Title
Appraising Timbre: Embodiment and Affect at the Threshold of Music and Noise

Author
Wallmark, Zachary Thomas

Publication Date
2014

UNIVERSITY OF CALIFORNIA

Los Angeles

Appraising Timbre:

Embodiment and Affect at the Threshold of Music and Noise

A dissertation submitted in partial satisfaction

of the requirements for the degree Doctor of Philosophy

in Musicology

by

Zachary Thomas Wallmark

2014

ABSTRACT OF THE DISSERTATION

Appraising Timbre: Embodiment and Affect at the Threshold of Music and Noise

by

Zachary Thomas Wallmark

Doctor of Philosophy in Musicology

University of California, Los Angeles, 2014

Professor Robert Fink, Co-Chair

Professor Nina Sun Eidsheim, Co-Chair

Timbre is crucial to the generation of musical affect and meaning. But despite its well-acknowledged importance, musicology remains largely "tone deaf"—timbre is in the peculiar position of being both vital to ordinary experience and invisible to analysis. In approaching timbre through the lens of embodied cognition, this dissertation aims to loosen the analytical impasse by advancing a flexible and dynamic model for understanding the material and affective dimensions of sound. I explore how timbral reactions and appraisals work in connection with the embodied mind to shape musical experience, particularly in the context of American popular music and jazz. Methodologically situated between the "two cultures" of the humanities and sciences—and drawing on results from original empirical studies using behavioral psychology, cognitive linguistic, and neuroimaging methods—I claim that timbre perception is a motor mimetic

process; we covertly mirror the bodily actions implied in the production of timbre when we listen. The larger implication of this finding is that cognition of timbre is a fundamentally social act. The dissertation is concerned, then, with the perceptual, social, and symbolic dynamics of timbre as experienced, and to best exemplify this linkage, I focus on contexts in which "musical" timbres bleed into "noise," both acoustically and epistemologically. As case studies, the dissertation considers musical contexts with visceral, polarized reception histories: the screaming saxophone in mid-1960s free jazz (as exemplified by "late" John Coltrane), and the highly distorted electric guitar and vocal timbres of contemporary extreme heavy metal.

The dissertation of Zachary Thomas Wallmark is approved.

Marco Iacoboni

Mark Johnson

Roger A. Kendall

Mitchell Morris

Robert Fink, Committee Co-Chair

Nina Sun Eidsheim, Committee Co-Chair

University of California, Los Angeles

2014

In loving memory of Anne Dhu McLucas.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF EQUATIONS

ACKNOWLEDGEMENTS

First must come the co-chairs of my committee, Bob Fink and Nina Eidsheim. Bob's inexhaustible intellect, creativity, and humor helped drive this work forward; his legendary editorial skills kept me honest as a writer; his openness to new ideas ("you want *me* to serve as PI for your brain imaging experiment?") allowed me to pursue my interests into whatever far-flung intellectual territory it eventually took me. I thank him for really pushing me as a thinker and teacher, for carving out the time to partner with Mindy and me on the OUP tone volume, and for the mound of recommendation letters he's penned in my support. Most of all, I want to thank Bob for his unwavering faith in my ability to actually *pull all this off*, even when I doubted myself. Bob has always treated me as a colleague, not a grad student, and for that I am deeply appreciative.

Bob's "Relentless Pursuit of Tone" seminar was one of two formative influences on the eventual direction of this dissertation. The other was Nina's "Musicology in the Flesh" seminar, in which I first starting playing with these ideas. Nina's impact can be felt on every page of this dissertation. Among many other things, she was the first to push me down the rabbit hole of timbre; she read, re-read, and meticulously edited virtually every word I wrote at UCLA; she served on all my committees, employed me as a student researcher, and advocated for me in every way; she acted as my trusted career counselor and was instrumental in landing me my first job; she was tremendously supportive when Remy was born, offering me everything from parenting advice to fresh-cooked meals to an apartment full of baby accoutrements. Nina has been present through every stage of this long strange trip; I cannot begin to adequately thank her for her many and deep contributions to my development as a scholar, teacher, and human being.

I am also grateful to the many contributions of Mitchell Morris. Mitchell is a force of nature, an intellectual polymath, even something of a guru. I will never forget that moment in his office: "So you want to do a *science* dissertation, huh?" I demurred. "That's a fantastic idea!" I thank him for his boundless enthusiasm and the sometimes strange but always fascinating new directions I have pursued under his guidance.

I am indebted to my outside committee member Mark Johnson, whose influence on this dissertation could not be exaggerated. Mark is the sweetest person one could hope to know, and meetings in Eugene changed the course of this project. I wish to thank him for originally hearing me out, for entertaining my hypotheses, and for accommodating my cross-fertilization of his ideas into the wacky context of timbre. In an important sense, this dissertation is a footnote to Mark's work.

Special thanks are also due to my committee's two science advisers, Marco Iacoboni and Roger Kendall. I came to the topic of this dissertation with a lot of enthusiasm but close to zero scientific training. Marco and Roger took me under their wing, patiently mentoring me in the hugely complex business of designing, funding, and carrying out empirical studies; showing me how to make sense of my data; and indulging me in all manners of naïve questions along the way. This project would have been impossible without their input and support.

Marco warmly welcomed me into his lab and gave me the encouragement—and audacity—necessary to carry out my own neuroimaging experiment. Exploring the brain during crash-course independent study meetings with Marco always had the effect of turning mine inside out. I thank him for believing in this project, for introducing me to neuroscience, for his support on the job market, and for the always-great conversation (usually about jazz).

Roger provided me with a bountiful wealth of practical know-how in the acoustics and psychoacoustics of timbre, including lab resources, guidance in analysis and statistics, and motivation to constantly refine my ideas and to *never* settle for the easy way out. His scientific integrity and attention to detail are without peer, and I have benefitted tremendously from his continuing mentorship.

Others along the way have provided invaluable big-picture guidance, editorial insight, letters, trouble-shooting help, computer programming, (unpaid) RA work, and encouragement regarding the empirical parts of this project. I thank Arnie Cox, Istvan Molnar-Szakacs, Eric Clarke, Lawrence Zbikowski, Katy Cross, Leo Moore, Jonas Kaplan, Bill Thompson, Richard Parncutt, Marita Meyer, Cody Kommers, Harish Arsikere, and Olivier Lartillot for their contributions. I also want to thank the many research subjects who helped lend a "head" to the project, and to Loren Nerell, Jerome Camal, Jeffrey Riggs, and Nina Eidsheim for their recording help.

A six-member committee was large enough, but it could have easily ballooned even larger thanks to mind-blowing interactions with the extraordinary scholars under the Schoenberg roof. Seminars, conversations, debates, and hallway banter with members of the UCLA Musicology and Systematic Musicology faculties shaped every aspect of this project, from inception to completion. In particular, I wish to thank Susan McClary, whose work, more than anyone else's, inspired me to become a musicologist, and whose early feedback and endorsement gave this dissertation legs; Elisabeth Le Guin, for introducing me to embodied ways of thinking about music; Timothy Taylor and Roger Savage, for productively challenging my ideas and pushing me toward a more critical understanding of science; Helen Rees, for helping me develop my grant-writing chops; and Olivia Bloechl, for inviting me to think about music as an ethical encounter. I also want to thank my mentors from the University of Oregon, particularly Anne Dhu McLucas, Marian Smith, Carl Woideck, and Steve Larson. Anne, you pushed me to think big and bold—your professional example will always guide my work. I continue to celebrate your life and mourn your passing. Marian and Carl—your avuncular wisdom and eager ear continue to sustain (and entertain) me. Thanks are also due to Deborah Schwartz-Kates for welcoming me into

the University of Miami music community, and Barbara Van Nostrand for her virtuosic management skills and uplifting personality. Staying calm and laughing are necessities in the dissertation process, you've helped me do both.

In addition to my committee chairs, gratitude for help with Chapter 4 is due to Ellen Waterman and Gillian Siddall, co-editors of *Sounding the Body*, whose careful prose edits and thoughtful suggestions pushed the chapter into unforeseen territories. Thank you, too, to David Ake for his keen jazz insights and steady editorial hand.

Chapter 5 benefitted from the astute and refreshingly frank comments of our editor at OUP, Suzanne Ryan, and from multiple anonymous reviewers.

My graduate student colleagues at UCLA (particularly the 2009 cohort) have provided not only good conversation, kind words, and helpful critique over these last five years, but a camaraderie that will outlive our short time at UCLA. Among many others, I want to thank Andrea Moore, Julius Carlson, Ben Court, Hyun Chang, Marissa Steingold, Alex Grabarchuk, Mike D'Errico, Oded Erez, and James Edwards. In addition, I wish to single out Mindy O'Brien for her incredible energy and enthusiasm, both as editor-in-arms of the tone book and as a friend; Mark Samples, who has been my dearest confidant and colleague since our days together at U of O, and fellow-traveler on the Taruskin Challenge adventure—his example, both in scholarship and as a family man, is *non pareil*; and Kevin Blankenship, my brother in systematic musicology solidarity, who *gets* what this dissertation is trying to do perhaps more so than anyone else. All of you inspire me deeply, and I am honored to count you as colleagues and friends.

Looking toward the future, I wish to thank the search committee (Jesús Ramos-Kitrell, Peter Kupfer, Sarah Allen, and Pam Pagels), Dean José Antonio Bowen, and Division Director Sam Holland at SMU for the extraordinary opportunity handed to me while still wrapping up

this dissertation. I could not ask for a kinder, more intelligent, dedicated, and welcoming team of colleagues than those I will be joining in Dallas this fall.

Thank you to Diana Elledge, Mark Kjos, and the rest of my colleagues at Neil A. Kjos Music Company. Besides keeping me sharp by putting me into daily contact with some amazing composers, freelance work as an editor and project manager for Kjos kept me solvent these last seven years.

A number of fellowships and grants from UCLA made this dissertation possible. Summer research mentorships with Bob Fink and Roger Kendall aided in the early development of the project; a Graduate Research Mentorship with Nina Eidsheim and a Dissertation Year Fellowship allowed me to focus entirely on the dissertation for two straight years; and a Transdisciplinary Seed Grant from the Office of the Vice Chancellor for Research enabled me to realize my dream of conducting a brain study on the neural processing of noisy timbre (with the huge help of Marco, Katy Cross, and the staff of the Ahmanson-Lovelace Brain Mapping Center, of course). At $600 a pop for scanner time, acquisition of this funding was instrumental to the successful completion of this dissertation.

To Bill Shozan Schultz, my *shakuhachi* teacher and something of a life coach, for his quiet wisdom, generosity, and insight. Although the shakuhachi chapter didn't ultimately make it into the diss, private studies with him since 2010 profoundly influenced my thinking about timbre.

To my family in Miami—Bertha, Batravil, Bernex, Rakelle, and Dean. Their good cheer and delectable *manje kreyol* kept me going through the long slog of dissertation writing on South Beach. *Mesi anpil*.

To my parents Gary and Linda, my brother Connor, and my Papa and Nana for a lifetime of love, laughs, and music. You implanted me with my passion for music, you nurtured it through years of private lessons, band practice, and noisy rehearsals in the basement, and you had faith that I could be happy *and* make a living doing this for a career (even when I seemed a perpetual student). Countless friends and teachers shaped me as a musician along the way, but lest these acknowledgements bulge to fill the remainder of this dissertation, I trust you know who you are. My most profound gratitude is owed to you.

And finally, to my wife Marie and son Remy. You are the reason for all of this, the prime motivation for all the effort that went into this dissertation. Marie, your wisdom, empathy, humor, and indefatigable patience continually amaze me; Remy, I have never known a human being who can so effortlessly make me smile. Thank you both for continually reminding me of what's most important.

VITA

2003	B.M., Music Performance (Bass), *cum laude* New York University
2006–07	Graduate Teaching Fellow Department of Music History/Ethnomusicology University of Oregon, Eugene
2007	M.A., Music History and Ethnomusicology Department of Music History/Ethnomusicology University of Oregon, Eugene
2007–14	Music Editor and Project Manager Neil A. Kjos Music Publishing Company San Diego, California
2009–10	Regents/Musicology Fellowship Department of Musicology University of California, Los Angeles
2010	Friends of Musicology Best Seminar Paper Award and Stipend Department of Musicology University of California, Los Angeles
2010–11	Teaching Associate GE Freshman Honors Cluster Program & Dept. of Musicology University of California, Los Angeles
2011	Graduate Summer Research Mentorship (w/ Robert Fink) Graduate Division University of California, Los Angeles
2012	Transdisciplinary Seed Grant (w/ Marco Iacoboni & Robert Fink) Office of the Vice Chancellor for Research University of California, Los Angeles
2012	Graduate Summer Research Mentorship (w/ Roger Kendall) Graduate Division University of California, Los Angeles
2012–13	Graduate Research Mentorship Fellowship (w/ Nina Eidsheim) Graduate Division University of California, Los Angeles
2013–14	Dissertation Year Fellowship Graduate Division University of California, Los Angeles
2014–	Assistant Professor of Musicology Meadows School of the Arts Southern Methodist University, Dallas

SELECTED PUBLICATIONS

Forthcoming	*The Relentless Pursuit of Tone: Timbre in Popular Music.* Co-edited with Robert Fink and Mindy O'Brien. New York: Oxford UP.
Forthcoming	"The Sound of Evil: Timbre, Body, and Sacred Violence in Death Metal." Chapter in *The Relentless Pursuit of Tone.*
Forthcoming	"Theorizing the Saxophonic Scream in Free Jazz Improvisation." Chapter in *Sounding the Body: Improvisation, Representation and Subjectivity*, eds. Ellen Waterman and Gillian Siddall. Durham, NC: Duke UP.
2013	"Noise, Music, and Power in the Industrial City, 1890–1930." *Music Research Forum* 28: 55-88.
2012	"Sacred Abjection in Zen Shakuhachi." *Ethnomusicology Review* 17 (online journal)
2008	"An Alternative Temporal Approach to Jazz Improvisation in the Music of Andrew Hill." *Dutch Journal of Music Theory [Tijdschrift voor Musiektheorie]* 13, no. 1: 69-75.

SELECTED PRESENTATIONS

2014	"Is Timbre a Metaphor?" Paper accepted for the national AMS/SMT joint meeting, Milwaukee. (Forthcoming)
2013	"Big Data and Musicology: New Methods, New Questions." Paper read at the national meeting of the AMS, Pittsburgh.
2012	"Embodied Listening and Musical Empathy: Perspectives from Mirror Neuron Research." (w/ Marco Iacoboni) Paper read at the national AMS/SEM/SMT joint meeting, New Orleans.
2012	"The Mirroring Mind: Embodied Cognition, Musical Empathy, and the Neuroscience of Aurality." Paper read at "Oral Traditions, Old and New" Symposium, University of Oregon, Eugene.
2012	"'The Most Powerful Human Sound Ever Created'? Embodied Cognition, Empathy, and the Ethics of Timbre in Free Jazz Reception." Paper read at "Music and the Body" Conference, University of Hong Kong, China.
2010	"'I Would Like to Die': Constructions of the Eroticized Body in the Monteverdi Madrigal." Paper read at "Material and Imagined Bodies" conference, Brown University, Providence.
2007	"An Alternative Temporal Approach to Jazz Improvisation in the Music of Andrew Hill." Paper read at annual Dutch-Flemish Society for Music Theory conference, Groningen, The Netherlands.

Introduction: The Problem of Timbre

> We should here recognize that timbre… is by far the most important aspect of tone and introduces the largest number of problems and variables.
>
> — Carl Seashore[1]

> Timbre is that attribute of auditory sensation in terms of which a listener can judge that two sounds similarly presented and having the same loudness and pitch are dissimilar.
>
> — American Standards Association, *Acoustical Terminology* (§6.05)[2]

> I think the definition of timbre by the American Standards Association should be this: "We do not know how to define timbre, but it is not loudness and it is not pitch." […] Until such time as the dimensions of timbre are clarified perhaps it is better to drop the term timbre…
>
> — Alfred Bregman[3]

Timbre is a shadow. It clings to sounds as a sonic silhouette; it mirrors them as they move. But when subjected to the bright light of analysis, it vanishes. We all know timbre contributes profoundly to musical experience, but it is hard to explain how it works, the nature of our affective responses. As Carl Seashore well understood, timbre is nothing if not problematic: we can recognize that it is "the most important aspect of tone," that the *way music sounds* affects what and how music means, yet we remain largely ignorant as to why. "A large number of problems," indeed.

---

[1] Carl E. Seashore, *Psychology of Music* (New York: Dover, 1938/1967), 21.

[2] *American Standard Acoustical Terminology*, (New York: American Standards Association, 1960), 45.

[3] Alfred S. Bregman, *Auditory Scene Analysis: The Perceptual Organization of Sound* (Cambridge, MA: MIT Press, 1990), 93-94.

One problem is the concept itself. To psychologist Alfred Bregman, the term timbre is "an ill-defined wastebasket category," a negative definition that specifies its object only in terms of what it lacks.[4] Frustration over this definitional ambiguity is likely familiar to anyone who has taught a "fundamentals of music" class: melody, harmony, and rhythm are easy enough to explain, but timbre provokes head scratching. Is it the texture of the orchestration? Is it the reverb on a Phil Spector track? Is it just a fancy French label for different instruments and voices? Yes, yes, and yes… Sort of. Lacking a uniform descriptive vocabulary to explain what exactly makes an oboe different from a clarinet, the beleaguered teacher may well resort to paraphrasing Justice Potter Stewart's famous ruling on obscenity: "you know it when you hear it."

Of course, we all know (if non-consciously) what timbre *is*—our innate ability to rapidly determine when two sounds are "dissimilar" is essential to language and, even more, survival.[5] The chief problem for the musicologist (and the central problem confronting this dissertation) is not primarily definitional or taxonomic; rather, we clearly don't understand with any specificity what timbre *does* to listeners. Indeed, the music-analytical tools at our disposal—and the ideologies that buttress them—were not designed with timbre in mind.[6] To generations of music scholars, timbre has been considered either irrelevant to analysis, invisible to it, or just so cussedly difficult to pin down that it is not worth the hassle. (As Robert Walser acknowledges, "timbre is the

[4] Ibid., 92.

[5] Timbre is the primary psychoacoustic ingredient in phonemic differentiation—spoken language without dissimilarities in vocal timbre is inconceivable. Further, timbre is fundamental to indentifying the sources of sounds, making it an essential evolutionary adaptation (to be discussed in Chapter 1). For more, see respectively: Aniruddh D. Patel, *Music, Language, and the Brain* (New York: Oxford University Press, 1999); Stephen Handel, *Listening: An Introduction to the Perception of Auditory Events* (Cambridge, MA: MIT Press, 1989).

[6] A cogent and singular critique of the ideological moorings of music theory's "timbre deafness" can be found in John Shepherd and Peter Wicke, *Music and Cultural Theory* (Cambridge: Polity, 1997). Although I do not address ideology in any depth in this dissertation, I recognize the critical importance of ideology critique in the establishment of any "new" analysis: see Robert Fink, "'Arrows of Desire': Long-Range Linear Structure and the Transformation of Musical Energy" (Ph.D. Diss., University of California–Berkeley, 1994).

least successfully theorized and analyzed of musical parameters.")[7] Instead, musicologists tend to focus on those parameters of sound that *do things*: pitch, harmony, rhythm, and form. Timbre thus remains in a double bind: it is either equated directly with a sound source (a *saxophone*) or a quality of sound (a *silky* saxophone), but it is not generally conceptualized as a *verb*, a prerequisite for earning a place at the table of music analysis. Cornelia Fales observes: "To the general listener, pitch and loudness are variable characteristics of sound, timbre is a condition; pitch and loudness are things a sound does, timbre is what a sound is."[8]

Endemic "timbre deafness" is the result. To many, timbre is seen as disclosing the world of things; it does not signify in the musical sense. Maurice Merleau-Ponty puts this baldly: "In music… meaning appears as linked to the empirical presence of the sounds, and that is why music strikes us as dumb."[9] Correction: that is why *timbre* strikes us as dumb. It is too close to the material, while music seems to demand the ideal; it is too direct, signifying nothing other than itself, while music needs to mean something beyond the wood, steel strings, tube amplifiers, and moving fingers that give it voice. To Emmanuel Levinas, the sound of the cello is just that: "celloness."[10] Timbre is objective aural evidence of physical presence, nothing more.

Or could it be the other way around? Fales emphasizes the dual—"paradoxical"—nature of the phenomenon: it originates as measurable vibrations in the "acoustic world," but does not turn into timbre until reaching the subjective "perceived world."[11] In this way, "perceived timbre

---

[7] Robert Walser, "The Body in the Music," *College Music Symposium* 31(1991): 122.

[8] Cornelia Fales, "The Paradox of Timbre," *Ethnomusicology* 46, no. 1 (2002): 58.

[9] Maurice Merleau-Ponty, *Phenomenology of Perception*, trans. Colin Smith (London and New York: Routledge, 1962), 219.

[10] Emmanuel Levinas, *Otherwise Than Being: or, Beyond Essence*, trans. Alphonso Lingis (The Hague and Boston: M. Nijhoff, 1981), 41.

[11] The term "paradox of timbre" comes from Fales: see Fales, "The Paradox of Timbre," 61.

exists in a very real sense only in the mind of the listener, not in the objective world."[12] Similarly, as John Hajda *et al.* explain, "variability among listeners means that stimuli *never* will be perceptually equal for every subject."[13] Seen from this angle, we are confronted with a different sort of problem: how do we know at all something so intensely subjective? Timbre is *both* the "dumb" consequence of material sound production, *and* the inner sound in our mind's ear when we perceive it. Being affected by it, then, is merely a matter of personal taste. Timbre thus seems to exist in limbo: too objective to signify, too subjective to be taken seriously. The paradox deepens.

Implicit in this paradox, too, is the Cartesian negative of Fales's comment—that timbre exists only in the *body* of the listener. Making sound, particularly the "grain" of vocal sound, can be an intimately corporeal experience, and its perception is similarly "internal."[14] This perspective drives timbre even farther from the rationalizing discourse of music: raw sound (or *phone*), as far back as the Greeks, has been seen as dangerously close to the flesh, sensuality, and the irrational.[15] Without the power to signify, "meaningless" *phone* is relegated to the fringes, associated with other outcasts from the discursive order of things (women, children, animals, etc.). In sum, we might say that timbre is a material girl living in an ideal world of Western music scholarship (to ludicrously mash up Madonna and Hegel).

So the practical question is: where do we begin in addressing the problem of timbre? Although I do not venture as far as Bregman in proposing that we drop the term, timbre is clearly long overdue for critical musicological treatment. In this dissertation I try to account for how

---

[12] Ibid., 61-62.

[13] John Hajda *et al.*, "Methodological Issues in Timbre Research," in *Perception and Cognition of Music*, ed. Iréne Deliége and John Sloboda (East Sussex: Psychology Press, 1997), 263. Italics in original.

[14] Here, of course, I am referencing Barthes's classic essay "The Grain of the Voice": see Roland Barthes, *Image, Music, Text*, trans. Stephen Heath (New York: Hill and Wang, 1977).

[15] Raw *phone* stands in contrast to *phone semantike*, sound that signifies (i.e., language). See Adriana Cavarero, *For More Than One Voice: Toward a Philosophy of Vocal Expression*, trans. Paul A. Kottman (Palo Alto, CA: Stanford University Press, 2005).

timbre *means* to listeners, how it affects us. Doing so requires an ontological, epistemological, and empirical reevaluation of the conventional wisdom.

Framing the Project

> I say sometimes to guys, don't play music. Play a sound… Because it's the tone, in some cases, not the music you play.
>
> — Neil Young[16]

> It's not what you play, it's how you play it.
>
> — Mary Lou Williams[17]

As every musician and every listener intuits, tone matters; yet the affective potency of timbre often evades conscious reflection.[18] The dynamic relationship between non-conscious perceptual experience and affective power serves as my point of departure for this dissertation. My desire to explore the "problems" so long associated with timbre arises from observation of everyday musical encounters—specific situations of listening with social and ethical valences—in which timbre plays a leading though largely invisible role. Consider, for instance, the collapse of empathy that can occur when other people "sound bad": the coastal urbanite with a contemptuous aversion to the "twang" of country music, the parent who hears a child's punk rock as "noise," the elderly couple covering their ears at the "acousmatically black" sound of bass-heavy hip-hop from a

[16] Quoted in Richard Echard, *Neil Young and the Poetics of Energy* (Bloomington, IN: University of Indiana Press, 2005), 84.

[17] Mary Lou Williams on Marian McPartland's "Piano Jazz," NPR, 1978.

[18] Writes Fales: "The very subliminality of these sensations is a clue to their power" (p. 77). Like ideology or religious ritual, timbre works because we *aren't aware* that it's working.

passing car.[19] In these ubiquitous moments, timbre structures experience in a way that reinforces difference.[20] If, as Fales suggests, timbre works with "secretive discretion" and "subterranean impact,"[21] then this presents us with a real epistemological problem: How do we rationalize an experience that operates sub-rationally? How do we think through a musical perception that is not generally *thought* about? The fact that timbre so often eludes conscious reflection makes it particularly potent in the formation of music-mediated judgments.

As shorthand for the distinction between the unconscious and the conscious sides of listening to timbre, I evoke David Huron's idea of *reaction* and *appraisal responses*.[22] Reactions have three characteristics: they are fast, unconscious, and self-protective. They are primarily, then, a matter of biology forged in the "deep history" of human evolution. Appraisals, by contrast, are slower, conscious, and largely determined by a concatenation of contextual and cultural factors. (The wince is the reaction; the feeling "that's just noise!" is the appraisal.) But appraisal should not be mistaken for "disinterested interest"; rather, it is grounded to some degree on the original bodily reaction. Thus in an instance of feedback, reactions influence appraisals, but appraisals—which are culturally conditioned—ultimately trump reactions. A central task of this dissertation is to disentangle the complex relationship between the two, asking where reactions and appraisals converge and where they diverge in the experience of listening to timbre.

---

[19] This last term comes from Mendi Obadike, "Low Fidelity: Stereotyped Blackness in the Field of Sound" (Ph.D. Diss., Duke University, 2005).

[20] If we assent to the definition of timbre proposed by the American Standards Association, indeed timbre is *about* difference: it allows us to hear two things as "dissimilar." How we then discursively map this perceptual difference onto the social, of course, is a matter far-afield from the acoustic definition. I will return to this question in various forms throughout the dissertation.

[21] Fales, "The Paradox of Timbre," 77.

[22] David Huron, *Sweet Anticipation: Music and the Psychology of Expectation* (Cambridge, MA: MIT Press, 2006), 13-15. Huron was not the first to evoke this terminology in exploring the distinction between these two response types. I will explore appraisal theory (and its precedents) in Chapter 1.

Just how essential is timbre to musical reactions and appraisals? To probe this question empirically, I carried out a simple psychoacoustic experiment (Appendix 4c, p. 376):[23] participants were asked to rate two sets of popular music excerpts, one consisting of 4-second clips and the other of much shorter (400-millisecond) versions of the same music. The longer condition provided the listener with enough musical context—melodic and harmonic elements, rhythm, groove, genre, and lyrics—to make an informed rating of how much they "liked" the music.[24] Most of the excerpted selections were well known, too, giving participants even more context to aid in their appraisal.[25] The short excerpts, by contrast, provided no tonal, rhythmic, or lyrical information—they were simply split-second blips of sound, enough time to convey essential aspects of timbre but nothing else.

The results surprised me. I expected longer excerpts would have higher scores,[26] but instead, I found no statistically significant difference between ratings in the two conditions: the appraisal was the same whether listeners heard an identifiable excerpt or just a brief timbral snapshot. What does this mean? If the additional musical context of the long version *did not* affect appraisal, it stands to reason that judgments of music were formed in that first split-second.[27] 400 milliseconds is enough time to form a "correct" appraisal of music; having access to more of the

---

[23] Thank you to collaborator Kevin Blankenship, RA Cody Kommers, and to members of the UCLA Workshop for Empirical Music Studies (WEMS) for *in absentia* help with this experiment.

[24] I used a standard semantic differential paradigm: listeners rated excerpts on a 0–100 scale from "strongly dislike" to "strongly like." See Appendix 4c for details and data analysis.

[25] For example, Hendrix's "Purple Haze," Led Zeppelin's "Whole Lotta Love," and Snoop Dogg's "Gin and Juice."

[26] Musical "liking" tends to increase with familiarity: see Josh H. McDermott, "Auditory Preferences and Aesthetics: Music, Voices, and Everyday Sounds," in *Neuroscience of Preference and Choice: Cognitive and Neural Mechanisms*, ed. Raymond Dolan and Tali Sharot (London: Academic, 2012).

[27] Bigand *et al.* (2005) reached a similar conclusion using 1-second stimuli. Since 400 milliseconds provides even less context, the fact that their result holds in this experiment underlines the affective immediacy of musical timbre. See Emmanuel Bigand *et al.*, "Multidimensional Scaling of Emotional Responses to Music: The Effect of Musical Expertise and of the Duration of the Excerpts," *Cognition & Emotion* 19, no. 8 (2005).

musical picture simply replicates the original snap judgment. In theory, the appraisal is predicated virtually entirely on the timbral reaction.

The implications of this counterintuitive result will ripple through the dissertation. In exploring this relationship between reactions and appraisals, I attempt to connect perception to action, poesis to aesthesis, lower-order psychobiological responses to higher-order cognition. The connective tissue that runs throughout this story, and the key to understanding timbre's "subterranean impact," I contend, is the body.[28] To account for the perceptual systems and cognitive processes linking the lived body with musical experience, I approach timbre from a vantage point that takes in the growing field of embodied cognition. Timbre is the sonic result of material engagement, imbued with the audible traces of bodies in motion, and, as recent neurological research has established, we motorically mirror (consciously or unconsciously) the fleshly circumstances of production when we listen.[29] By connecting timbral reactions and appraisals to their embodied-cognitive grounding, I hope in this dissertation to outline a general analytical framework for musical timbre by focusing on the psychodynamics of listening.

These dynamics can be seen clearest at the extremes, in affectively polarizing, "noisy" musical contexts. The project, then, concerns itself fundamentally with *timbre as experienced*, particularly in musical situations in which "musical" timbres bleed into the timbre of "noise," both

---

[28] Musicological works influential to my thinking on musical embodiment include: David Lidov, "Mind and Body in Music," *Semiotica* 66, no. 1 (1987); Susan McClary, *Feminine Endings: Music, Gender, and Sexuality* (Minneapolis: University of Minnesota Press, 1991); Elisabeth Le Guin, *Boccherini's Body: An Essay in Carnal Musicology* (Berkeley and Los Angeles: University of California Press, 2005); Tomie Hahn, *Sensational Knowledge: Transmitting Japanese Dance and Music* (Middleton, CN: Wesleyan University Press, 2007); Nina Sun Eidsheim, "Voice as a Technology of Selfhood: Towards an Analysis of Racialized Timbre and Vocal Performance" (Ph.D. Diss., University of California–San Diego, 2008).

[29] A review of the literature relevant to this important conclusion appears in Chapters 1 and 2.

acoustically and epistemologically.[30] The contextual threshold of music and noise is an "acid test" for hypotheses on embodied timbre perception: indeed, the mechanics of the process are laid bare most vividly in situations where music is heard as noise. I focus on this "noise" appraisal in the context of two cases wherein timbre plays an essential (perhaps *the* essential) role in affective response: free jazz, which was heard by some critics of the mid-1960s as a "tumult of noise," and extreme heavy metal, which has been described as "nothing more than a bunch of noise."

To do so, the dissertation builds on the work of music scholars exploring the intersection of embodiment and timbre, and four models in particular directly influence my broader thesis. Robert Walser's *Running with the Devil* (1993)—specifically the 3-odd pages dealing with guitar and vocal timbre—prefigures one of the main arguments in the dissertation, that timbral qualities of vocal exertion (i.e., distortion) condition our affective responses to other similar sounds, which are in turn conventionalized into signs.[31] In some way, this dissertation could be considered an extended working-through of this original insight. Significant, too, is the rare work of ethnomusicologist Cornelia Fales, particularly her exceptional 2002 article "The Paradox of Timbre."[32] While not directly concerned with embodied cognition *per se*, her deft blending of psychoacoustics and music analysis serves as an exemplar for how to fuse empirical and musicological discourses. In his "mimetic hypothesis," music theorist Arnie Cox (1999, 2001, 2011) provides another com-

---

[30] It is the slippage between acoustical and epistemological understandings of noise that are of most interest to me in this project. Thus, while I draw upon the classic theoretical sources on the idea of "noise" (e.g., Jacques Attali's *Noise* [1985]), I try to connect the discourse of "noise" always to the *sound* of noise.

[31] Robert Walser, *Running with the Devil: Power, Gender, and Madness in Heavy Metal Music* (Hanover, NH: University Press of New England, 1993). The section in question can be found on pages 41-46.

[32] Fales, "The Paradox of Timbre." See also "Short-Circuiting Perceptual Systems: Timbre in Ambient and Techno Music," in *Wired for Sound: Engineering and Technologies in Sonic Culture*, ed. Paul D. Greene and Thomas Porcello (Middleton, CN: Wesleyan University Press, 2005); "Listening to Timbre During the French Enlightenment" (paper presented at the Conference on Interdisciplinary Musicology, Montreal, Quebec, Canada, 2005); Cornelia Fales and Stephen McAdams, "The Fusion and Layering of Noise and Tone: Implications for Timbre in African Instruments," *Leonardo Music Journal* 4(1994).

pelling model for thinking through music as a motor-mimetic phenomenon, one with special links to human vocality.[33] Finally, Nina Eidsheim's work on racialized vocal timbre opened the door for this project by demonstrating both the bodily dimensions of timbre perception, and the social stakes of listening to timbre.[34] Eidsheim shows that appraisals of timbre are always about a whole lot more than just sound. (Since Chapter 1 serves as a more detailed literature review, this greatly truncated account is to be continued.)

Methodology

> For half a century the humanities have examined human difference, thickly describing individual instances of it, theorizing it, and exploring its political, ethical, and aesthetic implications. Meanwhile, it is probably fair to say, the broadest import of the humanistic enterprise in the modern university has remained underdescribed and undertheorized. The careful linking of evolutionary studies (and cognitive studies, too) to the precise, eminently humanistic delineation and evaluation of local human behaviors can reinvigorate this broader discourse, reconciling the two propositions bound up in the word *humanities* itself: that there is a common humanness at stake and that it is always and everywhere plural.
>
> — Gary Tomlinson[35]

Timbre research is inherently interdisciplinary. It has historically been taken up in departments all over the university, from physics to philosophy, psychology to electrical engineering, musicol-

[33] Arnie Cox, "The Metaphoric Logic of Musical Motion and Space" (Ph.D. Diss., University of Oregon, 1999); "The Mimetic Hypothesis and Embodied Musical Meaning," *Musicae Scientiae* 5, no. 2 (2001); "Embodying Music: Principles of the Mimetic Hypothesis," *Music Theory Online* 17, no. 2 (2011).

[34] Eidsheim, "Voice as a Technology of Selfhood: Towards an Analysis of Racialized Timbre and Vocal Performance."; "Synthesizing Race: Towards an Analysis of the Performativity of Vocal Timbre," *TRANS-Transcultural Music Review* 13(2009); "Sensing Voice: Materiality and the Lived Body in Singing and Listening," *Senses & Society* 6, no. 2 (2011).

[35] Gary Tomlinson, "Evolutionary Studies in the Humanities: The Case of Music," *Critical Inquiry* 39, no. 4 (2013): 674. Italics in original.

ogy to neuroscience. Yet scholarly treatments attempting to synthesize these diverse perspectives into an integrated approach remain rare.[36] Instead, timbre research tends to be divided into two camps roughly mapping onto the old "two cultures" divide: humanities scholars acknowledge the historical, cultural, aesthetic, semiotic, and discursive significance of timbre while seldom exploring its perceptual and cognitive grounding;[37] scientists document its underlying acoustic forms, perceptual systems, cognitive mechanisms, and neural wiring while rarely attempting to place empirical results into musical and cultural context.[38] Thus, while both sides have developed sophisticated tools for conceptualizing timbre within their own epistemological purview, there exists a deep lacuna where these perspectives overlap. And it is precisely in this overlap—at the convergence of reaction and appraisal—where timbre becomes affectively meaningful.

The methodological challenge I set for myself in this project (as evidenced in the experiment reported above) is to begin uniting multiple disciplinary threads into a perceptually-based model for analyzing the affective role of timbre in musical appraisal. The dissertation will rely on insights and methods from a wide range of disparate fields, including those traditionally associated with the humanities—music analysis, reception history, theory, hermeneutics—and those associated with the physical and biological sciences, including acoustics, psychoacoustics, and cognitive neuroscience. My method is thus syncretic at a level beyond the simple blurring of disciplinary borders: in combining very different types of knowledge claims, the project might be characterized as *inter-epistemological*, drawing on empirical research to inform my analyses and in-

[36] The first major attempt at an interdisciplinary account of timbre is the classic work of Helmholtz: see Hermann von Helmholtz, *On the Sensations of Tone as a Physiological Basis for the Theory of Music*, trans. Alexander Ellis (New York: Dover, 1867/1954).

[37] See Shepherd and Wicke, *Music and Cultural Theory*; Emily I. Dolan, *The Orchestral Revolution: Haydn and the Technologies of Timbre* (Cambridge: Cambridge University Press, 2012).

[38] Cf. Reinier Plomp, *Aspects of Tone Sensation: A Psychophysical Study* (London: Academic Press, 1976); Handel, *Listening: An Introduction to the Perception of Auditory Events*.

terpretations of case study materials. In focusing on the experience of listening, moreover, it aspires to "empirical phenomenology" (in Mark Johnson's formulation), and at certain points, "neurophysiologic phenomenology" (in Marco Iacoboni's); although I do not engage in great depth with the classic phenomenological literature, the dissertation aims to chart a "geography of human experience" regarding musical timbre.[39] Since the Appendices serve as a detailed description of these empirical methods, here I will focus on the broader inter-epistemological terrain of this geography.

