term to the gestural palette of song. As a complement to overvocalization, however, Kramer suggests the term “songfulness:”

Songfulness is fusion of vocal and musical utterance judged to be both pleasurable and suitable independent of verbal content. It is the positive quality of singing-in-itself: just singing. . . . [V]oice addresses itself in a sensuous and vibratory fullness to the body of the listener, thereby offering both material pleasure and an incitement to fantasy.”97


97 Kramer, Musical Meaning, 53-54.
Using Lacan, Kramer relates the enveloping voice of song to the infant’s enrapture with the mother’s face and voice. However, he does not elaborate on the specifics of listener embodiment, preferring to emphasize that songfulness “makes meaning extraneous, if not downright superfluous.” While Kramer’s efforts to elicit imaginative listening outside of language is commendable, his overall concept suggests some form of analytical end game, where music is beyond interpretation. Even he himself mentions in more recent writing that “[w]ith music, as with anything else, speech, language, may be deferred but it cannot be avoided. Boroughs’ linguistic virus seems all the more appropriate here. While the “infinite shades of the voice” may always exceed meaning, this does not preclude us from making some worthwhile observations along the way.

Musicologist Jonathan Dunsby sees vocality as “those qualities of music and text that enable one to identify it as articulating narrative, mood, the times of tenses, associations, grammatical tropes such as the interrogative, visual images, persons and landscapes, the mundane and the divine.” As he continues, “[v]ocality concerns everything that a replete analysis of music and text ought to explain, and ought not to neglect” Dunsby’s definition is ambitious and open-ended, however, his analyses center on works by Schubert, Brahms, Copland, Berberian, and Schoenberg. Focusing mainly on lieder and choral music, Dunsby neglects any form of digitally-mediated

98 ibid., 63.


voice. Moreover, textual analysis is an imperative component of his definition. Leaving the text portion aside, Dunsby’s description seems ripe for electroacoustic application: the listener’s direct material engagement with a variety of referential and non-referential sounds incites a temporal unfolding of landscape imagery and other extramusical associations.

Utilizing Charles S. Peirce’s theory of semiotics, Naomi Cumming posits that the concept of vocality revolves around the idea that sounds may be heard as signifying: “timbral quality,” in this sense, becomes “a ‘sign’ in giving rise to a mode of listening that goes beyond its reception as an acoustic fact.” This stresses that without interpretation, material sound produced by an acoustic instrument cannot be taken as a voice. Cumming focuses on the acoustic performer’s realization of vocality through score interpretation, with particular attention given to the violinist, and how this can inform listener-based perception and subjectivity. While she concentrates on the song-like qualities of instrumental performance—those that evoke “an operatically trained singer”—she does, however, touch on the potentials of electroacoustic sound:

[T]echniques of electroacoustic alteration . . . may create “voices” which give the illusion of “presence,” or perhaps of an ethereal transcendence of their “materiality”—by covering the sounding evidence of physical creation. . . . [T]o the extent that a listener retains even the idea of a possible “voice” as its content, he or she begins to negotiate unfamiliar possibilities in subjective experience. . . . If a listener is, then, willing to entertain the idea of a voice or action carrying subjective connotation, while also being aware of its falsehood, he or she is in a position to explore “subjectivities” that are the virtual creations of modified sound. It is

---

101 Unless of course he uses recordings to analyze his chosen works. Even so, recordings of notation-based works, more often than not, try to re-present the uniform space of the performance venue.


103 ibid., 130.
as if an identification with quality and an appreciation of its “otherness” have come together, to allow an identification with that which is “other.”

Similar to Kramer, Cumming suggests that vocality may be something that extends beyond the human body and into the virtual realm of fantasy. Moreover, she describes a process of giving voice to non-vocal sounds that is not unlike the animism granted by our distant ancestors. This lending of voice acts as a self-reflexive, virtual prosthesis—one that casts off from the listener’s body but, through the act of interpretation, boomerangs back “like the frog’s tongue.” For this empathic exercise to work, however, the listener has to uphold a pretense of “make-believe,” where “I believe I am really hearing what I would like to hear as a promise of pleasure,” thus turning timbre and its vocal shapes into an object of desire.

**Artificial Vocality**

Advances in human voice synthesis—Homer Dudley’s Vocoder (1939), Atal and Hanauer’s introduction of linear predictive encoding (1971), and John Chowning’s experiments with FM voice synthesis (1980)—have produced innovative timbres that despite sounding overly familiar today, still shape much of the electroacoustic canon over the past 40 years. With the variety of digital signal processing techniques

---

104 ibid., 130-131.


available today, boundaries between the natural and the artificial are not always clear cut. French musicologist Bruno Bossis uses the term “artificial vocality” to establish “a new link between the vocal quality of a sound event (its vocality) and technology (its artificiality)” within electroacoustic music.\textsuperscript{109} Bossis expands on this further:

Due to its remarkable, complex acoustical quality, and its obvious relationship to the body and to verbal communication, artificial vocality is a prime representation of the difficulty musicologists have in confronting music which is not written down. . . . As vocality concerns the vocal quality of a sound event, it is based on a perceptive analogy with the natural voice. According to this definition, nothing opposes the idea that sound can be treated as vocal, despite having no original link to the phonation organ. The voice can be produced either by pure simulation or by a non-vocal timbre, which sounds unintentionally like a voice.\textsuperscript{110}

Bossis focuses on vocality as a sonic trace in electroacoustic music; a silhouette of a voice that may or may not be present. As Simon Emmerson adds, “even within ‘abstract’ discourse where no sounds are directly identifiable as a result of real world sound events, there may remain direct reference to the behavior characteristics of such events.”\textsuperscript{111} A group of non-vocal sounds may resemble a speaker’s prosody, song-like gestures, or even postlinguistic interjections and babble. Trevor Wishart depicts this succinctly: “[j]ust as we can imagine a landscape containing an utterance, so we can imagine an utterance containing a landscape.”\textsuperscript{112}

\begin{thebibliography}{99}
\bibitem{110} ibid., 91.
\bibitem{111} Simon Emmerson, \textit{Living Electronic Music} (Burlington: Ashgate, 2007), 14.
\end{thebibliography}
The electroacoustic medium affords a plasticity of sound that “encourages composers to make smooth connections between vocal and non-vocal material.”\textsuperscript{113} American composer Paul Koonce demonstrates such fluidity in his work \textit{Hothouse} (1992), in which non-verbal utterances, environmental sounds, and acoustic instruments twirl, leap, and morph into one another, evoking a surrealist game of hide and seek, where the very nature of ‘having voice’ is constantly brought into question. A composer’s ability to simultaneously layer, filter, and interpolate amongst a variety of sounds can make listener-based distinctions between voice and not-voice nearly impossible. In this sense, the composer becomes akin to a \textit{ventriloquist}: one who creates aural and vocal illusions by ‘throwing the voice’ onto inanimate objects, potentially reinvigorating vocal myths of the past. The listener, therefore, revels in a differential dance of referential undecidability—“a feeling of strange familiarity and familiar strangeness”—continually recasting a game of desire for sound-objects that are at once both absent and present.\textsuperscript{114}-\textsuperscript{115} Although a potential headache for musicologists, this aesthetics of ambiguity opens the listener to a more fluid and \textit{symbiotic vocality}: one that disseminates the heterogeneity of our identities, “where boundaries shift and continually renewed definitions are the only constant.”\textsuperscript{116}

According to Michel Chion, “every sound to which one listens for a long time


\textsuperscript{114} Timothy Morton, \textit{Hyperobjects: Philosophy and Ecology after the End of the World} (Minneapolis: University of Minnesota Press, 2013), 84.

\textsuperscript{115} a \textit{sound-object} is “a sound with a discrete spatial reference point, differentiated from the sound of the enclosure or resonator within which the object is sounding.” See Natasha Barrett, “Spatio-Musical Composition Strategies,” \textit{Organised Sound} 7, no. 3 (December 2002): 313n1.

\textsuperscript{116} Jonty Harrison, \textit{Articles Indéfinis} (Montréal: Empreintes Digitales, 1996), 3.
becomes a voice. Sounds speak.”\textsuperscript{117} How can we phenomenologically account for this? Given that almost any acousmatic sound can be perceived as voice-like, might we consider an electroacoustic work as a simulated environment of semiotic discharges, filled with the paralinguistic outpourings of virtual bodies, where “sounds metaphorically represent instinctual drives”?\textsuperscript{118} Or perhaps it is the reverse—in our entanglement with sound, we conjure up these drives \textit{in ourselves} via the imitative, inner voice of \textit{subvocalization}? As I will demonstrate in Chapter 3, vocality is something we \textit{enact} through cross-modal simulations of virtual echolocation; our \textit{paralinguistic embodiment} imprints itself onto the sonic material that confronts us, opening an “inter-subjective space where ‘I am listening’ also means ‘listen to me.’”\textsuperscript{119}

\begin{quote}
\textsuperscript{117} Michel Chion, \textit{Sound}, 51.
\textsuperscript{118} Lecercle, \textit{Philosophy Through the Looking Glass}, 144.
\textsuperscript{119} Barthes and Havas, \textit{The Responsibility of Forms}, 246.
\end{quote}
CHAPTER 3
THE LISTENING VOICE

In order to understand our tendency towards animism and its relationship to music, as well our voice’s ability to inhabit intersubjective acoustic spaces, we need to examine how our body operates on sound, as well as how sound operates on our body. While Chapter 2 investigated the cultural, historical, mythological, and identity-based connotations of voice and music, Chapter 3 examines the following question: what is the voice’s role in listening perception and how does it influence our understanding of timbre? As I will argue, vocal embodiment extends much further than listening to other human utterances, whether mediated through technology or not; our voice helps navigate many aspects of timbre, irrespective of sound-source. To begin unpacking this question, it is important to probe the concept of mimesis and its indispensable role in human behavior.

**Mimesis, Mimicry, Imitation**

Imitation is a vital component of any given culture. Our desire to understand and be understood by others fuels our need to imitate. We routinely mimic the gestures, vocalizations, and postures of those who surround us. This behavior is apparent in any infant-parent relationship, where the two trade babble, facial expressions, and movements to build rapport and foster empathy. When others around us laugh, we tend to join in as well. Likewise, we may respond to a sad voice by feeling sad ourselves. Imitation also plays a role in understanding different animal sounds and movements, as well as inanimate objects such as machinery. Many of us can recall running around as young children and impersonating such things, where there was no limit to our capacity for make-believe:
Children’s games are interlaced everywhere with mimetic modes of behavior, and their range is not limited at all to what one human being imitates from another. A child not only plays at being a grocer or a teacher, but also at being a windmill or a train.¹

Such games, as Walter Benjamin describes, enlist our voice and body to imitate anything and everything that excites our imaginations—a superhero’s commands, a jet airplane’s takeoff, a light saber’s swoops, or even a bunch of stones falling. This pleasure “in being or passing for another” highlights a process called mimesis.²³ The term mimesis originates from the Greek word of the same name, meaning “to imitate.”⁴ Broadly speaking, mimesis specifies two major areas of scholastic inquiry: one is the representation of nature in art, music, and literature, while the other is the direct imitation of the actions and/or utterances of other people, animals, and objects.⁵ The desire to emulate others, according to anthropologist Michael Taussig, “involves the rehearsal of the practices of the body associated with primitivism.”⁶ Likewise, as Benjamin suggests, mimesis is “a rudiment of the powerful compulsion in former times

---

³ For more on understanding other beings and objects using metaphor and third-person subjectivity, see Thomas Nagel, “What is It Like to Be a Bat?,” The Philosophical Review 83, no. 4 (October 1974): 435-450.
to become and behave like something else.” These primordial notions follow Theodor Adorno’s emphasis on the “biological context” of imitation, in which he considers “mimicry to be a prehistorical or zoological version of mimesis.” Cahn discusses Adorno’s implications further:

Mimicry mediates between two states: life and death. The chameleon perfects mimicry to a point of unintelligibility: always taking on the color of its surroundings, this animal never seems to be “itself.” Other animals, man among them, feign death in the face of danger, they delude their enemy into believing that they are a stone, a leaf, or already dead. For these animals mimicry is a means of survival. In psychoanalytic terms, they survive because they are able to almost totally identify with something other than themselves, with dead, lifeless material.

As Roger Caillois adds, “what mimicry achieves morphologically in certain animal species,” we experience as “depersonalization by assimilation to space.”

**Ecological Perception and Embodied Music Cognition**

Such an organism/environment interaction suggests an ecological model of perception, “which studies the cognitive and perceptual system in the service of survival and orientation in the environment.” In this theory, first put forth by James Gibson, organisms actively search for information already part of the environment that provides perceptual significance. Central to Gibson’s view is the concept of affordances: “[t]he

---


9 ibid., 32-33.


affordances of an environment are what it offers the animal, what it provides or furnishes, either for good or ill."\textsuperscript{12} Phrased in another manner, "affordances are primarily understood as the action consequences of encountering perceptual information in the world."\textsuperscript{13} Gibson expands on this further:

\begin{quote}
An important fact about the affordances of the environment is that they are in a sense objective, real, and physical, unlike values and meanings, which are often supposed to be subjective, phenomenal, and mental. But, actually, an affordance is neither an objective property nor a subjective property: or it is both if you like. An affordance cuts across the dichotomy of subjective-objective and helps us to understand its inadequacy. It is equally a fact of the environment and a fact of behavior. It is both physical and psychical, yet neither. An affordance points both ways, to the environment and to the observer.\textsuperscript{14}
\end{quote}

Affordances, moreover, are organism dependent. For example, a bookshelf affords me a place to store books, magazines, and other memorabilia. For a cat, however, the same bookshelf affords a place to climb, sleep, and hide.\textsuperscript{15} In this regard, to echo Maurice Merleau-Ponty, our perception attunes itself towards possibilities for \textit{doing}:

\begin{quote}
What counts for the orientation of the spectacle is not my body as it in fact is, as a thing in objective space, but as a system of possible actions, a virtual body with its phenomenal ‘place’ defined by its task and situation. My body is wherever there is something to be done.\textsuperscript{16}
\end{quote}


\textsuperscript{14}Gibson, \textit{The Ecological Approach to Visual Perception}, 129.

\textsuperscript{15}The bookshelf, for a human, also affords throwing it out the window, but social factors preclude this from becoming a significant affordance.

Although primarily developed by Gibson to understand visual phenomena, many scholars use ecological models to comprehend aural perception and musical experience.\(^\text{17}\) Such an approach to listening emphasizes that perception and action are interdependent: perception initiates action and action reshapes perception. While not specific to music, nor an ecological framework, Merleau-Ponty argues a similar, reciprocal relationship between movement and perception:

Sensations, ‘sensible qualities’ are then far from being reducible to a certain indescribable state or quale; they present themselves with a motor physiognomy, and are enveloped in a living significance. . . . [S]ensations have a ‘motor accompaniment’, that stimuli set in motion ‘incipient movements’ which are associated with the sensation of the quality and create a halo round it, and that the ‘perceptual side’ and the ‘motor side’ of behaviour are in communication with each other.\(^\text{18}\)

This action-perception coupling forms the basis of embodied music cognition, in which “music perception is tightly linked with body movement, action and environmental interaction,” and helps explain our affective and imaginative responses to music, as well as the expressivity we attribute to it.\(^\text{19-20}\) Music, from this standpoint, “is something that

---


\(^{20}\) This is not to deny that symbolic representations play a significant role in musical meaning. However, representational systems have been previously associated with the classic view of cognition, where action and perception are completely separate from one another. Gibson, on the contrary, contends that symbols are grounded in our perceptual relationship with the environment: “There have to be modes of stimulation, or ways of conveying information, for any individual to perceive anything, however abstract. He must be sensitive to stimuli no matter how universal or fine-spun the thing he
induces a kind of *motor resonance* that prompts . . . listeners to experience . . . sounds as if they are involved in their production."\(^{21}\) One of the ways we understand music is through a simulation of movement, which grounds itself in bodily mimesis: we “spontaneously (and largely involuntarily) mentally imitate the movements that we see other people making, as well as the movements that we assume other people are making in cases where we cannot actually see their movements.”\(^{22}\) In this regard, our perception of timbre, instead of being merely a quality we hear and often describe with adjectives, is a *simulated action* that enlists cooperation amongst all sense modalities; we apprehend timbre *performatively* as listeners by “imagining the actions that are apprehends. No symbol exists except as it is realized in sound, projected light, mechanical contact, or the like.” Thus, from an action-oriented perspective, it is more productive to think of symbolic representations as *performative re-presentations*. See James J. Gibson, *The Senses Considered as Perceptual Systems* (Boston: Houghton Mifflin Company, 1966), 26; For more on the classical model of cognition, see Susan Hurley, “Perception and Action: Alternative Views,” *Synthese* 129, no. 1 (2001): 3-40.


implied in manipulating the perceived objects.”23 And, as Derrida claims, “it is especially
to the vocal organ that this supplement of motor activity finds itself entrusted,” for our
voice is our most intimate and potent imitator of sound, supplying us with the ability to
impersonate and reconstruct who or what we can be.24

Metaphor and Vocal Embodiment

Crucial to understanding what it’s like to be something else is metaphor:
“metaphoric understanding calls on the resources of imagination mediated by the
phenomenology of embodiment.”25 To philosophers George Lakoff and Mark Johnson,
“the essence of metaphor is understanding and experiencing one kind of thing in terms
of another.”26 More explicitly, “[w]e understand experience metaphorically when we use
a gestalt from one domain of experience to structure experience in another domain.”27
Last, as Johnson argues, “virtually all of our conceptualization and description of music
uses metaphors whose source domains are drawn from sensorimotor experience.”28
Our vocal embodiment of sound—where “the body converts a certain motor essence
into vocal form”—facilitates musical understanding through metaphorical exertion; our
vast experiences of paralinguistic utterance and echolocation and their relationship to

University Press, 2005), 147.
26 George Lakoff and Mark Johnson, Metaphors We Live By (Chicago: The University of Chicago
Press, 2003), 5.
27 ibid., 230.
our body’s various states of metabolic arousal, allow us to empathize with acoustic phenomena, create virtual imagery, and assign expressivity to music.\textsuperscript{29-30} Music theorist Arnie Cox argues that “because most of us have made vocal sounds everyday since birth, the voice becomes our central basis for understanding the sounds made by other humans.”\textsuperscript{31} While this certainly may be the case, it is my conjecture that the voice becomes our central basis for understanding all sounds, whether stemming from the human body, acoustic instruments, electroacoustic fabrication, or our everyday environments: “[u]ltimately we incorporate, even as adults, every sound we hear as a sort of internal vocalization.”\textsuperscript{32}

Electroacoustic music, with its removal of direct body imagery and use of ambiguous sound events that move in and out of the human body’s sound-producing capabilities, only amplifies the voice’s performative role in listening. Given the concept of affordances, our auditory perception is “an adaptation to the physical constraints of our own actions.”\textsuperscript{33} Our voice is the best tool for aural adaptation because 1.) we all

\textsuperscript{29} Merleau-Ponty, The Phenomenology of Perception, 211.