Wishing to account for timbral affect by investigating both sides of the equation (reaction and appraisal), the dissertation is situated at the convergence of two related philosophical tensions: first, between the "soft" humanities and the "hard" sciences, and second, between the material and the ideal (in the Hegelian sense). I enter with full awareness of the cultural baggage associated with these tensions, and I make no attempt to resolve them in any facile way. I will not, therefore, be drawing on empirical research in order to provide an "authoritative" grounding to my musical analyses; neither will I set up empirical work as a straw man in a "science studies"-styled critique. Rather, the project attempts to enact a conversation between these diverse perspectives in which each voice is equal, while also keeping the inherent tensions present throughout. Models for this sort of inter-epistemological method can be found in Eric Clarke's *Ways of Listening* (2005), which refracts diverse musical case studies through the lens of ecological psychoacoustics; Lawrence Zbikowski's *Conceptualizing Music* (2002), which explores the cognitive bases of music listening and analysis; Judith Becker's *Deep Listeners* (2004), which juxtaposes eth-

[39] See respectively Mark Johnson, *The Body in the Mind: The Bodily Basis of Meaning, Imagination, and Reason* (Chicago: University of Chicago Press, 1987), xxxvii; Marco Iacoboni, *Mirroring People: The Science of Empathy and How We Connect with Others* (New York: Picador, 2009), 17. The term "geography of human experience" comes from Johnson.

nographic observations of musical trance with its associated neurophysiology; and Jonathan De Souza's recent dissertation (2013) on musical instruments and grounded cognition.[40]

Edward Slingerland's *What Science Offers the Humanities* (2008) has proven a singularly valuable (and validating) resource in my navigation of this potentially perilous in-between.[41] Slingerland, an Asian Studies professor, sees the natural sciences not as providing a privileged, more "real" epistemic alternative to traditional humanistic modes of inquiry, but simply as a necessary, lower-order link in the explanatory chain of human culture and behavior. Reciprocally, Slingerland argues that humanists should have a special stake in the production of *scientific* knowledge. He summarizes his case for consilience:

> Bringing the humanities and the natural sciences together into a single, integrated chain seems to me the only way to clear up the current miasma of endlessly contingent discourses and representations of representations that currently hampers humanistic inquiry. By the same token, as natural scientists begin poking their noses into areas traditionally studied by the humanities—the nature of ethics, literature, consciousness, emotions, or aesthetics—they are sorely in need of humanistic expertise if they are to effectively decide what sorts of questions to ask, how to frame these questions, and what sorts of stories to tell in interpreting their data.[42]

To Slingerland, a more scientific humanities goes hand-in-hand with a more humanistic science. Both sides benefit from the exchange.

---

[40] Eric F. Clarke, *Ways of Listening: An Ecological Approach to the Perception of Musical Meaning* (Oxford: Oxford University Press, 2005); Lawrence M. Zbikowski, *Conceptualizing Music: Cognitive Structure, Theory, and Analysis* (New York: Oxford University Press, 2002); Judith Becker, *Deep Listeners: Music, Emotion, and Trancing* (Bloomington: University of Indiana Press, 2004); Jonathan De Souza, "Musical Instruments, Bodies, and Cognition" (Ph.D. Diss., University of Chicago, 2013).

[41] Edward Slingerland, *What Science Offers the Humanities: Integrating Body and Culture* (Cambridge: Cambridge University Press, 2008). Thank you to Mark Johnson for steering me in the direction of this book.

[42] Ibid., 9.

Some music scholars are coming to agree with Slingerland's position.[43] As the author of the epigraph to this section evinces, I am clearly not the first musicologist to traverse the "two cultures" divide and dabble in the sciences, though most scholars of this type could be fairly characterized (as I think they would agree) as "consumers" of empirical information rather than "producers" of it. Taking Slingerland's statement seriously, I felt it was important in this dissertation that I not only incorporate perspectives from the sciences, but that I myself *add* to the burgeoning scientific conversation on embodied music cognition. In addition to drawing deeply on the musicological literature on timbre, then, I take the (as yet) somewhat unusual step of designing and carrying out original empirical experiments to query issues of timbre, noise, and affect both at the level of perception (using techniques from experimental psychology) and neurophysiology (using fMRI).[44] The Appendices and associated Glossary, which report these experiments, thus serve as my own primary literature in the analyses that follow.

Uniting the diverse methods and intellectual orientations of this dissertation is the inherently interdisciplinary research program of *embodied cognition.*[45] To quote developmental psychologist Esther Thelen's definition:

---

[43] To name a few that come to mind: Gary Tomlinson, Ingrid Monson, Arnie Cox, Judith Becker, Robert Walser, Anne Dhu McLucas, Lawrence Zbikowski, Michael Bakan, Theresa Allison, and Kevin Blankenship.

[44] Music scholarship that deals with timbre in a thoroughgoing and comprehensive manner includes: Robert Erickson, *Sound Structure in Music* (Berkeley and Los Angeles: University of California Press, 1975); Robert Cogan, *New Images of Musical Sound* (Cambridge, MA: Harvard University Press, 1984); Wayne Slawson, *Sound Color* (Berkeley and Los Angeles: University of California Press, 1985); Fales, "The Paradox of Timbre."; "Short-Circuiting Perceptual Systems: Timbre in Ambient and Techno Music."; Eidsheim, "Voice as a Technology of Selfhood: Towards an Analysis of Racialized Timbre and Vocal Performance."; David K. Blake, "Timbre as Differentiation in Indie Music," *Music Theory Online* 18, no. 2 (2012); Rebecca Leydon, "Clean as a Whistle: Timbral Trajectories and the Modern Musical Sublime," ibid.; Robert Fink, Mindy LaTour O'Brien, and Zachary Wallmark, eds., *The Relentless Pursuit of Tone: Timbre in Popular Music* (New York: Oxford University Press, forthcoming).

[45] For a good general primer on embodied cognition, see Lawrence Shapiro, *Embodied Cognition* (London and New York: Routledge, 2010). My understanding of the topic has been shaped by the following (now seminal) sources: Francisco J. Varela, Evan Thompson, and Eleanor Rosch, *The Embodied Mind: Cognitive Science and Human Experience* (Cambridge, MA: MIT Press, 1991); Andy Clark, *Being There: Putting Brain, Body, and World Together Again*

> To say that cognition is embodied means that it arises from bodily interactions with the world. From this point of view, cognition depends on the kinds of experiences that come from having a body with particular perceptual and motor capacities that are inseparably linked and that together form the matrix within which memory, emotion, language and all other aspects of life are meshed.[46]

Including music. By connecting the contingencies of the sensing body with the broader material and cultural environment, embodied cognition offers an expansive framework for understanding the role of human perceptual systems in the process of producing, hearing, and making sense of timbre. It also helps to account for the space between "musical timbre" and "noisy timbre." As I will argue, acoustical noise is an index of bodily exertion; perceiving and conceptualizing noisy timbre is thus also, via mechanisms of act–sound coupling, *resonating* with the physical exertion of others. Hearing timbre is doing timbre.

The central ontological revision performed in this dissertation relates to timbre's largely passive status in contemporary music scholarship, its often-assumed direct link to source *identity* over source *behavior*.[47] Using tools from embodied cognition and frameworks from musicology, I suggest that timbre is a more active phenomenon: it is a verb, something we *do with our bodies*,

---

(Cambridge, MA: MIT Press, 1998); George Lakoff and Mark Johnson, *Philosophy in the Flesh* (New York: Basic Books, 1999); Alva Noë, *Action in Perception* (Cambridge, MA: MIT Press, 2004); Andy Clark, *Supersizing the Mind: Embodiment, Action, and Cognitive Extension* (New York: Oxford University Press, 2008). Music scholars have compellingly placed this theoretical framework into dialogue with music: see Vijay Iyer, "Microstructures of Feel, Macrostructures of Sound: Embodied Cognition in West African and African-American Musics" (Ph.D. Diss., University of California–Berkeley, 1998); Cox, "The Mimetic Hypothesis and Embodied Musical Meaning."; Marc Leman, *Embodied Music Cognition and Mediation Technology* (Cambridge, MA: MIT Press, 2007).

[46] Esther Thelen *et al.*, "The Dynamics of Embodiment: A Field Theory of Infant Perseverative Reaching," *Behavioral and Brain Sciences* 24(2001): xx.

[47] Psychoacousticians, on the other hand, have long recognized the link between timbre and the ecological, material circumstances of sound production. Nonetheless, psychological timbre studies that take an ecological approach remain rare compared to those that take an "information processing" perspective. For a helpful review of these two paradigms, see Stephen McAdams, "Recognition of Sound Sources and Events," in *Thinking in Sound: The Cognitive Psychology of Human Audition*, ed. Stephen McAdams and Emmanuel Bigand (Oxford: Clarendon Press, 1993).

both in production and in perception.[48] These bodily, experiential structures of timbre influence how we ultimately think about it.

ASPECS: Towards an Analytical Model for (Noisy) Timbre

This dissertation advances a "vertically integrated" approach for understanding the interrelationship of timbral reactions and appraisals.[49] The model—which I am giving the tentative acronym ASPECS—is less of a codified method *per se* than a loose conceptual framework for thinking through timbre, and particularly the "noisy" sounds that excite polarizing responses.[50] In this spirit, I propose treating the phenomenon as consisting of six distinct but overlapping moments, in more or less their order of phenomenal unfolding: *Act, Sound, Perception, Experience, Concept,* and *Sign* (Fig. 1):

---

[48] In drawing attention to the active character of music—a verb rather than a thing—Christopher Small's "musicking" concept is fundamental to the basic premise advanced in these pages: see Christopher Small, *Musicking: The Meanings of Performing and Listening* (Middletown, CN: Wesleyan University Press, 1998).

[49] "Vertical integration" is a major preoccupation of Slingerland's book. He views the vertical integration of knowledge as a "rough explanatory hierarchy, with the lower levels of explanation (such as physics) setting limits on the sorts of explanations that can be entertained at the higher levels…" (p. 261). To Slingerland, the humanities sit at the apex of this hierarchy, but they are buttressed (or, put more provocatively, "grounded") by all the lower levels.

[50] The ASPECS model applies equally to *all music*, not just the isolated attribute of timbre. I developed this model with timbre in mind, however, to emphasize the successive cognitive stages that go into its experience: if indeed timbre is uniquely occluded from conscious reflection, then it is my hope that such a systematic, step-by-step procedure will aid in elucidating an experience that may otherwise go unexamined.

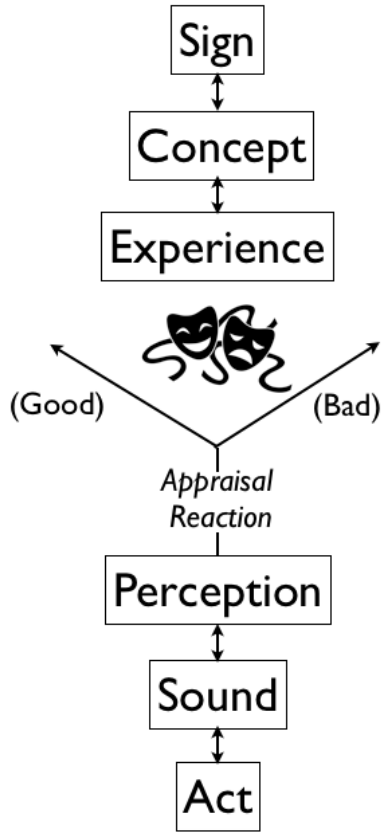


FIGURE 1: ASPECS model

As I will argue, timbre is first and foremost the consequence of movement and interaction, the byproduct of energy transfer between material objects (*act*). This process results in the creation and propagation of vibratory energy (*sound*), which could be understood as bodily acts transduced into acoustics. Sound is then picked up by the sensorimotor systems of the listener (*perception*). In sum, these first three stages can be considered as fundamentally reactive: they comprise the basic psychobiology of sound generation and listening, and are best understood through empirical explanations.

The appraisal process, however, adds higher-order cognition to the mix, beginning with the most basic appraisal of all: Is it good or bad (or neutral)? This is a categorical bifurcation (as addressed in Chapter 2), one that colors everything that sits atop it with "thick" moral, aesthetic, social, and emotional meanings.[51] To address the remaining three stages in turn: Perception takes on a different phenomenological quality when imbued with gestalt structure (*experience*). As philosopher Mark Johnson has argued, sensorimotor engagement with the world is image schematic; bodily experience forms a coherent set of pre-conceptual patterns that influence all thought.[52] The *experience* level thus enables and constrains the ways we understand timbre as an abstract idea (*concept*), and the conceptualization stage is facilitated to a large degree by metaphor. Finally, at the top of the vertically-integrated hierarchy, conceptualization enables timbre to do cognitive work outside of its immediate indexical moment—it can refer to something else entirely (*sign*).[53] When qualities of timbre are conventionalized into signs, meaning becomes polymorphous: the "twang" of country can reference rural Others (and their politics, class, race, etc.), in-group cultural solidarity, authenticity, or anything else, depending upon the interpretant.[54] The timbral *sign* is thus also the social level of timbre, where sound can come to represent real or imaginary groups of people beyond its original indexical link to moving bodies. It is also the level where timbre can deny meaning entirely. Noise is a preeminent instance of this: by vexing (musical) meaning, it opens us onto other interpretive possibilities (the Sublime, the "wholly other," divinity?). Spiritual readings of the threshold between music/noise are widespread in the reception of certain "noisy" sounds, as we will see in the case studies.

---

[51] The terminology "thick" and "thin," which I will use throughout, is borrowed from Clifford Geertz's account of cultural interpretation.

[52] See Johnson, *The Body in the Mind: The Bodily Basis of Meaning, Imagination, and Reason*. For example, the bodily experiences of "up/down" or physical balance influence how we conceptualize other, non-physical experiences. Chapter 3 will address this hypothesis in depth.

[53] My terminology is Peircian in the broadest sense only; what I mean by "sign" is more appropriately Peirce's "symbol." (Peirce used "sign" to refer to *all* sign-like things, i.e., icons, indices, and symbols.) For a helpful overview, see Thomas Turino, "Signs of Imagination, Identity, and Experience: A Peircian Semiotic Theory for Music," *Ethnomusicology* 42, no. 2 (1999).

[54] This fleeting example comes from Jocelyn Neal, "The Twang Factor in Country Music," in *The Relentless Pursuit of Tone: Timbre in Popular Music*, ed. Robert Fink, Mindy LaTour O'Brien, and Zachary Wallmark (New York: Oxford University Press, forthcoming).

Together, the stages of ASPECS will serve to structure much of the dissertation, but—as evinced in the two-way arrows in Fig. 1—I will continually frustrate the linear relationship between them. Reactions and appraisals are interdependent: a particular interpretation of a timbral *sign* can work backwards (or, downwards) to modify its *concept*, or its *perception.* The threshold between reaction and appraisal is a porous one, and slippage between levels of ASPECS—for example, hearing symbolically "noisy" people (*sign*) represented in the acoustical noise (*sound*) they produce—is an important part of this story.

Overview of Chapters

The dissertation is subdivided into two parts. In Part I, I lay out the theoretical and empirical groundwork for an embodied theory of timbre, fleshing out the details of the ASPECS model through a synthesis of the current empirical literature (Ch. 1), presentation of results from my own experiments (Ch. 2), and a cognitive linguistic theory of timbre's metaphorical properties (Ch. 3). Part II will apply this model in musical context, with analyses of the affective role of timbre in two polarizing, "noisy" musical genres: free jazz (Ch. 4) and extreme heavy metal (Ch. 5). The dissertation closes with a brief epilogue exploring the social dimensions of listening to timbre.

Chapter 1 develops a theory of timbre that takes into account its embodied aspects, tracing the relationship between affect generation, embodiment, human vocality, and act–sound coupling in perception. I explore the dynamics of vocal exertion to show a basic bioacoustic correlation between the "sound of arousal" in the voice (i.e., high-frequency energy, inharmonic noise, and auditory roughness) and high-arousal, usually negative affect, even in relation to non-vocal timbres. This correlation, I argue, plays a significant role in the way noisy timbres are perceived in musical context, and helps to explain how qualities of timbre can be "somatically marked." The chapter also situates the project in the literature and explores points of convergence and divergence between different disciplinary perspectives.

In Chapter 2, I share findings from three original empirical studies designed to explore the role of acoustic noise in listeners' reactions and appraisals of timbre. Experiments 1 and 2 use behavioral methods to assess certain somatic, semantic, and affective dimensions of timbre; Experiment 3 uses neuroimaging to explore what happens in the brain when we hear "noisy" timbral qualities. I conclude with the data-driven assertion that listeners rapidly divide timbral percepts into "good" and "bad" types, with a range of associated appraisals falling quickly into place under this fundamental perceptual binarism. Neurophysiologically, positive appraisals are reflected in activity in vocal motor areas, while negative appraisals manifest themselves in activity in motor and limbic areas. Together, these results point to a role for the human mirror neuron system in the evaluation of musical timbre: *perceiving* timbre, it seems, involves some level of motor resonance with the *acts* that produce sound.

Chapter 3 follows upon these empirical conclusions by exploring how the bodily *experience* of sound production influences the way we talk about and ultimately *conceptualize* timbre. The approach I take is cognitive linguistic, drawing heavily on the work of Mark Johnson and George Lakoff. I argue that timbre can be understood as a perceptual manifestation of certain simple experiential meaning structures; these structures constrain the metaphorical language commonly used to describe qualities of sound (e.g., "bright," "dull," "warm"). Timbre descriptions are not all arbitrary and subjective, then, but often reflect shared bodily experience. I go on to focus this cognitive linguistic model on noisy timbre, which is predominantly described using tactile metaphors (e.g., "rough," "harsh"). Returning to Experiment 3, I show that processing noisy timbre calls upon the same regions of the brain involved in both tactile sensation and tactile metaphor.

Having charted a path through the empirical and conceptual terrain of ASPECS, Part II puts theory into practice by employing it as a framework for analyzing two case studies. In Chapter 4, I turn to the polarized reception history of free jazz saxophone playing, particularly the strident timbral gesture I call the *saxophonic scream*. Using reception of John Coltrane's "late period" (1965–67) recordings as an example, I explore the cognitive and discursive landscape of saxophonic screaming, arguing that the notoriously divided reception was driven by the drastic corporeal demands of this vocally-mimetic timbre. Playing at the threshold of music and noise, Coltrane courted a range of competing, visceral reactions and appraisals, from spiritual ecstasy to "death-to-all-white-men wails."[55]

Chapter 5 examines the extreme heavy metal subgenre of death metal. The "brutal" timbres of death metal—including intensely distorted, detuned electric guitar and growled vocals—perform an essential psychological function in the concert-ritual of the genre, serving as a sonic simulacrum of sacred violence (in the sense first proposed by theorists René Girard and Jacques Attali). Musicians in this genre deliberately manipulate innately negative timbral reactions for symbolic purposes, summoning chaos only to bring it under meticulous control (sacrifice). "Brutal," "evil" timbres in death metal sound out the painful extremities of human embodiment, but their sense of danger is largely faked.

The epilogue briefly takes up a recurrent though underdeveloped thread running throughout the dissertation: the social and ethical stakes of timbre. As the case studies attest, tim-

---

[55] This description comes from poet-jazz writer Philip Larkin, *All What Jazz* (New York: Farrar, Straus & Giroux, 1970), 179.

bre can function as a sign of difference. I close with a discussion of how perceptions of timbre can contribute to the facilitation (or breakdown) of empathy with others.

Chapter 1

The Bodily Basis of Timbral Reactions and Appraisals

> For the present we judge the artistic effects [of timbre] only by feeling. How all that relates to natural sound we do not know … but we do write progressions of tone colors without a worry, and they do somehow satisfy the sense of beauty. What system underlies these progressions?
>
> — Arnold Schoenberg, *Harmonielehre*[1]

Introduction

Schoenberg's 1911 assessment of timbre exhibits a good deal of acknowledged mystification. Timbre (*Klangfarbe*), he concedes, conveys "artistic effects," but it stands outside conventional canons of musical meaning: timbre is aesthetically efficacious but invisible to analysis, potent but unintelligible. Making sense of timbre's contribution to musical experience relies on one faculty and one faculty alone: the emotions. We can *feel* it but we cannot *explain* it.

In this chapter, I explore the perceptual and cognitive implications of Schoenberg's speculation. Why is timbre thought to be so directly connected to the emotions? How do we "judge" emotional reactions evoked by timbre? Is timbre's role in musical affect even explainable, or is it bound to remain an enigma? In what follows, I will ultimately affirm Schoenberg's central point, but with a significant modification: as the audible trace of physical action and ma-

[1] Arnold Schoenberg, *Theory of Harmony*, trans. R.E. Carter (London: Faber, 1911/1978), 421.

terial engagement with the world, timbre is understood not only "by feeling," but specifically with the *feeling body*. The account offered here aims to emphasize the bodily basis of how we react to and ultimately appraise timbre; the feelings that stand as a basis for judgment are not purely subjective, but a reflection of shared human embodiment (our "system" of perceiving "natural sounds"). Timbre perception is fundamentally *motor mimetic*; listening to musical timbre, both vocal and instrumental, demands that we take up a "bodily attitude" consistent with the corporeal articulation of each particular timbral quality.[2] As Thomas Clifton explains: "[…] tone quality, as a phenomenon, is not to be confused with a stimulus impinging on my body. It is my body which produces its effects on tone quality, and it is only because of my body that there can be any talk of quality at all."[3] Thus, with Clifton, I propose that key to understanding timbre's connection to *emotion* is the bodily *motion* it requires, both to produce and (sympathetically) to perceive.

This set of relations becomes clearest at the outer limits of both embodiment and musical timbre, the point at which high levels of physical exertion—reflecting high-arousal, high-potency affects—translate into sound. This liminal interface leaves its mark on timbre in certain identifiable and consistent sonic forms, and the simple bioacoustic correlation between types of bodily/affective arousal and their effects on timbre—the close relation between *acts* and *sounds*—influences both how sound and the actors behind the sound are perceived, experienced, conceptualized, and ultimately inscribed with symbolic meaning.

My initial approach to these ideas will be more concerned with breadth than depth, traversing a range of theoretical and empirical literatures—particularly insights from cognitive psychology, psychophysics, acoustics, and neuroscience—to build toward an embodied theory of

---

[2] The term comes from Merleau-Ponty, *Phenomenology of Perception*, 249.

[3] Thomas Clifton, *Music as Heard: A Study in Applied Phenomenology* (New Haven: Yale University Press, 1983), 68.

musical timbre. (This provisional model will be fleshed out and revised in the chapters to come.) Although this literature is relatively sparse (and studies adopting an "embodied" approach sparser still), certain fundamental premises are, nevertheless, not unique to this dissertation. Psychologists Stephen Handel (1989) and Alfred Bregman (1990) were among the first to read musical timbre into a larger ecological framework, demonstrating that perception of timbre is always intimately associated with the material mechanisms and broader environmental field involved in sound production; Richard Middleton (1990) and Shepherd and Wicke (1997) developed sophisticated theoretical frameworks for acknowledging the materiality of timbre as a potent force in musical meaning; Arnie Cox (1999, 2001 & 2011) proposed that all music perception, including timbre, is grounded in "motor mimetic imagery"; David Huron (2001) developed an evolutionary model of timbre perception emphasizing the adaptive value of source "identification" and "state cues" derived from timbre; Rafael Ferrer (2009 & 2011) was the first to explicitly apply insights from embodied cognition to timbre perception; and Nina Sun Eidsheim (2008 & 2011) successfully theorized the role of human embodiment in our ability to derive meaning from vocal timbre.[4] Building on the work above, I will advance four interrelated claims:

(1) *Timbre is a verb*. Understanding musical timbre as the acoustic byproduct of motor exertion and viewing its perception as similarly "action-oriented," I argue that we perceive

---

[4] See Handel, *Listening: An Introduction to the Perception of Auditory Events*; Bregman, *Auditory Scene Analysis: The Perceptual Organization of Sound*; Richard Middleton, *Studying Popular Music* (Berkshire, UK: Open University Press, 1990); Shepherd and Wicke, *Music and Cultural Theory*; Cox, "The Metaphoric Logic of Musical Motion and Space."; "The Mimetic Hypothesis and Embodied Musical Meaning."; "Embodying Music: Principles of the Mimetic Hypothesis."; David Huron, "Toward a Theory of Timbre" (paper presented at the 12th Annual Conference of Music Theory Midwest, Cincinnati, OH, 2001); Rafael Ferrer, "Embodied Cognition Applied to Timbre and Music Appreciation: Theoretical Foundation," *British Postgraduate Musicology* 10(2009); "Timbral Environments: An Ecological Approach to the Cognition of Timbre," *Empirical Musicology Review* 6, no. 2 (2011); Eidsheim, "Voice as a Technology of Selfhood: Towards an Analysis of Racialized Timbre and Vocal Performance."; "Sensing Voice: Materiality and the Lived Body in Singing and Listening."

timbre in terms of implied bodily actions.[5] Evidence for this position comes from cognitive psychology, ecological acoustics, and mirror neuron research, among other areas.

(2) *We understand many dimensions of timbre via mimetic similarities to vocal expression.* It has long been acknowledged that aspects of musical timbre (vocal and instrumental) resemble qualities of human vocality; similarly, timbre is often perceived with the accompaniment of covert motor mimesis (subvocalization). I affirm Herbert Spencer's (1857) account of the voice's primacy in musical experience—if timbre perception is motoric, then more specifically, it can be conceived as something we do with our voices.

(3) *Physical exertion leads to timbral anomalies that are often heard as "noisy."* The motor quality of timbre perception can be seen most vividly *in extremis*, in bodily acts of high-arousal and energy. Such actions, in the voice and many instruments, leave an acoustic trace in the form of *high-frequency energy* (or "brightness"), *inharmonicity and a flat spectrum* (heard as "noise"), and *roughness*. Taken together, I call these acoustic elements *noisy timbre*, or "a-noise." Noisy timbre is perceived as an index for heightened bodily arousal and exertion.

(4) *The "somatic markings" of timbre enable and constrain affective response.* Following an embodied theory of emotion, I argue that timbre is "somatically marked" with different bodily meanings and affective implications. In its noisier forms, timbre can function as a trigger for certain negative psychobiological reactions. However, the reaction does not ultimately determine the judgment. We react to timbre reflexively based on the self-protective con-

---

[5] This ontological reassessment will be challenged and expanded in Chapter 3, which takes up the metaphorical basis of (noisy) timbre.

tingencies of our perceptual systems, but we appraise timbre according to a much more complex set of situationally-, personally-, and culturally-defined values.

In advancing the position that timbre perception is "embodied," I am not claiming that this one parameter of music is somehow a privileged site of musical embodiment. However, since timbre is not as obviously motional as melodies and rhythms, and musical elements that exhibit the most movement tend to be read with greater enthusiasm into an embodied framework, scholars of music and embodiment often elide timbre as a source of motor engagement. Thus, for example, Roland Barthes, whose "body in a state of music" seems to consist of nothing more than rhythm: "that which beats the body, or better: that body which beats."[6] To be sure, "the beat" and other musical parameters are often conceived as the actors in music/body engagement, with timbre in a subservient, adjectival role, a descriptive quality draped "over" musical structures but not actually involved in how the structures move.[7] This chapter takes the opposite position, placing the originating sound-generating (inter)actions prior to the (psycho)acoustic.

Ultimately this chapter is concerned with the connections between body, sound, and affect at the first three most primary stages of the ASPECS model. To weave these areas together, the course I chart takes us from embodied accounts of emotion to the expression of emotion in the human voice; from the role of vocality in musicking to the vocal aspects of instruments; from a discussion of arousal and physical exertion to an acoustical reading of how arousal effects timbre; and, finally, with an account of the motor dimensions of hearing (and the auditory dimensions of doing). Together, this account gestures toward a possible answer to the question from the

---

[6] Quoted in Middleton, *Studying Popular Music*, 265.

[7] As Shepherd and Wicke state: "Music in the functional tonal tradition has always played down the role of timbre in sounds, and so the material presence of sounds themselves. Timbres in functional tonal music have tended to become standardized as 'pure' and therefore to be conceived as 'neutral' or 'transparent' in their articulative role." See Shepherd and Wicke, *Music and Cultural Theory*, 154.

epigraph: the "system" underlying timbre's affective power is intimately connected to the bodily reactions and appraisals engendered by "natural sound."

Affect and the Body

Let us begin where Schoenberg leaves off: with *feelings*, the question of musical affect and emotion. Generalizing broadly, for much of Western intellectual history emotions were considered a trivial, subjective, and irrational part of human experience.[8] At best, emotions represented a nuisance to be overcome through self-control, and at worst, they were considered a dangerous malady, linked, it should be noted, with women and other less-than-fully-rational beings. To many thinkers of the past (and present), emotion is reason's Shadow: where reason helps guide us to truth through logic and deduction, emotion derails rational thought; where reason is transcendent, emotion keeps us bound to our animal nature.[9]

A large part of the denigration of emotion in Western culture stems from seeing it as an epiphenomenon ungrounded in anything "real." We can observe the bodily byproducts of emotional states, but these reactions, in this dated view, are built upon a mirage of purely subjective sense impressions. The view of emotions in this account—perhaps still the conventional wisdom—holds that perceptions lead first to private "feelings," which are then followed by corresponding and observable body states and actions. If somebody insults you, you are overcome by anger "inside," and as a consequence of this feeling, your pulse quickens, muscles tense, and so

[8] Antonio Damasio, *The Feeling of What Happens: Body and Emotion in the Making of Consciousness* (New York: Mariner, 2000), Chapter 2.

[9] This view of emotion was obviously informed by the ideology of difference circulating during this period: thus, emotion was seen as primitive, childish, and unredeemingly feminine vis-à-vis the more highly evolved, masculine quality of reason.

forth. Physiological reactions are thus the byproduct of mental states—the emotion itself is an unreachable interior experience that receives external "expression" through the body.

In the second volume of *The Principles of Psychology* (1890), William James challenged this conventional view with an audacious thesis: instead of *causing* bodily changes, emotions are *caused by* the body. By inverting the conventional understanding of how emotions work, he placed human embodiment front and center in the process of affective experience:

> […] We feel sorry because we cry, angry because we strike, afraid because we tremble, and not that we cry, strike, or tremble, because we are sorry, angry, or fearful, as the case may be. Without the bodily states following on the perception, the latter would be purely cognitive in form, pale, colorless, destitute of emotional warmth. […] *If we fancy some strong emotion and then try to abstract from our consciousness of it all the feelings of its bodily symptoms, we find we have nothing left behind*, no "mind-stuff" out of which the emotion can be constituted, and that a cold and neutral state of intellectual perception is all that remains. […] A purely disembodied human emotion is a nonentity.[10]

In the Jamesian view, an affectively weighted perception is registered first as a pre-organized, innate bodily reaction, and only secondarily as a conscious, felt emotion. Affective experience is a *post hoc* intellectual justification for what our bodies are already telling us.

Jamesian theory has influenced many contemporary psychological accounts of affect, but for the purposes of this dissertation, its utility is somewhat limited. To be fair, James made the vital step of acknowledging the primary role of the body in affective response, and offering a preliminary way to "ground" affect in material reality. But he does not tell us *how* physiological reaction adds up to phenomenological experience. An insult might cause my pulse to quicken, and I might interpret the changing state of my body to indicate that I am angry, but why exactly do I jump to this specific affective conclusion (and why might somebody else jump to a different one)?

---

[10] William James, *The Principles of Psychology*, vol. 2 (New York: Dover, 1890/1950), 450-51. Italics in original. A similar theory of the emotions was generated by Danish psychologist Carl Lange at around the same time as James's formulation; hence, today it is often labeled the James-Lange theory.

How is autonomic response translated into mental self-knowledge? Felt emotion is not simply a disinterested tally of physiological changes, but a texture of inner life far more dynamic and complex than its constituent parts. How do we leap from the actions of crying, striking, or trembling to *being affected* by such actions?

This question has been taken up by many psychologists since James.[11] One particularly compelling framework for understanding how bodily reactions translate into complex emotion is the psychological theory of appraisal, which hypothesizes that affect arises in response to evaluations, or appraisals, of our affectively weighted perceptions.[12] Returning to James's example, we might illustrate this by saying that *trembling itself* does not directly induce the emotion of fear; rather, we feel afraid when we appraise both an external percept (the tremble-inducing situation) and the resultant internal bodily reaction (the trembling) as *constitutive of* the experience of fear. In short, *pace* James, the trembling in itself is not causing us to feel afraid—our appraisal of the whole situation (cause and effect) leads to the emotion. In emphasizing the evaluative dimension of experience, appraisal theory tries to account for certain conundrums posed by James's theory, including the fact that the same emotion can be evoked by very different stimuli, and that the same affective situation can differ by individual and by moment.

The appraisal process is endowed with considerable nuance, but at its core is a simple binary scale, positive to negative.[13] Appraisal is fundamentally self-protective: we evaluate our environment in terms of what is *good* for us, *bad* for us, or somewhere in between. Even if overlaid

---

[11] This literature is far too vast to review with any specificity. To begin, see the sources listed in note 4, Chapter 7, Antonio Damasio, *Descartes' Error: Emotion, Reason, and the Human Brain* (New York: Penguin, 1994), 283.

[12] For a comprehensive overview of appraisal theory, see Klaus R. Scherer, Angela Schorr, and Tom Johnstone, eds., *Appraisal Process in Emotion: Theory, Methods, Research* (New York: Oxford University Press, 2001).

[13] See Klaus R. Scherer, "Appraisal Considered as a Process of Multilevel Sequential Checking," in *Appraisal Processes in Emotion: Theory, Methods, Research*, ed. Klaus R. Scherer, Angela Schorr, and Tom Johnstone (New York: Oxford University Press, 2001).

with subtler shades of affective resonance, appraisal is very much about maximizing pleasure and minimizing pain. Of course, the line between the two is porous: trembling in fear while watching a horror movie is a different phenomenological experience than trembling in fear while being stalked by an actual killer. But this is precisely where appraisal theory comes in handy—depending on what we perceive about the environment, the same physiological reaction can lead to very different appraisals (and thus different affects, from fearful titillation to outright terror). In the first case, a positive appraisal is predicated, to some degree, on forgery: we know intellectually the movie represents no real existential danger, but the body lacks such discrimination. (It is to our adaptive advantage to treat all potential threats *as* threats.) We will return to this phenomenon of "contrastive valence" later;[14] but suffice it to say that even ambivalent situations drive home the basic binarism of the appraisal process: although appraisals might not always be consistent with bodily reactions, both are founded upon the scale of implied pleasure *versus* implied pain.

Complementing the basic premise of appraisal theory from a neural perspective is Antonio Damasio's influential somatic marker hypothesis.[15] The essential thesis here is that emotions are dedicated, at the most ancient level, to the maintenance of homeostasis: they serve as an "automated alarm system" warning us of the pain or pleasure in store if we continue down a particular path.[16] Every event or situation, argues Damasio, is consciously or non-consciously evaluated for its implications to the thriving of the organism, and is "marked" accordingly with specific neurochemical responses. Thus, bad things for our genome—a rapidly approaching bear, a

---

[14] This term is David Huron's; see Huron, *Sweet Anticipation: Music and the Psychology of Expectation*, 21.

[15] See Damasio's books for detailed non-technical explanations of this hypothesis: Damasio, *Descartes' Error: Emotion, Reason, and the Human Brain*; *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*.

[16] *Descartes' Error: Emotion, Reason, and the Human Brain*, 173.

cheating lover, the taste of spoiled food—make us "feel bad" via specific internal changes (neural, musculoskeletal, visceral, etc.) which then, taken together, manifest themselves as something we might respectively label "fear," "anger," or "disgust." Though conscious appraisal is not required for these innate mechanisms to kick into gear, the positive/negative axis remains the most fundamental basis for somatic markers: good events and situations contribute to the flourishing of the organism and are rewarded with positive somatic markers, while bad events and situations are potentially harmful and accompanied by aversive markers. In contrast to earlier views of emotion as an unnecessary, arbitrary, and usually deleterious part of being human, Damasio's thesis affirms the central importance of (embodied) emotion in decision-making, rationality, and consciousness.

This truncated account of how emotions work should be intuitive for something as obviously harmful as a charging bear, or as obviously pleasant as food and sex. But this dissertation is not about bears, food, or sex. Now that I have traced in broad strokes the connection between perception, bodily response, and affect, we can ask what these accounts have to tell us about how we hear musical sound. Can timbre be somatically marked just like bears and sex, even though—like a horror movie—it really has nothing to do with actual survival?

Sounding (E)motion: The Vocal Grounding of Timbral Affect

> If we instinctively respond to a rising melodic pitch by a feeling of increased tension and hence of heightened expression, or a falling pitch by the opposite sensation; if an increase in intensity of sound intensifies our dynamic response to the music, and vice versa, it is because we have already in our vocal experiences—the earliest and most primitive as well as later and more complicated ones—lived through exactly the same effects. A raising of pitch or an increase in volume is the result of an intensification of effort, energy, and emotional power in the crying child just as truly as in the highly-evolved artistry of a Chaliapin or an Anderson.

— Mark Johnson[17]

The affect theories introduced above confirm what etymology might suggest: *e-motion* is intimately connected to *motion*. If we define "bodily reaction" broadly, indeed, all of the physiological instantiations of emotion are really movements of the body, the rush of dopamine and the chill of galvanic skin response just as much as overt acts of crying or trembling. Turning to music, it is patent that motion—of performers' bodies, of listeners as they sway and dance—is the essential medium of expression. The motional character of music has long been considered isomorphic to emotional experience: music is thought to be capable of "moving us" emotionally because it (and the body, in performance and listening) *moves.* To give one ancient example, Aristotle lays out this connection in his *Problems*:

> Why do rhythms and melodies, which are composed of sound, resemble the feelings, while this is not the case for tastes, colors, or smells? Can it be because they are motions, as actions are also motions? Energy itself belongs to feeling and creates feeling […] These motions … are active, and action is the sign of feeling.[18]

There are a number of entry points to the consideration of how "actions," via sound, might translate into "feeling," many of which have been explored in great depth.[19] In order to

---

[17] Mark Johnson, *The Meaning of the Body: Aesthetics of Human Understanding* (Chicago: University of Chicago Press, 2007), 237.

[18] Aristotle, *Problems* 27 & 29, quoted in Helmholtz, *On the Sensations of Tone as a Physiological Basis for the Theory of Music*, 251. It should be noted that the Greeks had no specific word for what we call today "timbre"—this is a coinage of the Enlightenment. Music (*mousike*) consisted of melody (*melodia*) and rhythm (*ruthmos*) arranged into a harmonious balance. But, of course, all melodies are "timbred" as well, and the timbre contributes to the motional character of the sound.

[19] The empirical and theoretical sources I have found most useful in conceptualizing the motional character of musical sound and its link to embodied affect include the following: Lidov, "Mind and Body in Music."; Bruno H. Repp, "Music as Motion: A Synopsis of Alexander Truslit's (1938) *Gestaltung und Bewegung in der Musik*," *Haskins Laboratories Status Report on Speech Research* 111-112(1992); Patrick Shove and Bruno H. Repp, "Musical Motion and Performance: Theoretical and Empirical Perspectives," in *The Practice of Performance: Studies in Musical Interpretation*, ed. John Rink (Cambridge: Cambridge University Press, 1995); Andrew Mead, "Bodily Hearing: Physiological Metaphors and Musical Understanding," *Journal of Music Theory* 43, no. 1 (1999); Rolf Inge Godøy, "Motor-Mimetic Music Cognition," *Leonardo Music Journal* 36, no. 4 (2003); Clarke, *Ways of Listening: An Ecological Approach to the Perception of*

develop an account of how timbre specifically can serve as an agent of affect, I will focus on a particular human capacity that, above all others, enables the affective experience of the individual to be sounded: *vocality*.

The notion that the voice has a privileged role in musical affect is nothing new—Plato, Rousseau (1781), Helmholtz (1867), and Darwin (1871), among others, have all weighed in.[20] One of the most compelling and detailed early accounts—and an important precursor to modern theories of embodied music cognition—comes from English philosopher Herbert Spencer, whose 1857 essay "The Origin and Function of Music" attempts to account for musical affect by invoking the physiological expression of the emotions.[21] To Spencer, different emotions (or "feelings") are associated with different patterns of embodiment, which in turn affect the acoustic output of the voice in different ways. He explains, "There is a direct connection between feeling and movement, the last growing more vehement as the first grows more intense."[22] Possible errors in causation aside—it could be argued, as does James, that movement drives feelings, not the other way around—Spencer rightly intuits the inseparability of human embodiment and emotion, and addressing music specifically, he roots this connection in the voice. He goes on: "All music is originally vocal. All vocal sounds are produced by the agency of certain muscles. These muscles,

---

Musical Meaning; Leman, *Embodied Music Cognition and Mediation Technology*; Rolf Inge Godøy and Marc Leman, eds., *Musical Gestures: Sound, Movement, and Meaning* (New York and London: Routledge, 2010); Steve Larson, *Musical Forces: Motion, Metaphor, and Meaning in Music* (Bloomington, IN: University of Indiana Press, 2012).

[20] We find Helmholtz, for instance, writing: "[…] Music in its initial state and simplest forms may have been originally an artistic imitation of the instinctive modulations of the voice that correspond to various conditions of the feelings." See Helmholtz, *On the Sensations of Tone as a Physiological Basis for the Theory of Music*, 371. For more on Platonic views of vocality, see Peter Kivy, *Introduction to a Philosophy of Music* (New York: Oxford University Press, 2002); Jean-Jacques Rousseau, *Essay on the Origin of Languages and Writings Related to Music*, trans. John T. Scott (Hanover, NH: University Press of New England, 1781/1998); Charles Darwin, *The Descent of Man, and Selection in Relation to Sex* (London: John Murray, 1871).

[21] I am not the first to acknowledge the importance of Spencer's approach to the contemporary discourse of music and embodied cognition; see Cox, "The Mimetic Hypothesis and Embodied Musical Meaning," 197.

[22] Herbert Spencer, "The Origin and Function of Music," in *Literary Style and Music* (New York: Philosophical Library, 1857/1951), 48.

in common with those of the body at 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."[23] For Spencer, musical affect is a more or less direct reflection of shared vocal embodiment; we can hear the affectively valenced bodily states of another in the sound of their voice, and these innately recognizable patterns of emotional vocalization translate into musical expression. Spencer concludes with the following piece of induction:

> 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 adjustment of the vocal organs; every different adjustment of the vocal organs causing a change in the sound emitted; it follows that variations of voice are the physiological results of variations of feeling. It follows that each inflection or modulation is the natural outcome of some passing emotion or sensation; and it follows that the explanation of all kinds of vocal expression must be sought in this general relation between mental and muscular excitements.[24]

Feelings can only "demonstrate themselves in sounds," however, if there is another person there to witness the demonstration. Spencer presciently concludes that vocal-based musical affects are intelligible to others due to *motor resonance*: we can decode the associated vocal signatures of different affects because we have voices too, and because perception is always on some level also (en)action. He writes:

> Having been conscious of each feeling at the same time that we heard ourselves make the consequent sound, we have acquired an established association of ideas between such sound and the feeling which caused it. When the like sound is made by another, we ascribe the like feeling to him; and by a further consequence we not only ascribe to him that feeling, but have a certain degree of it aroused in ourselves; for to become conscious of the feeling which another is experiencing is to

[23] Ibid., 49.

[24] Ibid., 49-50.

have the feeling awakened in our own consciousness, which is the same thing as experiencing the feeling.[25]

Drawing connections between human embodiment, vocality, and affect, Spencer contributes four major insights to the development of an embodied theory of timbre. First, in insisting that qualities of timbre are predicated upon "different adjustment(s) of the vocal organ," he recognizes that timbre is never simply an acoustical matter. Rather, in the language of ecological psychology, it reflects an *affordance* of the body, a way of moving and acting upon the world. Timbre is the sonic consequence of material interaction. Second, it follows that timbral *perception* is also grounded in shared human embodiment. If timbre is a perceptual aspect of tone and perception is inseparable from subject/environment interaction and sensorimotor engagement, then there is always already a material, corporeal quality to timbral perception, no less than in timbral production. Third, by relating "feelings" to bodily states and motor exertion, Spencer demonstrates the bodily basis of timbral affect. Timbre is able to convey expressive meaning, in other words, not due *exclusively* to arbitrary association or semiosis, as some authors have suggested, but rather in relation to "established associations" between different bodily states and actions, and their related affects.[26] The emotionally-inf(l)ected body leaves traces in the sound of the performer that are detected and reenacted in the moment of listening. Fourth, in locating the conveyance of affect and meaning in an intersubjective exchange between sound producer and listener—an exchange enabled through motor-mimetic properties of embodied cognition—Spencer highlights the fundamentally social nature of timbre perception. Timbre is not a purely

---

[25] Ibid., 56-57.

[26] For instance, Jean-Jacques Nattiez considers music "meaningful" only insofar as it brings to mind something other than itself. In the case of embodied meaning in musical timbre, the bodily acts themselves—and associated affective states—may, in fact, be the source of meaning over and above any extrinsic reference. See Jean-Jacques Nattiez, *Music and Discourse: Towards a Semiology of Music*, trans. Carolyn Abbate (Princeton: Princeton University Press, 1991). This is not, of course, to say that arbitrary association can not *also* play an important part in the transmission of affect.

subjective, interior quality of sound, then, but rather intercorporeal, *a shared experience* uniting listening subject and sonic object in common body-mediated affect.

I have dwelled on Spencer because his short, prescient essay is perhaps the earliest articulation of the embodied, motor mimetic account of timbral affect. The basic premise of his argument is the starting point for the motor-mimetic theory advanced in this chapter. Before moving to the next stage of this argument, however, we can flesh out Spencer's account with some more modern perspectives on the connection between musical and vocal expression.

Music theorist Arnie Cox has argued strongly for a Spencerian approach to musical meaning. Drawing on a wide range of complementary empirical and theoretical literatures, including developmental psychology, imitation studies, cognitive linguistics, imagery studies, neurophysiology, gesture research, and subvocalization studies, Cox's model for understanding the contribution of embodiment to musical experience—the "mimetic hypothesis"—proposes that the voice plays a starring role in the generation of meaning.[27] Through overt and covert motor mimesis, Cox argues that the dynamics of lived vocal experience determines a significant part of how we make sense of all musical expression, instrumental as well as vocal.

How does this work? Of the research areas adduced above, the phenomenon of *subvocalization* is the most directly relevant to this question. To illustrate, focus your mind's ear on the Happy Birthday song, or the "V-for-Victory" motif of Beethoven's Fifth, or Miles Davis's solo on "So What?" You will likely notice that imagining these tunes requires you to *sing along* to them. Subvocalization can be described as covert vocal imitation, either conscious or non-conscious.[28] The phenomenon does not involve voicing anything aloud, even at a low volume—this would

---

[27] The Mimetic Hypothesis was first proposed in Cox's Ph.D. dissertation, then refined in subsequent publications: see Cox, "The Metaphoric Logic of Musical Motion and Space."; "The Mimetic Hypothesis and Embodied Musical Meaning."; "Embodying Music: Principles of the Mimetic Hypothesis."

[28] "The Mimetic Hypothesis and Embodied Musical Meaning," 200.

simply be vocalization—but rather consists of the interior rehearsal of the voice, a form of motor imagery in which musical sound (pitches, rhythmic patterns, inflections, and timbre) is traced out by the voice. This phenomenon likely begins as direct imitation in infancy, illustrating a basic heuristic tendency in human learning: we comprehend through doing. To Cox, therefore, subvocalization is not something optional that we might elect to do on occasion—it is fundamental to the process of music perception and cognition, and comprehension ultimately depends on it. Subvocalization has been shown to play an important role in the comprehension of language (the so-called "motor theory of speech perception"),[29] and in the perception of vocal music.[30] (It also plays a role in instrumental music, to be discussed later.)

The mimetic hypothesis holds that perceiving the singing voice involves the covert rehearsal of the motor acts involved in singing. Listening to a soaring operatic soprano involves vicarious motions within the listener, motions that are laden, as previously discussed, with affective consequences.[31] Seen from this perspective, let us make an important revision to Spencer's model: musical affect is predicated not only on the "established associations" between different feelings and qualities of vocal sound, but also on the fact that these sounds are *literally generating*

---

[29] See A.M. Liberman *et al.*, "Perception of the speech code," *Psychological Review* 74, no. 6 (1967); Giacomo Rizzolatti and Michael A. Arbib, "Language within Our Grasp," *Trends in Neuroscience* 21(1998); Luciano Fadiga *et al.*, "Speech Listening Specifically Modulates the Excitability of Tongue Muscles: A TMS Study," *European Journal of Neuroscience* 15(2002); Stephen M. Wilson *et al.*, "Listening to Speech Activates Motor Areas Involved in Speech Production," *Nature Neuroscience* 7, no. 7 (2004); Friedemann Pulvermüller, "Brain Mechanisms Linking Language and Action," *Nature Neuroscience* 6(2005); Stephen M. Wilson and Marco Iacoboni, "Neural Responses to Non-Native Phonemes Varying in Producibility: Evidence for the Sensorimotor Nature of Speech Perception," *Neuroimage* 33(2006).

[30] Alan Baddeley and Robert Logie, "Auditory Imagery and Working Memory," in *Auditory Imagery*, ed. D. Reisberg (Hillsdale, NJ: Lawrence Erlbaum, 1992); J.D. Smith, D. Reisberg, and M. Wilson, "Subvocalization and Auditory Imagery: Interactions between Inner Ear and Inner Voice," ibid.; J.D. Smith, M. Wilson, and D. Reisberg, "The Role of Subvocalization in Auditory Imagery," *Neuropsychologia* 33, no. 11 (1995); M. Vaneechoutte and J.R. Skoyles, "The Memetic Origin of Language: Modern Humans as Musical Primates," *Journal of Memetics—Evolutionary Models of Information Transmission* 2(1998); Daniel E. Callan *et al.*, "Song and Speech: Brain Regions Involved with Perception and Covert Production," *Neuroimage* 31(2006).

[31] These vicarious motions are, as we will see in the next chapter, evidenced in activity in larynx control areas of the brain.

vocal activity in the listener (overtly or covertly). Read into the theories of emotion introduced earlier, experiencing affect relies on motor resonance to musical sound. But the process does not stop at the reaction stage. One person might love to take on the voice of an opera singer (i.e., subvocalize to it), while another might feel uncomfortable and another completely indifferent. *Pace* Spencer, vocal-mimetic reactions to music do not appear to translate directly into specific affective responses, but rather are subject first to appraisal. The reaction (subvocalization) is innate and direct, while the appraisal (affect) is mediated by many factors that we will tease out as the dissertation progresses.

An inquisitive reader might demand at this point: What exactly do you mean by "vocal sound"? How do embodied affective states translate into the voice of the speaker or singer in order to generate reactions and appraisals from the listener? Thus far I have spoken only in generalities about the various affective dimensions of the voice, and a comprehensive survey falls outside the scope of this chapter.[32] In musical terms, all of the major parameters of sound—pitch, rhythm, speed, contour, loudness, etc.—correlate cross-culturally with aspects of vocal expression: one does not find vigorous war songs set to placid, legato melodies, nor, as Philip Tagg once quipped, people "yelling jerky lullabies at breakneck speed."[33] Timbre, too, plays a crucial role in signaling affective states. In the next section, we will delve into the specific physiological, acoustic, and psychoacoustic qualities that register timbral affect in the voice, which will involve us in the affective extremes of human vocality.

---

[32] For a review, see Tom Johnstone and Klaus R. Scherer, "The Effects of Emotions on Voice Quality," *Proceedings of the 14th International Congress of Phonetic Sciences, San Francisco* (1999).

[33] Philip Tagg, "Towards a Definition of 'Music'." Institute of Popular Music: University of Liverpool, 2002.

The Sound of Arousal: Physiology, Acoustics, and Psychoacoustics of Exertion

Eek Eek Eek Eek!

— Marmot alarm call[34]

Spencer identifies various vocal parameters that transmit affect, with "quality or timbre" being number two on the list; but he elides specific correlation of timbral qualities and affective-bodily states. He does, however, conceptualize vocal embodiment usefully as a scalar function of *intensity*: "there is a direct connection between feeling and movement, the last growing more vehement as the first grows more intense."[35] Spencer proposes a linear correlation between "vehemence" of bodily activity and "intensity" of affect, and this correlation, mapped onto the voice, implies that affect leaves its mark through increasing or decreasing levels of physical exertion, strain, or energy. Effectively communicating a high-intensity, high-arousal, high-potency emotion such as anger, therefore, requires greater exertion than a low-intensity, low-arousal emotion such as sadness. Everyday embodied experience confirms this basic assertion: the "higher-faster-louder" dictum holds in most situations, from screaming babies to heated arguments to banging your finger with a hammer. I will argue in this section that the scale of low-to-high vocal exertion maps onto the (psycho)acoustic qualities of *brightness*, *noise*, and *roughness*. Generally speaking, these qualities of vocal sound indicate higher levels of physical energy driving the vocal production system. The consequences of this simple bioacoustic correlation, I maintain, plays a profound role in how we appraise the percept of "noise" more generally.

---

[34] Daniel T. Blumstein and C. Récapet, "The Sound of Arousal: The Addition of Novel Nonlinearities Increases Responsiveness in Marmot Alarm Calls," *Ethology* 115(2009). I will be using Blumstein's term "sound of arousal" to refer to the collection of acoustic and psychoacoustic correlates of arousal.

[35] Spencer, "The Origin and Function of Music," 48.

So what does vocal exertion sound like? Let's begin with a thought experiment. You are walking down the street and you see a robbery in progress; you also happen to have noticed a policeman just a couple blocks down the street. Your body in that moment is charged with what psychologists call "ergotropic arousal," a state of preparatory tension and heightened attention that accompanies situations where extremely rapid, "fight of flight" reactions are imminent. You open your mouth to alert the police. What does your voice sound like? How does it differ from your normal speaking or singing voice? For one, you would probably notice that your voice is higher in pitch and louder than it is under regular circumstances. Additionally, you would annunciate your words with greater speed, and possibly with more tonal inflection than usual. It is hard to imagine that anyone would drop into a lower register, lower his or her voice level, and speak in a plodding monotone during high arousal, dangerous situations.

It is difficult to picture such a voice in this example because we have all been in similar, if less dramatic, situations, have seen them represented countless times, and know what it feels and sounds like to be highly aroused. Indeed, we have known this with our bodies since infancy, and—from the perspective of our genes and the more phylogenetically ancient regions of our brain—since before we were even human.[36] High arousal, both in terms of bodily disposition and associated affective state, changes the quality of vocal production: the voice tends to get higher, louder, faster, and more pitch-variant.[37] But what happens to the timbre of the voice in such conditions? Returning to the imaginary scene of the crime, you would probably let out a shrill, harsh yell, not a harmonious, smooth tone. In short, your voice would be *noisy*.

There are a couple of reasons for this: first, your primary goal would be to alert the police, and you know intuitively—again, from living your life in a human's body with a human voice—that a harsh, grating, noisy, rough (pick your adjective) voice does a better job of signaling peril than the alternative. The policeman is more likely to notice you when you call out in a noisy timbre. Of course, you will not rationally deduce such an insight in the moment, modulating

---

[36] If we take the evolutionary view, the association of particular sonic qualities with heightened physical exertion—and the affective traces of such audible embodiment—has a deeper history than the musical contexts that form the core of this dissertation. Physical arousal is instinctively associated with certain acoustic features and perceptual *qualia* among a range of bird and mammalian species. Consider a few examples. Many rodents—including shrews, marmots, meerkats, and prairie dogs—let out shrill, rough warning calls when danger is imminent. These timbrally "nonlinear" calls are capable of conveying information both about the kind of predator and its behavior, information of obvious adaptive advantage to the group. Moving up the food chain, dogs and cats instinctively know what the timbre of a growl "means": both making the growl sound and hearing this quality of timbre causes the activation of deep abdominal muscles necessary for rapid action. The vocalizations of primates, our closest cousins, are also known to follow similar acoustic patterns, with greater brightness, noise, and roughness—features common to both marmot screams and dog growls—reserved for moments of greatest physical exertion, arousal, excitement, and danger. And as any parent of a young child instinctively knows, the harshness of a baby's crying grows proportionately with his need: by modulating brightness, noise, and dissonance, a newborn infant is able to project innately intelligible cues about his changing physical and emotional state. For more on the ethology of acoustically noisy timbre, see the following sampling of sources, respectively: W.T. Fitch, J. Neubauer, and H. Herzel, "Calls Out of Chaos: The Adaptive Significance of Nonlinear Phenomena in Mammalian Vocal Production," *Animal Behavior* 63(2002); Blumstein and Récapet, "The Sound of Arousal: The Addition of Novel Nonlinearities Increases Responsiveness in Marmot Alarm Calls."; Chen-Gia Tsai *et al.*, "Aggressiveness of the Growl-Like Timbre: Acoustic Characteristics, Musical Implications, and Biomechanical Mechanisms," *Music Perception* 27, no. 3 (2010); Marc D. Hauser, "The Sound and the Fury: Primate Vocalizations as Reflections of Emotion and Thought," in *The Origins of Music*, ed. Nils J. Wallin, Björn Merker, and Steven Brown (Cambridge, MA: MIT Press, 2000); W. Mende, H. Herzel, and K. Wermke, "Bifurcations and Chaos in Newborn Infant Cries," *Physics Letters A* 145(1990).

[37] For more on these vocal qualities and their affective and physiological correlates, see Klaus R. Scherer and James S. Oshinsky, "Cue Utilization in Emotion Attribution from Auditory Stimuli," *Motivation and Emotion* 1, no. 4 (1977); Klaus R. Scherer, "Expression of Emotion in Voice and Music," *Journal of Voice* 9, no. 3 (1995); Johnstone and Scherer, "The Effects of Emotions on Voice Quality."; Klaus R. Scherer and Marcel R. Zentner, "Emotional Effects of Music: Production Rules," in *Music and Emotion: Theory and Research*, ed. Patrik N. Juslin and John A. Sloboda (Oxford and New York: Oxford University Press, 2001).

your voice accordingly; rather, your voice will just *sound this way*. This is the case because, as Spencer observes, physiological arousal and exertion—and the affective states with which arousal is associated—leave a mark on the timbre of the voice: the aroused body *itself* causes the timbral shifts, with no conscious mediation by the higher cognitive faculties required. And just as the production of effortful vocal timbre is largely reflexive, so is its audition. The quality of your voice in this instance is perceptually salient in large measure because the policeman's auditory system, like yours, evolved to be exquisitely sensitive to timbres that sound out the body in a state of danger and alert. Such timbres are somatically marked to engage a whole network of autonomic nervous system responses.

How specifically does this aroused body translate into noisy vocal timbre? Voice scientists have developed sophisticated models for the relationship between the vocal physiology, acoustics, and psychoacoustics of arousal, and a few relevant findings from this literature warrant review.[38] Under normal circumstances, vocal production comprises an elegantly balanced relationship between *source* (the vibrating vocal folds in the larynx) and *filter* (the resonance cavities and tissues of the vocal tract).[39] But in situations of arousal, when the voice is pushed beyond its normal operating parameters, the energy input exceeds the physical capacity of the source and the filter is altered as a result of increased muscle tension and other factors. This overstressed system manifests two particular timbral anomalies.[40] First, due to abrupt glottal attack, increased larynx and phar-

[38] For an invaluable voice science resource on this and many other issues, see Jody Kreiman and Diana Sidtis, *Foundations of Voice Studies: An Interdisciplinary Approach to Voice Production and Perception* (Malden, MA: Wiley-Blackwell, 2011). Another helpful source on the physiology of the (singing) voice is the now-classic book by Johan Sundberg, *The Science of the Singing Voice* (DeKalb, IL: Northern Illinois University Press, 1987).

[39] I have borrowed the "source-filter" terminology from the work of pioneering voice acoustician Gunnar Fant. It has also been used in music research: see Slawson, *Sound Color*; Handel, *Listening: An Introduction to the Perception of Auditory Events*.

[40] Klaus R. Scherer, "Vocal Affect Expression: A Review and a Model for Future Research," *Psychological Bulletin* 99, no. 2 (1986).

ynx tension, increased tautness of vocal tract walls, and decreased lubrication of the tract (the "dry mouth" that accompanies ergotropic arousal), the timbre of the aroused voice skews toward high frequency energy, experienced as a "metallic" or "piercing" vocal quality. This skew is distinct from the rise in fundamental pitch discussed earlier; in addition to higher pitch, then, arousal tends to push more of the total energy of the voice into higher partials, leading to a *brighter* timbre. Second, as a result of increased laryngeal tension and supraglottal friction causing a disturbance in vocal fold vibration, aperiodicity—i.e., *noise*—is common in the tonal spectrum of the over-exerted voice. In the literature this quality is often called "harsh voice."[41] The perception of noisiness is related to inharmonicity (vibrational energy outside the harmonic series) and a flat spectrum (aperiodic, nonlinear components). Taken together, increased brightness and noise correlate to increased physical and affective arousal in both speech and non-speech vocalization.[42]

The same acoustic qualities are also present in the over-exerted *singing* voice. Push any voice beyond its material limits and it will manifest acoustic signs of strain and wear. Such effects can also be deliberately cultivated: growl-like singing, or "distorted voice," is common to many musical styles.[43] This timbral quality is produced by simultaneously vibrating the vocal folds and supraglottal structures (particularly the supraglottic mucosa), overdriving the voice with high sub-

---

[41] In medical terms, this timbre is described as follows: "The exaggerated laryngeal tension in harsh voice is a combination of extreme adductive tension and extreme medial compression, brought about by over-contraction of the muscle systems responsible for these two parameters in modal voice." J. Laver, quoted in ibid., 152.

[42] This statement of cross-cultural universality may raise the eyebrows of some, but research in emotion, vocal science, and (psycho)acoustics of the voice have converged on certain features of human vocal expression that are not culturally constituted. Paul Ekman was the first modern emotions researcher to systematically demonstrate the six basic categories of emotion (happiness, sadness, disgust, anger, fear, and surprise), and each emotional category—as Klaus Scherer and others have found—relate to specific features of the voice.

[43] Examples include many Afro-diasporic singing styles, punk, metal, Japanese *noh*, and Beijing opera. See Tsai *et al.*, "Aggressiveness of the Growl-Like Timbre: Acoustic Characteristics, Musical Implications, and Biomechanical Mechanisms," 210.

glottal pressure. The supraglottic mucosa, which vibrates about three times slower than the vocal folds, obstructs the airflow, introducing nonlinearities into the timbre.[44] The same biomechanics that are subconsciously involved in aroused speech and non-verbal vocal sound are consciously manipulated in certain musical contexts to produce a distorted singing timbre.

To visualize the relationship between physiological arousal and the resultant sound, Fig. 1.1 shows spectrographs of a female voice singing the same pitch in two conditions (equalized for loudness): the first involves a low degree of vocal tension, while the second corresponds to much more physically tense, effortful means of production. Note the increased strength of high frequency energy in the aroused condition, as well as the blurring of energy across frequencies consistent with inharmonicity and a flat spectrum. We will explore these acoustic features shortly, but at this point, let us consider what the mechanisms of shared vocal embodiment outlined here might mean for other, non-vocal musical timbres.

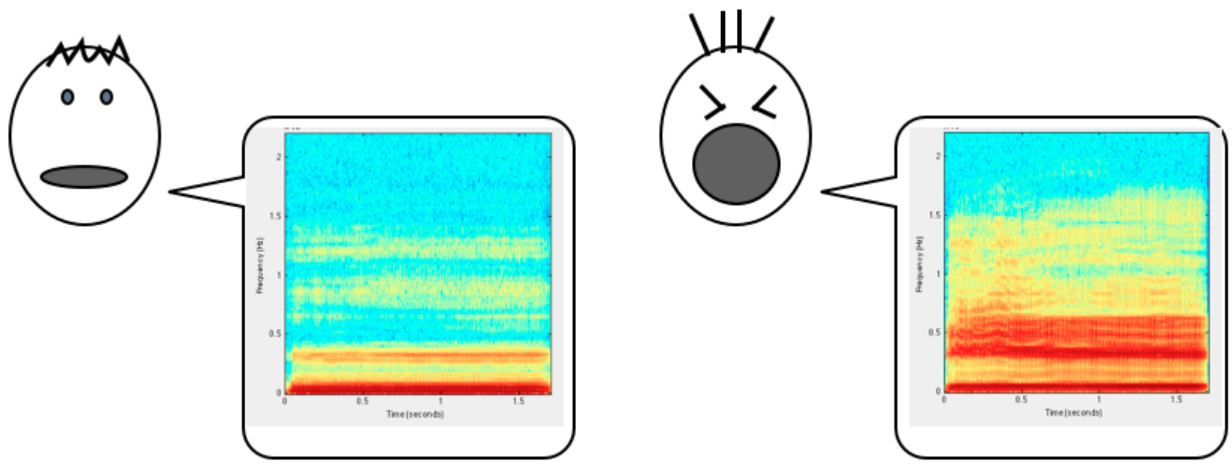


FIGURE 1.1: The un-aroused and aroused voice[45]

[44] D.Z. Borch *et al.*, "Vocal Fold Vibration and Voice Course Aperiodicity in 'Dist' Tones: A Study of a Timbral Ornament in Rock Singing," *Logopedics Phoniatrics Vocology* 29(2004).

[45] Spectrographs come from Experiment 1, to be reported in the next chapter: see Appendix 1 for complete parameters of the analysis. *N.B*: Spectrographs can be essential tools in the visualization of the physical attributes of timbre,

Shared Perceptual Mechanisms for Vocal and Instrumental Timbre

I have argued so far in this chapter that a line can be drawn from affective states, via the bodily actions that materially comprise them, to specific timbral outcomes in the voice, increases in high frequency energy and noise. But does this embodied process of timbral affect apply to musical timbres that are not vocal in origin? Can these same physiological mechanisms of arousal in the voice—and the perceptual systems to which they are coupled—translate into instrumental timbre as well through mechanisms of vocal mimesis? Direct empirical work on this question remains somewhat lacking (hence the intervention of the next chapter), but there are reasons to suggest that such an inference is supportable.

There are three interacting, scalar levels of the vocal mimesis theory of instrumental timbre to consider (roughly corresponding to the ASP- half of the model). First, the production of sound on some instruments overlaps significantly with vocal production, while other instruments do not at all, and the material means of production intrinsic to each source can govern the "voicelikeness" of the acoustic product.[46] I will call this level *actual vocal mimesis*:[47] instruments (generally aerophones) that rely on bodily *acts* also associated with the production of voice, such

---

but it will be important to keep in mind that they do not in themselves tell us anything about how the captured sound is *heard*. The words associated with the timbres of arousal—"harsh," "piercing," "bright," "distorted," etc.—refer to the experience of hearing a particular quality of sound, and not to the acoustical qualities that give rise to this sensation. This will not be the last time this distinction is made throughout the dissertation; the meatiest musicological insights to be gleaned from the style of timbral close reading advanced here, I believe, is to be found in the interaction between these two frames of reference rather than in approaching either frame separately. Although they are absolutely interdependent—timbre is about *both* the sound itself *and* our perception of it—we would do well to keep in mind this vital distinction. For a thoroughgoing review of psychoacoustic timbre research with a keen eye to this distinction, see Roger A. Kendall, "The Role of Acoustic Signal Partitions in Listener Categorization of Musical Phrases," *Music Perception* 4, no. 2 (1986).

[46] Wind instruments can be placed along a spectrum of voicelikeness in terms of their means of production, from the double-reeds on one end to mouth-resonator instruments like the didgeridoo and Jew's harp on the other. Non-vocal instruments—percussion and strings—do not involve the vocal tract, though the resultant sound may be quite voicelike.

[47] I use the term "actual" here not to imply that it is more *real* than the others, but rather to emphasize the "act" that goes into it.

as movement of the vocal tract, tongue, and lips. Second, an instrument can have an acoustic profile that closely resembles a voice, what we will call *sonic vocal mimesis*. For example, the inharmonic elements of an overblown saxophone might, in the physical world, resemble the inharmonicity of an aroused voice. The connection between vocal expression and instrument timbre could map onto perception as well, though this need not always be the case. On the third level, *perceptual vocal mimesis*, the relative degree of either *actual* or *sonic vocal mimesis* may match perfectly or become dissociated with how "vocal" the resulting timbre is perceived to be. Thus, for example, the Indian *sarangi* is often said to be highly voice-like, though it is a bowed string instrument; at the *actual* level, then, it is not at all voice-like, though it may well be at the *sonic* and *perceptual* levels.[48] Conversely, the flute might be relatively voice-like in the *actual* sense, not resemble the voice *sonically*, but invoke vocality in *perception* due to aspects of visual or kinesthetic imagery of flute playing that accompany perception, parameters of experience not visible on a spectrograph. When mapping vocal timbre onto instruments, we need to disambiguate where, exactly, the *voice* is located in the equation: an instrument can be voice-like in production, voice-like in sound, and voice-like in perception, but they might not interact in predictable, linear ways. It is outside the scope of this chapter to go into too much depth on this point—I will be returning to these issues throughout the case studies—but it needs mention here nonetheless, if only to highlight the complexity of the matter. Table 1.1 displays some characteristics of each of these three aspects of vocal mimesis.

[48] The oft-noted voicelikeness of the *sarangi* is due in part to the soft wood of its construction, which, like the soft vocal tract, enables the formation of strong formants. Other examples of instruments that are perceptually but not actually voice-like include the Indian tabla and the West African batá drum (though both of these instruments imitate pitch inflections of the voice, not timbre *per se*).

TABLE 1.1: Modes of vocal mimesis of instrumental timbre

Actual mimesis	*Sonic mimesis*	*Perceptual mimesis*
Physical means of sound production involve the vocal tract (i.e., larynx, pharynx, oral cavity, tongue, lips, etc.)	Acoustic output shares features that resemble aspects of vocal timbre (e.g., "singer's formant," ergotropic arousal effects, etc.)	Timbre is *heard as* voice-like irrespective of actual or sonic characteristics. Many factors may contribute to appraisal (e.g., imagery, visual association, personal experience, etc.)

The most direct approach to considering the perceptual link between vocal and instrumental timbre is through intuition and observation of everyday musicking. Consider the ubiquitous act of "singing along" to a favorite song, for instance. Next time you do this, note that you probably do not just imitate the pitch and rhythms of the singer's voice, but also (to varying degrees of fidelity) his or her vocal timbre as well, or at least its broad timbral character (e.g., a ringing operatic voice, a plangent blues voice, a gruff metal voice, etc.). Thus, singing along to Bob Dylan might cause you to take on a more nasal vocal character than is ordinary to you, and singing along to Louis Armstrong may elicit your own idiosyncratic version of his inimitable gravelly tone.

Now consider singing along with an instrument.[49] Chances are there is some aspect of timbral imitation going on here as well, though it might not be as immediately obvious. For example, you probably would not imitate Miles Davis's harmon-muted trumpet with the broad, "open-O" vowel—this would be far too dull a tone for the sharp, piercing quality of the harmon mute (which functions as a high-pass filter). Common onomatopoeic instrument vocables—

[49] For a singular empirical look at the phenomenon of imitating non-speech sounds with the voice, see Rodrigo Segnini Sequera, "On the Vocalization of Non-Speech Sounds: Implicit Mechanisms and Musical Applications" (Ph.D. Diss., Stanford University, 2006).

including the "bow-chicka-wow-wow" wah-wah guitar imitation that has come to serve as a sign of (raunchy) sex and the vast panoply of vocables used in scat singing—are ways to map instrumental timbre onto the voice. As Steven Feld *et al.* explain, this strategy of iconic vocal mimicry of non-vocal timbres is common to the discourse of the modern recording studio.[50] Animal sounds are another form of vocal onomatopoeia of non-vocal sounds. Moreover, many common descriptions of timbre are vocal in origin:[51] for instance, playing the oboe does not involve any more use of the nasal cavities than the clarinet, but the timbre is often judged more "nasal" owing to the similarities between oboe timbre and a "nasal" voice.[52] The "growling" saxophone and "screaming" trumpet are similarly nods to their vocal equivalents. To be clear, overtly imitating an instrument with the voice is not the same thing as timbral imagery or the simple act of *hearing* an instrument; however, since action and perception overlap to a significant degree (as we will see), the ways we commonly imitate different timbral qualities with the voice might provide an important clue as to how we perceive them.

In addition to the ubiquitous linkages above, there is a small but growing body of empirical literature supporting the claim that the voice (via subvocalization) is involved in the perception of instrumental timbre.[53] In psychology, perhaps the most common approach to this topic can be found in studies on *timbre imagery*. Results from this research are somewhat ambiguous, however, owing to a productive philosophical confusion over what precisely is meant by "imag-

---

[50] Steven Feld *et al.*, "Vocal Anthropology," in *A Companion to Linguistic Anthropology*, ed. Alessandro Duranti (Maldon, UK: Blackwell, 2004), 324.

[51] Cox, "The Metaphoric Logic of Musical Motion and Space," 89-98. The issue of instrumental sound being described in vocal terms is also discussed in ""Embodying Music: Principles of the Mimetic Hypothesis," 30 & 39.

[52] This specific example comes from Jamshed J. Bharucha, Meagan Curtis, and Kaivon Paroo, "Varieties of Musical Experience," *Cognition* 100, no. 1 (2006): 147.

[53] Of course, I should hedge this by acknowledging that a host of factors make such a claim difficult to generalize; for example, degree of attention, listening biography, and musical background (specifically, training in the instrument) all effect perception and cognition of instrumental sound.

ing." Phrased in the form of a *koan*: Can you hold a sound in your head without reproducing it? In one of the earliest studies, Crowder (1989) found that listeners performed same/different pitch judgments faster when paired stimuli were the same timbre, suggesting that imagery of non-vocal timbre facilitates pitch recognition. Although he rules out any role for subvocalization (or motor resonance more generally) in this process, arguing that since "humans are utterly incapable of reproducing physically any but the grossest dynamic or spectral features of timbre … an appeal to sensory rather than motor imagery is justified,"[54] this dismissal—echoed in Virpi Kalakoski's assertion that timbral imagery is a matter of the "inner ear" not the "inner voice"[55]—is founded on some problematic assumptions, most notably the contention that we can only subvocalize to sounds that we can overtly *reproduce*.[56] But just because we cannot with perfect fidelity vocally reproduce the wah-wah effect on the electric guitar, for example, does not mean that we cannot and do not *try* (both overtly and covertly).

I would suggest, moreover, that no matter how mangled the imitation if forced to sing aloud, this act of "trying"—subvocalizing to (or overtly vocalizing) instrumental timbre—helps enable us to perform two imperative cognitive functions: *memory* and *description*. To illustrate the first, try summoning the tone of a wah-wah guitar, or a saxophone growl, or a harmon-muted trumpet; it is probably difficult to conceive of the sound of an instrument without somehow subvocally (or otherwise sub-motorically) *reenacting* its performance. Is there any way timbre can be

---

[54] Robert C. Crowder, "Imagery for Musical Timbre," *Journal of Experimental Psychology: Human Perception and Performance* 15, no. 3 (1989): 478. The author makes a similar assertion in "Auditory Memory," in *Thinking in Sound: The Cognitive Psychology of Human Audition*, ed. Stephen McAdams and Emmanuel Bigand (Oxford: Clarendon Press, 1993), 134.

[55] This formulation comes from Virpi Kalakoski, "Musical Imagery and Working Memory," in *Musical Imagery*, ed. R.I. Godøy and H. Jørgensen (Lisse: Swets & Zeitlinger, 2001).

[56] This critique is echoed by Baddeley and Logie: see Cox, "The Metaphoric Logic of Musical Motion and Space," 82.

retained in memory and mentally represented *without* some level of covert bodily involvement?[57] Granted, this is something of an abstraction, since all instrumental timbre is also at the same time pitched and temporally dynamic, not to mention highly sensitive to all other manners of context (including room acoustics, make of instrument, player, etc.). However, the very fact that the "sound of the saxophone growl" exists as a perceptual gestalt independent of these complexities indicates a unified timbral impression, one that I am arguing likely has a vocal correlate.[58] I am not claiming that vocal mimesis is the *only* way to keep different timbral entities distinct in memory—a guitarist will hear the timbre of a guitar with greater motor empathy of the fingers than a non-player, for instance—but it does play an important role, even when listening to the instrument you play. Vocal mimesis also plays a role in timbre *description*, as discussed previously (and to be continued in Chapter 3): communicating details of musical timbre often relies on reference to the voice, either in the form of onomatopoeia or vocal metaphors. If motor resonance is involved in timbre perception, as suggested here, perhaps the "inner ear"—shorthand for imaged auditory percepts—cannot be kept so resolutely separate from the "inner voice" after all.[59]

Support for the theory of mimetic subvocalization of timbre also comes from neurophysiology studies. Halpern *et al.* (2004) found activity in the supplementary motor area (SMA) when

---

[57] This calls to mind James's theory of emotions: "without the bodily states following on the perception [of emotion], the latter would be purely cognitive in form, pale, colorless, destitute of emotional warmth." Similarly, can we really imagine a timbre without tracing out the bodily actions that stand behind it?

[58] To account for the ability of human perceptual systems to recognize a unified relationship between instrument and timbre despite variances in timbre across the range of each instrument, Gregory Sandell proposes the idea of "macrotimbre" as the basis for source identification. See Gregory J. Sandell, "Macrotimbre: Contribution of Attack and Steady State," *Journal of the Acoustical Society of America* 103, no. 5 (1998).

[59] Interpreted this way, Crowder's results—as well as that of Logie and Edworthy (1986), Baddeley and Logie (1992), Smith, Reisberg, and Wilson (1992), Pitt and Crowder (1992), and Bailes (2007), using similar behavioral paradigms—suggest that vocal mimesis may indeed play a role in the perception of instrumental timbre, insofar as these studies show that timbre imagery is a "real" phenomenon with quantifiable behavioral correlates.

people listen to non-vocal timbre.[60] Although this result does not point specifically to subvocalization (SMA is involved in *all* imaged bodily action), it is highly suggestive in light of the above; if timbral imagery is predicated upon vocal mimesis, activity in SMA may well be attributed to subvocalization. Other studies have pointed to such a possibility as well, including neuroimaging research showing the functional overlap of vocal- and musical timbre processing in the superior temporal sulcus.[61] But for the present purposes, perhaps the most suggestive study to demonstrate vocal-motor involvement in non-vocal music comes from Koelsch *et al.* (2006), who found that regions of the brain responsible for larynx control—the motor areas of the voice, particularly a structure called the Rolandic operculum—are active when we hear "pleasant" music.[62] Though this study does not isolate the role of timbre in these activations, it would seem to indicate that vocal areas are operative in musical appraisal in general. Considering that timbre plays a major role in affective response,[63] it stands to reason that motor regions coding for vocal production are involved in (positive) timbral appraisals, even for non-vocal timbres. In the next chapter, we will wade into the question with some new empirical findings.