\textsuperscript{30} Patrik Juslin defines ‘arousal’ as “physical activation of the autonomic nervous system. Physiological arousal is one of the components of an emotional response, but could also occur in the absence of emotion (e.g., due to exercise). Arousal is often reflected in the ‘feeling’ component (i.e., the subjective experience).” See Patrik N. Juslin, “From Everyday Emotions to Aesthetic Emotions: Towards a Unified Theory of Musical Emotions,” Physics of Life Reviews 10 (2013): 236.

\textsuperscript{31} Arnie Cox, “The Mimetic Hypothesis and Embodied Musical Meaning,” Musicae Scientiae 5, no. 2 (Fall 2001): 197.


\textsuperscript{33} Leman, Embodied Music Cognition and Mediation Technology, 61.
have this instrument and are intimately familiar with it and 2.) it can approximate any sound we give our attention to. So how does this brand of mimetic participation work?

**Motor-Mimetic Participation**

As adults, there is certainly a limit of outwardly imitative practices in public spaces; if one were to run through the supermarket with arms spread like wings and make sounds that emulate a jet airplane, he/she might be asked to leave. As Theodor Adorno and Max Horkheimer say, “[u]ncontrolled mimesis is proscribed.” However, imitation never ceases during adulthood; rather, it “becomes more subtle and covert as we mature” and “is seldom conscious.” Cox further differentiates between the two general types of mimetic enaction:

When we overtly imitate someone or something, we represent the observed behavior in our own skeletal-motor system and in associated neural activity and blood chemistry. When we covertly imitate someone or something . . . we represent the observed behavior in roughly the same way, except that the executions of the motor actions are inhibited, and the changes in other systems are attenuated.

The cultural context of listening has a significant impact on our level of overt or covert participation. If we are at a party, we may favor active bodily involvement with music, in

---


36 ibid., 4 (emphasis in original). Covert imitation is not an entirely new concept, as we can see traces of it in the work of Sigmund Freud: “When, now, I perceive a movement . . . of greater or lesser size in someone else, the securest way to an understanding (an apperception) of it will be for me to carry it out by imitation. . . . An impulse of this kind to imitation is undoubtedly present in perceptions of movements. But actually I do not carry the imitation through, any more than I still spell words out if I learnt to read by spelling. Instead of imitating the movement with my muscles, I have an idea of it through the medium of my memory traces of expenditures of similar movements.” See Sigmund Freud, *Jokes and their Relation to the Unconscious*, trans. James Strachey (New York: W.W. Norton and Company, 1960), 237.
the form of dancing, singing, or shouting. In the “Western art music tradition,” however, our listening involves “different varieties of more or less concealed or sublimated active engagement.”

In these situations, our outward movements subdue in favor of concealed imagery and subvocalization:

Mimetic subvocalization includes any motor imagery and motor activation related to the vocal musculature. Bear in mind that it need not be intentional or conscious, and it can be very much like singing along or only indirectly like singing, involving simply exhaling in time, or various kinds of exertions in the throat, and/or exertions of the tongue and/or lips. For myself, when I am aware of it (intentional or otherwise), I seldom find that it is exactly like singing along and that instead it is something more like wanting to sing along, where “singing along” is manifest in some combination of exertions of the throat, chest, and abdomen, all of which occur in actual vocalization but which in subvocalization are more generic or abstract.

Subvocalization of other speakers or singers seems rather straightforward—as humans, given a healthy body, we share similar vocal capacities, so imitation becomes more or less isomorphic. However, subvocalization is a significant aspect of all embodied music cognition, not simply mirroring another person’s voice. For example, our ability to ‘hear’ music internally while reading a traditionally-notated score, irrespective of instrumentation, depends on it: “notational audiation is a process engaging kinesthetic-like covert excitation of the vocal folds linked to phonatory resources.”

---


38 This is not to deny that other parts of the body play a significant role in the mimetic comprehension of music, such as simulating the hand and finger movements in playing the guitar or piano. However, subvocalization has been shown to also accompany such imagined movements. See Warren Brodsky et al., “The Mental Representation of Music Notation: Notational Audiation,” *Journal of Experimental Psychology: Human Perception* 34, no. 2 (2008): 427-445.


ability to re-present sounds that are non-vocal in origin—acoustic instrument or otherwise—emphasizes a cross-domain understanding of timbre and is realized through a neurophysiological process called motor equivalence.\textsuperscript{41}

Mimetic vocalization and subvocalization involves making the ‘same’ sound with a different set of motor actions, so that what is being imitated is not the specific exertions of, say, the violinist or the trombone player, but rather the pitches, durations, intensity, and to some extent the timbres of the sounds produced.\textsuperscript{42}

Cross-domain, vocal imitation is vital component of musical comprehension and an important tool for communicating timbral attributes to other listeners. Anthropologists Steven Feld et al. explain that studio engineers and musicians often use “[s]poken/sung ‘vocables’” as a means “to iconically mimic in vocalization the timbral features of the musical sound(s) under consideration.”\textsuperscript{43} For example, some may use their voice to mimic the ‘sizzle’ of various cymbals, placing the tongue between the teeth to obstruct airflow, producing a noise-based timbre. Likewise, the concept of the ‘human beatbox’ has grown out of the reflexive imitation of certain recording technologies—drum machines and MIDI sequencers that play heavily quantized percussion loops found on countless dance and rap records over the past 30 years. Furthermore, we often use voice-based, verbal metaphors to describe the expressiveness of certain sounds: a ‘squealing saxophone,’ a ‘growling synth-bass,’ or the ‘hissing wind.’ Each metaphor


\textsuperscript{42} Cox, “Embodying Music: Principles of the Mimetic Hypothesis,” 8-9 (my emphasis).

relates to vocal experiences that incorporate various amounts of air, copious amounts of muscular tension and relaxation, tongue, teeth, and jaw movements, as well as bodily postures and movements necessary for various levels of exertion. Our voice plays a significant role in cross-domain mappings like these because there seems to be no limit to what our voice can imitate: the more acoustic phenomena there is in our environment (whether from industrial machines or creative sound design made from recordings of such machines), the more opportunities there are for our voice to imitate. In other words, new material entanglements afford us new ways to vocally imitate; any set of vibrations that fall into our range of hearing is game for our voice to emulate.44

Psychologists Steven Pinker and Ray Jackendoff believe that “[h]umans are not notably talented at vocal imitation in general.” This is a valid claim, for if we compare our vocal-mimetic skills to those of a lyrebird, then we will most certainly fail.45 However, it is in these failures that mimesis becomes not a process of isomorphic replication, but a creative act that relies on “the free and pure productivity of the imagination.”46 As sound producing mechanisms, our bodies have significant limitations that influence vocal output, including, but not limited to, our range of pitch/spectrum, duration, and dynamics. Therefore, in coping with an electroacoustic work’s sonic environment, we move from something we can plausibly mirror, to something we can only imagine.

44 The human range of hearing at birth is approximately 20-20,000 Hz, but this range diminishes as we age, particularly in our ability to hear high frequencies.

45 David Attenborough narrates the lyrebird’s uncanny imitation of other species of birds, a camera shutter, a car alarm, and a chainsaw in one of his many wildlife documentaries. See http://www.bbc.co.uk/nature/life/Superb_Lyrebird (accessed May 29, 2015).

Electroacoustic music affords listeners the experience of leaping from the possible to the impossible and back again through the motor resonance of the voice.

**Subvocalization and Timbre Appraisal**

However, Arnie Cox argues that music does not just “afford” mimetic participation; it “demands” it: “mimetic participation is compulsory once we give our attention to music . . . musical engagement feels something like responding to an invitation, if not to a command.”[^47] If this is the case, do we subvocalize all sounds we listen to, and hence, the notion of giving voice is automatic? Or, does a sound have to be ‘voice-like’ first before we can vocally emulate it?

Studying neural-based music perception, Koelsch et al. demonstrated that “subjects coded vocal sound production (without actual movement) while they perceived . . . pleasant musical signals.”[^48] The authors found that the Rolandic operculum—the region of the brain that controls the movement of the larynx—activates when we listen to mainly consonant music signals, whether vocal in origin or not. Put simply, when we hear something we like, which according to the authors is consonant music, the vocal motor system operates as if we are singing along. However, as Hermann von Helmholtz declares, “the boundary between consonances and dissonances has frequently changed … and will still further change.”[^49] Fortunately, Koelsch et al. note that the

[^47]: Cox, “Embodying Music: Principles of the Mimetic Hypothesis,” 11 (emphasis in original). Cox supports his claim with mirror neuron research, arguing that subvocalization is innate and automatic when we give our attention to musical stimulus.


pleasantness of musical stimuli is not fixed, as “cultural experiences can modify judgments about the pleasantness of certain dissonances.” Moreover, the music selected for this experiment relies heavily on harmonic timbres from acoustic instruments. Does the vocal motor system respond automatically to inharmonic or noise-based signals? Electroacoustic composers, given the integration of listening into the composition process, repeatedly expose themselves to such timbres, leaving them well-adjusted to acoustic qualities that inexperienced listeners might call ‘harsh.’ Many composers even want to listen to such sounds, so how might we account for this?

Koelsch et al. underline the significance of timbre appraisal in music perception. Music theorist David Huron contrasts an appraisal response, one that engages conscious thought and a slower evaluation of an environment, with a much faster and unconscious, reaction response:

Extensive research has established that there are two types of responses to the advent of events. One type of response is very fast. The other type of response is more leisurely. The fast response represents a “quick-and-dirty” assessment of the situation followed by an immediate somatic (bodily) response. The second response represents a more “thoughtful” assessment of the situation—a response that takes into account complex social and environmental factors. I propose to call the fast response a reaction response, and the more complex slower response an appraisal response.

From an evolutionary perspective, surprise mediates our reaction responses: we favor being able to predict positive outcomes of situations, and surprise “represents a failure

50 ibid, 240.

51 Noise and inharmonicity also play important roles in rap and electronic dance music, which are both extremely popular globally.

of expectation.” As a result, surprise is “perpetually negative,” as our reaction to it assumes a worst case scenario. Thus, in a dangerous situation, those who can quickly “form accurate expectations about future events” have “significant biological advantages.” In a musical context, an extremely fast event onset, marked by an exponential rise in dynamics, will set off a reaction response in any listener. Attend an electroacoustic concert and watch people jump out of their seats, reach to cover their ears, or dash for the exit doors when they encounter such a musical gesture or when microphone/speaker feedback revs up instantaneously. Likewise, sustained loudness, which indicates high levels of physical energy that could pose danger, also excite reaction responses. However, while reaction responses are always negative, ensuing appraisal responses may be positive. Huron calls this phenomenon contrastive valence, and in music, such limbic reversals can act as amplifier that boosts listeners’ positive emotions.

Appraisal responses, on the other hand, are very complex, as individual listeners and/or groups of listeners may derive pleasure from certain sounds, whereas others

---

53 ibid., 21.
54 ibid., 22.
55 ibid., 3.
56 However, when a composer repeatedly uses such gestures over short periods of time, their predictability desensitizes the listener from having such strong reaction responses.
57 Huron points out that even our reaction responses ultimately lead us to conclude we are listening to music: “Unlike the growling of a bear, the sounds do not represent an imminent danger. There is no need to run, flee, or hide from these sounds.” Many people enjoy horror movies for this very reason: experiencing heightened surprise/arousal in a make-believe setting. Huron, however, does not account for high amplitude listening situations. In these cases, those interested in protecting themselves from hearing damage may flee. See Huron, Sweet Anticipation: Music and The Psychology of Expectation, 37.
58 ibid., 39.
may find them distasteful. Part of our appraisal of timbre depends on what the music invites us to do and, to paraphrase musicologist Suzanne Cusick, participate or not as we choose. Adds Cox, “a listener may like or dislike a given invitation, and this liking/disliking of what the music invites us to do will contribute to aesthetic evaluation.” In the case of Koelsch et al., the participants subvocalized because they liked the music’s invitation to sing along—an activity that requires, for many people, familiar patterns of bodily movement and exertion. In the case of noisy, non-harmonic timbres, composers may require the listener to do uncomfortable actions—something perverse or something that makes him/her feel dizzy—and many simply do not want to participate in such activities, for whatever social and/or personal reasons. For example, Iannis Xenakis’ electroacoustic work La Légende D’Eer (1977-78), from approximately 13:00 to 19:00, summons us to, among other actions, perform sustained, high-frequency vocal glissandi that often release into shout or scream-like gestures, despite the fact that the sound materials do not originate from a human voice. For some

---

59 This is apparent in various university composition programs or other groups of composers, where one school may champion a certain aesthetics of digital signal processing, and another group frowns upon the same techniques. For example, in certain “glitch” musics, composers seek out analog and digital artifacts that expose the limitations of the technology. In other circles, composers avoid these types of sound material as much as possible. There are also many other factors that influence negative appraisals, such as obvious plagiarism, disliking the composer’s previous works or his personality, or having unhappy memories triggered, etc.


62 The sound materials in La Légende D’Eer stem from three sources: instrumental sounds, including small Japanese hourglass drums called Tzuzumis; various noises, such as scraping against cardboard; and electronically generated sounds. See Iannis Xenakis, Xenakis: Electronic Music I, DVD Mode 148 (New York: Mode Records, 2001).
listeners, performing such vocalizations may produce anxiety or other uncomfortable associations related to previous paralinguistic experiences. Other listeners may enjoy covertly acting out in a primitive fashion, offering their movements as a means to meet halfway with the composition’s sound-objects.

As another example, sounds or phrases that do not fit or seem subpar in a composition can easily rupture the pretense of make-believe upheld by the listener. Curtis Roads comments further on this:

Inappropriate sound objects and dangling phrases inserted into a structure from another context tend to work against their surroundings. If they are not pruned out of the composition, their presence may be detrimental to the structural integrity of the piece. Even a small anomaly of this kind can loom large in the mind of listeners. An inappropriate sound object or phrase can cancel out the effect of surrounding phrases, or call into question the effectiveness of an entire composition.63

Often times, listeners give negative appraisals to sounds whose technological imprint is too strong: aliasing artifacts from pitchshifting too far, too much hiss in a recording, aggressive use of the phase vocoder, or using basic oscillators with little to no parameter modulation are all examples that can diminish the integrity of a work.64 Perhaps we tire of these sounds quickly because they involve rehearsing the same set of motor actions over and over again; we lose our desire to assimilate to a given timbre because it has already been incorporated into our motor repertoire and enacted one too many times. Regardless, negative appraisals of timbre still elicit a vocal response from the listener; even if we do not imitate a sound, we produce another non-mimetic, vocal

---


64 By basic oscillators I am referring to sine, square, sawtooth, and triangle waveforms.
sound derived from our appraisal of that initial sound—whether it be interjections of “ugh” or “meh,” an obscenity, or simply returning to our voice of consciousness to entertain us until the work has finished (provided we haven’t fallen asleep).