Koelsch *et al.*'s implication of the Rolandic operculum in processing "pleasant" music opens onto another question about the voice/instrument relationship: if subvocalization plays a role in the perception of instrumental timbre, how does this role relate to affective response conveyed through timbre? In her analysis of the performativity of vocal timbre, Nina Eidsheim pro-

---

[60] Andrea R. Halpern *et al.*, "Behavioral and Neural Correlates of Perceived and Imagined Musical Timbre," *Neuropsychologia* 42(2004).

[61] Pascal Belin *et al.*, "Voice-Selective Areas in the Human Auditory Cortex," *Nature* 403(2000); Vinod Menon *et al.*, "Neural Correlates of Timbre Change in Harmonic Sounds," *Neuroimage* 17(2002); Jean-Pierre Chartrand and Pascal Belin, "Superior Voice Timbre Processing in Musicians," *Neuroscience Letters* 405(2006). For a review of some of this research, see Patel, *Music, Language, and the Brain*, 344-51.

[62] Stefan Koelsch *et al.*, "Investigating Emotion with Music: An fMRI Study," *Human Brain Mapping* 27(2006).

[63] Isabelle Peretz, L. Gagnon, and B. Bouchard, "Music and Emotion: Perceptual Determinants, Immediacy, and Isolation after Brain Damage," *Cognition* 68(1998); Bigand *et al.*, "Multidimensional Scaling of Emotional Responses to Music: The Effect of Musical Expertise and of the Duration of the Excerpts."

poses a bipartite schema of *timbre sonic* and *timbre corporeal*. The first category refers to the traditionally defined "acoustic world" (Fales 2002) of timbre; *timbre corporeal*, on the other hand, acknowledges that musical timbres are actions first, and "sonic" second: "We experience vocal timbre ultimately through frequency and amplitude, but first through muscle recognition."[64] "Corporeal recognition" is central to vocal affect; vocal timbre communicates embodied meanings because the listener "recognizes," from a lifetime of experience living in an en-voiced body, what it feels like to produce different qualities of sound. She explains:

> Recognition, recollection and 'resonance' with vocal sounds are all individual subjective *translations* of the vocal sounds in music, the production of which is also based on various and specific levels of tension and release. Cognition, such as hearing, may be translated back to the basic level of the recognition of tensions and releases generated in the execution of corporeal mechanical functions.[65]

Following this model, if we view all musical timbre as similarly charged with corporeal implications—both in the specific motor mechanics of performance, and in overt or covert vocalization—then it is not difficult to extend *timbre corporeal* to instrumental sound as well. As Eidsheim puts it, "A deep experience of the timbre of instruments, such as intense emotion, is a manifestation of the corporeal experience stimulated by, and as an extension of, the sound of the singing voice."[66]

In the same vein but from the perspective of cognitive psychology, Patrik Juslin and Petri Laukka apply "Spencer's law" to musical instruments to suggest that certain acoustic cues, based in a shared, kinetic knowledge of human vocal expression, communicate affective qualities owing

---

[64] Eidsheim, "Voice as a Technology of Selfhood: Towards an Analysis of Racialized Timbre and Vocal Performance," 259.

[65] Ibid., 251. Italics in original.

[66] Ibid., 251-52.

to their iconic similarities to the voice.[67] According to this view, since the perception of emotion in vocal expression and music is mediated by special-purpose neural modules that do not "know" the difference between vocal and other acoustic stimuli, similar acoustical qualities from either domain would trigger the same reaction.[68] Juslin and Laukka dub this the *superexpressive voice theory* of instrumental affect:

> […] what makes a particular 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.[69]

These two theoretical perspectives—timbre corporeal and vocal superexpressivity—combine into a powerful explanatory tool for timbre-mediated affect. Returning specifically to the vocal manifestations of arousal, moreover, acoustic properties associated with high arousal states—boosts in high-frequency energy and noise—have been found to share similar affective functions in the realm of instrumental music. To cite three examples: In one novel study, Chen-Gia Tsai *et al.* (2010) found that both producing and listening to the timbre of a vocal growl has

---

[67] The first publication to theorize along these lines is 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). The idea is further developed in Petri Laukka, Patrik N. Juslin, and Roberto Bresin, "A Dimensional Approach to Vocal Expression of Emotion," *Cognition & Emotion* 19, no. 5 (2005); Patrik N. Juslin and Daniel Västfjäll, "Emotional Responses to Music: The Need to Consider Underlying Mechanisms," *Behavioral and Brain Sciences* 31(2008).

[68] The influential theory of brain modules was developed by cognitive scientist Jerry A. Fodor, *The Modularity of the Mind* (Cambridge, MA: MIT Press, 1983). It remains a controversial theory, and the jury is still out on its neuroscientific feasibility; nevertheless, even some skeptical researchers have localized Fodor's "modules" in subcortical areas: see J. Panksepp and J.B. Panksepp, "The Seven Sins of Evolutionary Psychology," *Evolution and Cognition* 6(2000).

[69] Juslin and Laukka, "Communication of Emotions in Vocal Expression and Music Performance: Different Channels, Same Code?," 803.

the effect of unconsciously increasing cortico-spinal stability, a necessary biomechanical precursor to fight-or-flight reflexes.[70] The authors suggest that this ergotropic arousal effect can be triggered by “growl-like” instrumental timbre as well, including distorted electric guitar and growled aerophones. The high-arousal vocal act and resulting sound can translate mimetically into the perception of instrument timbre, leading in audition to a similar affective/bodily state as in the act of production. Similar conclusions were reached by ethologist Daniel Blumstein *et al.* (2012), who found that increases in timbral nonlinearities (i.e., noise) in non-vocal musical excerpts lead to increased affective arousal in listeners, prompting the authors to conclude: “Most of the perceptual and cognitive mechanisms underlying music processing likely evolved for a variety of reasons not related to music—many related to emotional vocal communication.”[71] And finally, in his statistically-driven study of expectation in classical music, David Huron (2006) argues that musical climaxes tend to be marked with timbral indices of *fear*, including the acoustic qualities of high-frequency energy (*brightness*) and inharmonic elements (*noise*). Timbre in moments of musical climax tends to be “intense or harsh rather than mellow or warm.”[72] The “contrastive valence” produced when high-arousal timbres resolve in the aftermath of a big climax are, according to Huron’s model of musical anticipation, key to the affective efficacy of these common musical gestures.

Taken together, evidence from embodied experience, timbre imagery studies, neuroimaging, theoretical models, as well as the examples above, suggest that shared human vocality, spe-

---

[70] Tsai *et al.*, “Aggressiveness of the Growl-Like Timbre: Acoustic Characteristics, Musical Implications, and Biomechanical Mechanisms.” Cox would call this phenomenon the “amodal” aspect of mimetic motor imagery, defined as abdominal relaxation or contraction as a result of the perceived exertion dynamics of sound: see Cox, “Embodying Music: Principles of the Mimetic Hypothesis,” Principle 10.

[71] Daniel T. Blumstein, Gregory A. Bryant, and Peter Kaye, “The Sound of Arousal in Music is Context-Dependent,” *Biology Letters* 8(2012): 4.

[72] Huron, *Sweet Anticipation: Music and the Psychology of Expectation*, 324.

cifically the affective states associated with scalar levels of vocal tension, can be viewed as an inbuilt interpretive framework for understanding the embodied meaning of both vocal and nonvocal musical timbre. Let me summarize: through a process of "corporeal recognition," affectively laden bodily acts are transduced through instruments into acoustic qualities that—via actual, sonic, and perceptual mimesis of the voice—transmit the bodily energies (and affects) involved in production to the listener, who in turn translates them back into their originating acts. This act of perceptual "translation" will be the final topic of exploration. But before going on, I would like to make a brief detour to explore in greater depth the acoustic qualities that mark the "sound of arousal."

Components of Noisy Timbre

In this section, we will dig a little deeper into the sound of bodily exertion. In addition to the two timbral attributes, brightness and noise, I am adding one more quality to the mix: perceived dissonance, also known as *auditory roughness*. Impressions of "noisiness" as a byproduct of physical arousal are often related to increases in the roughness of a timbre, though this phenomenon is not explicitly addressed in any of the vocal studies previously cited. I will also begin sharpening definitions of the acoustical properties and perceptual *qualia* that will play a starring role throughout the dissertation, including the experiments that make up the core of the next chapter. (Additional acoustic information, including formulae, can be found in Appendix 1.)

Brightness (Spectral Centroid)

After the relative strength of upper partials, the perceptual quality of *brightness* is probably the most thoroughly researched aspect of timbre, with a history in the psychoacoustic literature going back to Helmholtz.[73] *Brightness* has been found to correlate most strongly with the acoustic property of *spectral centroid*, the average center of spectral energy in a sound (or the fulcrum point balancing high and low frequencies). For our purposes, acoustic *brightness* can be defined as a perceptual quality of timbre associated with a relatively high ratio of high to low frequency energy. *Bright* timbres, in short, are rich in high-frequency components.[74]

---

[73] Helmholtz, it should be noted, does not use the word "brightness," though he was referring to its acoustic correlate—spectral centroid—when he writes: "When partial tones higher than the sixth or seventh are very distinct, the quality of tone is *cutting* and *rough.*" See Helmholtz, *On the Sensations of Tone as a Physiological Basis for the Theory of Music*, 119. Major empirical sources on *spectral centroid* (and its perceptual correlate, "brightness," "nasality," or "sharpness") include the following: G. von Bismarck, "Timbre of Steady Sounds: A Factorial Investigation of Its Verbal Attributes," *Acustica* 30(1974); John M. Grey, "An Exploration of Musical Timbre" (Ph.D. Dissertation, Stanford University, 1975); Plomp, *Aspects of Tone Sensation: A Psychophysical Study*; John M. Grey, "Multidimensional Perceptual Scaling of Musical Timbres," *Journal of the Acoustical Society of America* 61(1977); David L. Wessel, "Timbre Space as a Musical Control Structure," *Computer Music Journal* 3, no. 2 (1979); McAdams, "Recognition of Sound Sources and Events."; Roger A. Kendall and Edward C. Carterette, "Verbal Attributes of Simultaneous Wind Instrument Timbres: I. von Bismarck's Adjectives," *Music Perception* 10, no. 4 (1993); "Verbal Attributes of Simultaneous Wind Instrument Timbres: II. Adjectives Induced from Piston's 'Orchestration," *Music Perception* 10, no. 4 (1993); Stephen McAdams *et al.*, "Perceptual Scaling of Synthesized Musical Timbres: Common Dimensions, Specificities, and Latent Subject Classes," *Psychological Research* 58(1995); Roger A. Kendall and Edward C. Carterette, "Difference Thresholds for Timbre Related to Spectral Centroid" (paper presented at the 4th ICMPC, Montreal, Quebec, Canada, 1996); Hajda *et al.*, "Methodological Issues in Timbre Research."; Emory Schubert and Joe Wolfe, "Does Timbral Brightness Scale with Frequency and Spectral Centroid?," *Acta Acustica* 92(2006); John Hajda, "The Effect of Dynamic Acoustical Features on Musical Timbre," in *Analysis, Synthesis, and Perception of Musical Sounds*, ed. James Beauchamp (New York: Springer, 2007).

[74] A brief note regarding terminology is in order. As with all adjectives describing qualities of timbre, it should be noted that the percept of *brightness* is influenced considerably by the connotations of the word itself; thus, the perception of *brightness* in a tone is contingent upon the language and cultural background of the listener. The sensitivity of perceptual responses to the precise language of the cue has led researchers to employ a range of words in describing spectral centroid, including "nasal" and "sharp"; indeed, there is a good deal of linguistic slipperiness when attempting to map the acoustics of timbre onto the psychoacoustics of timbre. For example, Kendall and Carterette (1993) found that the word "sharp," an oft-cited perceptual correlate of high spectral centroid from German studies (*scharf*), does not actually hold up in the English translation. Similarly, the synaesthetic description "bright," which is borrowed from the visual domain, carries vision-based connotations and symbolic resonances into assessments of timbral *brightness* among English speakers, as concluded in my experiments (to be reported in the next chapter). The linguistic dimensions of timbre will be taken up in greater detail in Chapter 3, but it is worth reminding ourselves here that linguistic categories of timbre perception are cultural constructs, not innate perceptual categories. Empirical researchers of timbre should always be sensitive to this fact so as to avoid shoehorning perceptual responses into perhaps arbitrary and non-representative linguistic categories. I opted for the word "brightness" in this text and in my experiments due to its ubiquity and intelligibility to study participants, not because it is necessarily a superior term to other perceptual correlates of spectral centroid. For more on the methodological and epistemological difficulties involved in relating acoustics to semantics, see the introduction to Kendall, "The Role of Acoustic Signal Partitions in Listener Categorization of Musical Phrases."

To illustrate spectral centroid visually, Fig. 1.2 compares spectrographs of clarinet and oboe playing the same pitch (C4), with dashed lines indicating the fulcrum of spectral energy. In the clarinet, as evinced by the thick, red, horizontal lines, we can see that most of the energy is concentrated in the lower partials (below 1 kHz), particularly the fundamental and the first couple harmonics. In contrast, the oboe distributes its energy liberally into the higher harmonics, providing the instrument with its characteristic "piercing," "clarion" quality. In terms of spectral centroid, the clarinet has a much lower center of gravity (776 Hz) than the oboe (1466 Hz); in fact, in many multidimensional mappings of orchestral instrument timbre, clarinet and oboe are virtually polar opposites in terms of spectral centroid. This correlates with very different "brightness," "sharpness," or "nasality" profiles.

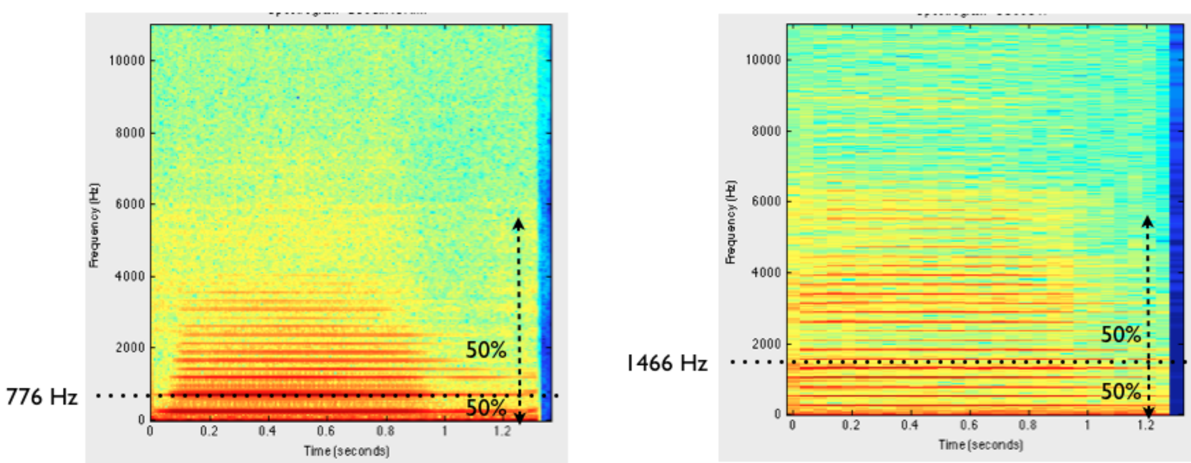


FIGURE 1.2: Clarinet (left) and oboe (right) playing C4 (266 Hz)

Brightness—and its acoustic correlate, spectral centroid—is perceptually important for a number of psychobiological reasons. To start, the human ear is quite sensitive to high-frequency energy; sounds in the 2–4 kHz range are perceived as being much louder than sounds outside this range (by as much as 30 decibels).[75] Screeching, scraping, fingernails on chalkboard—aversive sounds like these are deemed irritants, if not painful intrusions, owing in no small measure to the psychophysiological mechanisms of the auditory system. Further, the primary auditory cortex—the region of the superior temporal gyrus that plays the starring role in processing sound—is *tonotopically* organized, meaning that different frequency ranges are processing in different cortical locations.[76] The region responsible for processing high frequency sound exhibits greater functional connectivity to subcortical areas involved in ergotropic arousal states and "fight or flight" reflexes; moreover, since brighter sounds encompass a comparatively larger range of frequencies, a larger portion of the auditory cortex is required to process them. The architecture of the brain, then, may play a role in our innate sensitivity to *brightness*.

What adaptive purpose does such sensitivity serve? From an evolutionary perspective, there are two answers to this. First, *brightness* is essential to sound source identification. We know that an approaching sound indicates the present of a bear and not a deer, for instance, in part because of the unique ratio of high-to-low frequency energy associated with the sounds of each animal. The same adaptive mechanisms apply to the perception of musical timbre as well. Researchers have found that removing the *brightness* from isolated tones and recordings of music—i.e., low-pass filtering the signal—drastically reduces our ability to correctly identify them.[77] Perceptually, we *need* all the upper partials in order to know what we are hearing.

---

[75] McDermott, "Auditory Preferences and Aesthetics: Music, Voices, and Everyday Sounds," 203.

[76] C.M. Wessinger *et al.*, "Tonotopy in Human Auditory Cortex Examined with Functional Magnetic Resonance Imaging," *Human Brain Mapping* 5(1997).

[77] One of the first experiments to suggest the essential role of spectral centroid in instrument identification was conducted by psychologist Kenneth Berger (1964), who found that normal, unaltered tones yielded correct identifications 59% of the time, while low-pass filtered tones were only correctly identified 18% of the time. Schellenberg *et al.* obtained a similar result using very brief (100–200 ms) excerpts of pop music recordings. For more, see Kenneth W. Berger, "Some Factors in the Recognition of Timbre," *Journal of the Acoustical Society of America* 36, no. 10 (1964); E. Glenn Schellenberg, Paul Iverson, and Margaret C. McKinnon, "Name that Tune: Identifying Popular Recordings

High frequency is also an acoustical index of physical proximity, giving the listener vital clues about the source's location and behavior.[78] Consider what it sounds like to talk to someone in another room of the house: their voice is rendered dull, and it is often difficult to understand due to the attenuation of high frequencies that result from the physics of sound diffraction.[79] Generally speaking, sounds that are materially obstructed from the listener are less "bright" than they would be in an open, unfettered space. Another dimension of the proximity phenomenon has to do with sound propagation over distance. High frequencies tend to be strong at their source, but fade away quickly as they travel through space. Physical distance, in a sense, thus functions as a low-pass filter. The ecological implications of this acoustic phenomenon are clear: *brightness* plays an essential role in enabling us to both identify and localize the source of sounds.[80]

Noise (Inharmonicity and Spectral Flatness)

Harmonic timbres focus the majority of their energy within a harmonic series of overtones (integer multiples of the fundamental frequency). Referring back to Fig. 1.2, we can see that clarinet and oboe are both largely harmonic instruments, and they follow this exact pattern in their spectral distribution: in spectrographic form, harmonic timbres are easily recognizable, with stable, evenly distributed bands of energy periodically spreading upwards. Like most instruments in the modern orchestra, the distinctive timbral qualities of these two instruments are a reflection of the

---

from Brief Excerpts," *Psychonomic Bulletin & Review* 6, no. 4 (1999). Further details can be found in Handel, *Listening: An Introduction to the Perception of Auditory Events*, 244.

[78] It is quite fitting, then, that guitarists (and their gear) often refer to *brightness* as "presence." For more, see J.C. Middlebrooks and D.M. Greenhaw, "Sound Localization by Human Listeners," *Annual Review of Psychology* 42, no. 1 (1991).

[79] For a simple explanation, see David M. Howard and James Angus, *Acoustics and Psychoacoustics*, 2nd ed. (Amsterdam: Focal Press, 2001), 48-51.

[80] For more on the acoustics and psychoacoustics of source identification, see Ch. 8 in Handel, *Listening: An Introduction to the Perception of Auditory Events*.

Enlightenment aesthetic ideal of timbral purity and "transparency," where extraneous noise is assiduously avoided.[81]

What is an interloper in one tradition, however, is an essential component to many musical cultures around the world, including much of American popular music and jazz. With historical and cultural roots in the "heterogenous sound ideal" of African and Afro-diasporic musical aesthetics,[82] many iconic pop music timbres—the overdriven electric guitar amp, honking saxophone, and screaming vocalist, for instance—showcase noise as the *sine qua non* of their expressive identity. But despite its ubiquity, noise in the context of musical timbre proves an elusive property. How do we quantify "noise"? Most agree that musical timbres with noise components eschew the mathematical regularity of harmonicity in favor of a less periodic, more chaotic spectrum. Unlike brightness, which is generally related to the physical attribute of spectral centroid, there is no one agreed-upon measure of acoustical noise; when pressed to define the experience, the psychoacoustic literature provides us with very little practical direction.[83] There are myriad different approaches to this question, but for our purposes here (and in the Chapter 2 experiments), I will map the acoustics of noise using two measurements: *inharmonicity* and *spectral flatness*.

Inharmonicity is most closely associated with the concept of "noise" in the psychoacoustic and musicological literature. Loosely put, it is defined as energy *outside* the harmonic series associ-

---

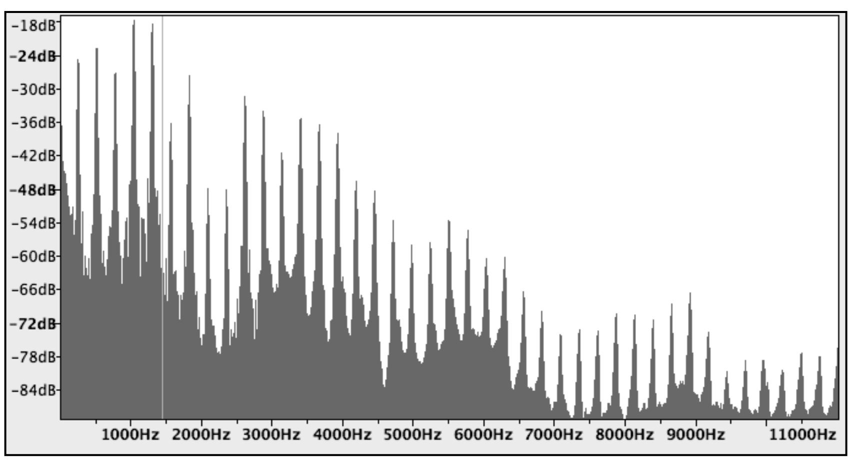
[81] For a fascinating history of this timbral ideal in the Enlightenment-era orchestra, see Dolan, *The Orchestral Revolution: Haydn and the Technologies of Timbre*.

[82] This term comes from composer Olly Wilson: see Olly Wilson, "The Heterogeneous Sound Ideal in African-American Music," in *Signifyin(g), Sanctifyin', and Slam Dunking: A Reader in African-American Expressive Culture*, ed. Gena Dagel Caponi (Amherst: University of Massachusetts Press, 1999).

[83] I am aware of only a few articles that attempt to ferret out the relationship between noise and harmonic elements in musical contexts. See B.C.J. Moore, R. Peters, and R.W. Peters, "Thresholds for the Detection of Inharmonicity in Complex Tones," *Journal of the Acoustical Society of America* 77(1985); Fales and McAdams, "The Fusion and Layering of Noise and Tone: Implications for Timbre in African Instruments."; Harris Berger and Cornelia Fales, "'Heaviness' in the Perception of Heavy Metal Guitar Timbres: The Match of Perceptual and Acoustic Features over Time," in *Wired for Sound: Engineering and Technologies in Sonic Culture*, ed. Paul D. Greene and Thomas Porcello (Middleton, CN: Wesleyan University Press, 2005).

ated with the fundamental.[84] This acoustic phenomenon is present in many Western instruments, including bells and even (to a much lesser extent) piano, but generally speaking it is a quality to be avoided. It is common to many of the "noisier" sounds of popular music.

Spectral flatness refers to the relative smoothness or spikiness of a signal. To illustrate, Figure 1.3 compares the spectra of oboe and a sung female "distorted voice."[85] With the oboe, we can see that most of the energy is located in discrete, harmonic frequency bands, which translate into a spiky spectrum—most of the total energy "spikes" on the frequencies that are harmonically related to the fundamental (spectral centroid is indicated by a faint vertical line). A "flat" spectrum, on the other hand, smears its energy aperiodically across a range of frequencies; note that beyond the first few partials, there are no harmonic spikes. (Also note the much higher spectral centroid in this noisier timbre [3066 Hz]). Often this complex wash of frequencies is experienced as "noise." (The fact that we call the random statistical distributions of acoustic energy "noise"—as in "white noise," "pink noise"—shows the connection between spectral flatness and noisiness.)



[84] It should be noted that this parameter is limited to single, isolated tones; it is meaningless when analyzing polyphonic music, since there is generally not one single fundamental.

[85] Signals are the same pitch (C4) and were equalized manually for loudness. The vocal signal will reappear in Experiments 1 and 3 as "Voice 2."

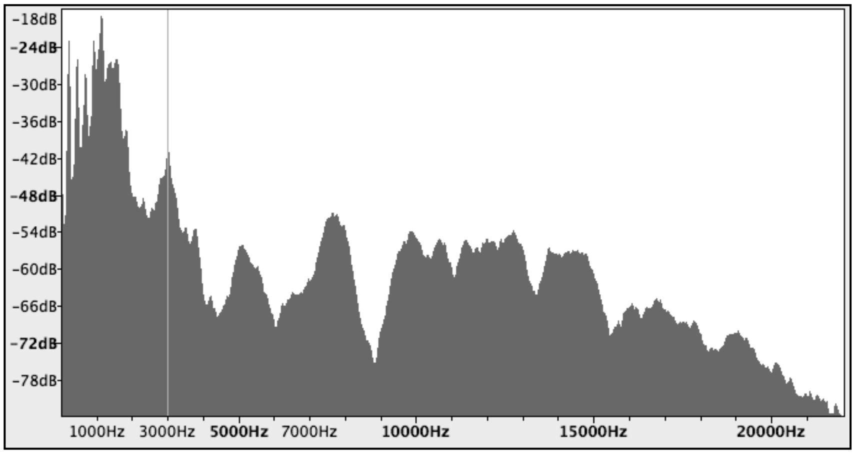


FIGURE 1.3: Spectra of oboe (top) and distorted voice (bottom) on C4[86]

Of course, it is unclear to what extent these three acoustic parameters relate phenomenologically to the perceptual world (if at all). What is "noise" to a signal processor might not be "noisy" to an actual human listener. Since I will be approaching the topic from both acoustic and perceptual angles in the pages ahead, to avoid confusion I will use *a-noise* to describe the physical qualities associated with this percept (Fales's "acoustic world"), and *p-noise* to reference the percept itself ("perceived world"). Our working definition of both a-noise and p-noise moving forward in the dissertation will be based in part on what the experimental data tell us in Chapter 2 and the Appendices. Therefore, rather than try to pin down a definition for this complex phenomenon *a priori*, we will return to the question of noise throughout.

Roughness (Sensory Dissonance)

Finally, physical exertion in the voice and many other sound generators tends to lead to increased sensory dissonance, which is related to the acoustic quality of auditory *roughness*. As anyone who has tuned a guitar or tried to sing in perfect unison with another singer knows, this phe-

---

[86] Spectra were generated using Audacity software. The analysis parameters were the same for both signals: 1024 frames with Bartlett window on.

nomenon, commonly known as "beating," occurs when two sinusoids lie very close in frequency—or in more specifically auditory terms, when two sinusoids lie within the same critical band—producing a sensation of rubbing and physical friction (in the form of amplitude fluctuation) as the two frequencies vie for control over the limited real estate of the inner ear.[87] The closer the two sinusoids get to the same frequency, the slower the beating and the "smoother" the resultant timbre; conversely, as the two frequencies diverge into separate critical bands, roughness will attenuate and the perception of two separate pitches will begin to predominate. Roughness is closely related to inharmonicity, since inharmonic components can interfere with critical bands. It has been proposed as a psychophysiological explanation for the perception of dissonant intervals and chords; it has also been shown to play a vital aesthetic role in many musical traditions of the world.[88]

Having reviewed the acoustic actors that play a starring role in the rest of the dissertation, we now turn our focus back to the perceptual processes involved in translating sound—and specifically, these three acoustic markers of arousal—into action. As I will be arguing, the sound of vocal exertion as manifested in *brightness*, *noise*, and *roughness* are not just matters of disinterested perception, but rather, due to the mechanisms of act–sound coupling, actually involve affectively loaded motor resonance on the part of the perceiver.

---

[87] For in-depth acoustic explanations of this phenomenon, see William Sethares, *Tuning, Timbre, Spectrum, Scale* (London: Springer, 1998); Pantelis N. Vassilakis, "Perceptual and Physical Properties of Amplitude Fluctuation and their Musical Significance" (Ph.D. Diss., UCLA, 2001).

[88] On the role of roughness in dissonance and musical tension, see Reinier Plomp and W.J.M Levelt, "Tonal Consonance and Critical Bandwidth," *Journal of the Acoustical Society of America* 38(1965); D. Pressnitzer *et al.*, "Perception of Musical Tension for Non-Tonal Orchestral Timbres and its Relation to Psychoacoustic Roughness," *Perception and Psychophysics* 62(2000). On the aesthetic dimensions of roughness, see Pantelis N. Vassilakis, "Auditory Roughness as a Means of Musical Expression," in *Selected Reports in Ethnomusicology: Perspectives in Systematic Musicology* ed. Roger A. Kendall and Roger W.H. Savage (Los Angeles: UCLA Ethnomusicology, 2005).

Hearing is Doing: Act–Sound Coupling

Returning to the scene of the imaginary crime outlined earlier, we can see how excessive physical tension manifests acoustic "abnormalities" in the voice. We can also see how similar abnormalities, when not actually vocal in origin, might trigger similar reactions in the listener. But what accounts for the perceptual salience of these specific abnormalities? How does noisy timbre translate into bodily meanings? What is going on in the policeman's head when he hears your screaming voice?

The motor theory of timbre perception advanced here is entirely predicated on this act of translation, so for the rest of this chapter we will be giving this crucial process a closer look. I will review two perspectives here that I think are particularly useful: ecological psychology and the neurophysiology of mirror neuron-enabled audio-motor interaction. Together, these bodies of research demonstrate that *hearing* is always on some level *doing*, at least with respect to man-made action sounds: the perceptual systems involved when we listen to sounds made by other humans are coupled to the sensorimotor acts we know are involved in the production of those same sounds. As Marc Leman puts it, "the human body can be seen as a biologically designed mediator that transfers physical energy up to a level of action-oriented meanings." In the rest of this section, we will explore how this transfer process works.[89]

Spearheaded by psychologist James J. Gibson, the ecological approach proposes that perceptual systems are closely attuned to the environment, providing unmediated access to the objects and events that are of particular relevance to human-level interaction. From this perspective, a sound is never simply a physical vibration, but rather an index for a material event. All sounds, simply put, are "sounds of things."

---

[89] Leman, *Embodied Music Cognition and Mediation Technology*, xiii.

This may seem obvious enough, but applied to music, the ecological approach is far from uncontroversial. The materiality of sound generation, indeed, can perhaps at times appear invisible in music listening: to some philosophers, perceptually transcending the mere world of stuff is an ontological prerequisite for music to *be* music.[90] From this perspective, we need to hear melodies, affective intentions, and formal structures, not gut strings, horsehair bows, and brass tubes in order for the perception of sound to transform into the richness of musical experience. Ecological acoustics, however, counters this position by arguing that hearing—even hearing Beethoven—can never be entirely abstracted from materiality. As William Gaver puts it, "sound provides information about an interaction of materials at a location in an environment," whether an oncoming car in "everyday listening" or a poignant oboe solo emerging from an ensemble passage in "musical listening."[91] In one study subjects were asked to freely describe a range of musical and non-musical sounds, and an overwhelming majority of responses (57%) referred to physical source, as opposed to genre (22%), acoustic quality (7%), emotion (5%), and other descriptions (<4%).[92] This conclusion is borne out by other empirical studies.[93] Psychologist Gerald Balzano points out that "the kinds of things we are capable of hearing that are important for timbre perception are events like pounding, blowing, plucking, rolling, whistling, screaming, and all sorts of physical processes that words can only hint at but which are nonetheless specified in the underlying dynamics of the signal, and therefore just as potentially 'available' to a perceiver as a Fou-

---

[90] Roger Scruton writes: "[…] the person who listens to sounds, and hears them as music, is not seeking in them some information about their cause, or for clues as to what is happening. On the contrary, he is hearing the sounds *apart* from the material world." See Roger Scruton, *The Aesthetics of Music* (New York: Oxford University Press, 1997), 221.

[91] William W. Gaver, "How Do We Hear in the World?: Explorations in Ecological Acoustics," *Ecological Psychology* 5, no. 4 (1993): 8.

[92] Nicola Dibben, "What Do We Hear, When We Hear Music? Music Perception and Musical Material," *Musicae Scientiae* 5, no. 2 (2001).

[93] For instance, see Handel, *Listening: An Introduction to the Perception of Auditory Events*; Bregman, *Auditory Scene Analysis: The Perceptual Organization of Sound*.

rier analysis."[94] It is a basic human perceptual orientation, then, to hear sounds as the presence of *things*. Musical listening, it appears, is founded upon everyday listening.

This insight broadens our discussion of embodied cognition considerably, as perception and cognition in this schema are rooted in more than just our own human body: they are also ineluctably connected to the environment. (Terms such as "extended cognition" and "grounded cognition" have been suggested to highlight this fact.)[95] In the ecological view our environment is perceived most directly in terms of its potential for bodily action; it is not just that we find our bodies thrown into a theater of things and events, but rather that the contingencies of human action facilitate and constrain what we are able to perceive in this theater. Thus, for instance, a chair is perceived as something we *sit in,* not as a collection of wood pieces in a particular shape; a pencil is something we *write with*, not a cylindrical structure of wood and lead.[96] This ecological view exemplifies the Gibsonian concept of *affordances*: percepts are processed according to the opportunities for human action they "afford."[97] Since we exhibit a perceptual bias towards action-relevant properties of the material world—in Leman's term, we possess an "action-oriented ontology"[98]—taking the ecological approach to music suggests that listening to musical sounds necessarily involves hearing the bodily implications of sound production.[99] A larynx, in this view,

---

[94] Gerard J. Balzano, "What are Musical Pitch and Timbre?," *Music Perception* 3, no. 3 (1986): 309.

[95] Respectively, Andy Clark and Lawrence Barsalou. See Clark, *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*; Lawrence W. Barsalou, "Grounded Cognition," *Annual Review of Psychology* 59(2008).

[96] It should be noted that affordances are entirely species-specific: to a termite the wooden chair affords not sitting but lunch, while the pencil to a chimpanzee might be a useful instrument for back-scratching.

[97] This term first appears in reference to vision in James J. Gibson, *The Senses Considered as Perceptual Systems* (Boston: Houghton Mifflin, 1966). The idea is developed further in *The Ecological Approach to Visual Perception* (Boston: Houghton-Mifflin, 1979).

[98] Leman, *Embodied Music Cognition and Mediation Technology*, 87.

[99] Examples of musical scholarship that productively adopt an ecological perspective include: Clarke, *Ways of Listening: An Ecological Approach to the Perception of Musical Meaning*; De Souza, "Musical Instruments, Bodies, and Cognition."; Simon Zagorski-Thomas, "The Spectromorphology of Recorded Popular Music: The Shaping of Sonic

affords us the ability to speak and sing (and swallow, etc.); a guitar affords string strumming; a saxophone affords blowing and fingering. None of these objects represent simply inert, abstract sonic entities, even in "passive" perception: they are opportunities for doing. Furthermore, in addition to instrumental affordances—i.e., the body/instrument interactions specific to production of sound on each instrument—I have been arguing in this chapter that most musical timbres can be also be considered affordances of the voice.

As the sonic dimension that tells us the most about the state of our physical environment,[100] timbre is ideally suited to an ecological treatment. In most situations we do not tend to hear timbre in the "acoustic world" abstracted from the source mechanics that give rise to the sensation, especially the interaction of excitation (hitting, scraping, etc.), materials (hard, soft, etc.), and the structure of the material excited (steel, nylon, conical, straight, etc.).[101] (Harmonic frequency distributions, by contrast, are commonly abstracted, at least in the West, to discrete pitch classes.) To our earlier definitions of timbre, therefore, we can add Stephen Handel's ecologically informed description: "[timbre refers to the] perceptual qualities of objects and events; that is, 'what it sounds like.'"[102] Concatenating these definitions, timbre turns out to be a verb: things sound like what we might do with them, the perceptual consequence of affordances for physical action, prior to their identity as nouns (saxophone, violin) or adjectives (sweet, crunchy). If timbre is the audible consequence of action, then its perception offers an invitation for embod-

---

Cartoons through Record Production," in *The Relentless Pursuit of Tone: Timbre in Popular Music*, ed. Robert Fink, Mindy LaTour O'Brien, and Zachary Wallmark (New York: Oxford University Press, forthcoming).

[100] Robert A. Butler, "The Relative Influence of Pitch and Timbre on the Apparent Location of Sound in the Median Sagittal Plane," *Perception and Psychophysics* 14, no. 2 (1973).

[101] See Bruno L. Giordano and Stephen McAdams, "Sound Source Mechanics and Musical Timbre Perception: Evidence from Previous Studies," *Music Perception* 28, no. 2 (2010).

[102] Stephen Handel, "Timbre Perception and Auditory Object Identification," in *Hearing*, ed. Brian C. J. Moore (San Diego: Academic Press, 1995), 426.

ied (re)enaction based both in individual experience (e.g., training on a musical instrument) and our common human embodiment.

Applying an ecological approach to the "sound of arousal," moreover, helps to explain why noisy timbre is so directly charged with affective, bodily meaning. Timbres that are perceived as "noisy" are somatically marked: they reflect a body in a heightened state of arousal by straining against the physical invariants that materially circumscribe the voice. If timbre sonifies the affordances of the voice (and our instruments, which can be considered an extension of the voice), then a noisy timbre showcases bodily action in its most effortful form—as action pressing up against the limits of the flesh. It can thus function as a timbral trigger for particular negatively-valenced reactions.[103]

Leman proposes that "perception can be seen as the creation of a motor image of the world that is based on sensory information."[104] In effect, perception is simulated action. From this perspective, what sort of motor image would "harsh voice" suggest? What does the timbre of arousal do to those who hear it? The concept of affordances may be too weak, indeed, for act–sound coupling. Cox suggests instead that music *demands* mimetic motor engagement; it does not simply afford us with opportunities for action, but rather thrusts the sensation of action upon us.[105] Different timbral qualities demand the listener to mimetically take up the actions they imply; noisy timbre thus presents us with an extreme, perhaps limiting, case. Sonifying the outer bounds of physical exertion, noise demands that same exertion from the listener. Put more strongly, noise perceptually flirts with the thresholds of the body.

---

[103] I explore the implications of this in the next chapter with the evidence-based suggestion that we can localize such negative reactions to noisy timbral triggers in limbic regions, specifically the amygdala.

[104] Leman, *Embodied Music Cognition and Mediation Technology*.

[105] Cox, "Embodying Music: Principles of the Mimetic Hypothesis," 48.

We will take up these points again in the case studies, but let me present here a preliminary sketch of the nuts and bolts of the "translational" process between acts and sounds in the brain. The most significant advance in understanding the neural basis for act–sound coupling comes from studies of the human mirror neuron system. Discovered in the brains of macaque monkeys in the mid-1990s and in humans shortly thereafter, mirror neurons cortically represent the perceived physical actions of one individual in the sensorimotor regions of another.[106] Though originally discovered only for vision, by the early 2000s mirroring mechanisms were extended to auditory perception as well; action-perception integration is involved not only in seeing physical actions, but in hearing them.[107] Allowing multiple individuals to experience the same motor events, mirror neurons have been theorized to play a key role in many essential human behaviors, including imitation, language, social cognition, and empathy.[108]

---

[106] Details about the mirror neuron system have been well rehearsed elsewhere, and a thorough review would obviously take us far afield of the present chapter. A helpful general overview can be found in two books: Giacomo Rizzolatti and Corrado Sinigaglia, *Mirrors in the Brain: How Our Minds Share Actions and Emotions*, trans. Frances Anderson (Oxford: Oxford University Press, 2008); Iacoboni, *Mirroring People: The Science of Empathy and How We Connect with Others*. For a shorter review, see Giacomo Rizzolatti and Laila Craighero, "The Mirror-Neuron System," *Annual Review of Neuroscience* 27(2004). The most significant early studies documenting the existence of monkey and human mirror neurons include G. di Pellegrino *et al.*, "Understanding Motor Events: A Neurophysiological Study," *Experimental Brain Research* 91(1992); Giacomo Rizzolatti *et al.*, "Premotor Cortex and the Recognition of Motor Actions," *Cognitive Brain Research* 3(1996); Vittorio Gallese *et al.*, "Action Recognition in the Premotor Cortex," *Brain* 119, no. 2 (1996); Marco Iacoboni *et al.*, "Cortical Mechanisms of Human Imitation," *Science* 286(1999); M.A. Umiltà *et al.*, "I Know What You are Doing: A Neurophysiological Study," *Neuron* 31(2001). The first direct evidence at the single-neuron level for human mirror neurons came in 2010 with a paper by Roy Mukamel *et al.*, "Single-Neuron Responses in Humans during Execution and Observation of Actions," *Current Biology* 20, no. 8 (2010).

[107] The essential papers here are Evelyne Kohler *et al.*, "Hearing Sounds, Understanding Actions: Action Representation in Mirror Neurons," *Science* 297, no. 846 (2002); Fadiga *et al.*, "Speech Listening Specifically Modulates the Excitability of Tongue Muscles: A TMS Study."; Christian Keysers *et al.*, "Audiovisual Mirror Neurons and Action Recognition," *Experimental Brain Research* 153(2003); Lisa Aziz-Zadeh *et al.*, "Left Hemisphere Motor Facilitation in Response to Manual Action Sounds," *European Journal of Neuroscience* 19(2004); Valeria Gazzola, Lisa Aziz-Zadeh, and Christian Keysers, "Empathy and the Somatotopic Auditory Mirror System in Humans," *Current Biology* 16(2006). For two comprehensive review articles on audiomotor mirroring, see Robert J. Zatorre, Joyce L. Chen, and Virginia B. Penhune, "When the Brain Plays Music: Auditory-Motor Interactions in Music Perception and Production," *Nature Reviews: Neuroscience* 8(2007); Salvatore Aglioti and Mariella Pazzaglia, "Representing Actions Through Their Sound," *Experimental Brain Research* 206(2010).

[108] For a wide-ranging discussion of the philosophical implications of human mirror neurons written for a lay audience, see Iacoboni, *Mirroring People: The Science of Empathy and How We Connect with Others*.

To these we can add music. Mirror neurons help us comprehend musical sound by way of *motor resonance,* allowing for an ecologically grounded association between bodily acts and the sounds they engender. (The term "motor resonance," for our purposes, is synonymous with activity in sensorimotor regions consist with mirror neuron involvement.) Neuroscientists Istvan Molnar-Szakacs and Katie Overy have been perhaps the most cogent theorists of music and mirror neurons in recent years. They summarize:

> […] Music has always been associated with motor activity. From drumming to singing to virtuosic sitar playing, the production of music involves well-coordinated motor actions that produce the physical vibrations of sound. The experience of music thus involves the perception of purposeful, intentional and organized sequences of motor acts as the cause of temporally synchronous auditory information. Thus, according to the simulation mechanism implemented by the human mirror neuron system, a similar or equivalent motor network is engaged by someone listening to singing/drumming as the motor network engaged by the actual singer/drummer; from the large-scale movement of different notes to the tiny, subtle movements of different timbres. This allows for co-representation of the musical experience, emerging out of the shared and temporally synchronous recruitment of similar neural mechanisms in the sender and the perceiver of the musical message.[109]

The authors extend this theoretical model to the question of affect generation, arguing that the emotional power of music is related precisely to the motor-driven "co-representation of the musical experience." Their Shared Affective Motion Experience (SAME) model holds that the motional qualities of music, via the mirror neuron system, lead to motor resonance in the listener consistent with the affective intentions of the performer.[110] This affectively colored motor activity is then translated into felt emotion through relay to the limbic system. Overy and Molnar-

---

[109] Istvan Molnar-Szakacs and Katie Overy, "Music and Mirror Neurons: From Motion to 'E'motion," *Social Cognitive and Affective Neuroscience* 1, no. 3 (2006).

[110] Katie Overy and Istvan Molnar-Szakacs, "Being Together in Time: Musical Experience and the Mirror Neuron System," *Music Perception* 26, no. 5 (2009).

Szakacs conclude: "the expressive dynamics of heard sound gestures can be interpreted in terms of the expressive dynamics of personal vocal and physical gestures."[111]

Although the precise relationship between music perception and mirror neurons is still far from clear—Molnar-Szakacs and Overy have theorized the relationship but not empirically tested it—researchers are beginning gradually to flesh out some of the more pertinent details. Motor resonance to rhythm has been the most thoroughly explored,[112] and a number of studies extend the mirroring properties of the motor cortex to a variety of additional musical contexts.[113] A crucial point here is the *bidirectional* character of act–sound coupling: hearing musical sound can activate motor regions of the brain, but the process also runs the other way—performing instrument-specific sound generating actions, such as fingering a violin or pressing the keys on a piano (or simply imagining these actions), can generate activity in the *auditory* cortex.[114] Movement can trigger imagined sound just as imagined sound can trigger covert movement. Hearing is a motor process as well as an auditory one.

Since none of the studies above isolate timbre as an experimental variable, it is impossible to disambiguate its specific role in neural motor resonance. But shifting the frame of reference in

---

[111] Ibid., 492.

[112] See M. Popescu, A. Otsuka, and A.A. Ioannides, "Dynamics of brain activity in motor and frontal cortical areas during music listening: a magnetoencephalographic study," *Neuroimage* 21, no. 4 (2004); Jessica A. Grahn, "Neural Mechanisms of Rhythm Perception: Current Findings and Future Perspectives," *Topics in Cognitive Science* 4(2012).

[113] For example, Drost *et al.* (2005) and Lahav *et al.* (2007) investigated the neuroplasticity of act-sound associations; Hickock *et al.* (2003) explored audio-motor similarities between speech and music perception; Buccino *et al.* (2004) and Calvo-Merino *et al.* (2005) demonstrated mirroring mechanisms behind music-related actions, e.g., fingering guitar chords and dancing; and Wallmark and Iacoboni (2012) found motor resonance in response to listening to familiar popular music excerpts.

[114] Imaging studies by Haueisen and Knösche (2001), Bangert and Altenmüller (2001), Haslinger et al. (2005), and Bangert et al. (2006) demonstrated that experienced pianists exhibit activity in motor regions when hearing piano music. Conversely, in an experiment by Lotze et al. (2003), experienced violinists showed activity in the auditory cortex when simply fingering a familiar Mozart concerto (no sound); and using a behavioral task, Repp and Knoblich (2007) found that "manual actions can make pianists *hear* tones differently" (p. 7). Taken together, then, it seems that instrument-specific movement can trigger imagined sound, just as instrumental sound can trigger covert movement in the form of motor resonance.

an ecological direction, to look at the material sources of different musical sounds, we can confirm that motor resonance is sensitive to timbre. Three representative studies illustrate this point. Pantev *et al.* (2001) used magnetoencephalography (MEG) to find that musicians responded with greater cortical activity to their chosen instrument (violin or trumpet) compared to another instrument or a sine tone.[115] Though the MEG readings were taken over the auditory (not motor) cortex, this result is suggestive nonetheless when one considers that auditory and motor areas tend to co-activate in timbre perception.[116] Using a clever behavioral task, Drost, Rieger, and Prinz (2007) found that pianists and guitarists showed "action–effect" coupling only to their instrument of training.[117] And Margulis *et al.* (2009) reached a similar conclusion in an fMRI study comparing violinists' and flutists' perception of music on their own instruments, which indicated greater syntax-processing and motor involvement when subjects listened to the instruments they play.[118] Taken together, these results suggest that components of musical timbre can trigger sensorimotor activity consistent with the background of the hearer.

But what about the voice, and (thanks to "Spencer's law") the vocal aspects of instrument timbre? It is still uncertain to what extent instrumental timbre evokes motor response consistent

---

[115] Christo Pantev *et al.*, "Timbre-specific Enhancement of Auditory Cortical Representations in Musicians," *Neuroreport* 12, no. 1 (2001).

[116] See Marc Bangert and E. Altenmüller, "Mapping Perception to Action in Piano Practice: A Longitudinal DC-EEG Study," *BMC Neuroscience* 4(2001).

[117] Ulrich C. Drost, Martina Rieger, and Wolfgang Prinz, "Instrument Specificity in Experienced Musicians," *The Quarterly Journal of Experimental Psychology* 60, no. 4 (2007). The researchers had pianists and guitarists play chords on their instruments when displayed on a screen (notated major and minor triads), with reaction time measured. Additionally, auditory stimuli of the chords—both congruent (same instrument) and incongruent (different instrument)—were given through headphones; subjects were instructed to ignore them and focus entirely on the visual stimuli. Results indicated that musicians were only distracted by the timbre of their own instrument, not chords played on other instruments.

[118] Elizabeth Hellmuth Margulis *et al.*, "Selective Neurophysiologic Responses to Music in Instrumentalists with Different Listening Biographies," *Human Brain Mapping* 30(2009). The researchers had violists and flautists listen to Bach solo partitas for both instruments while being scanning. Violinists showed greater activity while listening to the violin, and flautists showed the opposite.

with vocal production—this hypothesis, to my knowledge, has never been explicitly tested.[119] However, there is at least some relevant evidence. Warren *et al.* (2006) scanned subjects as they listened to nonverbal emotional vocalizations, finding activity in premotor areas consistent with mirror neurons.[120] Motor resonance to vocal timbre, moreover, involved more than just covert imitation: it modulated according to the affective qualities of the vocalization, with positive valence and high arousal driving motor activity.

This is a highly suggestive finding: for one, it demonstrates a motor component in listening to vocal timbre, and since nonverbal sounds were used, possibly non-vocal sound as well. But more than this, it shows that act–sound couplings are contingent upon the specific bodily meanings (affects) being conveyed. Motor resonance to vocal timbre does not reflect merely biomechanical reaction, but also emotional appraisal: voices expressing the positive emotions of "triumph" and "amusement" drive greater activity in motor areas than "fear" and "disgust." We mirror actions that are marked with positive somatic implications.

Furthermore, the motor activity demonstrated in this study overlaps with regions responsible for making and perceiving emotional *facial* expressions: hearing an emotionally-valenced voice activates the same areas of the brain responsible for expressing that emotion facially (and seeing those facial expressions in another). Consistent with the SAME model, motor resonance to timbre is driven by "shared affective motion" communicated intercorporeally between people. As the authors conclude: "The greater propensity for positive-valence communications to automatically activate motor representations may be a crucial component in the formation of empathic

---

[119] As the primary literature is somewhat lacking on this question, I will return to this topic in the next chapter with results from my experiments, which respond in the affirmative.

[120] Jane E. Warren *et al.*, "Positive Emotions Preferentially Engage an Auditory-Motor 'Mirror' System," *The Journal of Neuroscience* 26, no. 50 (2006).

responses."[121] This implies that listening to vocal timbre is a social act that relies on the same neural networks responsible for interpreting the emotions of others.

It is not my intention to follow the rich implications of this conclusion here, but it is important to mention in conclusion nonetheless, as it illustrates a central point. Act–sound coupling, in its broadest implication, not only helps us to navigate our environment; it also facilitates social interaction, creating a context of shared experience.[122] At the level of timbre, moreover, it implies that different qualities of sound evoke different motor responses according to their social associations. Since to hear another's timbre is, in a sense, to make that timbre, it matters tremendously *what* the bodily actions mean and *who* is making them. The moment of timbre perception is thus both ethical and political.

We are presented with perhaps conflicting explanations for how these two levels interact (the bodily actions themselves and the broader affective and social context for those actions). On the one hand, I have been arguing that the "sound of arousal" drives greater motor resonance due to act–sound coupling of high-exertion bodily actions. In this telling, we respond to the demand of another's effortful sound *because* the sound conveys effort. This would represent a pure reaction effect. But on the other hand, Warren *et al.*'s results point to another interpretation: motor resonance would be driven not by perceived physical effort *per se,* but by the type of expression. This response, which reflects more of an appraisal than a reaction, is modulated both by affective valence (positive or negative) and social empathy (feeling rapport with the individuals making the sound).[123] We will follow up on this unresolved issue in the next chapter.

---

[121] Ibid., 13074.

[122] We should note that mirror neurons seem to be specific to highly social primates.

[123] The role of interpersonal rapport in the unconscious imitation of others is well documented. See Tanya Chartrand and John A. Bargh, "The Chameleon Effect: The Perception-Behavior Link and Social Interaction," *Journal of Personality and Social Psychology* 76, no. 6 (1999).

Let me close by trying to weave together some key points from this section. Sounds in most situations of listening are not just physical waveforms but indicate the presence of ecologically-relevant objects and events. When a sound comes from another person, moreover, a sound is a social stimulus that provides clues about the emotional state and intentions of the other. Due to act–sound coupling at the neurophysiological level, hearing the sounds of another forces the hearer to respond in kind, covertly mirroring the bodily acts involved in sound production. When the timbre in question is the "sound of arousal," however, the demands on the listener are greater than usual, and the reaction is particularly vehement and visceral. Though such inherent reactions influence how we ultimately make sense of what we are hearing, the motor demands of noisy timbre do not always translate in predicable ways into appraisal.

Returning to Schoenberg's epigraph, we might agree that listeners react to somatically marked timbre with the emotions; but judgment (or appraisal) is not equivalent to this reaction. There is room for considerable slippage between reaction and appraisal, although the latter is predicated on the former. The threshold between the two—like the physical thresholds of bodily ability, the acoustic thresholds between "pure" and "noisy" timbre, and the perceptual/symbolic thresholds between music and noise—is a dynamic and malleable field indeed, one that plays out very differently in diverse contexts of musicking.

Conclusion

In this chapter I have fleshed out a motor theory of timbre perception by focusing primarily on what the empirical literature tells us about the role of human embodiment in the production and perception of timbre. Timbre, in this revised, ecological account, is active in character: it is both *something we do* and, thanks to act–sound sensorimotor coupling, *something done to us.* I have argued that listening to timbre can involve motor resonance on the part of the listener, and more specifically, motor resonance of the *voice* (subvocalization). Though the line separating perception of vocal and non-vocal timbres (including musical instruments) is far from crisp, I contend that mechanisms of vocal-based motor resonance play a role in musical timbre processing in general, and further argue that this motor quality may provide part of the explanation for its affective powers. The intimate connection between bodily action, affect, sound production, and perception can be seen clearest in acts of high-exertion, high-arousal vocalizations (and their instrumental correlates). These moments—and the "noisy timbre" they produce—place great motor demands on the listener, triggering negatively-valenced bodily reactions consistent with the affective/bodily states involved in production.

But read through the framework of appraisal theory, reactions are not the same as emotions; they influence affective response, but that response is ultimately the product of appraisal, which is a considerably more complex and socially-determined process. The motor exertion required for a particular timbral quality drives motor resonance in the listener, then, but so too do the affective, behavioral, social, cultural, political, biographical, situational, symbolic, and aesthetic associations of a particular timbre, matters that far exceed the biological. A generalized empirical account of embodied timbre perception and cognition, such as that presented here, therefore, can only get us so far in our efforts to understand the affective powers of timbre. *Acts*, *sounds*, and *perceptions* ground the higher levels of cognition—that is, they provide the material basis for timbre—but this grounding does not capture their full phenomenological richness. This remains the work for the central chapters of the dissertation.

Furthermore, for all the literature can tell us, I remain unsatisfied with the empirical account sketched here. Timbre is an elusive percept to pin down, and I have had to approach this body of evidence somewhat indirectly, with a fair dose of conjecture and liberal interpretation. In the next chapter, I aim to fill some of these holes with results from three perceptual studies. It is my hope that Chapter 2 will sharpen the focus of the motor theory of timbre introduced here, while also presenting us with new enigmas to consider as we move forward.

Chapter 2

Behavioral and Neuroimaging Evidence for Embodied Timbre Cognition

Introduction

To this point, I have presented evidence for the embodied nature of timbre perception and cognition by tracing a line of thinking that leads to the following four assertions: (1) emotion is embodied, and affect is sonified through the voice; (2) high physical and affective arousal leave acoustic traces in vocal timbre in the form of *brightness*, *noise*, and *roughness*; (3) embodied experience allows us to glean pre-attentively the meaning of timbral characteristics via motor resonance; and (4) the timbral behavior of the affectively marked voice provides a template for the way we hear and make sense of all musical timbre.

But how do acts and sounds congeal in the moment of perception to generate *meaning*? To be sure, the account above is light on specifics, particularly in regard to the appraisal-forming process. Despite rich sources of suggestive evidence, we are confronted with a dearth of direct, finely-tuned corroboration in the empirical literature. This chapter will narrow the focus by investigating the material specifics of embodied timbre perception and cognition. To that end, I have conducted three original experiments using behavioral and neuroimaging methods designed to test the hypotheses advanced in Chapter 1.

The experiments presented in this chapter are not without precedent, despite the paucity of timbre research in music studies. Scientific work on the topic tends to fall into one of three camps: acoustic, psychoacoustic, or neurophysiological studies. The acoustic approach to musical timbre deals only in the physical dimensions of sound, often using signal processing techniques. The psychoacoustics literature, discussed above, has produced many detailed accounts connecting acoustic features to perception, but rarely in relation to "less musical" qualities outside the context of the Western orchestra. Moreover, issues of affect and meaning fall beyond the purview of most purely empirical studies of timbre. But there are some notable exceptions: recent publications from researchers at leading European systematic musicology labs—including Leman (2008), Ferrer (2009, 2011), Hailstone *et al.* (2009), Alluri and Toiviainen (2010, 2012), Halmrast *et al.* (2010), and Eerola, Ferrer, and Alluri (2012)—venture into new terrain for timbre research, placing timbre perception into dialogue with theories of affect, embodied cognition, and ecological psychology.[1] These studies most directly inform the investigations in this chapter.

Additionally, a number of brain studies over the last fifteen years have revealed important information regarding the neural underpinnings of timbre perception, and the present chapter builds on this body of research. A few of the more relevant conclusions from this literature (in addition to those reviewed in Chapter 1) can be very briefly summarized: Studies by Tervaniemi, Winkler, and Näätänen (1997), Platel *et al.* (1997), Goydke *et al.* (2004), and Koelsch and Siebel

---

[1] Leman, *Embodied Music Cognition and Mediation Technology*; Ferrer, "Embodied Cognition Applied to Timbre and Music Appreciation: Theoretical Foundation."; Julia C. Hailstone *et al.*, "It's Not What You Play, It's How You Play It: Timbre Affects Perception of Emotion in Music," *The Quarterly Journal of Experimental Psychology* 62, no. 11 (2009); Vinoo Alluri and Petri Toiviainen, "Exploring Perceptual and Acoustical Correlates of Polyphonic Timbre," *Music Perception* 27, no. 3 (2010); Tor Halmrast *et al.*, "Gesture and Timbre," in *Musical Gestures: Sound, Movement, and Meaning*, ed. Rolf Inge Godøy and Marc Leman (New York: Routledge, 2010); Ferrer, "Timbral Environments: An Ecological Approach to the Cognition of Timbre."; Tuomas Eerola, Rafael Ferrer, and Vinoo Alluri, "Timbre and Affect Dimensions: Evidence from Affect and Similarity Ratings and Acoustic Correlates of Isolated Instrument Sounds," *Music Perception* 30, no. 1 (2012); Vinoo Alluri and Petri Toiviainen, "Effect of Enculturation on the Semantic and Acoustic Correlates of Polyphonic Timbre," ibid.29, no. 3.

(2005) demonstrated that timbre is processed *quickly* and *automatically*, with little contribution from higher cognitive faculties and a reaction threshold much lower than pitch and rhythm processing.[2] This extreme rapidity has been theorized to play a significant role in our ability to make split-second affective evaluations of music.[3] Samson *et al.* (1994, 1997), Menon *et al.* (2002), Caclin *et al.* (2006, 2008), and Alluri *et al.* (2012) uncovered great neuroanatomical specificity to the mechanisms involved in timbre processing, finding, for instance, that temporal features are left-lateralized while spectral features are right-lateralized.[4] Kumar *et al.* (2012) demonstrated the primacy of high-frequency energy in evaluating the affective valence of timbre.[5] And Patel (1999) and Belin (2006) showed a functional connection between timbre and voice processing (both linguistic and non-linguistic).[6]

This impressive though relative small body of work lays the foundations for the questions I am asking here, but it hardly exhausts them. How does timbre contribute to musical meaning?

---

[2] Mari Tervaniemi, I. Winkler, and R. Näätänen, "Pre-Attentive Categorization of Sounds by Timbre as Revealed by Event-Related Potentials," *Neuroreport* 8(1997); Hervé Platel *et al.*, "The Structural Components of Music Perception: A Functional Anatomical Study," *Brain* 120(1997); Katja N. Goydke *et al.*, "Changes in Emotional Tone and Instrumental Timbre are Reflected by the Mismatch Negativity," *Cognitive Brain Research* 21(2004); Stefan Koelsch and Walter A. Siebel, "Towards a Neural Basis of Music Perception," *Trends in Cognitive Sciences* 9, no. 12 (2005).

[3] Peretz, Gagnon, and Bouchard, "Music and Emotion: Perceptual Determinants, Immediacy, and Isolation after Brain Damage."; Schellenberg, Iverson, and McKinnon, "Name that Tune: Identifying Popular Recordings from Brief Excerpts."; Bigand *et al.*, "Multidimensional Scaling of Emotional Responses to Music: The Effect of Musical Expertise and of the Duration of the Excerpts."

[4] Severine Samson and Robert J. Zatorre, "Contribution of the Right Temporal Lobe to Musical Timbre Discrimination," *Neuropsychologia* 32, no. 2 (1994); Severine Samson, Robert J. Zatorre, and James O. Ramsay, "Multidimensional Scaling of Synthetic Musical Timbre: Perception of Spectral and Temporal Characteristics," *Canadian Journal of Experimental Psychology* 51, no. 4 (1997); Menon *et al.*, "Neural Correlates of Timbre Change in Harmonic Sounds."; Anne Caclin *et al.*, "Separate Neural Processing of Timbre Dimensions in Auditory Sensory Memory," *Journal of Cognitive Neuroscience* 18, no. 12 (2006); Anne Caclin *et al.*, "Interactive Processing of Timbre Dimensions: An Exploration with Event-related Potentials," ibid.20, no. 1 (2008); Vinoo Alluri *et al.*, "Large-scale Brain Networks Emerge from Dynamic Processing of Musical Timbre, Key, and Rhythm," *Neuroimage* 59(2012).

[5] Sukhbinder Kumar *et al.*, "Features versus Feelings: Dissociable Representations of the Acoustic Features and Valence of Aversive Sounds," *The Journal of Neuroscience* 32, no. 41 (2012).

[6] Patel, *Music, Language, and the Brain*; Pascal Belin, "Voice Processing in Human and Non-Human Primates," *Philosophical Transactions of the Royal Society B* 361(2006).

How does it convey affect? How is perception of it "embodied"? What role does *noise*—acoustically, perceptually, epistemologically—play in timbre perception? The behavioral and neurophysiology literatures avoid these (admittedly) broad and abstract questions. A next step in this project, then, is to explore the phenomenon of embodied timbre cognition from the vantage of the research questions specific to this dissertation; and to do so, I will need new empirical data. The three pilot experiments reported here—two using techniques from behavioral psychology, one using neuroimaging (fMRI)—aim to contribute to our evolving understanding of how reactions and appraisals of timbre interact in "noisy" situations of listening, both at the behavioral and the neural levels.

Findings from the experiments converge on a few key conclusions, which will form a cornerstone of the musical analyses and interpretations of Part II. First, it appears that appraisals of timbre are structured by an underlying perceptual binarism, what I call the *appraisal factor*: they are either "good" or "bad," with a series of other perceptual attributes all lining up behind each categorical evaluation. Such judgments occur very quickly, and are governed just as much by pre-attentive reactive mechanisms as conscious cognition. Second, the appraisal factor is consistent in both "acoustic" and "perceived worlds" (Fales 2002); parsing of timbre into "good" and "bad" types relies on the interaction of acoustic qualities and their affective and bodily implications. Third, the perception of physical exertion in a musical timbre is related to the actual exertion required in a sound's production, but it can also be driven purely by the "sound of arousal" irrespective of actual bodily effort. In short, the impression of exertion—and all its accompanying affective implications—can be faked. Fourth, appraisal of timbre involves conflicting activation patterns in two motor regions of the brain: "good" appraisals preferentially call upon activity in specifically *vocal* motor areas, while "bad" appraisals engage more general-purpose motor areas. And finally, motoric responses to musical timbre appear to share a functional connection with emotional processing (via the limbic system). These results, taken together, build on Chapter 1 to provide evidence both for the embodied (i.e., motor mimetic) nature of timbre perception, including the role of motor resonance in appraisal; the primacy of the voice in the perception of all musical timbre; and the connection between embodiment and affect in the process of listening to timbre.

But before delving into the findings, I want to lay out a few important caveats. For the sake of readability I will be glossing over most of the methodological details and essentially all the analysis that went into the design and execution of these experiments. The main goal of this chapter is to cut straight to the results; but of course, any results are predicated upon the particulars of experimental design and data analysis. The reader is therefore strongly encouraged to review the associated Appendices for a complete account of everything reported here. A comprehensive understanding of these scientific details, however, is not required to proceed; the discussion should stand alone without the need to peep "under the hood." I am also aware of the alienation and, in Judith Becker's felicitous term, "aesthetic distaste" that can easily accompany the reading of scientific writing by humanistic audiences—all those "Greek letters and numbers that often pepper the prose," as Huron notes—and have thus decided to keep the proceeding pages completely letter-and-number free, and with only minimal graphs and brain images.[7] Readers will find all accompanying tables and graphs in the Appendices.

One final note: The Appendices are meant to function as a parallel text for readers coming to this dissertation from the sciences. But in keeping with my translational aims, they are structured to be accessible to humanistic audiences. To facilitate this translation, the reader may refer to the non-technical Glossary for explanations of the major statistical concepts used to

---

[7] Becker, *Deep Listeners: Music, Emotion, and Trancing*, 7; David Huron, "The New Empiricism: Systematic Musicology in a Postmodern Age" (paper presented at the Ernest Bloch Lectures, University of California, Berkeley, 1999), 16.

ground these arguments, as well as for some broad ruminations on the epistemological challenges and opportunities that come along with synthesizing humanistic and empirical methods of music research. The Glossary should serve as an enabling "key" to the Appendices.

Variables: Bodily Exertion, Affective Valence, Semantics, and Emotion Conveyed

The experiments were framed by a number of interrelated research questions, including the following: (1) Is the bodily exertion that goes into the production of musical timbre consciously *perceived* as such by the listener? (2) Is perceived exertion related to perceived "noisiness" (or p-noise)? (3) Do certain "noisy" acoustic qualities (or a-noise), particularly *brightness, noise*, and *roughness*, convey a sense of heightened exertion even if little actual exertion is involved in their production? (4) Is *motor resonance*—the covert rehearsal of movement in response to musical sound—involved in timbre perception, and if so, (5) is it related to a timbre's degree of perceived exertion, liking/disliking, a-noise, p-noise, and affective connotations? (6) Does the voice play a preferential role in motor mimetic timbre perception? (7) *How* does timbre convey affective qualities? And (8) What are the cognitive mechanisms involved in *appraising* timbre?

In order to address these questions empirically, I opted to focus on the perceptual attributes of four main variables: (1) *bodily exertion*, (2) *affective valence*, (3) *semantic associations*, and (4) *"emotion conveyed"* by timbre. To explain each of these variables in turn: as argued previously, possibly the most fundamental perceptual scale for timbre reaction and appraisal consists of the relative degree of *bodily exertion* implied in the production of a sound, both vocally and through other physical means. For our purposes, therefore, "embodiment" is operationalized as a linear axis of perceived low to high exertion and arousal. Timbre is affected by *all* states of the body, the low-arousal states of boredom, depression, and sadness just as much as anger, fear, and joy. However, since *exertion* most explicitly manifests qualities of a-noise in the voice, it is this aspect of embodied experience that most clearly frustrates the threshold between "musical" and "noisy" timbre.[8]

Next, the experiments explored the *affective valence* of different timbral qualities (and again, timbres with high levels of a-noise first and foremost). In cognitive psychology, the term "affective valence" refers to the intrinsic attractiveness or aversiveness of a stimulus or an emotional state; for example, the emotions of fear and anger are "negatively valenced" while happiness is "positively valenced." In music studies, this dimension of perception has been shown to correlate strongly with *preference*: "strongly liking" and "strongly disliking" are opposite poles of the affective valence spectrum.[9] This dimension of timbre perception is similarly "embodied," as argued in the last chapter, since the perceptual systems and cognitive mechanisms that mediate affective response are intimately connected with organism/environment interaction and changing body states.[10] The *affective valence* scale aims to capture how strongly listeners like or dislike a particular timbral quality.

I also aim to explore how timbre is conceptualized semantically, a topic that will be taken up in greater depth in the next chapter. Many adjectives and metaphors are used to describe qualities of sound, but I selected two for this study: "bright" and "noisy." "Brightness," as discussed earlier, is closely connected in the psychoacoustics literature with spectral centroid (measured here using two different computational approaches). This quality, and the closely related though not identical cognates "sharpness" and "nasality," have been found to be the most per-

---

[8] In addition to "bodily exertion," I performed another experiment using the same design with the word "intensity" included as a ratings option (*not intense-very intense*). I found that this expression correlated so closely with both *bodily exertion* and "noisiness" as to make it a highly redundant measure; thus, I opted not to report this experiment here. For more, see Appendix 4b.

[9] Eerola, Ferrer, and Alluri, "Timbre and Affect Dimensions: Evidence from Affect and Similarity Ratings and Acoustic Correlates of Isolated Instrument Sounds."

[10] James, *The Principles of Psychology*, 2; Damasio, *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*.

ceptually salient feature of timbre in numerous studies.[11] "Noisiness" is addressed here for two main reasons: first, despite its utter experiential ubiquity, little work has been done exploring the perceptual correlates of impressions of "noisiness" in musical timbre; second, and for the same reason, I wanted to investigate what specific acoustic properties are present in this percept. The goal in these studies was to put the "acoustic" and the "perceived worlds" into conversation to better understand how a-noise (acoustically noisy timbre) and p-noise (the perception of "noisiness") interact in the experience of listening.

And finally, the experiments set out to determine how timbre—as defined through the properties above—might convey emotion. Research into music and emotion is a burgeoning field with many competing methodologies,[12] but for the present purpose I elected to use the five "primary" emotion categories: Happiness, Sadness, Anger, Fear, and Tenderness.[13] This is not a measurement of emotion *induction*, or what emotion a timbre causes the listener to feel, but rather a simple evaluation of which primary category best describes the "emotion conveyed" by each timbre. Table 2.1 summarizes all the behavioral and acoustical variables explored in the experiments.

---

[11] For a sampling, see von Bismarck, "Timbre of Steady Sounds: A Factorial Investigation of Its Verbal Attributes."; Grey, "Multidimensional Perceptual Scaling of Musical Timbres."; Kendall and Carterette, "Verbal Attributes of Simultaneous Wind Instrument Timbres: I. von Bismarck's Adjectives."; "Verbal Attributes of Simultaneous Wind Instrument Timbres: II. Adjectives Induced from Piston's 'Orchestration."; McAdams *et al.*, "Perceptual Scaling of Synthesized Musical Timbres: Common Dimensions, Specificities, and Latent Subject Classes."; Schubert and Wolfe, "Does Timbral Brightness Scale with Frequency and Spectral Centroid?." The most thorough review of psychoacoustic research on this topic to date, to my knowledge, remains Hajda *et al.*, "Methodological Issues in Timbre Research."

[12] For a review of this vast and growing literature, see Tuomas Eerola and Jonna K. Vuoskoski, "A Review of Music and Emotion Studies: Approaches, Emotion Models, and Stimuli," *Music Perception* 30, no. 3 (2013).

[13] "Tenderness" is not one of the five primary emotions—this honor goes to "disgust"—but is often used in music studies as a substitute for the latter. These five categories are used in a number of major music psychology studies, and have been shown to correspond to the emotions that performers on a range of instruments can consistently convey to listeners in brief improvisations. For more see Alf Gabrielsson and Patrik N. Juslin, "Emotional Expression in Music Performance: Between the Performer's Intention and the Listener's Experience," *Psychology of Music* 24, no. 68 (1996); Patrik N. Juslin, "Communicating Emotion in Music Performance: A Review and Theoretical Framework," in *Music and Emotion: Theory and Research*, ed. Patrik N. Juslin and John A. Sloboda (Oxford and New York: Oxford University Press, 2001); Carol L. Krumhansl, "Plink: 'Thin Slices' of Music," *Music Perception* 27, no. 5 (2010).

TABLE 2.1: Summary of experimental variables

Embodiment	*Affect*	*Semantics*	*Emotion Conveyed*	*Acoustics*
Bodily exertion	Valence (i.e., preference)	"Brightness" "Noisiness"	Happiness Sadness Anger Fear Tenderness	Spectral Centroid-1 Spectral Centroid-2 Inharmonicity Spectral Flatness Zero-Cross Rate Roughness

<u>Experiment 1</u>: Perceptions of Embodiment and Affect in Isolated Timbres [Appendix 1]

In the first experiment, I had thirty-six subjects listen to a range of isolated 2-second instrument and vocal timbres on a computer then provide ratings exploring the variables above.[14] The timbres consisted of four different sound generators—electric guitar, tenor saxophone, Japanese *shakuhachi* flute, and female singing voice, plus two "test" signals (sine wave and white noise)—and each generator was represented by three separate tones scaled for increasing levels of a-noise, for a total of twelve differential timbres and two control timbres. Thus, Timbre 1 was the "low-noise" condition, Timbre 2 was "medium-noise," and Timbre 3 was "high-noise" (e.g., Guitar 1: "clean," Guitar 2: moderate distortion, and Guitar 3: extremely distorted). All timbres were the same pitch (B♭3), and were manually equalized for loudness.

I used these four specific sound generators in order to represent a range of different material means of production, including electro-acoustic (electric guitar), reed (saxophone), air-stream

[14] The "semantic differential" method is a mainstay in behavioral psychology. Using a 0-100 ratings scale, subjects are asked to select where a stimulus sits between two bipolar terms (e.g., *not noisy-noisy*). For details, see the seminal account: C.E. Osgood, G.J. Suci, and P.H. Tannenbaum, *The Measurement of Meaning* (Urbana: University of Illinois Press, 1957).

(shakuhachi), and vocal production. If timbre perception is to be approached through the lens of embodied cognition, as suggested here, the means of production for each timbre are hardly of trivial concern: singing in a raspy, growled voice, for instance, is more effortful than playing a raspy, growled tone on the guitar with the aid of a distortion pedal. (And the guitar, in turn, requires more physical exertion than the synthetic test signals.) Though the acoustic byproduct might have certain similarities, the *act* is entirely different, though it remains to be seen how this difference plays out perceptually. On a pragmatic level, moreover, this particular assortment of experimental signals reappears in the case studies.[15]

Let us move now to some results. First, I wanted to know how all the individual acoustic parameters related to one another. As mentioned before, *brightness, noise,* and *roughness* tend to accompany moments of physical and affective arousal in the voice, though this remains to be seen for other sound sources. Correlating the acoustics of these twelve timbres revealed significant *covariance*; that is, *brightness, noise*, and *roughness* tend to move up and down together regardless of the source. Thus, we can think of the acoustic qualities associated with exertion as a unit: when you get one—whether in vocal, guitar, saxophone, or shakuhachi timbre—you also tend to get the others. Although we cannot generalize out to *all* instruments from these four sound producers, it seems that the "sound of arousal" in the voice applies to many other sound generators as well.

The first order of business in any empirical study is to determine whether the different experimental conditions elicited statistically significant differences. The results here (and for the other experiments) were positive: each timbre prompted a consistent, systematic appraisal in all four variable categories. This means that the data probably reflect a "real" pattern of perception rather than dumb chance. Comparing average ratings for the different timbres, moreover, inter-

[15] The *shakuhachi* was included in this sample set vestigially, as data for a "Chapter 6" that will appear as a separate publication.

esting patterns emerge: first, it is notable that the human voice produced far-and-away the most clearly etched perceptual profile, generating both the greatest range of responses and the most dramatic shifts in mean scores between conditions. The "high-noise" signal for the voice (Timbre 3) is both the *least liked* and the *most "noisy"* of the natural timbres. It would seem that appraisals of this particular sound are more potent and polarized than any of the instrumental sounds tested. Following the principle of act–sound coupling, hearing this harsh, effortful vocal timbre also involves the tacit rehearsal of the same motor acts that go into its production.