In this regard, thus far, our voice seems to play a vital role in critically shaping our understanding of timbre in three possible ways: 1.) we automatically subvocalize every sound we attend to; positive appraisal is giving our attention to the ‘demand’ of the music, and thus, subvocalization is a means by which we empathize with a given timbre, whether we like it or not; 2.) we subvocalize only what we like or desire to be like after a positive appraisal is made: “the desire to know the music as a means physically to be that music;” and 3.) we reject a certain timbre with an ambivalent or defiant vocal reply.65

Although appraisal theory gives some clues as to how our voice may operate in aesthetically evaluating timbre, it is far from definite. Despite this, some interesting questions and creative implications arise. In Steven Connor’s overview of the history of ventriloquism, he describes two different forms of the practice: 1.) “the power to speak through others” and 2.) “the experience of being spoken through by others.”66 In Chapter 2, I argued that our subjectivity is heterogeneous, as we carry many competing and overlapping voices. In this sense, our body is always being spoken by other voices that we cannot fully claim as our own; our voice is “founded in interiority but that the very interiority is always already social.”67 If covert vocal imitation is automatic once we

---

65 Cusick, Queering the Pitch, 75.


accept a piece’s invitation or demand, are we not, in a sense, *ventriloquized* by the sound-objects we attend to? Does this give the composer the power to ‘pull the listener’s strings,’ as he/she would a puppet? If, in coping with a sonic environment, we simulate the actions involved in manipulating perceived objects involuntarily through subvocalization, might we consider listening as a form of virtual glossolalia in which ‘speaking in tongues’ is replaced by *speaking in objects*?

**Expression, Exertion, Emotion**

Computer music researcher Perry Cook claims that “much of our musical expression is also linked closely to the vocal organ.”68 How do we account for this relationship, what role does our subvocal imitation play, and how does this affect the listener’s emotions? Musicologist and philosopher Peter Kivy offers some preliminary clues:

> We must see a visual pattern as a vehicle of expression—a face or a figure—before we see expressiveness in it. Likewise, we must hear an aural pattern as a vehicle of expression—an utterance or a gesture—before we can hear its expressiveness. . . . Indeed, far from being difficult to hear or see things as animate, it is, apparently, difficult not to.69

What are these acoustic patterns or features that draw our attention—are we subconsciously searching for phonemes and formant-based timbres that resemble the speaking and/or singing voice?70 Whose face or voice are we looking for—our

---

68 Perry R. Cook, “Voice Physics and Neurology,” in *Music, Cognition, and Computerized Sound: an Introduction to Psychoacoustics*, ed. Perry R. Cook (Cambridge: The MIT Press, 1999), 105. Cook argues that our understanding of musical expression develops out of vocal imitation and the verbal metaphors that result. He uses the “cuckoo bird” and the “laughing hyena” as examples, as well as the words “smooth” and “round,” which “mimic in articulation the textures and shapes they describe.” See p. 105. Synaesthetic metaphors are further discussed in the ‘Cross-Modal Vocality’ section of this chapter.


70 Computer music pioneer Max Matthews defines formants as “spectral peaks.” In human speech or song “the envelope of the spectrum has peaks at three principal formant frequencies that are
mother’s? Our own? Certainly we look for a semblance of pitch in sounds, since the concept of pitch is an affordance of the human body: “human vocalization is a principal source of periodic sound energy to which human beings have been chronically exposed over both evolutionary and individual time,” providing “a first approximation of the universe of tone-evoking stimuli.” Non-periodic rhythms and pitch contours that resemble the linguistic residue of prosody also attract our attention. In terms of rhythmic character, adds Ray Jackendoff and Fred Lerdahl, “[a] varied flow may correspond more closely to expressive speech, with its clusters of syllables of varying length and interspersed pauses, or in other cases, to a sense of hesitating movement.” What about all the other sounds our voice can produce—all the strange consonants, murmurs, interjections, and other oral peculiarities that poke into our discourse? Are we searching for similarly noisy patterns and anomalies that are acoustically and semantically difficult to quantify? Finally, how do we account for space in all of this? Far too often, scholars use only spectral analysis to depict timbre, describing sounds purely in acoustical terms. However, our understanding of timbre, whether we are aware of it or not, is a multisensorial activity, and the various sensations it produces are “the


perceptual modification[s] of an acoustic signal." Frequency and amplitude are vital components of a given sound envelope, however, these parameters do not exist in a two-dimensional field. Instead, it is important to discuss these parameters in relation to perceived size, scale, contour, texture, and proximity of sound-objects to our listening bodies, all of which our inner voicing of sound may influence. But first, how might the voice contribute to our affective responses while listening to music?

As I argued in Chapter 2, a person's voice is the main carrier of affective signals; we can infer someone's emotional state simply by listening to her paralanguage. To quote Immanuel Kant: “expression has in its context a tone appropriate to its meaning. This tone indicates, more or less, an affect of the speaker and in turn induces the same affect in the listener too.” Do sounds outside the human voice produce a similar mirroring of affect in the listener?

Philosopher Herbert Spencer is perhaps the first scholar to draw connections between music, voice, expression, and affect. In his 1857 essay, “The Origin and Function of Music,” Spencer claims that “music has its germs in the sounds which the voice emits under excitement,” arguing that all music is intimately related to the vocal expression of emotions:

All music is originally vocal. All vocal sounds are produced by agency of certain muscles. These muscles, in common with those of the body

---


large, are excited to contraction by pleasurable and painful feelings. And therefore it is that feelings demonstrate themselves in sounds as well as in movements. . . . And it is from this cause that in human beings bodily suffering expresses itself not only in contortions, but in shrieks and groans—that in anger, and fear, and grief, the gesticulations are accompanied by shouts and screams—that delightful sensations are followed by exclamations—and that we hear screams of joy and shouts of exultation. We have here, then a principle underlying all vocal phenomena; including those of vocal music, and by consequence those of music in general.  

For Spencer, the physiological basis of musical affect directly correlates with vocal embodiment. He believes a “general law that feeling is a stimulus to muscular action;” “[d]ifferent emotions produce different patterns of physiology, which produce different voice patterns.”  

From birth onwards, we develop associations between the sound our voice produces and the corresponding emotion that enacts it. This repository of connections, between feeling states and the muscular movements that precipitate vocal utterances, help us detect similar emotional patterns in music, which, in turn, enact our vocal musculature as “a means of exciting our sympathy with such emotions.” Finally, Spencer brings attention to the subtle, sonic differences our voice can produce:

The muscles that move the chest, larynx, and vocal chords, contracting like other muscles in proportion to the intensity of the feelings; every different contraction of these muscles involving, as it does, a different

76 ibid., 293-294 (my emphasis).

77 ibid., 297. In his studies of touch, which he calls Sentics, scientist and musician Manfred Clynes found that across cultures we systematically identify certain emotional patterns in sounds derived from touch. See Manfred Clynes, Sentics: The Touch of Emotions (New York: Anchor Press, 1977).


80 ibid., 297.
adjustment of the vocal organs; every different adjustment of the vocal organs causing a change in the sound emitted.81

From Spencer’s remarks, we might infer that recognizing different shades of timbral color, in part, depends on the micro adjustments of the vocal organs. If listening is a form of vicarious doing, then our voice provides a means to navigate the subtle nuances and inflections of a given spectral envelope, using our subvocal movements of the tongue, lips, larynx, and velum to mimic various acoustic stimuli.

Spencer’s theories continue to be influential today, as many scholars use them as a basis for their own ideas, with psychologists Patrik Juslin and Petri Laukka’s “superexpressive voice” theory being one such example. The scholars surmise that because music contains acoustical patterns similar to those found in vocal expression, we become aroused by music’s voice-like features, which lead us to internally mimic the perceived emotions:

[W]hat makes a particular music performance of, say, the violin, so expressive is the fact that it sounds a lot like the human voice while going far beyond what the human voice can do (e.g., in terms of speed, pitch range, and timbre). Consequently, we speculate that many musical instruments are processed by brain modules as superexpressive voices. For instance, if human speech is perceived as angry when it has fast rate, loud intensity, and harsh timbre, a musical instrument might sound extremely angry in virtue of its even higher speed, louder intensity, and harsher timbre. The “attention” of the emotion-perception module is gripped by the music’s voicelike nature, and the individual becomes aroused by the extreme turns taken by this voice.82

81 ibid., 294.

82 Patrik N. Juslin and Petri Laukka, “Communication of Emotions in Vocal Expression and Music Performance: Different Channels, Same Code?” Psychological Bulletin 129, no. 5 (2003): 803. Reaching beyond Spencer’s speculations, the authors derive their theory from examining 104 clinical studies on vocal expression and 41 clinical studies on musical performance. Nevertheless, Juslin’s more recent research maintains that “any credible theory of vocal expression must be based on Spencer’s law, and thus must be at least somewhat along the lines of this model. It is hard to find any other credible explanation of the connection between voice and affect.” See Patrik N. Juslin, “Vocal Affect Expression: Problems and Promises,” in Evolution of Emotional Communication: From Sounds in Nonhuman Mammals to Speech and Music in Man, eds. Eckart Altenmüller, Sabine Schmidt, and Elke Zimmermann
Likewise, Huron, citing psychologists Scherer and Oshinsky, shares a similar sentiment regarding the relationship between music, bodily arousal, and vocal output:

[A]nger and fear are associated with high pitch level, ascending pitch contours, and fast event sequences. In vocal production, all of these acoustical characteristics are associated with high metabolic arousal. When we are excited we speak faster, we speak at a higher pitch level, and we are more likely to glide upward than downward in pitch. These same features are often evident in musical climaxes. . . . When we are relaxed, both the tension in our vocal cords and the activating pneumatic energy are relatively low, and so the resulting spectrum contains relatively little energy in the upper partials. Conversely, when we are aroused or anxious, increased tension in the vocal cords and increased pneumatic energy tend to evoke a subjectively “harsher,” less mellow timbre.\(^{83}\)

Finally, Jackendoff and Lerdahl give their take on vocal embodiment and affect:

Not only human but much mammalian communication modulates vocal pitch, volume and timbre to convey threat, reconciliation, fear, excitement, and so on. . . . These modulations can be carried over into musical performance, sometimes in the character of melodic contour, but often also in the performer’s manipulation of vocal or instrumental tone production. . . . Listeners respond affectively to such manipulations in the same way as they respond to the corresponding vocal communications—that is why we can speak of “sighing violins.”\(^ {84}\)

Both Juslin and Laukka and Huron give examples of a voice at heightened levels of arousal; a body whose acoustic output contains a high spectral centroid and that articulates at a rapid pace.\(^ {85}\) Many of us can recall what it is like to create such a

---


\(^{84}\) Jackendoff and Lerdahl, “The Capacity for Music: What is It, and What’s Special About It?,” 63.

\(^{85}\) Spectral centroid is the center of amplitude-weighted frequency distribution and is most often correlated with perceptual accounts of “brightness,” “sharpness,” or “nasality.” See Roger A. Kendall, Edward C. Carterette, and John Hajda, “Perceptual and Acoustical Features of Natural and Synthetic Orchestral Instrument Timbres,” *Music Perception* 16, no. 3 (1999): 327-63.
sound—the physical energy required and the resulting postures, movements, and facial expressions. Of course, acoustic instruments, both individually and in combination, can naturally produce the ‘harsh’ timbres that Huron describes, depending on how they are played and the listener’s proximity to the instrument(s). However, composers and musicians can also fabricate and emphasize such overtone-rich sounds using a signal processing technique called *distortion*.\(^6^6\) Musicologist Robert Walser draws parallels between distorted, electric guitar timbres and the human voice at its brink:

> Not only electronic circuitry, but also the human body produces aural distortion through excessive power. Human screams and shouts are usually accompanied by vocal distortion, as the capacities of the vocal chords are exceeded. Thus, distortion functions as a sign of extreme power and intense expression by overflowing its channels and materializing the exceptional effort that produces it.\(^6^7\)

Such aggressive sounds excite listeners because they flirt with the limits of the human body, and in other cases, extend beyond what a human can do. Although distorted guitar timbres may be falling out of favor in popular music production, high-energy, harmonically-saturated sounds, whether derived from acoustic sound-sources or synthesis, are still prevalent in today’s music.\(^6^8\) We can observe such sonic power in the bass sounds of electronic dance music, where extremely low fundamental frequencies are layered with bright oscillators from FM, additive, and/or subtractive synthesis and

---

\(^6^6\) Distortion adds even and/or odd harmonics to a signal, depending on the shape of the wave it is being multiplied by: hyperbolic tangent, arcus tangent, linear hard clip, etc.


\(^6^8\) There are variety of means to harmonically distort an audio signal: overdriving analog circuits found in various tube and solid state amplifiers is one such way; using digital signal processing for waveshaping and soft or hard clipping audio is yet another.
often modulated with formant filters, exhibiting growl-like acoustic characteristics. For example, Skrillex’s *Scary Monsters and Nice Sprites* (2010) contains numerous phrases that alternate between siren-like synthesizer gestures and growl-like utterances in a call and response fashion. The growl-like sounds have a fundamental frequency that glissandi slightly upward starting at 49Hz (G0), but they also contains stacks of high-amplitude, inharmonic partials that extend up to 6 kHz and beyond.

Figure 3-1. Sonogram and waveform of Skrillex’s *Scary Monsters and Nice Sprites*, 0:41–0:51.

---

Chen-Gia Tsai et al. further discuss these types of timbres:

In addition to a loud dynamic level, the growl-like timbre acoustically is characterized by the presence of inharmonic components that are distributed among adjacent harmonics in a spectrum. According to psychoacoustic models of roughness, these inharmonic components add a rough quality to a growl-like sound because they interfere within critical bands. While the unpleasant quality of the growl-like timbre can be explained by psychoacoustic models of roughness, we argue that its aggressive characteristic is closely related to the *muscle recruitment* for growl production.\textsuperscript{90-91}

Growl-esque sounds, like those found in electronic music, may be disagreeable for certain listeners, but for others, they enhance “the gesture of rush, rage, excitement, aggression, pain, fear, anger, emotional chaos and ecstasy.”\textsuperscript{92} In this case, synthetic sound-objects afford listeners simulated vocal exertions that would require an immense amount of physical energy to produce, and possibly from a body much larger than a human figure.\textsuperscript{93} Likewise, noise music, such as the work by Japanese musician Merzbow, distorts sounds to the point of listener fatigue and pain. In the piece *Uzumushi* (2015), Merzbow layers bass ostinati with sustained, high-frequency material,


\textsuperscript{91} Neuroscientist Josh H. McDermott defines *roughness* as “the perceptual correlate of fluctuations in energy (intensity) that occur over time, analogous to the fluctuations in surface depth that determine the roughness of an object to the touch. Fluctuations at rates between ~20-200Hz are those that determine roughness; any lower, and the fluctuations can be heard individually rather than contributing to a sound’s timbre.” See Josh H. McDermott “Auditory Preferences and Aesthetics: Music, Voices, and Everyday Sounds,” in *Neuroscience of Preference and Choice: Cognitive and Neural Mechanisms*, eds. Raymond Dolan and Tali Sharot (London: Academic, 2012), 229-230.


\textsuperscript{93} Psychologists Belin, Fecteau, and Bedard note that formants “carry important identity information: they are directly related to the size of the vocal tract and can therefore provide estimates of body size.” See Pascal Belin, Shirley Fecteau, and Catherine Bedard, “Thinking the Voice: Neural Correlates of Voice Perception,” *TRENDS in Cognitive Sciences* 8, no.3 (March 2004): 130.
with particular emphasis in the region of 1-4 kHz, most of which falls within our most sensitive area of hearing.\textsuperscript{94}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{image}
\caption{Sonogram and waveform of Merzbow’s \textit{Uzumushi}, 0:00–2:40.}
\end{figure}

\textit{Uzumushi’s} rough acoustic properties and high-frequency saturation may elicit reaction responses out of listeners; many of the gestures exhibit qualities found in natural and artificial alarm signals that are useful for quickly evaluating danger.\textsuperscript{95}

\begin{flushright}
\textsuperscript{94} Computer music pioneer Max Matthews says “the auditory system is most sensitive at a frequency at about 2000 Hz.” The auditory canal that terminates in the eardrum is a closed tube that has resonant frequencies in this region, boosting signal levels up to 30 decibels. See Max Matthews, “What is Loudness?,” in \textit{Music, Cognition, and Computerized Sound: an Introduction to Psychoacoustics}, ed. Perry R. Cook (Cambridge: The MIT Press, 1999), 71-78.

\end{flushright}
For most if not all of *Uzumushi’s* distorted gestures, the physical load required to produce comparable vocalizations would result in bodily harm, such as the formation of vocal nodules. Although many consider noise music as masochistic, some listeners enjoy the catharsis they receive from covertly screaming and shouting with the sounds.