Second, turning to perceived *bodily exertion*, one might hypothesize that scores would reflect the actual exertion required for the production of each specific timbral quality. Thus, the noisy vocal and saxophone conditions, for example, would generate higher scores than the noisy guitar and white noise. Interestingly, I found that while this works for the former set of signals—subjects gave noisy vocal and saxophone timbres the highest *exertion* ratings—guitar signals also generated a similar pattern: higher levels of *exertion* were heard in the distorted timbres, even though the physical means of producing guitar distortion do not require increased physical effort. It could be argued that the acoustic qualities under observation somehow "trick" the listener into hearing increased bodily exertion, although no more work is required to produce distorted electric guitar than a clean tone. If the voice is primary in mapping timbral qualities onto embodied experience, then perhaps listeners were responding not to the perceived pick hitting the string, an act involving fingers, steel, and transducers, but rather to what this timbre *might feel like* to produce vocally. Distorted guitar might *simulate* bodily exertion; in Juslin and Laukka's terminology, it may function as a "superexpressive voice."[16] Further, although both test signals were computer-generated, white noise was given a higher *exertion* score than the sine wave, probably (like the gui-

[16] Juslin and Västfjäll, "Emotional Responses to Music: The Need to Consider Underlying Mechanisms."

tar) as a result of the presence of certain acoustic qualities that are often correlated in human experience with higher levels of physical exertion. White noise *sounds more effortful* than the dull sine wave, even though it is not connected with human bodies in any ecological sense, therefore, perhaps owing in part to mechanisms of vocal-based motor resonance.[17] These results can be interpreted to indicate that the *sound* of arousal matters more to perceptions of bodily exertion than the actual thing.

A similar pattern is at play with the saxophone: intensity of added growl drives up *exertion* ratings. In this circumstance, however, the exertion is real and not simulated, though the significant role of the vocal tract in the production of saxophone growls, altissimo screams, and multiphonics blurs the distinction between vocal exertion and other forms of instrument-specific bodily effort—the vocal mimesis here is *actual*, in other words. (This will be explored in depth in Chapter 4.) Listeners' personal backgrounds would be a decisive factor in this distinction: if you know how hard it is to growl on the saxophone because you yourself are a saxophonist, you might experience this timbre with greater motor empathy than a listener with no background on the instrument. (As noted in Chapter 1, this interpretation is supported by extant neurophysiological literature.)[18]

Figure 2.1 summarizes these data by timbre condition rather than instrument. When collapsed across all instrument categories, note that Timbre 3 (the "high-noise" condition) is heard as requiring the most *bodily exertion*, with Timbre 2 following behind. This conforms to the hy-

---

[17] This result resonates in counterintuitive ways with the literature. Numerous neurophysiological studies of both monkeys and humans have found that motor mirror neurons only activate when we hear *action* sounds; the sound of waves, or the wind, do not drive mirroring. This behavioral result would seem to contradict these brain studies. For more see Kohler *et al.*, "Hearing Sounds, Understanding Actions: Action Representation in Mirror Neurons."; Aziz-Zadeh *et al.*, "Left Hemisphere Motor Facilitation in Response to Manual Action Sounds."

[18] Pantev *et al.*, "Timbre-specific Enhancement of Auditory Cortical Representations in Musicians."; Margulis *et al.*, "Selective Neurophysiologic Responses to Music in Instrumentalists with Different Listening Biographies."

pothesis. In the next perceptual quality condition, *affective valence,* we can see the order reversing: subjects tend to find Timbre 1 to be the most positively valenced (i.e., they like it the most), while Timbre 3 received lowest mean ratings. To summarize, high a-noise timbres indicate a greater amount of bodily effort going into their production, and they are viewed more negatively than low a-noise signals. The p-noise condition (perceived "noisiness") is also quite clear, with a spread of over 20 points between Timbre 1 and Timbre 3, with Timbre 2 right about in the middle. As expected, timbres of isolated instruments that are scaled in ecologically valid ways (i.e., without manipulating acoustic parameters systematically) into ordinal levels based on different degrees of a-noise are perceived as such, despite varying means of production and vastly different acoustic profiles.

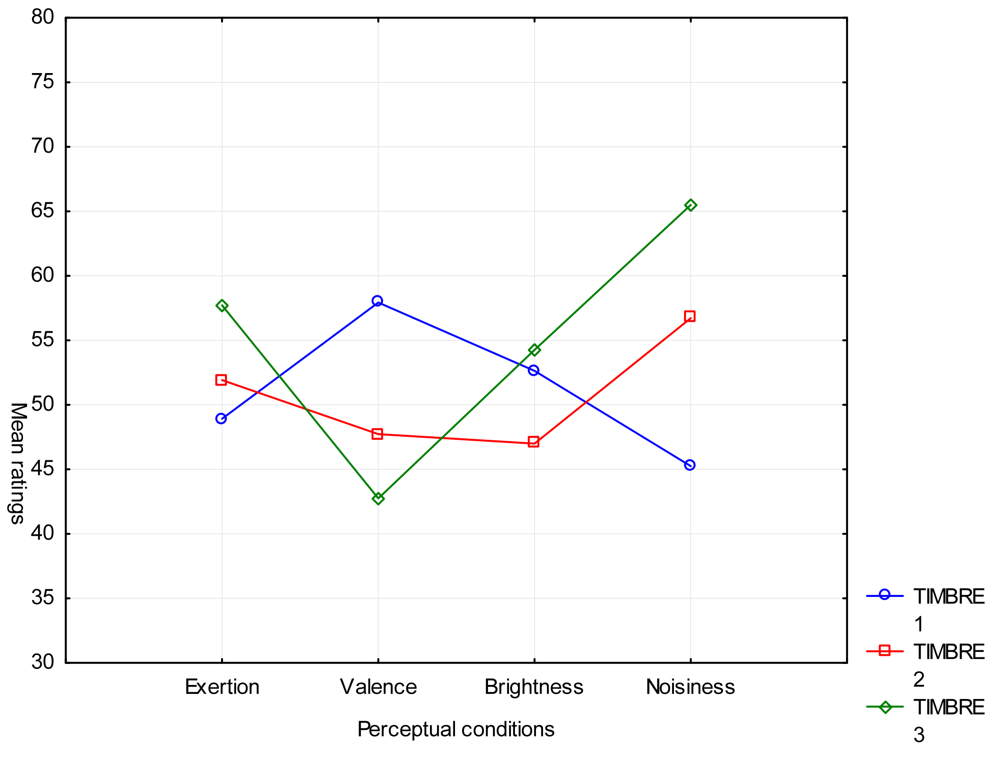


FIGURE 2.1: Average perceptual ratings by timbre condition

Raw averages can only tell us so much, however, even if the differences between conditions are substantial. To explore what these numbers actually mean in relation to the questions they are meant to address, I correlated data for all variables. Results indicate a binary perception schema based on the poles of *positive-negative* (or, to use a more affective binarism, *good-bad*). Listeners hear different timbral qualities as somehow belonging to one group (good) or the other (bad), with all associated perceptual qualities (embodiment, affect, semantics, emotion conveyed) falling into the same category. In their definitive 1957 account of the semantic differential procedure, psychologists Osgood, Suci, and Tannenbaum label this the "evaluative factor," and determine it to be the single most significant factor in determining meaning. The evaluative factor colors words with morally-, aesthetically-, socially-, and emotionally-valenced meanings along the axis of positive-negative.[19] Evidently listeners also divide timbral percepts along a positive-negative axis: "good sounds" are positively correlated with high *affective valence* (i.e., people like them), "brightness," *happiness*, and *tenderness*; "bad sounds" are associated with *bodily exertion*, "noisiness," *anger,* and *fear.* The primary emotion of *sadness* sits ambiguously between the poles of the evaluative factor.

In an important sense, what we find in this dataset is the vexed boundary between reaction and appraisal. If we maintain that the first order of timbre perception—the reaction—is innate, calling upon mechanisms of unconscious motor resonance to "mirror" the physical acts required of its production, then appraisal comes about when these reflexive perceptions are conceptually evaluated. Recall that reactions are largely self-defensive and driven by primitive regions of the brain—they do not comprise cognition *per se.* But they do appear to ground the more complex, associative cognitive tasks in their immediate wake. Appraisal comes when a reaction is

[19] See Chapter 2 (particularly pages 70-71) in Osgood, Suci, and Tannenbaum, *The Measurement of Meaning*.

interpreted, when listeners engage in the active process of making meaning out of timbre, justifying their reactions and connecting their interpretations to similarly valenced concepts. In this experiment I selected a small number of variables by which to evaluate perception of noisy timbres, but one can imagine many other binary pairs that would follow a similar pattern if subjected to experimental inquiry: *rough-smooth*, *clean-dirty*, *beautiful-ugly,* and so on. Appraising timbre is the act of hearing timbre *as* something—as "rough," "clean," "bright," "hollow," name your adjective or metaphor—but also of investing these qualities with moral, aesthetic, social, and emotional meanings. In the schema of Osgood *et al.,* "evaluation" might also be labeled the *appraisal factor*.

The clear presence of an appraisal factor in the perception of timbres is commonsensical for certain relationships, and perhaps surprising for others. For instance, *bodily exertion* is a negative percept linked with all the other bad qualities ("noisiness", *anger, fear*). This might contradict expectations, since so much of the pleasure of music to many people in our culture (and many other cultures) is tied to physical showmanship and spectacle: the daredevil virtuoso who plays with effortless mastery or, in popular music, the "hardest-working men in show business" who foreground the blood, sweat, and tears that go into performance. As Simon Frith writes:

> … one of the recurring pleasures of popular culture is the difficult or spectacular act, the drama of which lies precisely in its liveness, in the resulting sense of risk, danger, triumph, virtuosity […] What's valued here is not (as in high culture) seeing something unique, but seeing something difficult, something that *takes work*. Far from wanting the means of production to be concealed, the popular audience wants to see how much has gone into its entertainment. Performance as labor is a necessary part of the popular aesthetic.[20]

The key distinction here is context. Bodily exertion in the context of actual performed music is physical work for a communicative *purpose*; with these isolated sounds, on the other hand, there is

[20] Simon Frith, *Performing Rites: On the Value of Popular Music* (Cambridge, MA: Harvard University Press, 1998), 207. Italics in original.

no larger aesthetic or social purpose other than the bodily meanings of the timbres themselves. Exertion without a clear cause is negatively valenced, possibly because of its role in more urgent, dangerous moments of subject-environment interaction.

How do perceptions of *exertion*, *valence*, "brightness," and "noisiness" relate to "objective" acoustic factors like spectral centroid, inharmonicity, flatness, zero-cross, and roughness? I correlated standard acoustic measures of the various test timbres with the perceptual group means to find that the appraisal factor obtains here as well: due to high covariance, all five measurements listed above tend to move together according to whether a timbral percept belongs in one appraisal category or the other. Thus, all correlate negatively with *affective valence*, "brightness," *happiness, sadness*, and *tenderness* (all the "good" qualities), while the same acoustic parameters correlate positively with "noisiness," *anger,* and *fear* (all the "bad" qualities). (Interestingly, *exertion* correlates only weakly with most of the acoustic parameters.)

In conclusion, Experiment 1 revealed certain suggestive patterns of embodiment, affect, and conceptualization of timbre, both in the "acoustic" and the "perceived worlds." To summarize four key findings: (1) The acoustic qualities most closely associated with arousal in the voice *covary* (i.e., go together) in all timbres; (2) Perception of timbral quality is mediated by an underlying appraisal factor that splits percepts into "good" and "bad" types; (3) the appraisal factor applies to both acoustical and perceptual variables; and (4) perceived *bodily exertion* is closely related to p-noise, even when hearing timbral qualities that are not actually the byproduct of increased human motor exertion. But using 2-second isolated timbres as stimuli presents us with many obvious limitations; it is unclear to what extent the findings reported here actually apply to the act of music listening. To address this concern, I carried out a parallel experiment using brief recordings of popular music (same as the Introductory chapter) as stimuli.

<u>Experiment 2</u>: Perceptions of Embodiment and Affect in Brief Popular Music Excerpts [Appendix 2]

Most of the published literature on timbre perception uses isolated sounds as stimuli (and my first experiment was no exception), but obviously this is not the way listeners actually experience timbre in musical context.[21] Polyphonic timbre presents the listener with a temporally dynamic, dense, and complex mix of different sound generators; and while individual sources can usually be easily distinguished and identified, there is also a larger perceptual field comprising the "global sound" of all timbral sources in sum.[22] Such "timbral environments," as Rafael Ferrer calls them, are a sonic composite of different interacting sound producers.[23] Yet despite the ubiquity of polyphonic timbre, few studies have systematically compared the psychoacoustics of monophonic and polyphonic timbre.[24] This experiment attempts such a project by pairing the same paradigm

---

[21] For a very small sampling, see W. F. Lichte, "Attributes of Complex Tones," *Journal of Experimental Psychology* 28(1941); Grey, "An Exploration of Musical Timbre."; "Multidimensional Perceptual Scaling of Musical Timbres."; Plomp, *Aspects of Tone Sensation: A Psychophysical Study*; Wessel, "Timbre Space as a Musical Control Structure."; Paul Iverson and Carol L. Krumhansl, "Isolating the Dynamic Attributes of Musical Timbre," *Journal of the Acoustical Society of America* 94, no. 5 (1993); McAdams *et al.*, "Perceptual Scaling of Synthesized Musical Timbres: Common Dimensions, Specificities, and Latent Subject Classes."; 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).

[22] J.J. Aucouturier, "Dix expériences sur la modélisation du timbre polyphonique" (Ph.D. Diss., Universite Paris 6, 2006).

[23] Ferrer, "Timbral Environments: An Ecological Approach to the Cognition of Timbre."

[24] There are notable exceptions to this: see Roger A. Kendall and Edward C. Carterette, "Perceptual Scaling of Simultaneous Wind Instrument Timbres," *Music Perception* 8, no. 4 (1991); "Verbal Attributes of Simultaneous Wind Instrument Timbres: I. von Bismarck's Adjectives."; "Verbal Attributes of Simultaneous Wind Instrument Timbres: II. Adjectives Induced from Piston's 'Orchestration.", Andrew H. Gregory, "Timbre and Auditory Streaming," ibid.12, no. 2 (1994).; and Roger A. Kendall and Pantelis N. Vassilakis, "Perception and Acoustical Analyses of Traditionally Orchestrated Triads vs. Non-Traditional Counterparts" (paper presented at the 2nd Pan-American/Iberian Meeting on Acoustics, Cancun, Mexico, 2010). However, these studies examine perception of dyads, instrumental "blend," counterpoint, and triads, respectively, and not timbre in the context of ensemble music. Goodchild (2012) explores global timbral attributes of recorded classical music, but is concerned with how timbral components temporally track with listener responses to perceptions of increasing tension, and not as much with the timbral parameters themselves: see Megan Goodchild, "Towards a Perceptually-Based Theory of Orchestral Gestures" (paper presented at the AMS/SEM/SMT Joint Conference, New Orleans, LA, 2012). The work that most directly informs this study are two *Music Perception* articles by Alluri and Toiviainen: Alluri and Toiviainen, "Exploring Perceptual and Acoustical Correlates of Polyphonic Timbre."; "Effect of Enculturation on the Semantic and Acoustic Correlates of Polyphonic Timbre," ibid.29(2012). Additionally, Alluri's neuroimaging research has

used in Experiment 1 with very brief recordings of actual music. The essential purpose of using very brief excerpts—to be precise, 400 milliseconds[25]—is to transmit information about the timbral environment of the music without allowing enough time for significant melodic, rhythmic, harmonic, or lyrical information to pass into perception. Such "thin slices" serve as a timbral snapshot, so to speak, of a given genre, artist, and song.[26]

I selected stimuli from six popular music genres: Rock, Electronic Dance Music (EDM), Hip-Hop, Pop, Heavy Metal, and Country. Each genre was represented by three excerpts, for a total of 18 stimuli, and each song was chosen based on the criterion of "genre representativeness" to include acoustic elements commonly associated with each genre.[27] Thirty-six subjects performed the same ratings tasks as detailed in Experiment 1.

So how do extremely short clips of popular music compare to isolated timbres when subjected to the same experimental design? Presenting findings in the same order as above, we can see that our acoustical parameters of exertion are all significantly interrelated (though *roughness* correlations are lower than the others). In other words, clips with higher spectral centroid—such as the heavy metal excerpts—also tend to be richer in inharmonic energy and roughness. (Conversely, country music exhibited the lowest levels of all three acoustic dimensions.) Judging from

---

helped frame my study methodologically: see Alluri *et al.*, "Large-scale Brain Networks Emerge from Dynamic Processing of Musical Timbre, Key, and Rhythm."

[25] The three most significant rapid cognition studies in music psychology are Schellenberg, Iverson, and McKinnon, "Name that Tune: Identifying Popular Recordings from Brief Excerpts."; Robert O. Gjerdingen and David Perrott, "Scanning the Dial: The Rapid Recognition of Music Genres," *Journal of New Music Research* 37, no. 2 (2008); Krumhansl, "Plink: 'Thin Slices' of Music." The approximate timeframe I use here—more precisely, 380 ms with 10 ms amplitude ramps on both ends of the signal—is the "medium-length" condition in these studies. Another compelling example of this "scanning the dial" phenomenon comes from in-class listening exercises recounted in Anne Dhu McLucas, *The Musical Ear: Oral Tradition in the USA* (Aldershot, UK: Ashgate, 2010).

[26] This term comes from Krumhansl, "Plink: 'Thin Slices' of Music."

[27] See Appendix 2 for a list of excerpted songs. To determine if these excerpts were actually *heard* as the genres they were intended to represent, I conducted an identification experiment on a different pool of subjects ($N = 28$). Results indicated that the "correct" genre was selected 82% of the time—pure chance would have yielded about 19%—suggesting that the excerpts are largely representative of their respective genres. See Appendix 4a for details.

two very different sets of stimuli in these first experiments, we might conclude that the "sound of arousal" is irreducible to any one specific acoustic component, but rather makes up an emergent composite of *brightness*, *noise*, and *roughness*.

Correlating behavioral responses also yielded similar results to the first experiment. Responses are grouped into the same categories based on the appraisal factor: "good sounds" correlate positively to *affective valence*, "brightness," *happiness*, *sadness*, and *tenderness*, while "bad sounds" correlate positively to *bodily exertion,* "noisiness", *anger,* and *fear*. In fact, the appraisal factor is even more pronounced in this dataset than Experiment 1: correlations are statistically stronger in this study, and some of the ambiguous variables from the previous experiment, such as *sadness*, are much clearer here. (*Sadness* is, incidentally, a "good" quality of musical timbre despite its negative connotation in everyday speech.)

Moreover, "emotion conveyed" responses resonate in interesting ways with the other data. For example, heavy metal, as we will discuss in Chapter 5, is the outlier genre in terms of polarity of response: a full 95% of all responses indicate the predominant emotion of *anger.* It is no shock to see, then, that metal is far and away the most *embodied* genre (in terms of perceived physical exertion required to perform it); the least *liked* genre; and the most "noisy." These data support the sociological, musicological, and social-psychological literature on heavy metal.[28] Also of note, Pop received the most *happiness* responses (73%), with very little of any other single emotion represented, and also the highest "brightness" score.[29] Country, on the other hand, has the

---

[28] See Deena Weinstein, *Heavy Metal: A Cultural Sociology* (San Francisco: Jossey-Bass, 1991); Walser, *Running with the Devil: Power, Gender, and Madness in Heavy Metal Music*; Bethany Bryson, "'Anything But Heavy Metal': Symbolic Exclusion and Musical Dislikes," *American Sociological Review* 61(1996); William Neil Gowensmith and Larry J. Bloom, "The Effects of Heavy Metal Music on Arousal and Anger," *Journal of Music Therapy* 34(1997).

[29] Another interpretation of this result might relate to Pop's perceived simplicity. This is supported by Aniruddh Patel's contention that *happiness* is generally considered a "simpler mental state" than *sadness*, and is borne out by empirical research demonstrating that people tend to associate simple melodies with *happiness.* See Laura-Lee

highest *sadness* and *tenderness* responses (32% and 50%, respectively) and the lowest *exertion* and p-noise ("noisiness") ratings. (See Appendix 2 for specific descriptive data pertaining to these categories.)

Figure 2.2 summarizes average ratings of the six genres in the four scalar perceptual conditions. As hypothesized, the genres with the highest level of a-noise and p-noise (Metal, Rock, and EDM) tend to be heard as imbued with greater bodily exertion than the other three genres (Pop, Hip-Hop, Country, in descending order of *exertion*). Further, as in Experiment 1, signals with higher average *exertion* scores tend to show lower *affective valence* numbers than signals with lower *exertion* scores. This can be seen plainly here: the three "high exertion" genres listed above—all of which received average ratings of over 60—exhibit a decline in *valence*, while the genres with lower *exertion* scores (c. 30–50) demonstrate an increase in *valence*. The slope of the change between these two conditions seems to be related to the distance in the *exertion* condition between the actual mean and some hypothetical "ideal" level of exertion (around 55). For example, Metal, with a mean rating in the 80s, falls precipitously in the *valence* condition even as Rock and EDM, with means in the low 60s, decline only moderately, with similar inverse patterns at play in the lower *exertion* genres.[30] Illustrating the strength of correlations rooted in the appraisal factor, we can see that the p-noise condition brings us back to essentially where we started: the lowest *exertion* genre (Country) is also the least "noisy" genre, the highest (Metal) is again the highest, and the other genres sort themselves into the same order as the *exertion* condition. In fact, the extreme variation in the Metal ratings between conditions exemplifies the correlations discussed earlier—*exertion* and p-noise are close to the same, as are *valence* and "brightness" ratings. The

Balkwill and William Forde Thompson, "A Cross-Cultural Investigation of the Perception of Emotion in Music: Psychophysical and Cultural Cues," *Music Perception* 17, no. 1 (1999); Patel, *Music, Language, and the Brain*, 313.

[30] The EDM *exertion* results are intriguing when we note that it is entirely electronically produced. This would seem to corroborate our earlier conclusion that perception of bodily exertion is largely driven by *sound*, not *acts*.

shape of this relationship between perceptual variables and signals resembles Fig. 2.1 in striking ways, just with greater spread and stronger correlations.

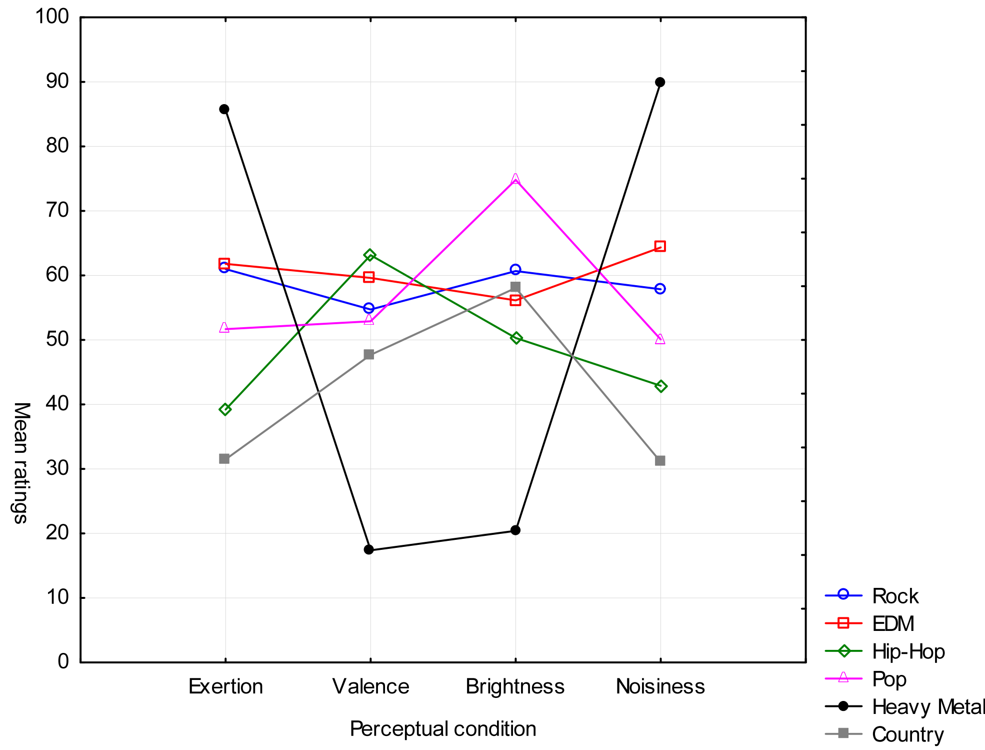


FIGURE 2.2: Average ratings for six genres in four perceptual conditions

There are additional findings to consider. First, I would like to briefly address what these first two experiments are saying about the semantic quality of "brightness," a variable that I have not yet discussed. As previously mentioned, "brightness" is perhaps the most studied timbral quality in the literature, and its association with spectral centroid has long been established. It is perhaps surprising, then, that "brightness" produced such ambiguous results in both of these experiments, showing no discernible relationship to the acoustic parameter associated with it (spectral centroid) and a *negative* correlation to p-noise, despite the central role of spectral centroid in evaluations of "noisy" timbre. In short, the semantic quality of "brightness" clearly has little to do with psychoacoustic *brightness* to subjects in these experiments. There are a number of possible reasons for this, most significantly the complex semantic, cultural, and even synaesthetic associations that are linked to the word "brightness." Heavy metal, for example, was rated very low in "brightness" despite having relatively high spectral centroid measurements, likely owing to its association with themes of darkness and evil and the penchant of adherents to wear black. Conversely, Pop—the "brightest" genre—might be labeled as such for its emotional connotations (*happiness*), commercial appeal, cheerful, "bubblegum" quality, and so on. There is a prominent social dimension to this evaluation, therefore, beyond the mere acoustic and psychoacoustic.

Though it complicates the dataset considerably, this disconnect over the concept of "brightness" exemplifies a central claim of this dissertation: appraisals of timbre are informed by innate motor-mimetic reactions, but even more by the "thick" contingencies of audition. Appraisal is never *just* about the sound quality being heard, or the "sounds themselves." Rather, as Osgood *et al.* remind us, appraisal of timbre is also at the same time a moral, aesthetic, social, and emotional judgment. In these experiments, the concept of "brightness" appears to have been perceptually incommensurable with noisy timbres, whether a growled saxophone or a short clip of heavy metal, even though these timbres were, in fact, psychoacoustically quite *bright*. It might seem a purely semantic distinction, but I think it illustrates both the linguistic slipperiness of timbre, and the importance of the broader cultural contexts in which all acts of music production and listening take place. Due to this ambiguity, moreover, I will henceforth refer to psychoacoustic *brightness* with the more neutral term *high-frequency energy*.

Returning to our results, I next correlated acoustic data with behavioral responses to confirm that the appraisal factor, as in Experiment 1, grounds perception of all parameters comprising the "sound of arousal." Taken in aggregate, the acoustic dimensions of the short sound examples correspond to a pattern that is now becoming familiar, negatively correlating with "good" sounds and positively correlating with "bad" ones. Moreover, statistical tests reveal that these five acoustic measurements alone—spectral centroid, inharmonicity, spectral flatness, zero-cross, and (most dramatically) roughness—account for approximately 80% of listeners' appraisals of *bodily exertion* and p-noise. While the broader cultural context of audition obviously plays a major role in appraising timbre, then, we should also not lose track of the fact that *sound*, at the end of the day, is the carrier of these meanings, even though the acoustic world is indelibly filtered through schemas of human embodiment. While tightly coupled, bodily acts and their acoustic consequences do not always move in perfect unison in the "perceived world." The blurriness between the two, in fact, can provide a powerful basis for creative musical (mis)understanding, as I will explore in the case studies that follow.

In sum, these results extend the major findings of Experiment 1, indicating that polyphonic timbre has much the same properties of perception as monophonic timbre. The two experiments revealed significantly similar relationships between the perceptual qualities and acoustic features of timbre, and these features were consistent whether subjects were listening to 2-second clips of isolated timbres or 400-millisecond clips of various popular music genres. It appears that for most perceptual categories, "thin slices" of actual music (Experiment 2) yielded more clearly defined, robust correlations than test tones and timbres. Even though these signals were far shorter than Experiment 1, reactions to timbre in musical context seem to be more associatively rich than to isolated instrumental and vocal sounds. Furthermore, these results suggest that the act of timbral appraisal is extremely rapid: just as quickly as we form a timbral percept it seems that we place it into a context of evaluation, judging it as "good" or "bad" with an alacrity that argues against the contribution of higher-level cognition. In an important sense, it might be argued that timbral reactions and appraisals are often, if not one and the same, at least highly redundant; the appraisal may well be a quick-and-dirty *post hoc* justification for the emergent bodily meanings that arise in the reaction.

Experiments 1 and 2 aimed to explore embodied timbre cognition by noting evaluations consciously available to the listener and quantifiable in the form of behavioral tasks (i.e., ratings, "best fit" word selections, etc.). This methodology has allowed timbre researchers to determine important relationships between the "acoustic" and the "perceived worlds"; it helped reveal the appraisal factor, and it allowed us to begin mapping the relationship between embodiment, affect, and semantic qualities in "noisy" musical timbre. Still, the semantic differential method has its limitations, most notably its reliance upon linguistic mediation. The appraisal factor indeed maps onto most if not all bipolar adjectival pairs; it is therefore quite obvious that, for example, "noisiness" would be a *bad* quality and "happiness" would be *good*. It is worth remembering, though, that linguistic designations—while undeniably slippery—are not totally arbitrary: if we accept the somatic markers theory of emotion, then the appraisal factor maps onto most (if not all) *percepts* as well. The "feeling of what happens," to quote Damasio, precedes our labels for it: meaning begins in the "theater" of the body, and linguistic appraisal (as observable through the semantic differential) is thus a behavioral manifestation of something going on within the "black box" of the body-mind. In the next experiment, I use neuroimaging to explore the dynamics of embodied timbre cognition from within the "black box."

<u>Experiment 3</u>: Timbre and Motor Resonance: An fMRI Study [Appendix 3]

To explore what happens in the brain—and specifically in motor areas associated with the mirror neuron system—when people listen to "noisy" timbres, I used functional magnetic resonance imaging (fMRI) on fifteen subjects as they listened to the twelve natural timbres from Experiment 1.[31] Subjects also completed the same behavioral ratings task outlined in the first experiment.

This final study was concerned with assessing the role of two specific brain regions—motor areas associated with the mirror neuron system, and the limbic system—in the processing of noisy aspects of musical timbre. Mirror neurons, as discussed in Chapter 1, have been shown to play a crucial role in imitation, language, social cognition, and empathy, and recent imaging research at the Iacoboni Laboratory at UCLA suggest they play a role in musical taste and aversive response.[32] However, it remains unclear to what extent motor resonance—the activation of motor mirror neurons during the observation (including by audition) of an action—is involved in timbre perception, if at all. Studies that address motor activation in music perception tend to focus primarily on rhythm and melody,[33] and while there is some evidence that one motor region (the SMA) plays a role in timbre imagery,[34] the contribution of motor areas to the perception of timbre is not well understood.

This raises three central questions: (1) Do mirror neurons play a role in hearing timbre? (2) If so, which specific mirror neuron regions show greatest activation? And (3) does a-noise (and its implied escalated bodily exertion) drive the auditor to respond with greater motor resonance than less noisy timbres? According to the model presented in Chapter 1, we might hypothesize the following answers: (1) Yes, mirror neurons play a role; (2) Vocal motor regions will show

---

[31] Although I will continue using the first-person pronoun in this section, I want to acknowledge the major contributions from my collaborators at the UCLA Ahmanson-Lovelace Brain Mapping Center, Marco Iacoboni and Katy Cross.

[32] For a preliminary discussion, see Zachary Wallmark and Marco Iacoboni, "Embodied Listening and Musical Empathy: Perspectives from Mirror Neuron Research" (paper presented at the AMS/SEM/SMT Joint Conference, New Orleans, LA, 2012).

[33] The most comprehensive review of auditory/motor interactions that I know of can be found in Zatorre, Chen, and Penhune, "When the Brain Plays Music: Auditory-Motor Interactions in Music Perception and Production."

[34] Halpern *et al.*, "Behavioral and Neural Correlates of Perceived and Imagined Musical Timbre."

greatest activation, since the psychoacoustics of arousal show greatest sensitivity to vocal or voice-like timbres, particularly timbral qualities that signal high-exertion states; and (3) Yes, a-noise will drive greater motor resonance for the reason outlined above. We shall return to these questions and hypotheses throughout the rest of the chapter.

Turning to the other region of interest for this experiment, the limbic system is the primary area associated with emotional processing, and its role in musical experience is well documented (though its contribution to timbre remains unclear).[35] Addressing the affective dimensions of timbre, we might reason that limbic areas will likewise play a role in reaction and appraisal. Experiment 3 sets out to explore functional connectivity between motor resonance and limbic response, if indeed such a relay system exists. One of the main claims of this dissertation is that embodiment and affect are inseparable in the timbre appraisal process—motor activity is therefore affectively weighted, just as affect is rooted in bodily actions and proprioception. The most compelling theory to date that attempts to put both of these two systems into conversation is the Shared Affective Motion Experience (SAME) hypothesis introduced in the previous chapter.[36] Overy and Molnar-Szakacs's model postulates a link between music-induced mirror neuron activation and the limbic system by way of the insula, but this idea has yet to receive any direct empirical attention. Experiment 3 aims to test this hypothesis.

(A few additional words of introduction are needed. It goes without saying that the brain is an extraordinarily complex system. Analyzing brain imaging data is, likewise, a highly nuanced, computationally intensive, statistically sophisticated process, a full explanation of which

---

[35] See Anne J. Blood *et al.*, "Emotional Responses to Pleasant and Unpleasant Music Correlate with Activity in Paralimbic Brain Regions," *Nature: Neuroscience* 2, no. 4 (1999); Isabelle Peretz and Robert J. Zatorre, eds., *The Cognitive Neuroscience of Music* (Oxford and New York: Oxford University Press, 2003); Koelsch *et al.*, "Investigating Emotion with Music: An fMRI Study."; Koelsch and Siebel, "Towards a Neural Basis of Music Perception."

[36] Overy and Molnar-Szakacs, "Being Together in Time: Musical Experience and the Mirror Neuron System."

could easily consume dozens of pages. For this reason, I am including a basic neuroimaging primer in the Glossary; this short introductory text explains in very simple terms the procedures and conventions that stand behind both this chapter and Appendix 3 so that the reader can proceed unencumbered by such details here.)

When comparing brain activation in response to individual timbres *vs.* silence, we see that every stimulus produces a unique neural activity signature. Table 2.2 lists some of the regions that are active when listening to the timbres, along with their primary functions. Included in this list are a number of areas that have been shown to play a role in action-perception coupling via mirror neurons (for example the motor areas, parietal lobe, visual cortex, and areas involved in speech production and perception). Further, as I will discuss in the next chapter, somatosensory (tactile) areas were active in the perception of certain timbres. We are thus justified in drawing the broad conclusion from this imaging data that a wide range of sensorimotor, limbic, subcortical, and other areas are involved in the processing of timbre, irrespective of temporal aspects of music, and that mimetic motor resonance likely plays a role in some of these activations.

TABLE 2.2: Regions of activity and primary associated function

Region	*Abbr.*	*Primary Function*
Primary auditory cortex	PAC	Auditory processing
Pre/primary motor cortex	PMC	Action planning and execution
Supplementary motor area	SMA	Movement control, motor imagery
Primary somatosensory cor.	PSC	Tactile processing
Primary visual cortex	PVC	Visual processing
Broca's area	BA	Speech production
Rolandic operculum	RO	Larynx control (vocal motor production)
Inferior parietal lobule	IPL	Body image, sensory interpretation, recognition of facial emotions, etc.
Superior parietal lobule	SPL	Spatial orientation, sensory-motor integration, etc.
Cerebellum	C	Motor learning, motor control, etc.
Thalamus	T	Relay between subcortical structures and cortex
Hippocampus	H	Memory, spatial navigation

Parahippocampal gyrus	PG	Memory
Amygdala	A	Emotional reaction
Anterior insula	AI	Emotion, homeostasis, relay between MNS and limbic system
Brainstem	S	Regulatory functions, autonomic nervous response, etc.

Why are motor areas active when subjects are passively listening to (noisy) timbre? Part of the explanation—and the reason that listening is never entirely "passive"—likely has to do with the presence of audiomotor mirror neurons in these regions. Recall that subjects were listening to these timbres in a highly restrictive sensory environment (the MRI magnet): they could not move, and they were instructed to stare at a fixation cross through goggles during the whole study (to reduce visual confounds).[37] Seeing this sort of activity in these regions during a passive listening task, therefore, points to mechanisms of act–sound coupling. Although it is difficult to parse how much of what we are seeing is strictly speaking *motor mimesis*—i.e., imitation of motor acts implied in the production of the sound—and how much is a motor reaction to each timbre, independent of actually mirroring its production—i.e., implicit, possibly imagined motor acts (such as the desire to cover one's ears)—it is relatively clear that motor resonance of some sort is involved when people listen to these noise-scaled isolated musical timbres.

But, while this addresses Hypothesis 1, we have said little about how the act of appraisal is processed in these regions. The most interesting question we ask in this experiment, indeed, has less to do with which parts of the brain light up when listening to which timbres, and more with how appraisals of timbre play out in the brain. Rather than drawing inferences about how a tim-

[37] A "confounding variable" is something that biases the results. In this case, not controlling for vision would introduce a range of neural activity into the data that would be impossible to dissociate from the experimental manipulation.

bre is experienced based on the patterns of neural activation that each timbre evokes,[38] we can put behavioral responses into conversation with imaging data to uncover relationships between conscious appraisal and neurophysiology. To this end, I correlated subjects' behavioral responses with the behavior of their brains, so to speak, to see what activated regions modulated with perceptions of *exertion*, *affective valence*, and "noisiness."[39] This approach cross-fertilizes the two paradigms of this chapter, empirically accounting both for the conscious manifestations of certain patterns of brain activity (behavioral response) and the brain activity underlying those conscious appraisals (fMRI).[40] Looking at the *interaction* between these two levels allows us to sidestep contentious "chicken or egg" questions as to which level—behavioral or neural—is ontologically prior; instead, both levels can be approached simultaneously as two sides of the same dynamic process of reaction-plus-appraisal.

This is where things start to get interesting. Let us begin with regions that show an increase in activity the more a person hears *bodily exertion* in a timbre. I found significant correlations when contrasting all Timbre 3 (maximally noisy) with all Timbre 1 (not noisy). Surprisingly, activations were virtually entirely contained to subcortical structures, cerebellum, and limbic areas, not to motor areas. In particular, activity in *amygdala* (both hemispheres, but particularly right) correlated highest with *bodily exertion* ratings: subjects who heard Timbre 3 as significantly more effortful than Timbre 1, in other words, also tended to exhibit greater activity in the

[38] This would be a classic case of "reverse inference," which is a common criticism of many fMRI studies.

[39] "Brightness" was omitted from the analysis due to highly ambiguous responses from Experiment 1. Further, since "emotion conveyed" responses were nominal and not continuous, they could not be added to the multivariate regression model.

[40] A model for this sort of fusion of behavioral and brain methods can be found in Petri Toiviainen *et al.*, "Timbre Similarity: Convergence of Neural, Behavioral, and Computational Approaches," *Music Perception* 16, no. 2 (1998); Alluri *et al.*, "Large-scale Brain Networks Emerge from Dynamic Processing of Musical Timbre, Key, and Rhythm."

amygdala, an important limbic structure for emotion processing.[41] Figure 2.3 shows activity in the amygdala accompanying perceptions of bodily exertion.[42] This holds when stimuli are grouped according to level of noisiness, not sound generator, implying that this amygdala activity is driven largely by the acoustics of the timbres and not the actual physical effort involved in their production, which varies from instrument to instrument. This result resonates with our finding from Experiment 1 regarding perceived exertion of guitar and white noise: people are "hearing" physical effort in noisy timbre, and this perception of exertion is related to the amygdala, not (as we might assume) the motor areas.

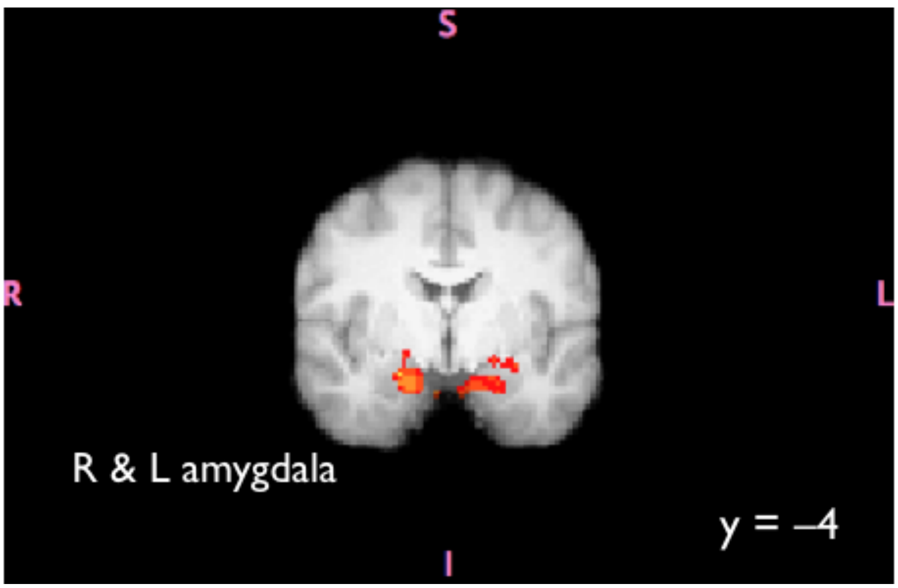


FIGURE 2.3: Timbre 3 > Timbre 1 correlated with perceived *bodily exertion*

---

[41] The literature on the amygdala's role in emotional experience is vast. For a cursory review, see Damasio, *Descartes' Error: Emotion, Reason, and the Human Brain*, 133-34. Ethnomusicologist Judith Becker, drawing on Damasio's work, has theorized about the role of this structure in "deep listening" and music-induced trance: see Becker, *Deep Listeners: Music, Emotion, and Trancing*.

[42] See the Glossary and Appendix 3 for basic information on how to read imaging data. Numbers in the bottom right corner indicate coordinates along *x*, *y*, and *z* axes.

Next, I correlated brain activity with *negative affective valence*. Contrasting noisy (Timbre 2) with not-noisy (Timbre 1) conditions revealed increased activity in a variety of motor and limbic regions, including right pre/primary motor cortex (PMC) and SMA. It is worthy of note that motor resonance in these areas in relation to disliking is entirely *right lateralized*; no significant difference in activation was found in the left, dominant hemisphere.[43] It has long been suggested that emotion processing is right dominant, and that processing of negative emotions in particular tends to be right-lateralized, so the right dominance of these results is consistent with this hypothesis.[44] We can also see significant action in other regions associated with negative responses to music.[45] Taken together, this result indicates that PMC—the main motor area and a crucial mirror neuron region—is involved in *negative* assessments of musical timbre.

Arnie Cox has suggested that liking or disliking music depends to a large degree on whether we like or dislike "what [the music] invites us to do."[46] Put in these terms, this result may indicate that noisy timbres are "inviting" us to *do* something physically effortful—something associated with high arousal states, activity, and potency—that we do not necessarily *want* to do. I would like to posit that this activation pattern represents a motor component of the generation of negative affect, possibly reflected in the limbic activity cited above, specifically the right anterior insula (AI). The insula is a well-documented emotion-processing area and a possible relay center between motor representations and limbic response.[47] It has also been theorized to play an im-

[43] All subjects were right-handed in order to achieve symmetry of hemispheric dominance.

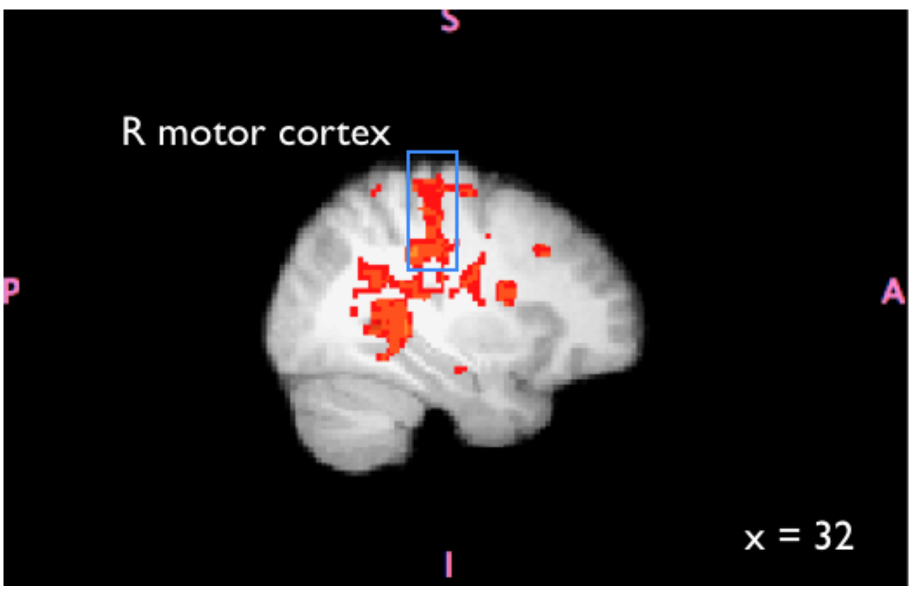
[44] See Damasio, *Descartes' Error: Emotion, Reason, and the Human Brain*, 140; K.M. Alfano and C.R Cimino, "Alteration of Expected Hemispheric Asymmetries: Valence and Arousal Effects in Neuropsychological Models of Emotion," *Brain and Cognition* 66(2008).

[45] Specifically superior parietal lobule and parahippocampal gyrus: see Blood *et al.*, "Emotional Responses to Pleasant and Unpleasant Music Correlate with Activity in Paralimbic Brain Regions."

[46] Cox, "Embodying Music: Principles of the Mimetic Hypothesis," 51.

[47] Laurie Carr *et al.*, "Neural Mechanisms of Empathy in Humans: A Relay from Neural Systems for Imitation to Limbic Systems," *Proceedings of the National Academy of Sciences of the USA* 100, no. 9 (2003); Bruno Wicker *et al.*, "Both

portant role in musical affect, relaying mimetic motor activations to the limbic system for higher-order affective processing and appraisal.[48] In addition to its involvement in *disliking* noisy timbre over not-noisy timbre when collapsed across sound generators, moreover, I found that the left anterior insula played a role in negative affective response to *vocal* timbre in particular. Together, these results support Overy and Molnar-Szakacs's SAME hypothesis, suggesting that the insula may function as a neutral conduit between the mirror neuron and limbic systems. This dataset also fills in details of the SAME model, implying a preferential role for the anterior insula in the generation of *negatively valenced* affect. This area of the brain may play a central role in the disliking of timbre, and possibly (by extension), in the timbral disliking of music in general. Figure 2.4 maps activity in the right PMC and SMA (top) and the insula (bottom) in response to disliked qualities of timbre.



of Us Disgusted in My Insula: The Common Neural Basis of Seeing and Feeling Disgust," *Neuron* 40(2003). For a review, see Overy and Molnar-Szakacs, "Being Together in Time: Musical Experience and the Mirror Neuron System," 492-93.

[48] Molnar-Szakacs and Overy, "Music and Mirror Neurons: From Motion to 'E'motion."; Overy and Molnar-Szakacs, "Being Together in Time: Musical Experience and the Mirror Neuron System."

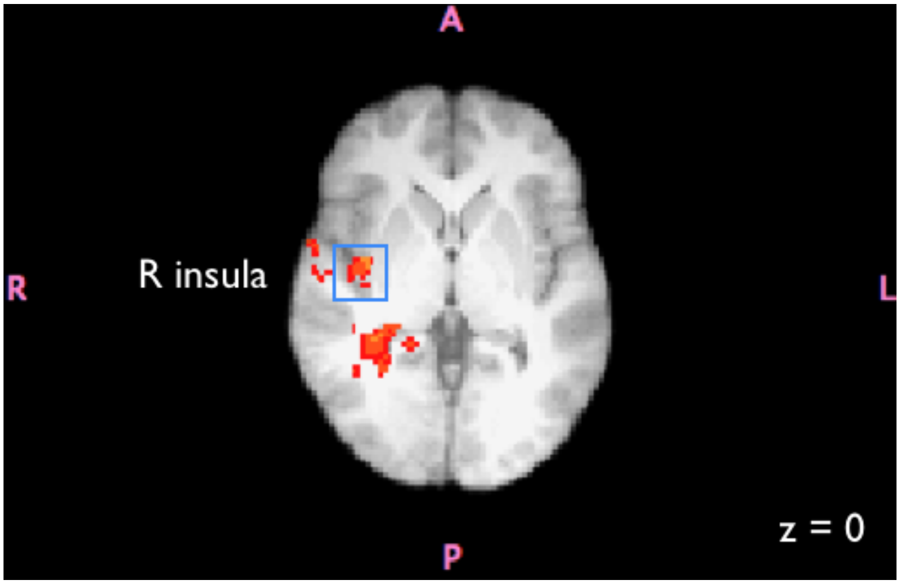


FIGURE 2.4: Timbre 2 > Timbre 1 correlated with negative *affective valence*

The final correlation looked at regions involved in perceived "noisiness" (p-noise). Results are a hybrid of salient features from both the *exertion* and negative *affective valence* images reviewed above. P-noise is significantly correlated *both* with activity in limbic areas—most notably the amygdala and hippocampus—*and* with right PMC and SMA. Figure 2.5 displays these activations: note also the activity in the primary somatosensory cortex (PSC), a matter we will take up in greater depth in the next chapter. This indicates a possible relationship between certain motor areas and limbic response in the perception of "noisy" timbre, suggesting a functional link between motor mirroring of this specific perceptual feature of timbre and affective response. It is further significant that the right amygdala, as in the *exertion* correlation (Fig. 2.3), is the most active limbic structure of the contrast. The importance of the amygdala in affective response has long been known, particularly its role in *fear* responses and other negatively-valenced, high arousal emotions, as well as its contribution to aversive reactions to music.[49] This result indicates a possible association between amygdala and motor areas in the perception of timbral noisiness: perceiving a sound as "noisier" than another seems to involve some interplay between these regions.

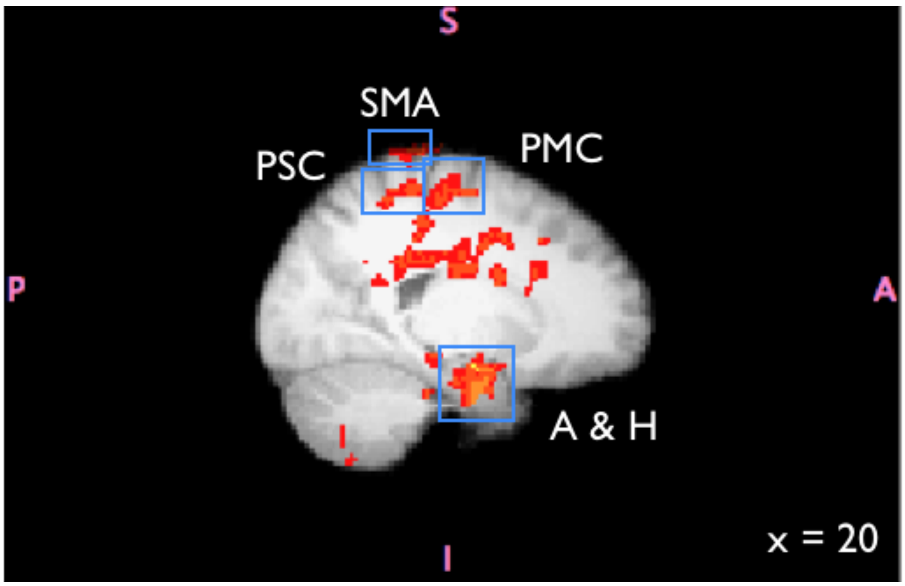


FIGURE 2.5: Timbre 2 > Timbre 1 correlated with p-noise

This result is consistent with the behavioral data collected and analyzed in Experiments 1 and 2. *Bodily exertion* and "noisiness" are high-arousal, negatively-valenced percepts—"bad sounds"—so the greater limbic activity associated with them (particularly right amygdala) is expected. As I noted earlier, however, it is also intriguing that behavioral measures of perceived *bodily exertion* would relate most clearly to limbic areas associated with emotional processing (specifically negative, high arousal emotions) rather than to areas that code for motor imagery, plan-

---

[49] For a great general summary, see Joseph LeDoux, *The Emotional Brain: The Mysterious Underpinnings of Emotional Life* (New York: Simon & Schuster, 1998); Koelsch *et al.*, "Investigating Emotion with Music: An fMRI Study." For evidence of the amygdala's importance in processing negatively-valenced vocal expression specifically, see S.K. Scott *et al.*, "Impaired Auditory Recognition of Fear and Anger Following Bilateral Amygdala Lesions," *Nature* 385(1997).

ning, and execution. This is perhaps evidence for a link between the conscious perception of the bodily implications of a particular timbre (i.e., responding that one hears a timbre as more or less imbued with physical effort) and the corresponding non-conscious affective reaction (i.e., showing activation of brain regions, like the amygdala, responsible for processing high arousal emotions). This (admittedly) hypothetical affective link bypasses motor systems to activate the more phylogenetically ancient regions of the brain.

With "noisiness" correlations, though, we see motor and amygdala activity together in the same image. In the behavioral experiments, p-noise correlated strongly with both *bodily exertion* and negative *affective valence,* and here we see corroborating neural evidence: the perception that a given timbre is "noisy" is associated with increased motor activity (PMC and SMA) as well as amygdala activity, possibly related to negative affective response. Although the perception of "noisiness" is related to these two qualities, moreover, it appears to rely on neural resources that encompass both motor mimesis and affective response. Hearing signals as "noisy" enacts mechanisms of motor resonance—possibly because a-noise, as shown in Chapter 1, is an indicator of high arousal affective states, which in turn are more ecologically urgent—and the limbic responses consistent with the audition and infectiousness of high arousal states. The data indicate that the perception of "noisiness" in musical timbre is both embodied (via mechanisms of motor resonance) and weighted toward negative affect (via right-dominance of limbic activity and prominence of amygdala). The perceptual systems involved in evaluating "noisiness" therefore point once again to an interconnection between motor resonance and emotion processing.

As a final step in this analysis, I journeyed back to the external "acoustic world" to determine how our chosen sonic parameters—spectral centroid, inharmonicity, spectral flatness, zero-cross rate, and roughness—relate to neural activity. The results were ambiguous, with few correlations reaching statistical significance. However, one small area called the Rolandic operculum showed a strong *negative* relationship with a-noise. This result is highly suggestive, since this precise region (see Table 2.2) is responsible for controlling the muscles of the larynx.[50] It appears therefore that the area of the brain responsible for vocality—the Rolandic operculum (RO)—modulates according to the a-noise present in a given timbre.

Chapter 1 introduced some of the discourse on the role of vocality in musical expression, but Experiments 1 and 2 did not specifically address vocal mimesis. Here, though, it seems we are confronted with direct evidence that subvocalization is relevant to timbre perception. The voice does not appear to be involved in the perception of all timbres equally; rather, it is *selectively involved in the perception of "good" timbral qualities*. To follow up on this theory, I sharpened the analytical focus to compare areas involved in vocal production with behavioral data using two additional analyses.[51] Both pointed to the same conclusion: RO and Broca's area (BA)[52] showed a linear *decrease* of activity between the "low-noise" and the "high-noise" timbre conditions, while the more general-purpose PMC (particularly in the dominant, left hemisphere) *increased* with the addition of a-noise, as discussed previously. Figure 2.6 graphs this inverse relationship (which is strongest, incidentally, between Timbre 1 and Timbre 2),[53] and Figure 2.7 plots vocal activity in relation to *affective valence*. These two results appear to converge on the following conclusion: RO seems to be sensitive to the relative degree of "liking" and a-noise attendant on a particular tim-

---

[50] Steven Brown, E. Ngan, and M. Liotti, "A Larynx Area in the Human Motor Cortex," *Cerebral Cortex* 18(2008).

[51] Details can be found in Appendix 3, but without getting too technical the thumbnail version is as follows: in the first analysis I extracted statistical data for individual vocal-area *voxels* (small three-dimensional squares that subdivide the space of the brain) in all three timbre conditions, then compared them; in the second analysis, I correlated these individual-voxel data with behavioral responses. The correlations discussed earlier were performed at the *whole-brain*, not the individual-voxel level, and the statistics behind each approach are quite different; hence, these vocal activations did not appear in the earlier analyses.

[52] Broca's area is another voice production/perception center, and the most direct human homologue to the macaque monkey mirror neuron area.

[53] To be clear, this result at the single voxel level is just shy of statistical significance—as I explain in the introduction to the Glossary, at less than a 7% probability of being the result of chance, this result is more "musicologically relevant" than "statistically significant" (set at the conventional 5% chance of error). See Appendix 3 for more details.

bre, activating most when people listen to timbres that fall on the "good" side of the appraisal factor. Conversely, PMC—the mirror neuron-rich region responsible for all physical action—showed the most robust activity when subjects listened to "bad" sounds (i.e., high implied *exertion*, negative *affective valence*, high p-noise).

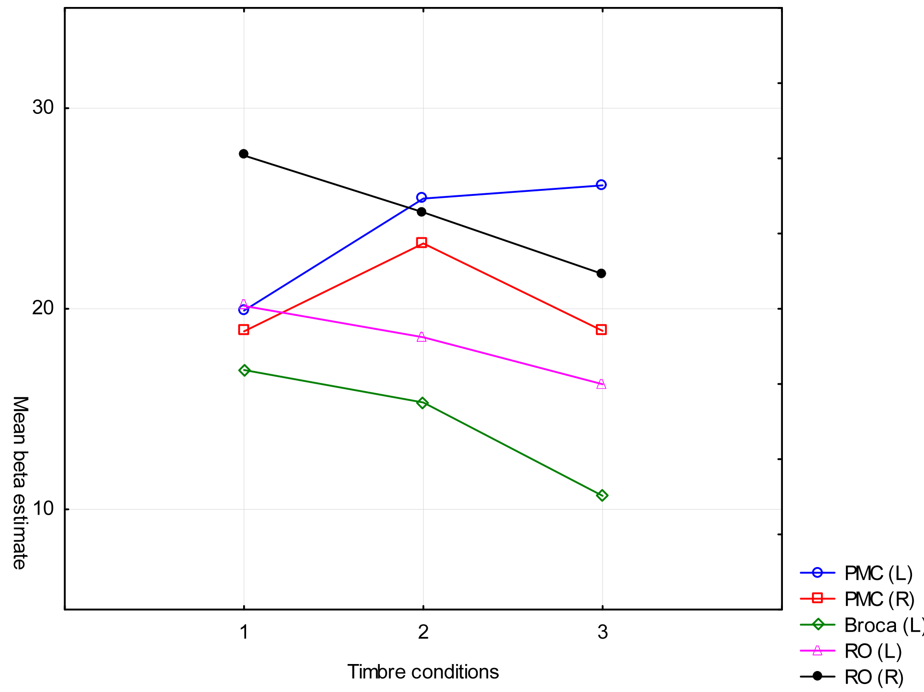


FIGURE 2.6: Effects of timbre on activation of vocal motor areas *vs.* PMC

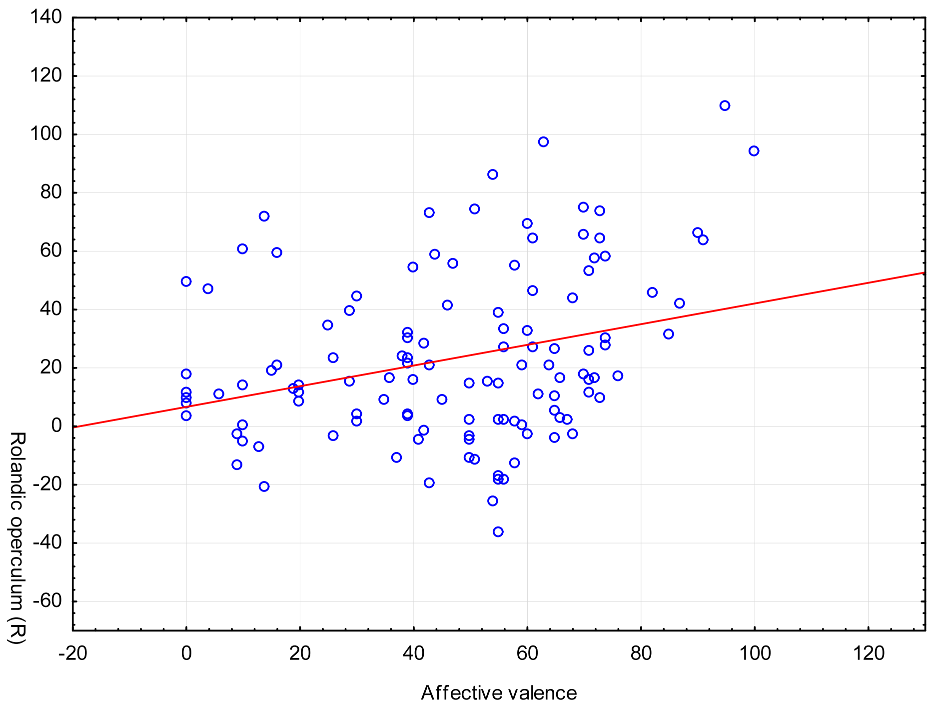


FIGURE 2.7: Positive correlation between RO and *affective valence*

This finding corroborates the influential 2006 study by Koelsch *et al.* showing a role for RO in listening to "pleasant" music, suggesting that we internally "sing along" to music we like.[54] However, Koelsch's study used recordings of classical and jazz music, stimuli that, with the addition of all the temporal features of music, would presumably be much richer in sources for mimetic vocal engagement than isolated 2-second monophonic timbres. Subvocalizing to a pleasant melody feels intuitive, but my results point to a role for this important vocal region in the mere audition of *pleasant timbre alone,* outside of any consciously perceived musical context. We seem to appraise timbre through the voice: engaging with a sound complex by vocally mirroring its qualities of production is, it seems, a prerequisite for its positive appraisal. Or, perhaps, the vocal mirroring *is* the positive appraisal.

---

[54] Koelsch *et al.*, "Investigating Emotion with Music: An fMRI Study."

To pull the various threads together and summarize, findings from Experiment 3 suggest three broad conclusions. First and most generally, different instruments and timbral qualities (from not noisy to noisy) produce different neural fingerprints, and components of the motor system—including mirror neurons—are likely involved in this differentiation. Limbic activity also suggests that perception of different timbral qualities is marked by a range of affective qualities. This might be interpreted to indicate that timbre is *not* an abstract musical building block that only acquires emotional meaning when structured into larger musical units; different timbral qualities, owing in part to motor mimicry, are inherently loaded with affective valences.[55]

Second, and most importantly for the thesis of vocal-based motor mimicry of timbre, it appears that motor areas associated with vocal production—RO and (to a lesser extent) BA—are sensitive to changes in timbre, even when the instrument category is collapsed. Listening to acoustically and perceptually "noisy" timbres—in electric guitar, saxophone, *shakuhachi*, and particularly the singing voice—seems to have a suppressive effect on motor mimetic properties of the vocal production system, specifically RO.

And lastly, Experiment 3 suggests an important role for the amygdala (particularly in the right hemisphere) in perceptions of *bodily exertion* and "noisiness," *versus* the insula in perception of negative *affective valence.* These structures, preferentially associated with the processing of fear, disgust, surprise, and other negative, high arousal emotions, appear to modulate significantly along with just these three perceptual variables. Connecting motor activation (PMC and SMA) with the limbic system in a manner consistent with the SAME hypothesis, these correlations may for the

[55] It should be noted that *all timbre*—not just musical sounds—appears to be governed by the appraisal factor. This is not surprising at all, considering the similar perceptual systems at work in all human audition. For more on perception of non-human-made timbre, see E.A. Björk, "The Perceived Quality of Natural Sounds," *Acustica* 57(1985).

first time suggest a functional link between these two neural systems during the negative appraisal of timbre.

Certain parts of these results problematize aspects of the motor mimetic hypothesis advanced in Chapter 1. Motor resonance does not appear to increase uniformly with increased p-noise, but rather follows a more complex, non-linear pattern. First, it is contingent upon the stimuli: for example, saxophone timbres induce a response consistent with the hypothesized model, while *shakuhachi* timbres do not. It is also contingent upon the specific motor region under observation: these results suggest a functional difference between specifically *vocal* motor areas and general-purpose motor areas. Different mirror regions appear to be involved in different appraisals. Averaged across all the stimuli employed in this study, however, "bad sounds" seem to attenuate vocal mimetic engagement, and thus weaken a central component of embodied music cognition. The cultural ramifications of these findings are significant, and I will address them in different musical contexts throughout the rest of this dissertation.

Hypothesis 3, which predicts that a-noise will lead to greater motor resonance, yielded inconclusive results. Correlations between brain activity and acoustic parameters failed to reach significance in many of the regions besides RO, which showed a negative correlation with all parameters. For vocal production areas, a-noise appears to attenuate resonance; but positive acoustic correlations with other motor areas, including PMC (which plays a role in disliking and "noisiness") were too weak to draw any conclusions. PMC and SMA, then, are involved in processing the *perceptual* attributes of "bad sounds" but not necessarily their associated acoustic cues.

This experiment raises many more questions than it answers: for instance, does the positive correlation between these vocal motor regions and "liking" indicate that the listener might somehow *choose* to resonate with liked, less noisy timbres while rejecting the invitation for motor mimicry with disliked, noisy timbres? What comes first, the neural pattern of motor mirroring consistent with a certain reaction, or the conscious appraisal of said reaction? (Or are they co-terminous?) Do these results hold in the cognition of timbre in "real-world" musical situations? These questions remain open for further study.

Conclusion

At this point I will spell out some of the major points of convergence between my three experiments, briefly summarize the evidence gathered therein for the embodied cognition of timbre, and theorize what these results might mean for broader questions of timbral embodiment and affect.

Figure 2.8 presents some of the key findings from these experiments. The top of the diagram recapitulates the first three stage of the ASPECS model (in reverse order from Fig. 1 in the Introduction). At the *perception* stage, the innate, self-protective reaction—an immediate and automatic process, we should remember, that is driven by all the "thin" psychobiological mechanisms outlined in Chapter 1—quickly transforms into an appraisal. The appraisal is the gateway into all the higher-order, "thick" meanings that we can attribute to timbre—levels of meaning that form the basis of the rest of this dissertation. But the most basic facet of the appraisal process consists of a simple act of binary categorical perception: appraisal can be conceptualized as a moment of bifurcation where percepts are grouped—extremely rapidly—into "good" and "bad" types. In a nutshell, "good" qualities promote the flourishing of the organism, while "bad" qualities discourage it; thus appraisal—the conscious evaluation of a quality of experience—is grounded to some degree upon the lower-order reaction, though it quickly becomes much more complex and socially determined than a mere reflex. Although I have been keeping these two concepts separate for the sake of clarity, reaction bleeds into appraisal, and we should really view them as different aspects of the same fundamental process of perception: reaction is more material and less culturally-mediated, and appraisal is the reverse, but they both contribute to the ultimate output.

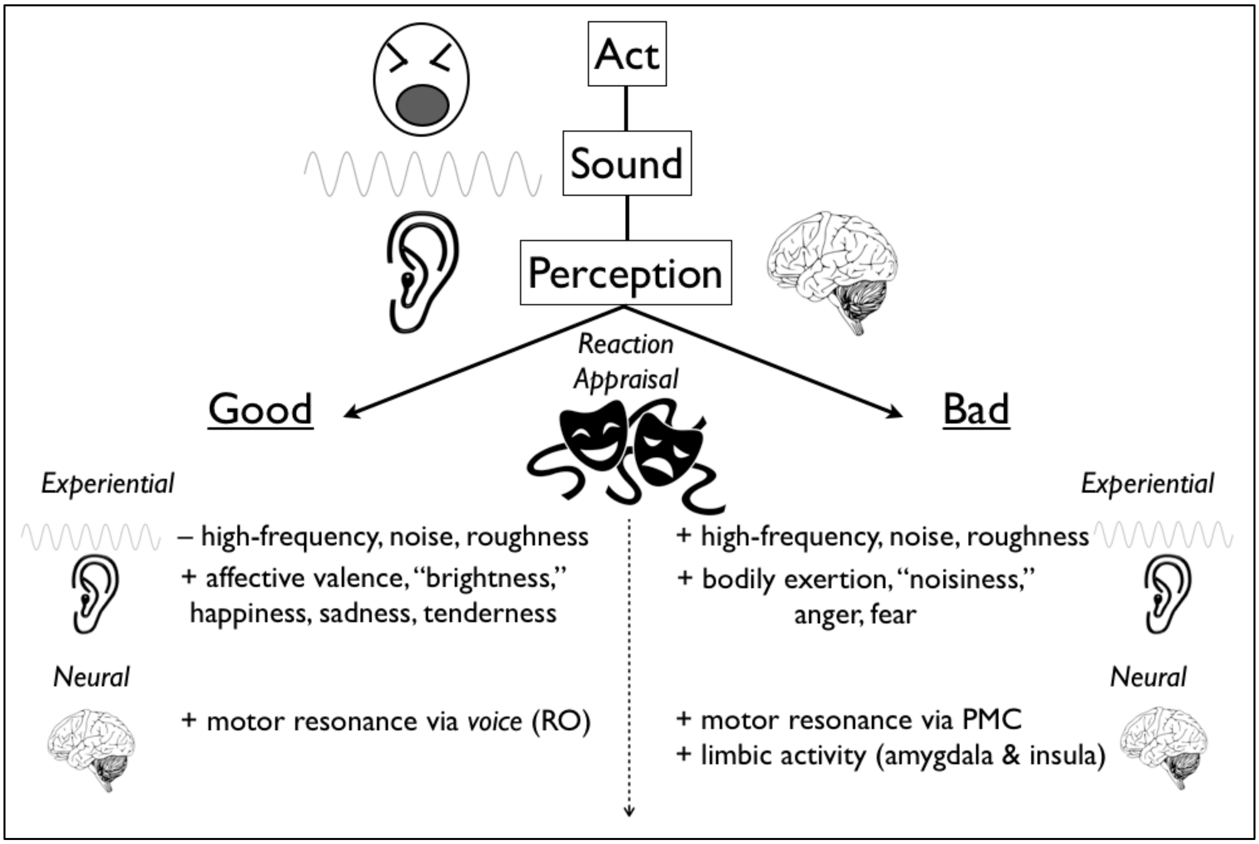


FIGURE 2.8: Diagram of conclusions

For our purposes, this "output" can be conceived as consisting of two levels: first, there is the *experiential* level, which was probed in Experiments 1 and 2; and second is the *neural* level, which I explored in Experiment 3. The convoluted relationship between the material brain and the subjective stuff of phenomenological experience—the so-called "qualia problem"—is a complex topic obviously far-afield from this dissertation, so I do not presume to explain how precisely these two levels relate to each other.[56] However, results from these experiments display notable overlaps between experiential and neural levels of good and bad appraisals.

"Good sounds" are associated with low levels of our three psychoacoustic markers of physical effort, *high-frequency energy, noise,* and *roughness.*[57] They are also associated with high *affective valence* (i.e., strongly liking), the semantic quality of "brightness," and the emotions of *happiness*, *sadness* (somewhat ambiguously), and *tenderness.* When we look at how the brain responds to "good sounds," we can see that specifically *vocal* regions—particularly RO—show activity consistent with the conclusions reached by Spencer, Cox, and Eidsheim regarding the primacy of the voice in music perception. This is true even for instrumental timbre, supporting the "superexpressive voice" theory of Juslin and Laukka.

"Bad sounds," on the other hand, are characterized by a different set of experiential qualities and neural activation patterns, and because this dissertation is about the threshold of music and noise, I will focus more on this half of the appraisal factor. First, all of the acoustic characteristics of physical effort in the voice are heard as "bad" qualities of timbre. These qualities correlate strongly with the other negative percepts from Experiments 1 and 2: *bodily exertion*, p-noise ("noisiness"), and the emotions of *anger* and *fear*.[58] This was predicted by the theoretical model developed in Chapter 1: vocal exertion leads to higher levels of the acoustic qualities listed above, which in turn, via processes of motor resonance, sonically refer back to the strained body that gave them voice. However, the perception of bodily exertion is not limited entirely to timbres whose associated acts actually require greater relative degrees of effort; *high-frequency energy,*

---

[56] I am highly skeptical of the notion that the brain provides the privileged *explanans* for the *explanandum* of human experience, though it is doubtless a decisive piece of the puzzle.

[57] As mentioned previously, I changed terminology from "brightness" to "high-frequency energy" here to avoid confusion: recall that "brightness," as an actual percept, did not correlate with spectral centroid in the signals.

[58] As determined in Appendix 4b, moreover, *intensity* is a "bad" quality of timbre.

noise, and *roughness* seem to convey exertion irrespective of a sound's mode of production. This suggests that timbral qualities associated with effort can simulate heightened bodily states without having to actually call upon high-arousal acts. Implied exertion, moreover, is connected to the emotions most commonly associated with arousal and potency, and all of these qualities are closely related to impressions that a timbre is "noisy." Taken together, then, "noisy timbres" have both acoustic and perceptual correlates consistent with negative appraisal: they are rich in arousal-indexical a-noise (i.e., high spectral centroid, inharmonicity, flat spectrum, zero-cross, and roughness); they are perceived as "noisy"; they sound physically effortful to produce, even if the effort is entirely simulated; and they sound representative of the bodily experience of *anger* and *fear* (i.e., they resemble a person's voice when angry or gripped with fear).

But exploring the neural correlates to the perceptions above adds further complexities to this "bad" appraisal. We can point to two significant, interrelated patterns of activation in response to noisy timbres, one motor, one limbic. First, activity in the PMC (and also to a lesser extent the SMA) appears to increase with negative appraisals. Bad sounds, it seems, are therefore embodied just like good sounds—they induce motor resonance in the brain of the listener—though the specific neural substrates differ: where good sounds preferentially draw upon mechanisms of action-perception coupling of the voice, bad sounds preferentially activate the more general-purpose motor cortex (particularly the right hemisphere). The perception of bad sounds thus involves a disjuncture of these two motor regions, an inverse pattern of activity. Tension between vocal resonance and broad motor resonance might be a significant part of the neural explanation for timbral appraisal. Motor regions of the brain are at odds with themselves: positive appraisal is strongly related to subvocalization (and the suppression of other motor areas), while negative appraisal is strongly related to activity in other motor areas (and the suppression of subvocalization). The response is ambivalent. Either way, the listener is motorically engaged with different qualities of timbre: it just depends on how the *experience* of this engagement—either vocal or more general—is appraised.

This suggestive result begs for some interpretation. To close out the chapter, therefore, I will briefly stray from the data to speculate a little more freely about what is going on. Merleau-Ponty suggests that perception is an act of "surrender," the moment of fleshly encounter where the perceiving self and the perceived world co-mingle and interpenetrate. Nowhere is this phenomenon more acute than in the realm of music. He writes: "[the] sensible takes possession of my ear … and I surrender a part of my body, even my whole body, to this particular manner of vibrating and filling space…"[59] Hearing is a uniquely vulnerable sense modality: you cannot easily shut yourself off from sounds the way you can with sights. And one can easily see how such an understanding of auditory perception, applied to music—and to timbre, more specifically—could disturb the agency and autonomy of the listener by challenging the borders between self and other, inner and outer. Ever since Plato, the notion of musical "surrender"—the idea that we are susceptible to the corporeal influence of sound—is at the root of questions of both musical affect, and the danger of music.

Timbre literally grips us whether we want it to or not, and this might help to explain its power, not to mention some deeply-rooted cultural anxieties. According to these experiments, we have little control over the extent to which we engage motorically with timbre. We *submit* to it as a reaction; conscious will has little purchase. Timbre thus possesses a certain vehemence, making us move one way or the other through motor resonance.[60] However, we may not *like* the way it makes us move, or the manner of embodiment it suggests. Appraisal—hearing a quality of sound

---

[59] Merleau-Ponty, *Phenomenology of Perception*, 246.

[60] The same can be said of motor empathy in the visual modality. Consider a scene in a film in which the bad guy has something violent happen to him: you still flinch when the knife goes in, even if you don't have any empathy for the character.

as "good" or "bad" (or neutral, for that matter) is predicated upon how you respond to the way such vicarious movement *makes you feel*. When you are listening to "good sounds," you covertly sing along to them: regions responsible for vocal control dominate the resonance. When you are listening to "bad sounds"—timbres you consider "noisy"—you mirror their movement just the same, but you may not like what they are making you do. And you certainly do not join in with something as intimate and personal as your voice! Put another way, in bad appraisals your embodied mind is responding to the motoric invitation of the timbre, while your better judgment and your viscera (embodied in the voice) are rejecting it. Bad sounds, in this scenario, are a form of dissonance between in-built mechanisms of motor resonance—the reaction—and complex, imaginative processes of appraisal mediated by a large number of cultural, historical, situational, and personal factors.

In order to address adequately the ways timbre is conceptualized, endowed with symbolic values, and consciously deployed as a carrier of specific musical meanings—topics of inquiry that will fill the rest of this dissertation—it was first necessary to investigate the simpler, psychobiological bases for these much thicker meanings. I will be referring back to the general theoretical terrain and empirical findings of these opening chapters throughout; but at this point, we cross the imaginary line separating "thin" from "thick," turning our focus to explore how the *acts*, *sounds*, and *perceptions* that form the early stages of ASPECS make their way up the cognitive chain to influence the meaning of timbre. The next stage in the unfolding analysis comes when timbre is transformed into a bodily *experience*, and from there, into an abstract idea, or a *concept*. In Chapter 3, I investigate how human embodiment enables and constrains this process.