![Figure 3-3. Roughness in Merzbow's *Uzumushi*, 1:51–1:53](image)

For these people, noise music provides “a means of gaining new or alternative sensuous experiences.”96 The music becomes a safe means to simulate extreme affect,

---

without inflicting harm on others or damaging the body.\textsuperscript{97} So, while \textit{Uzumushi} may not actually use a screaming voice as its sound-source, we re-present the heavily-distorted sounds, which exhibit varying degrees of bright and rough acoustic properties for sustained periods of time, as scream-like through emulated vocal exertion.\textsuperscript{98}

Inharmonic and noise-based sounds can also operate in a reverse direction, bringing about delicate movements on the verge of being as hidden and secret as our subvocalization. Exertion is pulled back, into the mutters, mumbles, and whispers that release a steady flow of breath. For the listener, these sounds do not \textit{need} to originate from another human voice; we can hear the whisper of the wind in the trees or in the bow against the body of a cello. We hear the murmurs of machinery dampened by the distance and the walls at home that enclose us, and the rain that lightly chatters across our car windshields. Natasha Barrett uses similar aural delicacies in her electroacoustic work \textit{Gentle Sediment (Barely: Part-3)} (2008). Between, 0:00-0:30, she uses overlapping, breathe-like gestures, which contain slow attack and release time envelopes, articulated in a cavernous space. Barrett reveals a variety of human vocalizations as the piece progresses, along with countless ambiguous sounds whose sources have been obscured or hidden, all of which trigger our paralanguage.

Although Spencer and his contemporary followers depict the more extreme poles of muscular movement and emotion—from calm stillness to flying anger—our imaginations can fill in the continuum between these opposites and observe how our inner, imitative voice acts as an \textit{empathic antenna} that fuels sonic animism,

\textsuperscript{97} Obviously, listening to such distorted timbres can harm one’s hearing if played too loudly.

\textsuperscript{98} \textit{Brightness} is a term often used to describe a sound’s proportion of high-frequency energy.
synchronizing itself with just about anything that moves. Nevertheless, the animism we experience through our vocal embodiment points to something significant: ultimately, it is we who assign expressivity to music through our mirroring responses. This counters the idea, as classical mimesis postulates, that music is an expression of something: a place, a person, a geological formation, etc. To do so would imply that music is an “organism equipped with intentional capabilities,” when, in fact, it is not.99 Rather, music seems expressive because the listener perceives her body’s arousal from the music—i.e. the music is sad, because we feel sad. Adds Peter Kivy, “expressive music only resembles expressive behavior. For in most cases it no more represents it than the face of the Saint Bernard represents a sad countenance.”100 In Kivy’s example, we perceive the dog as feeling sorrowful through our own subjective mirroring of the face. In imitating the dog’s countenance, the corners of our mouth turn downward and our eyebrows raise up, creating a gesture that we associate with being unhappy and project onto the animal. Similarly, music’s ability to express depends on the listener’s mimetic response and observation of this response, thus emphasizing the reciprocity between action and perception.

**Mishearing and Vocal Reflexivity**

Recall Bossis’ claim that “nothing opposes the idea that sound can be treated as vocal, despite having no original link to the phonation organ.”101 Although there may be


certain acoustic features that mark traces of vocal behavior more so than others, there is no definitive set of parameters which makes something voice-like. Moreover, following De Certeau, given our obstinate pursuit to find meaning in everything, the very activity of 'looking for voice' may turn into an auditory confirmation bias—where probing for a pattern that resembles vocality, even when non-existent, reinforces that pattern. Human hearing has a propensity for mishearing acoustic signals, with the restorative process of auditory induction being one such example:

Auditory induction appears to be a quite useful perceptual phenomenon permitting a highly selective reinstatement of sounds which would otherwise be lost through masking. The listener can thus establish a simpler and more stable interpretation of his auditory environment than the intermittent extraneous sounds present in our noisy world would otherwise permit.102

Our auditory perception often acts as a gap-filler, adding information that might not have been there to begin with. Regardless, adds Fales, “[w]hether or not the auditory system makes actual mistakes in processing acoustic information, there is clear evidence that it shapes a percept beyond the actual features of the signal according to priorities of its own.”103 If so, then we may certainly hear a voice in white noise if we gaze into the 'acoustic mirror' long enough.

In Chapter 2, I argued that human beings privilege the sound of voices. In the spirit of Michel Chion, “[c]all this vococentrism if you will. Human listening is naturally

---


103 Fales, Wired for Sound, 163.
vococentrist.” From an ecological perspective, we privilege all sounds that resemble human vocality because they are an affordance of our body. If we actively search for stimuli in our environment that provide perceptual significance, would this not mean, in some sense, that we bend all sound into the condition of voice? If so, then perhaps the self-reflexivity of Derrida’s s’entendre parler (to hear oneself speak) infiltrates our entire listening process, not simply our voice of consciousness. Citing audiologist Alfred Tomatis’ concept of the “audio-phonatory loop,” Chion claims that “one can only emit vocally what one hears,” meaning, “someone who can no longer hear certain frequency zones can no longer attain them with her voice.” Perhaps the inverse of this statement is equally valid: we can only listen to that which we can vocally approximate. If true, then are we not behaving similar to Narcissus, enamored by sounds that reflect our voice back to us? If so, to cite Marshall McLuhan, is our potential for digital vocality “in the true Narcissus style of one hypnotized by the amputation and extension of his own being in a new technical form?” Or, perhaps the acousmatic image, with its potential for rapid, spatial counterpoint, can continuously deflect our gaze, helping us to listen outside our own reflection, and instead, immerse us into experiences of

---


105 The notion of s’entendre parler contaminating the entire listening experience can be found in Chion, *Sound: An Acoulogical Treatise*, 15. However, I arrived at this idea independently, prior to reading his text. Nevertheless, I take the discovery of Chion’s hypothesis as a good indicator of its validity.


Cross-Modal Vocality

Musicologist Rafael Ferrer says “it is evident that individuals’ embodiment of timbre is mostly visual and tactile.” Referring to our sense of touch, music psychologists Zohar Eitan and Inbar Rothschild argue that “tactile metaphors for musical sound are neither coincidental nor subjective, but relate systematically to basic qualities of sound . . . a rich, yet consistently applied, array of cross-modal connotations underlies the perception of auditory stimuli.” Regarding our vision, musicologist Simon Emmerson suggests that “sound has the power to create its own visual response in humans—one which is sometimes not accounted for by visual artists—a sense of place, of aural landscape.” Likewise, for anthropologist Tim Ingold, “it is the very incorporation of vision into the process of auditory perception that transforms passive hearing into active listening.”

Discursive descriptions of sound such as ‘wide,’ ‘sharp,’

---

108 Electroacoustic composer Denis Smalley defines the acousmatic image: “From sound alone I may constitute, in my imagination, the configurations of a scene or narrative, or a spatial climate; the visual and physical appearances of objects and textures; the motions of entities and shapes in spectral space. These may have a certain reality if I think I can identify possible sources or causes of sound, but they may equally be fantasy, a kind of ‘vision’ where the auditory sense calls on companion senses to participate in interpretation. This conjunction of imagined sensory phenomena may be captured in the idea of the ‘acousmatic image.’” See Denis Smalley, “The Acousmatic,” in The Cambridge Companion to Electronic Music, eds. Nick Collins and Julio d’Escrivian (New York: Cambridge University Press, 2007), 79-80.

109 Rafael Ferrer, “Timbral Environments: An Ecological Approach to the Cognition of Timbre,” Empirical Musicology Review 6, no. 2 (2011): 68. Ferrer asserts that our most common descriptions of sounds are visual and tactile metaphors, such as “colorful - colourless” and “dull - sharp.” See p. 68.


‘warm,’ ‘thin,’ ‘smooth,’ and ‘bright,’ are all *synaesthetic metaphors* that refer to other sensory domains. Moreover, these descriptions are not simply passive features; they are active, virtual engagements with sound via motor resonance. How might our vocality enable us to leap into the perceptual domains of touch and vision?

To neuroscientists Vilayanur Ramachandran and Edward Hubbard, synaesthesia is “a curious condition in which an otherwise normal person experiences sensations in one modality when a second modality is stimulated” and is perhaps “the neural basis for metaphor.”¹¹³ One commonly cited example is the experience of a certain color while listening to a particular sound. Synaesthesia, however, can entail different combinations of all of our senses. Ironically, such *cross-modal* perceptions seem like “a curious condition,” to follow Ramachandran and Hubbard, only because of science’s discretization of the senses.¹¹⁴ Merleau-Ponty explains further:

> Synaesthetic perception is the rule, and we are unaware of it only because scientific knowledge shifts the centre of gravity of experience, so that we have unlearned how to see, hear, and generally speaking, feel, in order to deduce, from our bodily organization and the world as the physicist conceives it, what we are to see, hear and feel. . . . I hear the hardness and unevenness of cobbles in the rattle of a carriage, and we speak appropriately of a ‘soft’, ‘dull’ or ‘sharp’ sound. Though one may doubt whether the sense of hearing brings us genuine ‘things’, it is at least certain that it presents us, beyond the sounds in space, with something which ‘murmurs’, and in this way communicates with the other senses.¹¹⁵

---


¹¹⁴ Cross-modal perception “denotes the cooperation of the different sense modalities.” For example, a certain set of acousmatic sounds, when spread across the stereo field, may make the listener imagine a visual texture that recalls our sense of touch with our hands and/or mouth. See Rolf Inge Godoy, “Motor-Mimetic Music Cognition,” *Leonardo* 36, no. 4 (2003): 317.

¹¹⁵ Merleau-Ponty, *Phenomenology of Perception*, 266.
We do not typically think of sound as something we can see, smell, touch, and taste; many think of sound as something we hear only with our ears. Numerous electroacoustic composers and scholars have written about the simulation of quasi-visual imagery while listening to works, and some even use these images as a compositional tool.\textsuperscript{116} Moreover, experiments such as Ramachandran and Hubbard's notable "bouba/kiki" experiment, originally developed by Köhler (1929, 1947) and Werner (1934, 1957), demonstrate that listeners may experience cross-modal imagery in a similar manner.\textsuperscript{117} The researchers found a "non-arbitrary synaesthetic correspondence between visual object shape . . . and sound contours represented in the auditory cortex."\textsuperscript{118} In the experiment, subjects listened to nonsense words and then assigned each word to a choice of visual figures. They found that subjects map nonsense words with rounded consonants and vowels, such as 'bouba,' to rounded shapes and nonsense words with unrounded consonants and vowels, such as 'kiki,' to sharp figures.\textsuperscript{119-120}


\textsuperscript{118} ibid., 21.

\textsuperscript{119} ibid., 3-35.

\textsuperscript{120} Luca Nobile’s more recent study (2015) in sound-shape symbolism also suggests that “different phono-articulatory features tend to be associated to different graphic-visual properties and that their acoustic characteristics play an important role in determining these associations.” The study relates oppositional pairs of acoustic features ([grave : acute], [continuous: discontinuous], [alveo-dental: palato-velar]), to articulatory features ([fricative, plosive],[sonorant: obstruent], [nasal, oral]), and graphic features ([large : small], [dark : bright], [thick: thin], [sparse, dense], [curved, angular]). See Luca Nobile,
Although a simple test, it demonstrates non-arbitrary links between subvocalization and auditory imagery; the kinematics of the mouth—the various movements of the tongue, lips, teeth, and palate—may help us simulate surface texture, shape, size, and scale of perceived sound-objects. This highlights another important feature of vocal embodiment: tactility. Motor-mimetic simulations from subvocal participation include vibrotactile adaptations to sound, which turn “hearing, listening, and the voice into kinds of touching, modalities of haptic approach or appropriation.”

Hyperproximity and Oral Entanglements

Advances in recording technology, amplification, and digital signal processing magnify the seductive qualities of voice and sound. Rather than listening to a vocalist from a fixed distance in a concert hall, singers can appear right beside our heads, intimately tickling our ears with whispers, breaths, and croons. We can periscope down


Derrida, On Touching—Jean-Luc Nancy, 147.
into the depths of the mouth, watching a tongue slosh around saliva as we plunge further towards the belly. We can also reach for a specter of a voice, whose smooth features resound at a distance that far exceeds the bounds of any performance space. Steven Connor comments further on the haptic and proximate qualities of sound that microphones help accentuate:

The sound of the microphone is a touched sound. It not only selectively amplifies the noises of the mouth’s own slidings, impacts, poppings and palpations, it also joins to these noises the sound of the tactile contact between the mouth and itself: the bangs, bumps and scratches. The distinction between classical and popular performance has come to be a distinction between self-projective singing at a distance from the microphone and close, wet and "dirty" proximity to it.¹²²-¹²³

Microphones, coupled with careful digital signal processing, allow for drastic alterations of sound-object scale: “where a large image size results in a close, wide sound, while a point source may hover, at any chosen point in the stereo width, ambiguously close or far, depending on its relation to other sounding objects or to room acoustics.”¹²⁴ If the composer is skilled enough, “the origin of the sound can appear to be closer to the listener than the point of the loudspeaker.”¹²⁵ The decrease in “the apparent spatial


difference between the sound location and the listener creates a tighter bond to the listener’s physical space.” On top of this, as Barrett argues, the “closer the sound-object to the listener, the larger its perceived sound image.” Finally, the closer a sound-object appears, the more accurately we can evaluate its surface roughness, or texture. Overall, sound recording and signal processing techniques create listening experiences of hyperproximity by (1) zooming into surface textures and fine details of sounds, much like a camera lens or photographic blowup technique does with visual images; and (2) creating close and wide sound-objects that jump into the foreground and synaesthetically occupy most of, if not all of, our visual field. This not only highlights the ability to inhabit new perspectives by enlarging and shrinking features that often lie outside the human scale of perception, but also helps create large disparities between foreground and background information.

---

126 ibid., 318

127 ibid., 314

128 Psychologists Roberta Klatzky and Susan Lederman argue that texture is a “multisensory phenomenon,” and that “[a]pects of texture such as surface roughness can be represented by means of touch, vision, and audition.” Attenuating a sound-object’s high frequency energy, through techniques such as low-pass filtering, digital reverberation, or distant microphone placements, decreases sensations of tactile roughness; textures grow smoother, much like a sharp-edged rock that becomes rounded and sanded down by the ocean’s tide. See Roberta L. Klatzky and Susan J. Lederman, “Multisensory Texture Perception,” in Multisensory Object Perception in the Primate Brain, eds. Marcus J. Naumer and Jochen Kaiser (New York: Springer, 2010), 226.

129 Hyperproximate sound-objects can be created by placing a stereo pair of microphones six inches or closer to the source being recorded. One can also emulate this with monaural recordings using time-based signal processing techniques such as delay (Haas effect) or modulating delay (stereo chorusing), as well as equalization. In addition, binaural recordings, in which two small, lapel condenser microphones are mounted on each side of eyeglasses, hats, or inside the ears, create an extremely wide stereo field due to their placement. With this technique, the listener experiences a similar aural perspective as the person doing the recording.

130 For example, an audio recording can be rate changed to twice or quadruple the length of its original rate, lowering extremely high frequencies into a more audible range for human hearing. Higher sampling rates of 96kHz or 192kHz, coupled with newer microphone technology, whose frequency range extends up to 40Hz and even 100Hz, allow for more drastic rate/pitch changes with fewer artifacts.
separation can operate in an almost confrontational way: high-resolution textures appear right before our faces, exciting vibrotactile sensations of rubbing and friction in our mouths, as if we are discharging them ourselves. For example, M39 Diffain (2013) by the British electronic duo Autechre elicits such sensations, and perhaps does so without using non-vocal sound-sources. Rob Brown, one half of the duo, describes some of the recordings used for the work:

[I]t was made up of disparate elements . . . and a lot of found sounds, a lot of field recordings going into it—rain, New York, bits of city life, bits of weather, direct recordings of rain on certain objects mic'd-up—things like that.

Although the work provides a myriad of ways for vocal simulation throughout, one particular moment stands out. From 4:10–4:13, an ambiguous sound object—one that pokes up sporadically in the previous three minutes of the work—rips through the foreground. This sound-object, whose width covers the entire stereo image, contains two important timbral features composited together: (1) a glissando that descends and re-ascends, from approximately D6, down to F#5, and back up to D6; and (2) an acoustically rough, noised-based texture that also evokes a quasi-visual texture containing many sharp, surface particles. Despite the fact that the sound probably did not originate from a human voice, this sound-object, nevertheless, provides us a virtual means to scream through material that resembles a coarse membrane stretching and on the verge of collapse.

---


As digital converters and amplification improve to almost negligible amounts of self-noise, sound seems to get closer and closer to us, with more clarity, more spatial precision, and thus, greater haptic presence. Moreover, such technological improvements increase the plasticity of digital audio—the composer can repeatedly layer and shape audio, juxtaposing and reconfiguring materials with fewer sonic

---

133 High frequency energy is an indicator of a sound’s proximity; high frequencies also enable more precise spatial localizations in both the horizontal and vertical dimensions, and thus, more accurate clues to a sound’s behavior. For more on sound localization, see John C. Middlebrooks and David M. Greenhaw, “Sound Localization by Human Listeners,” Annual Review of Psychology 42, no. 1 (1991): 135-159.
consequences. \(^{134}\) Hyperproximate sound-objects are at once seductive and repulsive: seductive because their shiny features lure us in to lose our everyday inner speech and touch the not-yet-symbolized, and repulsive because they encroach into our physical space and ignore the boundaries of our body. Furthermore, this heightened sense of proximity creates an erotics of presence that catalyzes experiences of hybridity—diffracting, scattering, and mutating our own image into the sounds we entangle ourselves with. \(^{135}\)

Hyperproximity highlights the significance of skin as a negotiating membrane between listening subject and sound-object, mediating virtual acts of material-based introjection and expulsion. We often associate touch with actions related to our hands and fingertips, but the most investigative portion of the skin may be inside the mouth. Didier Anzieu describes the significant partnership between skin and mouth:

The primary function of the skin is as the sac which contains and retains inside the goodness and fullness accumulating there through feeding, care, the bathing in words. Its second function is as the interface which marks the boundary with the outside and keeps that outside out; it is the barrier which protects against penetration by the aggression and greed emanating from others, whether people or objects. Finally, the third function—which the skin shares with the mouth and which it performs at least as often—is a site and a primary means of communicating with others, of establishing signifying relations; it is, moreover, an ‘inscribing surface’ for the marks left by those others' body. \(^{136}\)

\(^{134}\) For example, a single, gestalt sound-object can be composited from a multitude of recordings and/or other sound-objects without worry of noise or hiss build up and a loss of clarity.


Our voice leaves from the mouth, but the mouth is also an opening where acoustic vibrations come in and get reshaped by muscular enactment. Asks philosopher Jean-Luc Nancy: “Isn’t the space of the listening body, in turn, just such a hollow column over which a skin is stretched, but also from which the opening of a mouth can resume and revive resonance?”

If subvocalization is automatic, then perhaps it is initiated first by acoustic vibrations encroaching and pushing up against the body, and in particular, the mouth. This may be the very 'demand' that sound enacts upon us, whether we are willing or not. Our mouth’s haptic contribution to timbral perception potentially brings new life to sounds we typically deem as abstract or non-referential. For example, composers’ fascination with microsound, from a listening perspective, suddenly seems more related to oral fantasies of chewing, chomping, biting, crunching, slurping, and spitting than it does to grain distributions and window types. As composer and musicologist Katharine Norman declares, “it’s almost like eating. Touching and being touched by sounds.”

---


138 Michael Filimowicz and Jack Stockholm note the difference between abstract and representational sounds: “the difference between an abstract and a representational sound is the difference between a sound which seems to have a source or to have been produced by some known thing (the referent—representation), while a sound that is abstract seems to generate its source. In other words, rather than being produced by its source, the abstract produces its source(s), which emerge in a play of tangents and possibilities of suggestiveness.” See Michael Filimowicz and Jack Stockholm, “Towards a Phenomenology of the Acoustic Image,” *Organised Sound* 15 (April 2010): 6.

139 The term *microsound* is often used to describe extremely small particles of sound in a composition and are often created through a technique called granular synthesis or sampling, although many everyday sounds exhibit these features, such as the flowing of water over rocks. Curtis Roads claims that microsound “extends from the threshold of timbre perception (several hundred microseconds) up to the duration of short sound objects (~100 ms).” See Curtis Roads, *Microsound* (Cambridge: The MIT Press 2001), 20-21.

touch; objects cannot touch one another, and rather, the sensations we feel when touching something are the repellent forces between objects:

Touch, for a physicist, is but an electromagnetic interaction. A common explanation for the physics of touching is that one thing it does not involve is . . . well, touching. That is, there is no actual contact involved. You may think you are touching a coffee mug when you are about to raise it to your mouth, but your hand is not actually touching the mug. Sure, you can feel the smooth surface of the mug’s exterior right where your fingers come into contact with it (or seem to), but what you are actually sensing is the electromagnetic repulsion between the electrons of the atoms that make up your fingers and those that make up the mug. Electrons are tiny negatively charged particles that surround the nuclei of atoms, and having the same charges they repel one another, much like powerful little magnets. As you decrease the distance between them the repulsive force increases. Try as you might, you cannot bring two electrons into direct contact with each other. . . . That is the tale physics usually tells about touching. Repulsion at the core of attraction. See how far that story gets you with lovers. No wonder the romantic poets had had enough.  

Perhaps our subvocalization acts as an acoustic resonance gatekeeper—a boundary negotiator, pushing back on sounds that get too close. Or, on the other hand, is it an indicator that our bodies do not necessarily “end at the skin?”  

Asks Merleau-Ponty, “[w]here are we to put the limit between the body and the world, since the world is flesh?” Our voice is always lurking to disrupt these boundaries we stubbornly insist on—we just need to simply listen.  

Nevertheless, the sensations of proximity, force, and friction we experience while coping with electroacoustic sound-objects reveal that vocal embodiment is not as simple


as listening to other humans vocalize, nor is it purely singing or shouting along with the
music; our vocality is a multisensorial, performative act that uses virtual, paralinguistic
approximation—including haptic and visual participation—to become and behave like
something else. As our audio technologies continue to mediate and remediate sound
into our everyday environments, our very definitions of voice can only follow suit. If our
perception of timbre is something our voice participates in while we give our attention to
music, then our vocality is only bound by the sheer number of material entanglements
that both the listener can imagine and the composer can reveal.
CHAPTER 4
TOWARD A THEORY OF HYPERVOCALITY

Dissolution of Voice and Re-enchantment

Today, the shock and awe of voices once heard from Wolfgang von Kempelen’s speaking machine (1769), the radio broadcast of H.G. Wells’ *The War of the Worlds* (1938), and Thomas Edison’s phonograph (1877) have washed away, as we rarely stop to consider how strange our technological ventriloquisms actually are.¹ Mladen Dolar discusses the commonplace effects of media technology and the human voice:

Radio, gramophone, tape-recorder, telephone: with the advent of the new media the acousmatic property of the voice became universal, and hence trivial. They all share their acousmatic nature, and in the early days of their introduction there was no shortage of stories about their uncanny effects, but these gradually waned as they became common, and hence banal.² Consequently, adds Steven Connor, we “allow the voice to be so easily replicated, redoubled and impersonated,” increasing “our capacity to live with so little disturbance in a world of voice-doubles.”³ While our technological ventriloquisms can revitalize the forgotten fantasy of omnipotent voices, in actuality, we seem to no longer care.

Moreover, despite the increasing material consequences of voice and sound, sadly,

---


there seems to be a decreasing perceptual awareness of the two. Media theorist Frances Dyson comments on this further:

Despite the fact that most of the word’s sense-making occurs through various technological devices and sounds within physical spaces, the relationship between the output device and the room in which it is heard in the making of sense is rarely questioned. In other words, the actual “sound” of the media is ignored, as are the conditions of hearing it.\(^4\)

For many, hearing a voice articulate clear language takes firm precedence over any other acoustic considerations; as long as we can imagine a human body and face uttering controlled and predictable speech, our alert system suspends, and we move on to consuming the next set of voices. In this sense, our audio technologies act as “a conduit through which the presence of the voice, in its metaphysical guise, can be reasserted.”\(^5\) Add to this the effect of our excessive intake of online voices, through social media or text messages, and we increasingly “lose capacity acoustically to respond to a reality which repeatedly addresses us.”\(^6\) Media theorists Rosalind W.


\(^5\) Frances Dyson, *Sounding New Media: Immersion and Embodiment in the Arts and Culture* (Berkeley: University of California Press, 2009), 51. Dyson’s claim that technologically mediated voices have merely become an extension of the voice of consciousness stems from her earlier concept of “the dominant radio voice.” “It does not mumble or stutter, it pronounces full and meaningful sentences, it says something . . . having a certain timbre and intonation that suggests a belief in what it is saying and a degree of authority in saying it. . . . Critics of the dominant radio voice have dubbed it ‘the voice of authority.’ Yet, although this voice appeared with the advent of radio and seems therefore to have sprung from the 20th century, it in fact has been a long time in formation and has accumulated a host of characteristics that connect it to the deepest symbolic and epistemological structures governing thought, speech, and media in western culture. See Frances Dyson, "The Genealogy of the Radio Voice," in *Radio Rethink: Art, Sound and Transmission*, eds. Daina Augaitis and Dan Lander (Banff: Walter Phillips Gallery, 1994), 167.

Picard and Jonathan Klein describe some of the consequences of our online interaction with digital devices:

Today’s systems limit your ability to see facial expressions, hear tone of voice, and sense those non-verbal gestures and behaviors that might otherwise help you disambiguate a hastily-sent, non-angry message from a genuinely angry one. To the extent that people spend more time communicating with each other via technology without sufficient affect channels, they may actually be reducing some of their emotional skills—a kind of `use it or lose it’ opportunity cost.  

As Dyson adds, “we believe that we understand a voice without resonance—or, more frankly, a voice without sound.” Furthermore, all we seem to care about is the racketing of our own voices back and forth—“shouting, whispering, crying, caressing, threatening, imploring, seducing, commanding, pleading, praying, hypnotizing, confessing, terrorizing, declaring”—our consumption of voices has turned into a vat of brown noise, where “all that is visible is the sticky problem of humanity’s own captivity within language.” And, contends David Appelbaum, “stoppered with phonemic domination, we never hear.” Our historical moment is vastly approaching deafness.

All of these issues place today’s composer in a difficult position. In addition to our culture’s decreasing ability to listen, sonic novelty in electroacoustic music has become increasingly rare. Esoteric sounds that once took countless hours to create are now

---

7 Rosalind W. Picard and Jonathan Klein, “Computers That Recognize and Respond to User Emotion: Theoretical and Practical Implications,” Interacting with Computers 14 (2002): 157. Granted, emoji use has helped alleviate some miscommunication, however, they often induce a hyperbolic, emotional rounding in the receiver of the message. Nevertheless, these icons are no substitute for sound’s ability to transmit affect.

8 Dyson, The Tone of Our times, 98.


10 Karen Barad, Meeting the Universe Halfway (Durham: Duke University Press, 2006), 137.

commonplace and easily replicated using today’s commercial software. Likewise, we hear robot-like voices processed through vocoders and autotune and it doesn’t even phase us. More than ever, the composer has to find ways to enchant the listener through compositional inventiveness that overrides the banality of many electroacoustic sounds. Political theorist Jane Bennett discusses the significance of enchantment:

> Without modes of enchantment, we might not have the energy and inspiration to enact ecological projects, or to contest ugly and unjust modes of commercialization, or to respond generously to humans and nonhumans that challenge our settled identities. These enchantments are already in and around us.

One important duty of today’s composer is to enchant listeners about the powers of sound and its relationship to our ever-changing vocality. More specifically, how can the composer enhance the listener’s “receptivity to ‘propositions’ not expressed in words,” whether they originate from a human voice or not? How can the composer illuminate “what sounds from a human throat without being language, which emerges from an animal gullet or from any kind of instrument, even from the wind in the branches: the rustling toward which we strain or lend an ear.” As I have argued, our listening vocality allows us to examine how things are entangled—humans, animals, and objects—into a

---

12 Autotune is digital signal processing, plug-in effect that was created to correct poor vocal and instrumental tuning. However, it became a means to quantize pitch between a few selected intervals, creating a distinct effect that is used heavily in popular music today.


“plurality of voices that are already linked to one another in resonance.”\textsuperscript{16} The idea that “no fundamental line of demarcation can be drawn between the sounds of nature and of human speech” offers a productive means for the composer and the listener to explore the vocal continuum that exists among all sounds.\textsuperscript{17} In doing so, we can renew the sense of uncanniness our ancestors once felt from a distant clap of thunder or the mutters of a reverberant cave—a feeling where “the familiar grows strange, and the strange familiar.”\textsuperscript{18} However, as Sigmund Freud notes, “[t]he better orientated in his environment a person is, the less readily will he get the impression of something uncanny in regard to the objects and events in it.”\textsuperscript{19} How, then, can the composer make audible “the irreducibility of in-between spaces, polyphonic hybridization, [and] multiple sonic interferences,” disturbing the relation between the self that listens and the surrounding environment?\textsuperscript{20}

\textbf{Attention, Distraction, Counterpoint}

Today, “we are in the midst of a generational shift in cognitive styles.”\textsuperscript{21} Literary critic N. Katherine Hayles further discusses this shift, distinguishing between deep attention and hyper attention:


Deep attention, the cognitive style traditionally associated with the humanities, is characterized by concentrating on a single object for long periods (say, a novel by Dickens), ignoring outside stimuli while so engaged, preferring a single information stream, and having a high tolerance for long focus times. Hyper attention is characterized by switching focus rapidly among different tasks, preferring multiple information streams, seeking a high level of stimulation, and having a low tolerance for boredom. . . Deep attention is superb for solving complex problems represented in a single medium, but it comes at the price of environmental alertness and flexibility of response. Hyper attention excels at negotiating rapidly changing environments in which multiple foci compete for attention; its disadvantage is impatience with focusing for long periods on a noninteractive object such as a Victorian novel or complicated math problem. In an evolutionary context, hyper attention no doubt developed first; deep attention is a relative luxury, requiring group cooperation to create a secure environment in which one does not have to be constantly alert to danger.²²

Are deep and hyper attention mutually exclusive, or can both coexist in a virtual realm? In the context of electroacoustic music, we can immerse ourselves into a sonic environment of make-believe, and yet, many high-detailed sound-objects can command and compete for our attention. This is one of the many rewards of listening to carefully-crafted works: encountering vast streams of information that we cannot completely absorb in one listening session. Despite the fact that an electroacoustic work is fixed, which prevents a listener from changing the temporal enfolding of events, each subsequent listening, nonetheless, provides new and different interactions. If the composer’s orchestrational skill is up to the task, he/she can elicit a deep listening experience that simultaneously provides high levels of stimulation.

Recall in Chapter 2, I discussed our biological need to visually locate sound in the service of assessing danger and finding meaning. Roland Barthes sums up these two modes of listening, which parallel Huron’s theory of reaction and appraisal responses:

[L]istening is that preliminary attention which permits intercepting whatever might disturb the territorial system; it is a mode of defense against surprise; its object (what it is oriented toward) is menace or, conversely, need; the raw material of listening is the index, because it either reveals danger or promises the satisfaction of need.\(^{23}\)

In terms of our reaction responses, according to psychologists Michael Kubovy and David Van Valkenburg, “[s]ound localization itself is influenced by the act of visual orienting.”\(^{24}\) If a sound event surprises us, our eyes move to locate it, and that event, however momentarily, becomes the object of our attention. Kubovy and Van Valkenburg further discuss the limits of our auditory perception:

[W]hen an auditory sequence breaks into two streams, we cannot pay attention to more than one of them at a time. For example, Bregman and Campbell (1971) presented observers with a repeating sequence of three high pitch notes (ABC) and three low pitch notes (123) in the order A-1-B-2-C-3. Observers typically reported the order of notes as A-B-C-1-2-3 or 1-2-3-A-B-C; they were able to attend to one stream or the other, but not both streams simultaneously. As in vision, whichever stream is being attended becomes the figure and the other the ground.\(^{25-26}\)


\(^{25}\) ibid., 103. Note that “figure” and “ground,” in this context, center on which information stream captures our immediate attention. In a electroacoustic work, a monaural sound that is heavily processed with large reverberant space will appear distant and contain smoother textural features. However, if our listening focuses on this sound-object, despite being background information in the acousmatic image, it becomes the figure of our auditory perception.

\(^{26}\) Psychologist Albert Bregman discusses auditory streams: “In hearing, we refer to the result of auditory grouping as an auditory object or a perceived sound (when it creates a single sound), and as an auditory stream (when it creates a sequence that unfolds over time). The perception of a stream is the
For the composer, this perceptual limitation becomes a significant means to distract, disorient, and lure away the listener’s focus to another sound-object or group of sound-objects. By doing so, concurrent sounds can develop, modulate, and transform into new sound-source identities, shapes, textures, pitch/frequency collections, etc. without being directly obvious to the listener and potentially ruining any developing or heightened sense of anticipation. Carefully orchestrated distractions realized through \textit{spatial counterpoint} disrupt our subjective grounding, cultivate extramusical connotations, cloud our memory and sense of time, and immerse us further into the composer’s game of make-believe.\footnote{Curtis roads defines \textit{spatial counterpoint}: “the choreography of sounds in an interplay between fixed and moving positions and between foreground and background elements.” See Curtis Roads, \textit{Microsound} (Cambridge: The MIT Press 2001), 223.} Distraction acts as an anti-form disruptor, making us lose sight of how we arrived in a composition’s present moment. Instead, we ask ourselves ‘how did I get here?’, to this state of pleasurable confusion, which may be one of the more exciting responses to have while listening to music. Through spatial counterpoint and techniques of hyperproximity, the composer can dangle sound objects in and out of the listener’s grasp, seducing him/her into a dance of ‘now it is here’ and ‘now it is not.’ As sounds recede to the background, evading our touch, we reach to bring them closer to us again—that is, unless our perception has moved onto another sound-object.

\footnote{Brain’s way of concluding (correctly or incorrectly) that sounds included in the stream have been emitted over time by the same sound source (e.g., a drum, a voice, or an automobile). . . . [T]here is a tendency for similar sounds to group together to form streams and that both nearness in frequency and in time are grounds for treating sounds as similar. The Gestalt psychologists had shown that, in vision, objects that are nearer in space (or more similar) tend to form tighter perceptual clusters. The same principle seems to apply to audition.” See Albert S. Bregman, “Auditory Scene Analysis and the Role of Phenomenology in Experimental Psychology,” \textit{Canadian Psychology/Psychologie Canadienne} 46, no. 1 (2005): 37.}
Spatial Counterpoint in *Walkabout*

*Walkabout* (1998) by Paul Koonce uses spatial counterpoint to challenge our attention economy through a form of *hyperlistening*, decentering our subjective grounding inside a familiar-yet-strange, strange-yet-familiar ecology of everyday sound-objects. The work contains a high turnover of change—an elusive tableau that is always in flux and predicated on an aesthetics of disappearance and reappearance, playing aural peekaboo for those who accept the composer’s invitation. We try to settle down into a particular scene or landscape, but the image is taken away from us and replaced with something else, wiping our memory until slivers of past sounds make surprise returns. *Walkabout* demands that we look left, then right, then far down the center, and then back, constantly shifting our focus from car horn variations, non-verbal vocalizations, acoustic instruments, and everything in-between. What just happened—was that the sound of an automobile or was it an acoustic instrument? Maybe it was both? The work forces us to move on and cope with another set of sound-objects, each containing perspectival entrances and exits that differ from previous gestures.²⁸ Child-like vocalizations remind us of the calls we once made while playing hide and seek, and the pleasures of hearing our own voice reflect outdoors. These calls, however, bend in pitch and panorama to the wide left and wide right portions of the image, as well as teeter-totter between the foreground and deep into large enclosures. We glide and spin around in a circle, following the reverberant decay of an utterance; we choose to follow

---

²⁸ Denis Smalley defines *perspectival space* as “the relations of spatial position, movement and scale among spectromorphologies, viewed from the listener’s vantage point.” *Spectromorphology* refers to “the interaction between sound spectra (spectro-) and the ways they change and are shaped in time (-morphology).” See Denis Smalley, “Space-Form and the Acousmatic Image,” *Organised Sound* 12, no. 1 (2007): 57; and Denis Smalley, “Spectromorphology: Explaining Sound-Shapes,” *Organized Sound* 2 (August 1997): 107.
this echo, making it the object of our perception, but lose track of what is happening concurrently in the foreground. We will have to listen again.

Pitch, noise, and inharmonicity fluidly intermingle into an egalitarian web of interlocking parts: a sonic environment in which no particular agent or physical object takes primacy over another. The piece, in the spirit of Merleau-Ponty, constantly gives our voice new things to do. We sing along with acoustic instruments, but also with car horns and sirens. We cough with a human voice, as well as with sputtering auto ignitions and garbage cans dropped at the curb. But these sound-objects do not exactly fit the prototypical images and spectral envelopes we are so accustomed to. Koonce uses subtle spatial, spectral, and temporal signal processing to modify these objects, making us do double takes that leave us with ‘almost but not quite’ feelings of uncertainty. For example, from 3:21-3:28, child-like utterances morph from disfigured yodeling to crunchy guitar solos and back. Likewise, at 4:39, we hear an impossible stew of microtonal piano, timer bell, and vocalizations dance in the background of a small concert hall enclosure, each disclosing their source ingredients before congealing back into an in-between neither.

The opening two minutes tease us with whiffs of the proximate; objects seem close, but not too close. At 0:32, a car siren flashes near our face from left to right, raising in pitch as it approaches our body and then lowering, producing a Doppler shift like those heard while standing on the side of a road.29 The sound of a vibraphone

---

29 The Doppler effect, according to engineer John Pierce, is “the shift in frequency of a sound source moving toward or away from the hearer” which gives “a sense of motion of a sound source.” See John Pierce, “Hearing in Time and Space,” in Music, Cognition, and Computerized Sound: an Introduction to Psychoacoustics, ed. Perry R. Cook (Cambridge: The MIT Press, 1999), 97.
captures our attention (1:51); with its raised dynamic level and almost monaural presence panned to the right, we localize it to a specific point in the stereo field. At 2:08, an unknown voice, highly fricative and with a high spectral centroid, cranks right in front of our face, in counterpoint to a car engine that coughs and sputters until it finally gives up, creating a hyperproximate magnification that jumps out of the speakers.

Figure 4-1. Sonogram of high-friction, hyperproximate voice in *Walkabout*, 2:08–2:13.

---

30 *Pan or panning* is audio jargon for the stereo position of a sound-object. If something is panned hard-left, it exists only in the left speaker of the stereo image. A stereo recording that is panned hard-left and hard-right maintains the original width captured by the two microphones. That same recording, with both left and right channels panned center, turns the image into a monaural one, eliminating its width and placing it in the center of the stereo field. This type of treatment, among others, can shrink the scale of a sound object.
This is a *trompe l’oeil* moment: the sound-object’s proximity expands the front to back depth of the image, breaking the frame we thought we were looking into for the previous two minutes.\textsuperscript{31} We hear some high-pitched material poke through—a smaller object, reminiscent of metallic clank of some sort—but the ignition’s aggressive interruption keeps drawing us back, as its roughness moves closer to our body.\textsuperscript{32}

Overall, listening to *Walkabout* becomes a performative re-enactment inside a rapidly changing, overlapping, and conflicting perspectival maze—a hyper-echolocative activity of identifying behavior, source, and placement of sound-objects that vanish up into smoke as we try to pin down any stable meaning. *Walkabout* dissolves notions of an authoritarian voice—there is no unitary ruling persona in the work, aside from our own subjective viewpoint. However, the more we eavesdrop into this auditory carnival, the more we joyfully entwine into a ball of wide-mouth smiles, pirouettes, and exaltations that resists unravelling. Outside the work, *Walkabout* encourages us to stop and listen wherever we may be and to inject wonder into everyday acoustic events that we often ignore in favor of our phonemic bubbles, whether we sit at a bus stop, prepare food in the kitchen, converse with a feline friend, or play an acoustic instrument. And this wonder reveals that sound is never as univocal as our eyes and inner voice want us to believe.

---


\textsuperscript{32} At 2:24, however, Koonce begins to shrink the car engine image by narrowing its stereo width. This reduces the scale of the image and places it more in the middle ground. We jump back and forth between the middleground and proximate images and think the sound-object will come even closer to us at 2:34, perhaps as it does at 2:14, but this arrival never comes to fruition.
Holistic and Biomorphic Space

As an electroacoustic work unfolds, the listener amalgamates a variety of sound-objects in memory. Some might be brief in duration and tiny in scale, yet exist close to our body, with sharp surface textures and more apparent sound-sources. In contrast, other distal sounds with smooth surfaces may sustain for longer periods of time, winding about the image much like a snake does on the ground. We can also take a mental snapshot of the acousmatic image at any given moment—a landscape assessment of the spatial relationships among sound-objects, much as we do while viewing a photograph. As these sound-objects accumulate, the listener starts to form larger quasi-visual structures that recapitulate the composer’s use of time and space, similar to my commentary on Walkabout. Outside the listening experience, when considering an entire composition, we often rapidly summarize its contents, compressing its duration down to a second or two. For our favorite orchestral or chamber work, we may quickly recall the climax before applying a harmonic reduction or encapsulating a jumble of score fragments, performance gestures, and instrument images together. Adds musicologist Nicholas Cook, “it is as if what is heard sequentially in the concert-hall were distilled into a single, heightened experience that embodies everything that is characteristic of the music.”33 For an electroacoustic work, our analysis may squeeze the piece down to a holistic space that becomes similar to our experience of a painting, sculpture, or architectural form.34 Denis Smalley further describes the listening phenomena of holistic space:

An analytical stance, realized by mentally amalgamating an array of spatial forms into a unified spatial view. . . . I can collapse the whole experience into a present moment, and that is largely how it rests in my memory. The temporal disposition of, and relations among, sounds serve to articulate and shape spectral and perspectival space, but even though my perception of sound is the product of time, I ultimately sideline time’s formative role. So space can be more significant than time, or at least we can profit by starting with the idea that time can be placed at the service of space rather than the reverse. Time becomes space.\textsuperscript{35}

For example, Paul Koonce’s electroacoustic works from the 1990s—*Hothouse* (1992), *The Flywheel Dream* (1994), *Pins* (1996), and *Walkabout* (1998)—call to mind holistic spaces akin to the surrealist paintings of Salvador Dali: bending and juxtaposing prototypical sound-shapes into dream-like absurdities.\textsuperscript{36} Likewise, recent work by Autechre—*tac Lacora* (2013), *M39 Diffain* (2013), and *feed1* (2016)—suggest spaces similar to the alien, monochromatic, and biomechanical paintings and sculptures of H.R. Giger.\textsuperscript{37} These type of analytical overviews provide a powerful means to think about sound cross-modally and observe how the orchestration of spectrum and space, and the extramusical connotations arising from it, coalesce into a mass of interdependent components.

Biomorphic visual art, especially the work of Chilean painter Roberto Matta, offers another analogy of how we might consider holistic space in an electroacoustic

\textsuperscript{35} ibid., 37-38. Smalley defines *spectral space* as “[t]he impression of space and spaciousness produced by occupancy of, and motion within, the range of audible frequencies.” See ibid., 56.


work, and, in particular, our vocality: as, to cite Timothy Morton, a *mesh* of interlocking, interdependent parts, that simultaneously contains holes and absences.\(^\text{38}\)

![Figure 4-2. Roberto Matta, *Elle Loge La Folie* (1970), biomorphic painting (image taken from http://www.matta-art.com/1970.htm).](image)

According to art historians Oliver Botar and Isabel Wünsche, the term biomorphism was never used to identify an artistic movement or unified school, but “to describe the fluid, organic shapes in the work of artists as diverse as Hans Arp, Constantin Brancusi, Jean Hélion, Wassily Kandinsky, Joan Miró, Henry Moore, Yves Tanguy, and others.”\(^\text{39}\) As a whole, biomorphic forms favor “ambiguous and organic shapes in apparent movement, with hints of the shapeless and vaguely spherical forms of germs, amoebas and

---

\(^{38}\) Morton defines *mesh*: “A mesh consists of relationships between crisscrossing strands of metal and gaps between the strands. Meshes are potent metaphors for the strange interconnectedness of things, an interconnectedness that does not allow for perfect, lossless transmission of information, but is instead full of gaps and absences.” Morton draws on theories of evolution and symbiosis to argue that boundaries between living and nonliving forms are permeable, and therefore, there can be no fixed identity in the system of life-forms. Morton’s line of thinking is very similar to Jacques Derrida’s deconstruction of the system of language. See Timothy Morton, *Hyperobjects: Philosophy and Ecology after the End of the World* (Minneapolis: University of Minnesota Press, 2013), 83; and Timothy Morton, *The Ecological Thought* (Cambridge: Harvard University Press, 2010), 27-50. Also see Jacques Derrida, “Structure, Sign and Play in the Discourse of the Human Sciences,” in *Writing and Difference*, trans. Alan Bass (London: Routledge, 1978), 351-370.

embryos.” Such forms parade a “stylistic heterogeneity” characterized by “evocative swells, curves and arabesques.” Moreover, their use emphasizes the “interdependence of form and space” and “the relationship between human and landscape forms.” Last, biomorphic forms “exist both as autonomous, organic wholes and as rhythmic voices in a larger decorative chorus.”

**Biomorphic Voice-Bodies in Red Snow**

Natasha Barrett’s electroacoustic work *Red Snow* (1999 rev. 2001) offers the listener a performative means to re-present biomorphic imagery. Barrett’s juxtaposition of voice and environmental recordings, coupled with extensive digital signal processing, reveals a fluid sound ecology: one in which we behave like permeable states of matter that are continually composited and reconfigured together. *Red Snow* gives us new ways of having or being a body, allowing us to empathize with something different than a prototypical human face or figure. The work suggests multiple, decentered, female personae with no firm identity or shape; the personae are ambiguous and malleable voice-bodies, injected with uncertainty because our motor repertoire does not have an exact set of actions to draw from. Our listening voice turns plastic to contort, stretch, and dismember ourselves so that we can become like these sounds, simultaneously creating voices out of bodies and bodies out of voices. Traces of human and animal voices disappear quickly, embed themselves into other sound-objects, or morph into

---


42 Lichtenstern, *Grove Art Online, Oxford Art Online*.

43 ibid., 9.
something new and unknown. As a result, our embodiment brings about a paralinguistic discharge, filling our voice with acoustic and motor resonance that defers our linguistic commentary. Barrett gives us moments to ground our subjectivity that lead us to ask, was that the sound of an everyday metal object? Was that a soprano vocalist coming out of that sound-object? These moments, however, are quickly stolen from us, as familiar sounds congeal into dubious masses and heteroglossic chimeras. We perform elongated hums, sighs, and secret whispers alongside the sound-objects, producing a semiotic residue within us that says nothing until we begin to describe what we are hearing. But what are we hearing? Red Snow does not give us many opportunities to reflect while listening, precisely because its use of rapid counterpoint demands we move on and imitate something else.

The piece decorates itself with many sustained, wind-like vocalizations: hybrid voices whose durations and spectral makeup extend well beyond what we can perform as humans (at 0:20, 1:55, 4:50, 5:50, 7:19, and 8:00). These sound objects take different roles as foreground and background material, and as solo voices, duets, and choirs. Numerous instances—with 1:00, 2:58, and 3:37 being a few examples—evoke imagery of various streams of light. These are purer tones, lacking the harmonic density needed to bring these objects into the closest foreground. Our oral imagination opens to cope with these resonant frequencies, which, in turn, reflect our gaze back to us. These overlapping resonances afford us new ways to sing, as we pick and choose different voice parts out of the spectral choir, upon each subsequent listening.

Carolyn Abbate defines chimera as “an object congealed out of false metonymic juxtapositions, and, as such, a physical tribute to the failure of logic.” See Carolyn Abbate, Unsung Voices: Opera and Musical Narrative in the Nineteenth Century (Princeton: Princeton University Press, 1991), 125.
At 2:29, we confront a large and wide sound-object with a low formant frequency around 27Hz. The object is close to us, yet not completely foregrounded, because its high frequency energy is rolled off.

Figure 4-3. Sonogram of wide voice-body containing low formants in Red Snow 2:29–2:45.

This sound-object, however, has a small companion sound that winds about the image and reifies the former’s size. Does this small whisper emanate from the larger sound-object, or is it a distinct entity? The fact that our perception delineates between the two indicates some form of individual autonomy between the sound-objects, however, their concurrent movements suggest they are working together. Taken as a whole, this
gesture entices us because we don’t know what it is or how to fully cope with it; it is something we can approximate with our voice, yet in doing so, we inevitably get it wrong. Moreover, our failure to linguistically describe what we are listening to draws us closer to these sounds, forcing us to describe even further. But at the moment we cannot do so, because if we are still listening, then resonance occupies us. Descriptions will have to wait, and when we get around to it, perhaps after the work has finished, it may already be too late. Like *Walkabout*, we will have to listen again.

From 9:23-9:41, we hear hyperproximate and aggressive footsteps in the snow that remind us of playing outdoors during the winter. We might enjoy the feeling of stomping on snow with our feet, but the true joy lies in the sounds’ resonance in our mouth, not at the top of our toes. In this context, snow gives us a new way to bark, crunch on, and devour materials with our voice, allowing us to feel a varied texture of packed grains with every vocal movement. At 10:00, we settle into a familiar landscape, filled with sounds that indicate the presence of birds, light raindrops, distal traffic, and church bells. But this realistic scene is not stable, as indicated by a small, whispering voice-body that glides from right to left across the image at 10:12, and by cymbal-voices in the background that slap delay from hard-left to hard-right of the image.\footnote{A *slap delay* produces a copy of a sound, typically 60-150ms from onset of the original sound. This type of treatment decorrelates a sound into two separate sound-objects. We typically perceive delay times lower than 60ms as a single sound-object.} We listen *into* these sounds; moving closer in our seats, with our heads pushing forward, we eavesdrop to pick out any familiar patterns that resemble memories from everyday life. We inevitably have to flee backwards, however, as the work surprises us at 10:19 with a
hyperproximate and ambiguous vocal gesture that, nonetheless, gives us a hollow “o” vowel to latch onto, followed by wisps of birds and bird-like creatures.

Other strange creatures play a role as well. At 4:36 and 4:56, a mutated, animal-like vocalization forms into a major second dyad, with the former being large in scale and width, and the latter small, monaural, and off to the right. From 5:58-6:08, proximate, hybrid voices, perhaps part animal and part cello, make us wonder what is being said. Finally, from 13:00-13:20, a small crowd of vocal hybrids trade off light interjections, making sure we recognize each entity.

Taken as a whole, Red Snow offers the listener a glimpse into a reconfiguration of material reality; a virtual world that contains just enough slivers of everyday acoustic reality—human and animal voices, metallic objects, and outdoor landscapes—to disorient our motor responses, allowing us to hide in shadows and masquerade as a body that has yet to find a form.

A Theory of Hypervocality

Both Walkabout and Red Snow offer lenses into a major affordance of the electroacoustic medium: listening experiences of hypervocality—a continual decentering and deferral of voice from its inner, metaphysical privilege. Hypervocality highlights the impossibility of pinning voice down to a stable definition or univocal source, and instead, revealing through spatial counterpoint, the indivisibility between subject and object, self and other, body and timbre, actual and virtual, action and perception, listening and speaking, composing and listening, and voice and not-voice.46

Moreover, the theory stresses a desire to move beyond human vocality through performative re-enactment, yet, the self-reflexive impossibility given our body’s limitations. This highlights an imaginative and often exaggerated vocality of madness and flight towards countless sound-objects that evade our grasp, unleash the semiotic, and allow our animism to masquerade in the many objects our vocal embodiment projects onto. Such a theory emphasizes a transcendence of our anthropocentrism; yet, sound, by its very definition, is anthropocentric:

[H]uman beings reside at the center of any meaningful definition of sound. . . . As a part of a larger physical phenomenon of vibration, sound is a product of the human senses and not a thing in the world apart from humans. Sound is a little piece of the vibrating world. 47

Likewise, once we consider something as music or sound we trap ourselves into the confines of human meaning-making, and inevitably language. Nevertheless, electroacoustic music supplies us with new ways to project our shortcomings onto sound-objects that more or less reflect back, albeit in fragmented manner, the proprioception of our own bodies, allowing us to confront our finitude and what it means to be human.

CHAPTER 5
REVEALING VOCAL ENTANGLEMENTS: CONSTRUCTING THE HYPERVOCAL

Electroacoustic composition affords me experiences of vocal plasticity—how does my body have to contort, dismember, and re-assemble into the postures and assemblages dictated by my vocal embodiment? The computer becomes a prosthetic device for vicarious echolocation—an intersubjective dialogue in which the composer copes with voices and/or vocal traces vis-à-vis motor resonance and metaphorical projection. Recall Barthes claim that voices are “necessarily erotic,” for “there is no human voice which is not an object of desire—or of repulsion.”1 If listening involves mimetic participation with our voice, then does not the electroacoustic composer, as both the sender and primary receiver of sound, reside in a self-reflexive dialog of desire and demand? Composition, in this sense, becomes akin to a psychoanalytic activity of unpredictable, free associations, bodily affects, and imagination mediated through the composer’s voice. I cast off sounds into virtual spaces and into the rooms I compose in, taking leave from my body and waiting for them to boomerang back, creating a constant mediation of expulsion and incorporation. Through software experimentation and algorithmic design, I can create unpredictable sonic results that await my body’s reply—sometimes the sounds are seductive; sometimes they are startling or disturbing; sometimes they are silly; sometimes they generate new quasi-visual forms; sometimes they make me want to eat; and sometimes they make me apathetic or even disgusted. How can my voice echolocate and entangle with objects and enclosures in ways it has

not yet done before? With what new objects, manufactured and rearranged through
digital assemblage, can my voice share resonance?

This ambiguity and undecidability of voice is an affordance of the electroacoustic
medium and a testament to its ontological inconsistency in everyday life. As a
composer, one of my aims is to reveal the symbiotic continuum of the voice: one that
cannot be simply pinned down to human bodies and faces, but rather, as an
entanglement of indeterminate objects. It is my hope that listeners, composers,
musicologists, and theorists, consider a more multisensorial vocality—one in which all
of our sense modalities navigate and potentially obfuscate a precarious boundary
between self and other, subject and object, and body and timbre.

Because I use the human voice as either 1.) recorded source material, 2.)
analysis data for other target sounds, or 3.) an intuitive model for signal processing non-
vocal sounds, my sound-objects often allude to vocal behavior. Electroacoustic
composer Barry Truax elaborates on how vocal traces are difficult to erase:

Given the human auditory system’s abilities to detect human vocal
qualities, even amidst noise and distortion, and even to ascribe those
qualities to sounds of non-human origin, it is unclear whether any sound of
vocal origin ever becomes entirely abstract. A great deal of processing
would be required for that to happen, but it is not impossible.²

On the other hand, I do not always wish to be completely transparent and want the
listener to revel in ambiguity and the uncanny. This could mean, for example, creating a
composite sound from a variety of non-vocal acoustic or synthetic layers, and burrowing
a tiny vocal sliver—a cough, cluck, or breath—into the gestalt. In doing so, my hope is

---

that listeners may ask, for the reasons outlined in Chapter 2 and Chapter 3, ‘is there a voice to be heard in this noise?’

As the primary listener, I aim to become a paralinguistic puppet—creating sound-objects that, through my appraisal responses, speak through me. Throughout the entire compositional process, I operate through a mimetic dialog with the virtual soundscapes I create, appraising every sonic detail along the way in hopes that other listeners participate in a similar manner. The sound-objects I create contain traces of my own bodily affect that afford other listeners, given their experience with paralinguistic expression, opportunities for vocal simulation and multisensorial imagery. Of course, I am aware that others have different listening strategies, backgrounds in music, and cultural and aesthetic biases that may preclude them from experiencing my music in this regard. However, I believe that vocality, as a conceptual and structural framework, is more universal than other compositional frameworks; while it is wonderful to have other composers and scholars as an audience, it is equally rewarding to connect with new and/or less experienced listeners through something to which everyone relates: the voice! The following sections explore my own compositional and signal processing methods for creating listening experiences of hypervocality. I intersperse this discussion with brief, personal responses in hopes that others may share these resonances with me.

**Reconstruction: Voices Through Mimetic Reciprocity**

I created my electroacoustic work *Reconstruction* (2011) using recordings of a cello played in an acoustically-dampened closet. My aim was to use the instrument to generate material with extremely rough acoustic properties, by playing it with friction producing objects. I used bows, coat hangers, pens, pot handles, and various balls to
apply as much force to the cello’s strings as possible, stopping short of damaging the instrument. In contrast, I also recorded very delicate bowing and touching of the cello body away from the strings with the same objects. The idea was to create a softer pallet of sounds to complement the harmonically-saturated, highly friction-based sounds. Along with a seagull effect glissando, I recorded only a few sustained pitches that were played with *mp, mf, f,* and *ff* dynamics.³

Object 5-1. *Reconstruction:* electroacoustic composition soundfile (.wav file 117MB)

While my first intentions were to employ digital signal processing lightly, using mostly equalization and creative editing to preserve the sound-sources’ intelligibility, I discovered later that using chance procedures to create composites drawn from both sound taxonomies (low friction and high friction) sparked my imagination. This highlights one of the major strengths of electroacoustic composition: the ability to take unexpected formal turns and to produce surprising sonic results at any stage of the creative process. Composer John Young shares a similar sentiment:

> The ease with which sound materials can be developed and mutated using computers presents exactly the potential for extensive probing, reshaping and recombining of sound objects. . . . Indeed, the electroacoustic medium’s gift of a heightened freedom to disassemble and reassemble sounds in conjunction with the realization that these sounds simply may not respond to processing in the way one imagines they will, or *need* to for the musical purposes envisaged for them.⁴

---

³ George Crumb’s *Vox Balaenae* is an example of a work that uses this extended technique.

My initial associations of the cello, which were visual image archetypes of the instrument (its shape, its color, and its curves and contour), as well as a performer's bodily, tactile relationship to the instrument (the movement of the limbs, hands, fingers, and sitting position), started to fade, as the sound-objects resembled a group of various-sized organisms flying around the stereo field. Unconsciously, I was trying to make the cello speak, creating nonsense utterances primarily out of wood, metal, hair, and the other objects the cello was entangled with during the recording process (plastic, metal, human bodies, closet walls and its blankets, the ceiling, the floor rug, etc.). The traditionally-bowed recordings, when pitch-modulated using interpolated noise, became seasick cries of joy in a cartoon-like manner (3:16-3:19); the rough materials, when pitch-modulated and layered together, yapped with anger (4:40); and the delicate bowing, through filtering and granulation, became hybrid breaths of repose and hushed whispers (2:21 and 4:01-4:35).

Throughout the work, there are numerous gestures invoking fusion and fission between sound-objects. For example, at 3:00, a voice-like sound-object emanates from the center middle-ground of the image. This object contracts outward at times, forming a wider stereo spread whose multiple voices then collapse back to a single, monaural object. Contrapuntal to these gestures are sound-objects that race by the foreground from left to right using Doppler-based signal processing. From 3:10-3:13, slightly grotesque, eating sounds poke out, with an emergent, “o” vowel pronunciation, the origin of which, be it from one or many, is unclear. The process of compositing and re-

---

5 Fusion is a process by which “multiple components fuse into a single sound,” and fission is a process by which “sounds split into multiple independent components.” See Curtis Roads, Composing Electronic Music: A New Aesthetic (New York: Oxford University Press, 2015), 328.
compositing filtered, fragmented, and re-spatialized sound materials masks certain acoustic features, while accentuating others, creating new, gestalt sound-objects that resist being pinned down representationally. What’s more, having selected processed recordings that could occupy the closest spatial proximity to my body, I confronted those that sounded voice-like; as a result, sound and touch were meeting in my mouth.

As a listener, I was more interested in imitating what the computer helped me create, forming a self-reflexive dialogue between myself and processed sound. Sound-objects that were voice-like animated my own perception of them; I found myself mimicking these sounds both covertly and overtly and selecting them based on these positive appraisals. Through prosopopoeia, the fragmentation of cello seems to be put back together by multiple, decentered voices—as if they were they were the ones doing the construction. But, in actuality, through my participation, I was trying to speak through the cello, enmeshing my vocality into its fricative paralanguage that became a chimera—–yet a chimera displaced by the countless, contrapuntal voice-like fragments from which it was constructed.

Is the piece a process of fragmentation and re-assemblage—one of reconstructing the cello into its archetypal images? Of course, the ending may suggest this, as I progressively reveal an unprocessed recording of the cello towards the climax (4:46) and finally reveal as the very last sound of the work (7:04). But isn’t there an equally opposite destruction occurring: one that erases the human face as it effaces the human body? In picking out patterns that afford human vocality, am I not trying to reconstruct a human face? Or, is my mimetic dialog with the sound-objects one of
inevitable imperfection and failure, that, instead, creates a *body-in-pieces*? Deleuze and Guattari comment on the effects of effacement:

> [W]hen the face is effaced, when the faciality traits disappear, we can be sure that we have entered another regime, other zones infinitely muter and more imperceptible where subterranean becomings-animal occur, becomings-molecular, nocturnal deterritorializations over-spilling the limits of the signifying system.⁶

Nevertheless, the work formed out of my own mimetic reciprocity with the computer’s output, creating a recursive feedback loop between myself and a model the computer helped me simulate: a virtual environment formed out of an impossible chimera of cello voice-bodies.⁷ My initial goal of preserving the recorded cello gestures became a yearning to create a variety of organisms out of these materials, which, in actuality, was my desire to animate and re-animate my vocal capacities in new ways, producing an acoustic mirror that refracts indefinitely through contrapuntal and paralinguistic interruption.

---


⁷ This calls to mind René Girard’s theory of *mimetic desire*: our desires derive from imitating models whose objects of desire become our objects of desire. Such an “appropriative” or “acquisitive” mimicry, as Girard argues, is the main source of violence among humans: “We are competitive rather than aggressive. In addition to the appetites we share with animals, we have a more problematic yearning that lacks any instinctual object: desire. We literally do not know what to desire and, in order to find out, we watch the people we admire: we imitate their desire. Both models and imitators of the same desire inevitably desire the same object and become rivals. Their rival desires literally feed on one another: the imitator becomes the model of his model, and the model the imitator of his imitator. Unlike animal rivalries, these imitative or mimetic rivalries can become so intense and contagious that not only do they lead to murder but they spread, mimetically, to entire communities.” Despite detailing desire’s destructive aspects, Girard believes it is positive force because it involves the opening out of oneself, forming a basis for love and imaginative relatedness. Girard’s theory has interesting implications within music communities, including the mimetic rivalry between composer colleagues, students, and teachers. More specifically, compositions that become the targets of others’ desires, creating social recognition. While beyond the scope of the dissertation, the intersection between the social and anthropological effects of mimesis and desire and embodied cognition, provide an appealing path for future research. See René Girard, “Violence and Religion: Cause or Effect?” *The Hedgehog Review* 6, no. 1 (2004): 8.
Figure 5-1. Sonogram of cartoon-like voice glissando in *Reconstruction*, 3:16–3:19.

**Vocal Silhouettes: Short of Touch and Chatterbox**

My electroacoustic works *Short of Touch* (2012) and *Chatterbox* (2012) developed out of the following question: how could I make an acoustic instrument sound more like a speaking voice, while at the same time, make a speaking voice sound more like an acoustic instrument?

**Object 5-2. Short of Touch: electroacoustic composition soundfile (.wav file 97MB)**
Object 5-3. *Chatterbox*: electroacoustic composition soundfile (.wav file 36MB)

I was looking to create an ambiguous *silhouette* that defers sound-source identity between, in my case, a human voice and a piano, resulting in a mutual acoustic entanglement or hybrid, where the piano speaks no less than the voice, and the voice speaks no less than the piano.

In order to achieve this, I developed a software program that tracks a voice’s prosody—tracing vocal pitch and rhythmic contours using a microphone input signal. Using this input analysis data, I mapped my voice’s first nine harmonics to a bank of equal-tempered piano samples, recorded with *mf* dynamics, sometimes rounding the frequency of each harmonic to the closest MIDI pitch, at times rounding the harmonics to specified intervals and transposing them anywhere from a minor second to several
octaves. An envelope follower tracked the amplitude of each vocal harmonic, independently triggering a piano sample once the amplitude crossed a certain threshold. Experimenting with different thresholds produced interesting results: high amplitude thresholds often required loud exclamations and shouting to trigger the virtual instrument, with only the fundamental, second, and third harmonics triggering events. For example, in the last section of *Chatterbox* (1:19-2:12), I introduce a barking counterpoint to the rapid prosody of minor third intervals, which was triggered with fairly aroused, nonsensical utterances. On the other hand, low amplitude thresholds allowed the triggering of nine polyphonic voices all at once. To compensate for this increased polyphony, I reduced the piano notes’ range of possible amplitude decay, sustain, and release times to create more pointillistic textures, rather than having a muddy overlap created by too many sustained envelopes. Furthermore, I routed each sample into its own band pass filter and panner, set to a unique center frequency, resonance strength, and spatial position. The use of these parameter modulations, at times, created greater separation between simultaneous events, and hence, a greater polyphonic clarity and semblance of surface texture. This also resulted in more subtle, *micro movements* on top of gestalt, *meso movements.* For the listener, this excites the concurrent movements of the tongue, lips, larynx, and velum that are required for rapid speech. I used such multi-parametric modulations to suggest rapid utterance in both works;

---

8 An analogy of this would be to compare a waving hand that moves across the torso, versus a waving hand that moves across the torso with all five fingers moving in a slightly different manner. In the latter case, the hand (or simply the entire limb or body) is the *meso movement*, while the coexisting finger actions are *micro movements*. In terms of electroacoustic sound, the antithesis of multi-parametric modulation is repeating the same instrument sample over and over without any modification to the sound. This lack of variation is partly responsible for the general MIDI effect that often plagues notation software. For more, see Rolf I. Godøy, “Understanding Coarticulation in Musical Experience” in *Sound, Music and Motion: Lecture Notes in Computer Science*, eds. M. Aramaki et al. (Berlin: Springer International Publishing, 2014), 535-547.
however, in *Short of Touch*, these modulations are more subtle and occur less frequently.

![Sonogram of barking piano interjection in Chatterbox, 2:04–2:05.](image)

Figure 5-3. Sonogram of barking piano interjection in *Chatterbox*, 2:04–2:05.

Taken as a whole, *Short of Touch* and *Chatterbox* are perhaps monochromatic due to the limited pallet of sound-sources and signal processing methods. However, this limitation allows the listener to focus on the hybrids that arise from the interlocking resonance between the two. As an afterthought, the listener may face a certain undecidability: is the human voice present or absent? Is a piano speaking through a body (or bodies), or is a body (or bodies) speaking though a piano?
Ultimately, the listener speaks through both. *Short of Touch* and *Chatterbox* highlight a covibratory relationship between the piano and the voice, bringing the two objects, the human body and the piano, closer to one another, magnifying a more proximate resonance the two have always shared together.

**Ventriloquy: Composer as Ventriloquist and Listener as Dummy, or Listener as Ventriloquist and Composer as Dummy?**

My electroacoustic piece *Ventriloquy* (2015) is a culmination of all the voices I have put forth in this document. Unlike my aforementioned works, *Ventriloquy* uses the
human voice as source material; sometimes I leave these voices relatively unprocessed, and other times, I distort, mutate, and composite vocal recordings beyond recognition. Regardless, all of these sound-objects afford the listener a platform for multisensorial vocal simulation.

Object 5-4. Ventriloquy: electroacoustic composition soundfile (.wav file 156MB)

The work invites the listener to sing, stutter, hum, shout, whisper, weep, crunch, plea, breathe, cheer, and growl along with the sound-objects, performing a variety of strange consonants, contortions, air flows, and oral stops. “[S]queaky-clean enunciation” is there, as Katharine Norman would say—vocal recordings picked, pruned, and equalized to be pristine and full of tactile texture—except these utterances are of utter nonsense. Amidst hushes and secret whispers, these voices say nothing, and because of this, we continue to ask ‘what are they saying?’ As Abbate declares, citing literary theorist Peter Brooks, “[o]bjects that are present but lack or resist meaning are fundamentally erotic. . . . Eros produces the desire to comprehend these objects, despite their resistance, as signifiers. Human bodies are thus semioticized by the desire to know their assumed secrets.”

Ventriloquy, in the spirit of Deleuze and Guattari, contains “a multiplicity of sound bodies and silhouettes” and “a launching of probe-heads in all directions.” Traits of faciality dance back and forth between appearance and disappearance, disfiguring, re-assembling, multiplying, and collapsing into new perspectival configurations. Female-


like voices constantly interject, rupturing the foreground and capturing our attention—at times, we may make out a familiar face, and at other times, we confront a puzzling hybrid. For example, from 0:47-0:52, an unknown voice-body rips though the foreground, briefly hovers to the left, and then spreads throughout the image, evoking a combination of human speech, paper ripping, and a bird fluttering all at once. Our vococentrism props up these female voices as lead roles in the dramaturgy, and, given the repetition and variation of certain utterances, as perhaps personae controlling the flow of events. Who are these personae—do they form part of my compositional voice? Are these characters a prosthesis I speak through? Or does my “fort-da game” of spatial and facial counterpoint, one whose rapid influx of voice-bodies continues to challenge my memory, attention economy, and interpretation of the work, preclude this as a possibility? ¹¹ As novelist and playwright Oscar Wilde says, “[m]an is least himself when he talks in his own person. Give him a mask, and he will tell you the truth.” ¹² Is Ventriloquy a veil that I hide behind, letting me speak at a distance and in my own absence? As a composer and listener, am I flirting with the anxieties aroused by the female voice devoid of logos, or am I merely giving voice to vocal detritus that is often eschewed? Is the work an ironic re-presentation of the siren song narrative, albeit in a hypervocal fashion? As a composer, I control these voice-bodies, locking them down to a fixed temporal trajectory, but is this not driven by my desire, as a listener, to be surprised and seduced by the semiotic overflows that the music demands my voice to


speak? The more answers I look for, the more questions that arise, the more chatter that accumulates, “with no remaining plumb except my gaze that shows me—blackly vivid in the mirror at the bottom of the well—my own face that observes me observing it.”

**Unveiling the Voices**

While I cannot map out all of the work’s parameters in this document, there are several things that I would like to highlight. One of the main instruments I created for the work consisted of recording the vowel sounds ‘Oo’ and ‘Ah’ for three octaves of equal-tempered semitones (C1-C4), with each semitone lasting for two seconds. I used pitchshifting to extend my voice down one more octave to C0 and up two more octaves, with C6 as my top pitch. Using granular-based signal processing, I triggered many small, overlapping grains from all six octaves, which pan randomly throughout the stereo image. As a result, I was able to create sound-objects reminiscent of crowd-noise, akin to dense cocktail party chatter. Furthermore, by convolving these sounds with impulse responses of cymbals, I created hybrid sound-objects that sound similar to crowd cheers (for example, at 1:45, 2:15, and 7:42). The idea was to create paradoxical illusions that allow the listener to rapidly jump from the singular, to the plural, and back, moving from a pinpoint voice-body, to a choral collective that forms a unified whole; from an individual voice that amplifies into a mass of the many. Likewise, convolving and cross-synthesizing my voice with various enveloped and equalized piano tones produced new singing voice-bodies which were tuned microtonally to one of several...
harmonic series and often further processed with resonant filters.¹⁴ Last, I built a
taxonomy of vocal noise using close mic’d recordings to magnify the inner workings of
my mouth (pops, saliva, tongue clucks, etc.) and composited them in various ways to
form garbage: heap-like impenetrable masses of sound-object debris. The opening
gesture (0:01), with its exponential rise in amplitude, is one instance of these stacks of
vocal waste. Sound-objects such as these cause the listener to defer reference and
cultivate vocal embodiments that take on additional significance where sound-objects
“open up a crack where fantasy comes flooding in.”¹⁵

One way I created rough textures that poke into the foreground was with
waveshaping synthesis.¹⁶ After heavily distorting nonsense recordings of my voice to
the point where my voice sounded like it was between stations on an FM radio, I used

---

¹⁴ Convolution and cross-synthesis are types of spectral domain, digital signal processing
methods that enable the superimposition of one sound source onto another by boosting similarities or
attenuating differences between two audio signal’s spectral envelopes. For example, verbal speech could
be cross-synthesized with a 50 ms, 90 Hz sawtooth waveform, creating a monotone voice that is one part
human and one part synthetic oscillator, while retaining the prosodic rhythm of the original speaker. One
form of cross-synthesis involves multiplying the amplitudes of two audio signals, while retaining the
phases of the first signal. Another form uses the amplitudes of one signal with the phases of the second.
Convolution, on the other hand, involves multiplying the amplitudes and adding the phases of two audio
signals. Fast convolution is often used for digital reverberation, in which impulse responses of actual
spaces, such as concert halls, churches, and rooms, are used to place a sound-object into the particular
enclosure. Furthermore, this technique works with smaller objects. For example, a human voice
processed with a metal bowl impulse response creates an ambiguous entanglement where the voice and
bowl mutually coarticulate one another; the voice equally speaks the bowl and the bowl equally speaks
the voice. For more on convolution and cross-synthesis, see Curtis Roads, *The Computer Music Tutorial*

no. 2 and 3 (2011): 133.

¹⁶ In waveshaping synthesis, a transfer function remaps the instantaneous amplitude of an audio
signal; each sample value of the audio signal, in the amplitude range [-1, +1], is used as an index to look
up a stored value in an array of numbers. Waveshaping can create smooth distortion effects, and thus
light harmonic generation, when remapped through a hyperbolic tangent transfer function. However,
using random and asymmetrical transfer functions, such as step, ramp, or quadratic noise, can produce
rough textures that indicate high friction and sharp, tactile detail. For more on waveshaping synthesis, see
the processed recordings as impulse responses for convolution and cross-synthesis with clean, vocal recordings. This created crunchy sounds that appear one part human voice and one part rough-textured matter. I wanted the listener to feel these sounds in his mouth, while at the same time, to visually imagine particles that might assemble these strange voice-bodies. These sharper textures begin to interject at 3:15, as brief shards of voice, and accumulate into larger, prosodic phrases up until 5:15. In general, the rough sound-objects provide a contrast to the smoother, vocal resonances that develop simultaneously in the background, teetering between vague, bodily figures and sustained streams of light that the listener sings along with.

Beginning at 5:55, my formal goal was to use the same recordings of my voice to interpolate from a pitched, female-sounding choir singing (6:00-6:25), to a dense, polyphonic mass of raucous crowd noise (6:25-6:42), and to a machine-voice hybrid that begins to fuse into a monophonic sound-object (6:41-6:45).\(^{17}\) A female-like voice object, however, interrupts this transformation in a plosive manner by spitting into the foreground, resulting in an expulsion of smaller voice-bodies. From 6:45-6:58, these monophonic sound-objects rhythmically unfold in an exponentially decreasing geometric series, with each entrance given a random pan position at either center, hard-left, or hard-right.

---

\(^{17}\) Trevor Wishart defines interpolation as the "[t]he process of moving gradually between two defined states." The composer can use interpolation to move from pitched materials to broadband noise and vice versa over time, as well as to modulate sound-object size and scale. As a time domain example, the composer could take a 30 second, panoramic stereo recording of a park, process it with a steep resonant filter at 440 Hz, and collapse the signal down to mono. Through automation, he/she can gradually attenuate the resonant filter and reintroduce the original stereo width over 30 seconds. The resulting interpolation moves from a small/narrow sound-object located in the center of the stereo-field, containing a defined pitch and lacking a source identity, to a wide, aural landscape containing full-spectrum noise. From the listener’s perspective, he/she moves from singing to imitating environmental sounds, such as bird calls, wind in the trees, and children playing. Various Fast Fourier Transform processes are also useful for interpolation. For more on interpolation, see Trevor Wishart, *Audible Design: a Plain and Easy Introduction to Practical Sound Composition* (York: Orpheus the Pantomime Ltd., 1994), 122.
hard-right. Moreover, as the decelerando gradually comes to a halt, the voice-bodies, as a result of the increase in the decay and sustain portions of the amplitude envelope, become larger, more sparse, revealing a ‘ca-choo’ vocalization. Concurrently, at 6:55, the machine-like voice comes to a halt, interpolating into a resonating cymbal and a sustained vocal gesture, reminiscent of an operatic soprano. One final entrance of the ‘ca-choo’ occurs at 6:58, but the scale and size of this sound-object is much larger. We stare right into this face, but the face gets bigger and bigger, with the mouth opening wider and wider, chuting us down the throat and toward the belly. However, the soprano voice-body, located in the middle ground, swallows up this cavernous hole, which becomes scream-like as it joins the soprano, morphing into pitched material and shifting our perspective once again.

At 7:45, I created another set of hybrid voices, using a combination of spectral noise gating, multi-tap delay, and reverberation with longer tail decay rates (five to ten seconds). It was my intent to create voice-bodies that whisper, talk, and sing, but which also leave a gaseous-like trail in their decay, as if they were one part human voice and one part cloud, expelling vaporous streaks in their movement across the image. These types of sound objects occur earlier in the work (2:33-2:46), albeit closer in the middle ground and with less overlapping polyphony.

Last, at 1:51 and 7:24, there is a homorhythmic duet between two descending glissandi in the background of the image. As I argued in Chapter 3, these may resemble weeping utterances because our mimetic participation raises our voice upward to reach either pitch, and then descends our face downward into a frown, dropping the folds of
the mouth and the countenance as whole, which is a gesture we typically affiliate with being sad.

Figure 5-5. Sonogram of weeping glissandi in *Ventriloquy*, 7:21–7:31.

The Ventriloquial Reply

Overall, *Ventriloquy* produces, within me, a heteroglossic mesh of crisscrossing voice-bodies, where even the smallest particle of sound, at the tiniest of scales, can take center stage in my perception, before re-forming back into the silhouettes of the choral collective. This effigy is part human, part landscape, part machine, part matter,
part soloist, part choir, and everything in-between. Each time I listen to the work, or consider it in silence, strange talking heads, propelled by plosive gibberish, pop out at me like serpents from Medusa’s crown, while shadows of utterance glisten and refract in the backdrop.

_Ventriloquy_, however, for me, still leaves many unanswered questions. Even if I map out the entire structure of the work, detail all of its acoustic parameters, and discuss every technological aspect involved in its making, ultimately, the listening voice, whether mine or someone else’s, will undermine, in some capacity, what I claim the work to be about. This is perhaps the true power of our vocality, which can, potentially, slip us into madness! The fantasy of sonorous omnipotence may have driven the creation of _Ventriloquy_, and many other electroacoustic pieces, for that matter, which is a fantasy about inhabiting all space and making it our own. But what comes back to me, as the creator of the work, is not the full exercise of power I had in mind, but the full submission to bodily affects, exertions, and imagery that leaves me laughing one moment and crying the next. My embodiment of _Ventriloquy_, paradoxically, is a re-interiorization of the externalization of my own interior space—a re-mediated space that, despite the viral operations to consolidate me into a single subject, I cannot fully claim as my own.\(^\text{18}\) Rather, the work exists as a bodily performance where many voices pass, fuse, and fissure, some of which stick around for awhile, and others, after their brief glow, never speak again. As Nancy details, “the subject of listening is always still

\[^{18}\text{A preliminary, interior space that is already social to begin with!}\]
yet to come, spaced, traversed, and called by itself, *sounded* by itself.”19 As a listener, first and foremost, the composer, therefore, always emerges from the ventriloquial loop as a subject with an ineradicable polyvocality.

Despite being extremely frustrating at times, this is one of the true joys of listening to music: the endless production of more results that resist being jammed into a straightjacket. Music’s ability to supply us with more questions than answers is why we continually return to the same pieces that fire us up, make us smile, make us dance, make us feel lost, and allow us to escape—all so that we can try to better understand why we move, imitate, and behave in the ways we do.

**Final Remarks**

Composing and analyzing *Ventriloquy* has led me to believe there is ventriloquial trace in all composing, listening, and utterance in general. Our voice is the key to our subjectivity, yet also to our deconstruction. And what a better way to chase ourselves in this never-ending hall of mirrors than by creating and listening to electroacoustic music, while increasing our wonder and compassion for all that we are not.

Our voice makes noise, but it also makes significance out of noise. Sound, after all, is what we make of it; it is a relational activity configured by our efforts to capture a precise moment that has now passed. And what we *actually listen to*, strangely enough, is dictated by the desires and fantasies that shape our listening. In this sense, sound always presents us with an undecidability of whether it exists internal or external to ourselves. Timbre is not an acoustic fact; it is the result of shared action that tethers

---

different bodies and objects together. Our voice is a virtual prosthesis; it’s our body’s “second incarnation”—an autonomous specter that allows us to transcend our limits and explore new forms of expression that move beyond song, beyond language, and beyond the human figure—“[an] attempt, never successful but never frustrating, to reach the unattainable by exploring the paths of the possible.”

Our vocality is smooth, rough, curvy, and jagged. It is tiny and huge, yet right next to us, but strangely out of reach. It expels and incorporates, it opens and encloses. It locates, animates, disrupts, and fails. It empathizes and rejects, it croons and cries. It is fluid, it is solid, it is gaseous, it is tasty. It is singular and plural, it is viral, it is repressed. It is childish, it is motherly, it secretes, it tethers. Our vocality contracts, expands, displaces, fragments, stutters, stretches, spits, screams, winds, twirls, coughs, sloshes, breathes, barks, pokes, propels, repels, murmurs, shouts and much more—all so that our imagination can occupy space in new ways.

Moving forward, composers have a difficult task: our culture, rather than actually listening to space, consumes infinite, impersonal descriptions of space through our digital devices. At a time when listening is sorely needed, composers have to find a way to harness this ongoing hypermaterialization of information to create music that satisfies our evolving cognitive style and re-injects the material aspects of voice that are quickly vanishing.

---


There will always be a trace of voice in every sound we listen to, for if we are listening, then we are speaking; as Barthes says, “listening speaks.” All voices are merely noise that have been given voice. Perhaps this is the essence of voice: a phantasmagorical leap that allows us to make something out of nothing:

It is manifest that behind the so-called curtain which is supposed to conceal the inner world, there is nothing to be seen unless we go behind it ourselves, as much in order that we may see, so that there may be something behind there which can be seen.

And for a brief moment, if you are still reading, I am your voice: I am the master behind the curtain who controls your vocalization exactly as I choose. This is your voice. And this is my voice. Or is it?

---


LIST OF REFERENCES


Riley, Denise. “‘A Voice Without a Mouth’: Inner Speech.” *Qui Parle* 14, no. 2 (Spring/Summer, 2004): 57-104.


BIOGRAPHICAL SKETCH

Born in Buffalo, NY, Andrew Babcock has been working in a variety of contexts with music and multimedia for over 15 years. Before earning his Ph.D. in Music Composition at the University of Florida, studying with Paul Koonce, Paul Richards, and James Paul Sain, Andrew completed an MA in Music Composition at the University at Buffalo, studying with Cort Lippe and Jeffrey Stadelman, and a BA in Music at Hamilton College, studying with Samuel Pellman. As a composer, Andrew's main interest lies in the transmission and perception of voice in the electroacoustic medium. His works have received numerous international prizes and honors including CICEM (Monaco, 2016), Città di Udine (Italy, 2016), Metamorphoses (Belgium, 2012), Música Viva (Portugal, 2015), Prix Destellos (Argentina, 2015), SCI/ASCAP (United States, 2016), and Sound in Space (United States, 2011) and have been featured at festivals such as ICMC, L'Espace du Son, NYCEMF, SEAMUS, Sonorities, and TIES.