Chapter 3

Conceptualizing Timbre

> The closest I ever got to the sound I hear in my mind was […] in the *Blonde on Blonde* album. It's that thin, that wild mercury sound. It's metallic and bright gold with whatever that conjures up. That's my particular sound.
>
> — Bob Dylan[1]

Introduction

In a later attempt to make sense of Bob Dylan's famous evocation of "that wild mercury sound," interviewer Ron Rosenbaum summarized Dylan's creative process: "It wasn't the words that drove him…; it wasn't the melody, it wasn't the ideas, it wasn't self-expression. Rather, it was *a sound in his mind*."[2] The fact that a specific timbral quality, or set of qualities, can exist "in his mind" means that the elusive "wild mercury sound," despite Rosenbaum's comment, is itself an idea (we might even call it an *idée fixe*). Dylan's "particular sound" is a particular *concept*. The language Dylan used to express his personalized concept of timbre, moreover, was central to the conceptualization process itself, allowing Dylan to "image" sound, transforming something previously experienced as amorphous into a unified mental structure; it also allowed him to convey an essential experiential component of his timbral ideal to others, even if ambiguous. His choice

[1] Ron Rosenbaum, "Bob Dylan Interview," *Playboy*, March, 1978.

[2] 'That Wild Mercury Sound': Tangled Up in Bob Again," *New York Observer*, 28 May, 2001. Italics mine.

of metaphors in this interview—thin, wild mercury, metallic, bright gold—kindle new ways of hearing *Blonde on Blonde*. And although his oft-quoted description is no substitute, of course, for simply listening to the album without metaphorical mediation, this particular conceptualization cannot help but shape how it is heard. To those who know of Dylan's elusive *idée fixe*, the album, indeed, sounds like wild mercury. (Or conversely, perhaps "wild mercury" sounds like the album.)

In the same 1978 *Playboy* interview, Dylan told Rosenbaum that, above all else, it was "the sound I'm trying to get across." But what's in a sound? Would a timbre by any other name sound as thin, metallic, and bright gold? As I will argue in this chapter, it would. Metaphors for timbre are not arbitrary, squishy things, but rather reflect deep, embodied, and fundamentally transformative characteristics of sonic experience. "Whatever that conjures up," in short, is not *carte blanche* for unrestrained conceptual abandon; a "thin, "metallic," "bright gold" sound paints a specific timbral picture, one not likely to be mistaken for a "thick," "wooden," and "dull" sound.[3] The coherence of meaning attendant to timbre descriptors, despite their apparent ambiguity, betokens certain consistent, shared patterns of underlying bodily experience. For how else do we know these descriptive qualities than with the senses? The way we talk about timbre both *discloses* aspects of the underlying physical dynamics of sound production and perception, and *enacts* the somatic implications of sound. Metaphor is thus performative. In a very important sense that we will explore throughout this chapter, Dylan's wild mercury sound is metallic and bright gold because he *says* it is.

[3] This statement might appear to directly contradict the decoupling of "brightness" from high spectral centroid, as reported in the previous chapter. The contradiction resolves itself, however, when the metaphor "bright" is interpreted in contexts other than the psychoacoustic (e.g., "she has a bright personality"), including its literal meaning. While "brightness" may map reasonably well onto spectral centroid in controlled, laboratory conditions, its meaning is clearly diffused when applied to actual music listening. Its ambiguity in the Chapter 2 experiments, I would argue, is not evidence for its *incoherence* as a descriptor; rather, it demonstrates that common vocabulary for timbre can interact in unpredictable ways with the acoustics of timbre, especially when cultural context is considered.

So let's talk about timbre. Or rather, let's talk about *how we talk* about timbre. I want to open the chapter with a question: Why do we use the words we do to describe timbre? Put differently, why do we describe qualities of sound as "thin" and "bright" and not, for instance, "jealous" and "bipedal"? Is it purely a matter of linguistic convention? To explore this question, I am going to draw primarily on the embodied cognition paradigm of Mark Johnson and the cognitive linguistics of Lakoff and Johnson; I will also close with a foray into the neuroscience of metaphor, including a surprising result from Experiment 3. Though there is a large body of literature on music metaphors in general, few writers explicitly incorporate insights from this rich corpus of cognitive research into discussions of timbre. This is a lamentable omission: since timbre lacks its own domain-specific vocabulary, it is preeminently metaphorical (i.e., it pervasively borrows from other sensory domains); a more thorough understanding of metaphor is thus, I argue, critical to developing a working explanatory model for timbre. Dylan's comment was not just whimsical and poetic turn of phrase (though it was both); it reveals something fundamental to the way we conceptualize musical sound and noise.

Experiential Structures of Timbral-Corporeal Events

In *The Body in the Mind* (1987), Mark Johnson posits a cognitive connection between the basic, recurring experiences of human embodiment—e.g., proprioception, somatosensation, and common motor programs—and higher levels of human understanding, including language, imagination, and reasoning.[4] This account, which aims to "put the body back into the mind,"[5] shows that

---

[4] Johnson, *The Body in the Mind: The Bodily Basis of Meaning, Imagination, and Reason*. I thank Susan McClary for initially pointing me in the direction of this work, and Mark Johnson for agreeing to sit on my committee and assist me in applying his ideas to an embodied framework for timbre.

human meaning is both facilitated and constrained by the experience of having a (human) body. For example, basic experiential structures such as physical balance, orientation in space (up/down, left/right), and physical containment (in/out) shape the way we think by connecting up with basic conceptual structures. This is accomplished through metaphor (or, in Lakoff and Johnson's formulation, *conceptual metaphor*), which maps bodily structures of meaning onto different domains.[6] In contrast to "objectivist" accounts of meaning, which focus virtually entirely on the logical structure of language, Johnson argues that preconceptual, nonpropositional bodily experience plays a dynamic role in the process of human meaning-making. Bodily experiences do "not merely form a background against which meaning emerges; rather, they *are themselves meaning structures*."[7]

Johnson designates experiential meaning structures *image schemata*: "In order for us to have meaningful, connected experiences that we can comprehend and reason about, there must be pattern and order to our actions, perceptions, and conceptions. *A schema is a recurrent pattern, shape, and regularity in, or of, these ongoing ordering activities*. These patterns emerge as meaningful structures for us chiefly at the level of our bodily movements through space, our manipulation of objects, and our perceptual interactions."[8] Image schemata thus have *gestalt* characteristics; they are "coherent, meaningful, unified wholes" with consistent internal structures. They are not "arbitrary and 'mushy'" forms, but rather comprise a commonly shared set of fundamental, recurring fea-

---

[5] Ibid., xxxvi.

[6] We will flesh out more of the details of this thesis as we progress, but to briefly illustrate the principle of conceptual metaphor in relation to the experience of *balance*: we commonly project this basic experiential structure onto our understanding of human *psychology* (e.g., "he is a stable guy"), *rational argument* (e.g., "I pile up arguments" and two ideas have "equal weight"), and *legal/moral balance* (e.g., "an eye for an eye"), among many others. See ibid., 74-98.

[7] Ibid., 48. Italics in original.

[8] Ibid., 29. Italics in original.

tures of subject/environment interaction.[9] We begin to form image schematic understandings of the world in infancy, when we are first learning to physically make our way through the environment, but these schematic structures reinforce themselves continually throughout our lives: the act of eating, for instance, draws upon the schema of *containment* (putting food *inside* your mouth); the act of walking calls upon *force* and *path* schemata (moving *quickly to* the store).

As one of our most pervasive and compelling behaviors for symbolically ordering reality, musical conceptualization is subserved, as well, by image-schematic structures. The bodily basis of musical metaphor has been discussed in depth in the literature,[10] so at this point, having reviewed the basic theoretical framework, I want to illustrate in generalizable terms some central schematic structures underlying the main themes of this project, particularly the relationship between sound production and bodily exertion at the point where timbre becomes "noisy." I will leave the grand theorizing for later.

The basic underlying experiential structure for understanding timbre, I propose, is Johnson's *force schema*.[11] The experience of exerting force upon the environment (and having force exerted upon you *by* the environment) is so primary that it may not seem *forceful* at all, especially in

---

[9] Ibid., 41.

[10] The majority of studies, however, focus on pitch, motion, and text-music relationships, not timbre. For a chronological sampling, see Janna Saslaw, "Forces, Containers, and Paths: The Role of Body-Derived Image Schemas in the Conceptualization of Music," *Journal of Music Theory* 40, no. 2 (1996); Cox, "The Metaphoric Logic of Musical Motion and Space."; Zbikowski, *Conceptualizing Music: Cognitive Structure, Theory, and Analysis*; Mark Johnson and Steve Larson, "'Something in the Way She Moves'—Metaphors of Musical Motion," *Metaphor and Symbol* 18, no. 2 (2003); Echard, *Neil Young and the Poetics of Energy*; Johnson, *The Meaning of the Body: Aesthetics of Human Understanding*; Lawrence M. Zbikowski, "Music, Language, and Multimodal Metaphor," in *Multimodal Metaphor*, ed. Charles Forceville and Eduardo Urios-Aparisi (Berlin: Mouton de Gruyter, 2009); "Musicology, Cognitive Science, and Metaphor: Reflections on Michael Spitzer's *Metaphor and Musical Thought*," *Musica Humana* 1, no. 1 (2009); Zachary Wallmark, "'I Would Like to Die': Constructions of the Eroticized Body in the Monteverdi Madrigal" (paper presented at the Material and Imagined Bodies Conference, Brown University, Providence, RI, 2010); Larson, *Musical Forces: Motion, Metaphor, and Meaning in Music*.

[11] For a detailed reading of *force*, see Chapter 3, Johnson, *The Body in the Mind: The Bodily Basis of Meaning, Imagination, and Reason*. I also want to acknowledge that I am not the first to relate *force* to questions of musical sound and affect; to my knowledge, the first to do this was Walser, "The Body in the Music."

instances where motor intentions translate perfectly into specific outcomes. We possess, to recall Leman, an "action-oriented ontology"; the world, then, can be conceived as a field of possible activities, a space in which the "I can" is continually rubbing up against the material presence of other objects, natural forces, and people.[12] We experience *force* most vividly when our intentions are somehow frustrated: you want to get to the top of a hill, but the slope exhausts you; you want to open a door, but it is jammed shut. Applied to the topic of this chapter, we can see that *force* is involved in virtually every timbre-generating situation; and via the dynamics of act–sound coupling discussed in Chapter 1, *force* is also involved, vicariously, in the *audition* of timbre. I will be calling this act–sound interaction (following Eidsheim's *timbre corporeal* coinage)[13] the "timbral-corporeal event": all natural timbres are acoustic byproducts of corporeal (inter)action in the musicking event, and all actions require force. Producing musical sound—singing, pounding a drum, blowing a horn—is thus structured at its deepest level by instantiations of *force*.

But the *force* schema is only a very general category of experiential gestalt. To sharpen the focus, I would propose four schematic sub-categories that apply in special ways to timbre: *enablement, blockage*, *exertion*, and *ultimate limit.* Figure 3.1 shows the interacting levels of schematic organization: *force* here is the umbrella category, with the other schemata representing specific types of force.

---

[12] This formulation is Merleau-Ponty's. He posits that, prior to the consciousness of "I am" (derived from *Cogito*), we formulate our being-in-the-world based on "I can," a series of negotiations with outside reality involving our willed physical capabilities as they rub up against the limits of the possible. See Merleau-Ponty, *Phenomenology of Perception*, 159.

[13] Eidsheim, "Voice as a Technology of Selfhood: Towards an Analysis of Racialized Timbre and Vocal Performance," 253.

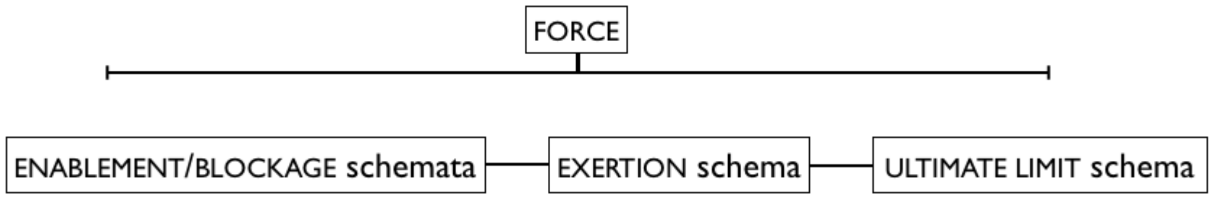


FIGURE 3.1: Image schemata of (noisy) timbre

Enablement is experienced preconceptually as the absence of constraints when performing a given action. Motor intentions, in this schema, are freely directed toward a given purpose without being frustrated by counteracting forces. In timbral-corporeal terms, this is the sound of singing with full, effortless power and mastery, or finding that "sweet spot" on an instrument; in such situations we might come to feel our voice or our instrument as a transparent extension of our will. Sound production is unfettered by material obstacles (Fig. 3.2).

FIGURE 3.2: ENABLEMENT

But the physical act of producing sound can as easily come up against forces that thwart the clear and direct movement of our goals. Johnson describes numerous such obstructive image schemata, but perhaps of most interest to timbre is the experience of *blockage*. Here we are confronted with a material impediment to the open and unobstructed production of sound, one that requires evasive maneuvering, such as might occur when a woodwind player is using a bad reed, or a singer is performing with a stuffy nose. When involuntary, it can be experienced as a frustration of intentions, and this feeling of inhibition can take forms beyond the purely material. (To a concert performer a dreadful sound technician can be experienced as a *blockage*.) But *blockages* are not uniformly negative: often we *want* to sonify the presence of certain material obstacles, and indeed, "blocked" sounds are ubiquitous in many musical cultures. A harmon mute on a trumpet is a desired blockage, for example. A *blockage* in the material sense does not lead to the cessation of sound, but it does mean that sound production is mediated by a physical obstacle; and this process of negotiation leaves a mark on the resultant timbre (Fig. 3.3). *Blockage* has an associated *sound* specific to each forceful timbral-corporeal event.

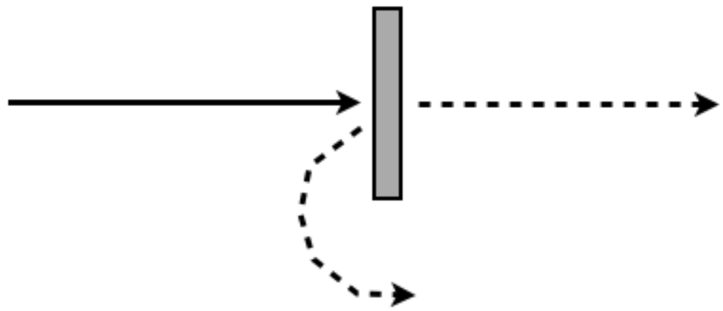


FIGURE 3.3: BLOCKAGE

I group these two together because *enablement* and *blockage* often work in tandem, but there are other force schemata at play in how timbre is typically produced and conceptualized. Cox proposes an *exertion* schema comprised of the strength, duration, and "shape" of a musical gesture.[14] Adding to this definition with a focus on timbre, we can think of the *exertion* schema (*a la* Herbert Spencer)[15] as a linear scale of *low* to *high* (Fig. 3.4). Some vocal acts require a minimal level of physical effort (such as humming in a comfortable, middle register), while others demand substantial output of energy (such as shrieking at the top of your lungs). The same is true for many—but not all—musical instruments. Loudness is a primary auditory dimension connected to the *exertion* schema, but increasing levels of bodily exertion also leave timbral traces, particularly in the form of a-noise (*high-frequency energy, noise,* and *roughness*). *Exertion* interacts in idiosyncratic, contingent ways with both *enablement* and *blockage*: you can produce sound using consider-

[14] Cox, "Embodying Music: Principles of the Mimetic Hypothesis," 45.

[15] As discussed in Chapter 1, Spencer viewed musical expressivity as a function of linear *intensity* based in the voice.

able effort but still feel unimpeded by material limits, which only manifest themselves at the outer edges of physical possibility. Consider a trumpeter playing with confidence, power, and ease at the stratospheric end of his instrument ("screaming"): this experience might call upon both a sense of *enablement* and high *exertion.* But *blockage* always threatens. Even the greatest high-note trumpeter must confront the material boundaries that demarcate the relationship between body and instrument, and as this limit approaches, timbral effects of *blockage* may appear (perhaps a pinched, wobbly, and thin tone). In addition to an obstacle, then, *blockage* in this case presents itself as an index of encroaching physical limits.

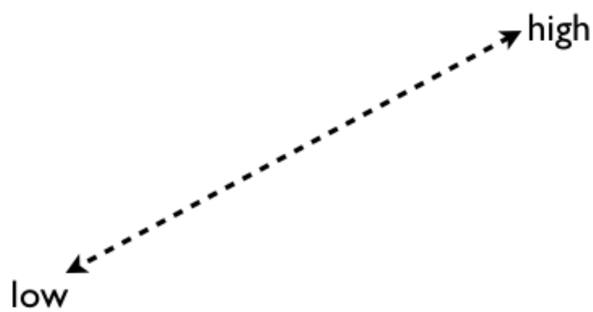


FIGURE 3.4: EXERTION

This sensed limit is grounded in what I am calling the *ultimate limit* schema. In contrast to these other force structures, an *ultimate limit* is, as the name implies, absolute: it is the experience of force running up against the boundaries of material possibility. In terms of sound production, this is the point where physical exertion and blockage threaten to go silent—a trumpet cannot perform past certain limits of loudness and pitch, all singers know intimately their vocal capacities and the notes they simply *cannot* hit, all guitar amplifiers (no matter the power) have their breaking point. These limits certainly change, but even when they do, new thresholds present themselves; it may be ever-receding, but all material sound production is circumscribed by the confines of the material (Fig. 3.5). This schema, unlike the others discussed here, is rarely actualized in normal experience. I am arguing that *enablement*, *blockage*, and *exertion* have sonic correlates, but *ultimate limit* is unsounded (or perhaps negative-sounded) because it represents a limit case of embodied force interactions, a point in sound production that is always theoretically possible but seldom (thankfully) breached. A singer who experiences a sudden and painful crash of the vocal apparatus in performance is one example; a blown amp and a trumpeter who squeezes so hard that the sound cuts out entirely are two others. For other instruments, like the piano, *ultimate limit* is a rarity given the limited force potential of human hands. Like the other schemata, *ultimate limit* is governed both by the physical invariants of the body and our sound-producing technologies, and by context.

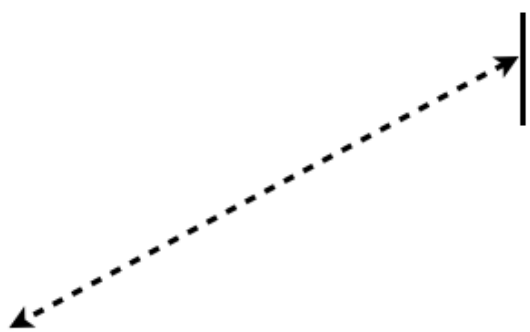


FIGURE 3.5: ULTIMATE LIMIT

In aggregate, these four schemata of force lay the basic bodily grounding for a variety of timbre conceptualizations. But retracing our steps with greater attention to the underlying force dynamics of tone production will hopefully clarify certain aspects of bodily experience, enabling us to place timbre into a rich metaphorical framework in the next section.

Let us begin with a different sort of diagram, one that captures the general essence of how sound-acts are often conceptualized while highlighting the interaction between the physical energy and materiality of sound production (Fig. 3.6).[16]

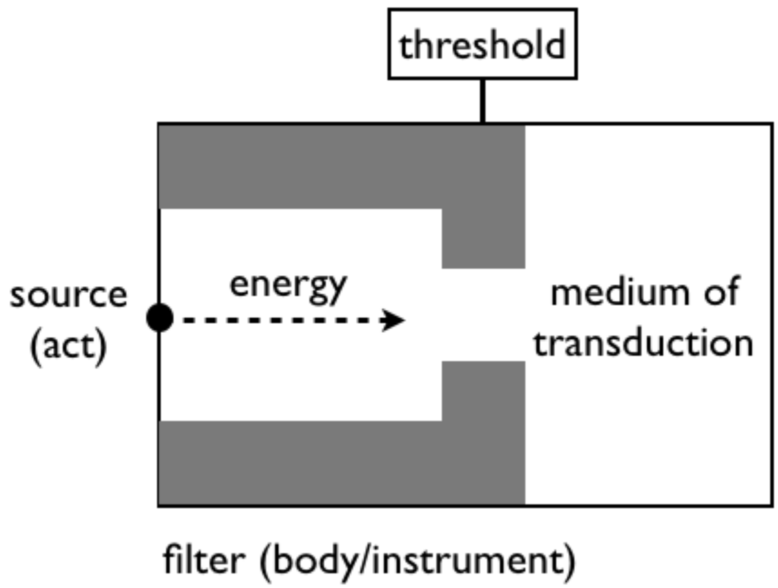


FIGURE 3.6: A timbral-corporeal event

First, following the lead of the source/filter model discussed in Chapter 1, we can conceive of the sound-producing act as the *source* of excitation: in the voice, this is the vibration of the vocal folds; in the saxophone, this is the buzzing reed. This act, which requires energy, also produces vibratory and (eventually) acoustic energy that propagates outward away from the source. We are still, at this point, well within the confines of the body/instrument system, or the *filter* (represented here by the gray gate-like structure). But most filters have a material threshold; I have visualized it here as the narrow opening of the filter, but this is obviously not the only physi-

[16] The obvious exception here is electronic sound, but even in this case it could be argued that we hear purely electronic timbres through the ecological lens of the various physical models our species with which our species have evolved. For an amazing analysis of the role of perceptual illusion in electronic dance music, see Fales, "Short-Circuiting Perceptual Systems: Timbre in Ambient and Techno Music."

cal form a threshold can take. A physical threshold in the sound production system can be conceived as the point at which source and filter cease to function in balance, leading to timbral changes. It can thus be understood in relation to the *blockage* schema as the point at which the physical strain of sound production becomes audible. Usually this situation involves greater energy from the source than the filter is able to handle without the addition of timbral anomalies, but it can also entail a change in the filter. This stage can also be the threshold between the sound production system (interacting source and filter) and the outside air (or other medium of transduction).

Now, let us picture three different types of basic timbral-corporeal events scaled for increasing degrees of *exertion*. In the first condition (Fig. 3.7), the force that goes into the source is low to moderate in intensity, and the energy is shaped by the filter in the way it is "meant" to be shaped; the energy is far from exceeding any thresholds (represented here with a blank square), and the sound emerges unscathed by any *blockage*. In short, the sound production system is functioning in perfect sync. Again, defining what exactly comprises a balanced and in-sync timbre production system is too situational and contingent to make sense in this context, but one can imagine many such situations: this is a condition of "normal" sound production in typical musical situations. Timbrally, there is nothing out of the ordinary. Represented as a formula, this event would be described as $E < T$, where E is the energy of the source and T is the energy threshold of the filter.

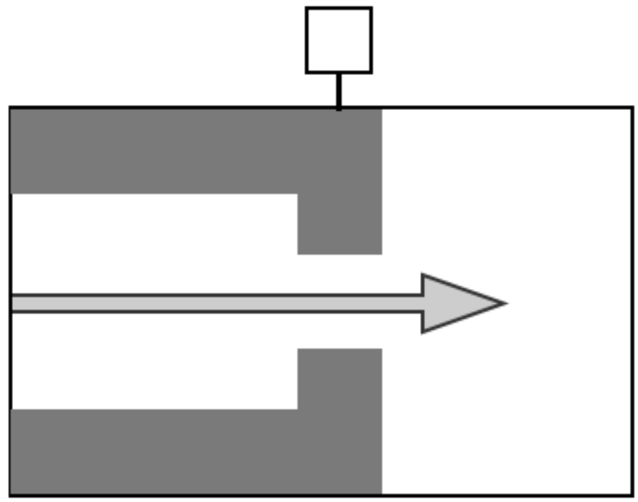


FIGURE 3.7: Type 1 (low exertion timbral-corporeal event)

Now consider an increase in the total energy entering the system through the source (Fig. 3.8). Here the energy—and underlying physical force—is considerably greater, but the filter is still able to transform this vibratory energy into sound without the addition of any timbral irregularities. The exertion is audible, but it does not present itself in the form of "noise." The threshold here remains untouched, though the anticipation of an encroaching breach might begin to assert itself phenomenologically (indicated in gray). As a formula, this would be expressed $E \leq T$.

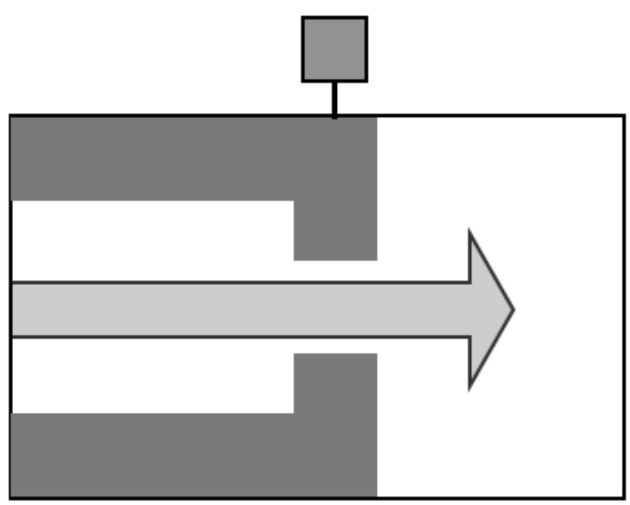
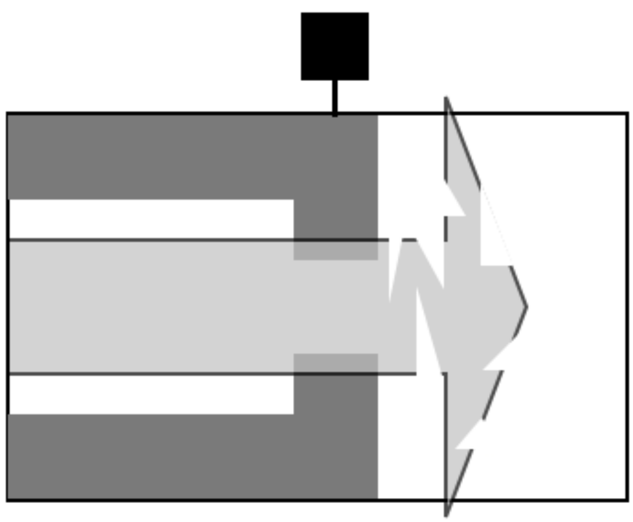


FIGURE 3.8: Type 2 (medium exertion timbral-corporeal event)

Next, let us imagine pushing even more energy into the sound production system (Fig. 3.9). This sort of timbral-corporeal event, in contrast to the others, reaches the material threshold of the filter and *exceeds* it (indicated in black). Timbre in this instance is marked by the "sound of arousal" (a-noise) as the materiality of the filter fails to match the energy of the source.[17] Again, this basic sensorimotor structure originates in vocal experience, but applies in key ways to many instruments, and even to certain aspects of electro-acoustic sound. Since the limits of the filter are effectively obstructing the energy of the source—as visualized in the overlap—this moment can be conceptualized in terms of the *blockage* schema; *something material* is standing in the way of faithful transmission, and friction is required to overcome it. This *blockage* can occur in couplings of corporeal with corporeal materials (such as the voice), corporeal/mechanical interfaces (such as playing an instrument), and mechanical/mechanical interactions (Table 3.1).



[17] It should go without saying that not all noisy timbre is the result of high states of exertion. On many instruments, "noisy" effects (such as trumpet growling) are simply a *technique* that may or may not be coupled to high exertion, and in electric and electronic sounds the connection to embodiment is clearly far from firm.

FIGURE 3.9: Type 3 (high exertion timbral-corporeal event)

This situation transmits the timbral consequences of extreme bodily force—*blockage* being a particular material manifestation of timbral-corporeal exertion—but the bodily meaning of this moment is determined entirely by context. Thus, in certain situations the appearance of "noisy timbre" functions as an alarm signal indicating immanent collapse; in others it may betoken a lack of control, or a sign of mechanical damage or physical frailty or incompetence; in others still (and most commonly in popular music) it signals extreme power, the ability of the musician to *surpass* the material limits of the filter. Though the image schema is rooted in human embodiment, the interpretation of what this liminal timbral moment *means* calls upon much more complex, symbolic orders of explanation. But as we will discuss later, conceptualizations of noisy timbre are still connected to these basic experiential gestalts, as evinced by the metaphors commonly used to describe them, just as appraisals are founded upon reactions. (Stated as a formula, this timbral-corporeal event would be represented as $E > T$.)

TABLE 3.1: Three types of *blockage* causing noisy timbre

Coupling Type	*Source of Blockage*	*Examples*
corporeal/corporeal	Source and filter are both parts of the human body. *Blockage* occurs at interface between two or more corporeal bodies, e.g., tongue blocking airstream, laryngeal contraction impeding vibration of vocal folds, etc.	- vocal noise: growling, screaming, harsh voice, "shhh" sound, etc.
corporeal/mechanical	Source is corporeal (e.g., airstream, fingers) and filter is mechanical (e.g., instrument). *Blockage* occurs anywhere in body/instrument system.	- "blatty" brass sound (i.e., overblown instrument) - saxophone multiphonics (i.e., vocalized pitch "blocks" fingered pitch) - *shakuhachi* noise techniques

mechanical/mechanical	Source and filter are both mechanical/electrical. Noisy timbre manifests in live sound reproduction (e.g., through loudspeakers or headphones) or recording medium. *Blockage* is typically related to voltage thresholds or physical thresholds of transducers.	- electric guitar distortion - mixing "in the red" (e.g., Public Enemy's Bomb Squad) - microphone clipping (e.g., Aretha Franklin's classic recordings)

There is an implied telos to these progressively higher-exertion schematic types, and indeed, the *ultimate limit* force schema indicates that every material source/filter relationship has a final, unsurpassable threshold (Fig. 3.10). In this scenario, the *blockage* is total, and there is no alternative passage for sound to reach the air. Such a condition, as we know from bodily experience, is not necessarily permanent; perhaps a trumpet player with busted chops can make a slight embouchure adjustment that allows her to keep playing, an example of the *removal of restraint* schema. But the essential point to understand here is that *ultimate limits* are usually experienced as *potentialities*, not realities; we experience the *ultimate limit* schema in timbral-corporeal contexts not usually by running headlong into it, but by sensing its unforgiving and seemingly absolute limits then staying clear. This structure can be expressed as $(E > T)/0$, where the nullifying denominator cancels the sound of the nominator.

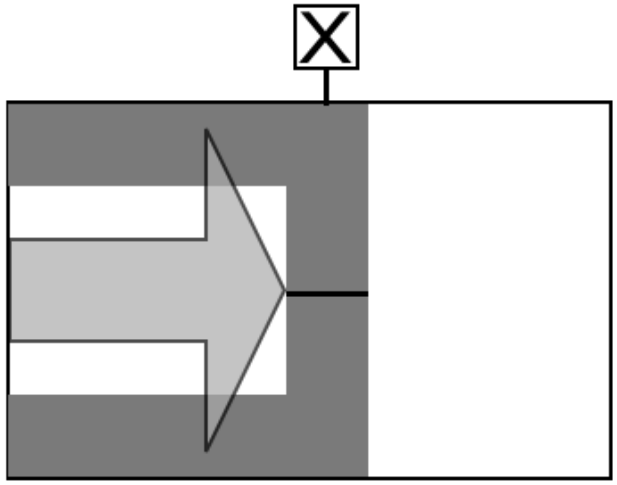


FIGURE 3.10: Type 4 (threshold condition)

Since there is such an elemental bioacoustic correlation between arousal/exertion and a-noise—a link that is also reflected in *perception*, as we saw in Chapter 2[18]—we can visualize the relationship between the event types above as a line (Fig. 3.11). As modeled here, a certain level of real and perceived exertion is required for sound production to be heard as "musical": performances described as *dull* or *flat*, for example, often betoken a "nonmusical" deficit of exertion. Once this requirement is met, however, a range of exertion is possible while still producing "musical" tone; but at some threshold, elements of noisy timbre begin to betray the physical labor involved in production. If exertion drives the source/filter system beyond this already liminal "noise" zone, moreover, we reach a real or imagined endgame.

[18] The correlation between perceived *bodily exertion* and p-noise was, averaged across Experiments 1 and 2, a statistically robust $r = .76$ (on a scale of 0 to 1). See Glossary for explanation.

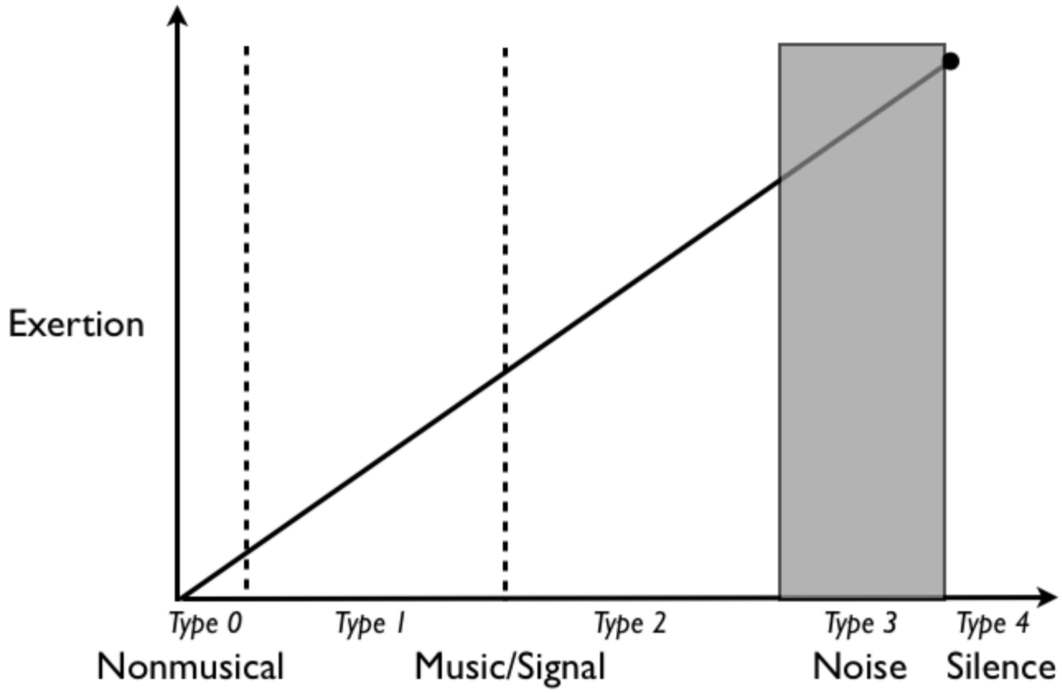


FIGURE 3.11: Linear model relating (noisy) timbre to *exertion*

The force dynamics of timbral-corporeal events have audible consequences that connect up to how these timbral qualities are conceptualized. But *bodily image schemata are entirely preconceptual:* they are physical, experiential structures of meaning, not abstract concepts. Schemata facilitate and constrain the conceptualization process, but they are not themselves mental structures. So how do these basic bodily structures structure cognition? Between the schema and the concept is a crucial middle term.

Is Timbre a Metaphor?

In their pioneering book *Metaphors We Live By* (1980), George Lakoff and Mark Johnson make the case that metaphor allows us to understand one domain of experience in terms of another. (For example, the metaphor PITCH IS VERTICAL SPACE enables the conceptualization of notes along a vertical axis; e.g., *rising* and *falling* pitches.)[19] As Johnson later explained, image schemata provide us with the fundamental structures of experience that serve as the metaphorical basis for abstract thought: "The projection of definite, and highly articulated, structure [is] always from the domain of the physical (as it is preconceptually patterned) onto the social, epistemic, and speech-act realms. Such projections of structure from one experiential domain to another domain *of a different kind* are metaphorical processes."[20] In this expanded definition, then, "a metaphor is not merely a linguistic expression (a form of words) used for artistic or rhetorical purposes; instead, it is a process of human understanding by which we achieve meaningful experience that we can make sense of."[21] Put differently, metaphors synthesize pre-organized sensorimotor experience into concepts about the world.

Bob Dylan was not the first to describe his "sound" metaphorically, of course—timbre has long been understood in terms of other domains of experience. Twelfth-century translator Dominicus Gundissalinus, for example, noted the paucity of vocabulary for describing differences in tone color, commenting that sound qualities "had no name of their own" and instead had to be described metaphorically with reference to the other senses.[22] This situation has continued to the present: throughout this dissertation we have identified *bright*, *metallic*, *nasal*, *rough*, and many other types of timbres without giving much thought to their underlying metaphorical structures. Descriptive terms like these, as common as they are, are not an essential part of the perception process—we can make sense of timbre without verbalizing it. However, "perceptualization"

---

[19] As has become the convention in cognitive linguistics, I will be using Lakoff and Johnson's SMALL CAPS to denote metaphors.

[20] Johnson, *The Body in the Mind: The Bodily Basis of Meaning, Imagination, and Reason*, 62-63. Italics in original.

[21] Ibid., 15.

[22] See Charles Burnett, "Sound in the Middle Ages," in *The Second Sense: Studies in Hearing and Musical Judgement from Antiquity to the Seventeenth Century*, ed. Charles Burnett, Michael Fend, and Penelope Gouk (London: University of London Press, 1991), 63.

(Fales 2002) is not the same as *conceptualization*; timbre becomes an abstract mental structure through the workings of language, when we *name* the quality we are hearing. The best sources for rich descriptive vocabulary of timbre, therefore, are contexts where people need to communicate in practical, concrete, and results-driven ways about tone: music pedagogy,[23] music criticism,[24] orchestration manuals,[25] sound technology/instrument magazines and trade journals,[26] music search and retrieval systems,[27] online music communities,[28] and the discourse of the recording studio,[29] for example.

As a window into the topic, I will begin with the useful taxonomy of timbre description offered by Steven Feld *et al.* in their ethnography of the modern American recording studio. The authors suggest that producers, musicians, and sound engineers trade in a codified descriptive language for timbre consisting of five general categories: (1) spoken/sung vocables, (2) lexical onomatopoeic metaphors, (3) "pure" metaphor, (4) association, and (5) evaluation.[30] Thus, a

---

[23] Eidsheim discusses the pedagogical dimensions of vocal timbre in Eidsheim, "Voice as a Technology of Selfhood: Towards an Analysis of Racialized Timbre and Vocal Performance."

[24] Metaphor is, of course, a core currency of music journalism and criticism. Roland Barthes writes: "If one looks at the normal practice of music criticism (or, which is often the same thing, conversations 'on' music), it can readily be seen that a work (or its performance) is only ever translated into the poorest of linguistic categories: the adjective. Music, by natural bent, is that which at once receives an adjective." See Barthes, *Image, Music, Text*, 184.

[25] Kendall and Carterette, for example, performed an exhaustive study of the timbre descriptions used by Walter Piston in his 1955 manual *Orchestration*: see Kendall and Carterette, "Verbal Attributes of Simultaneous Wind Instrument Timbres: II. Adjectives Induced from Piston's 'Orchestration."

[26] Robert Fink has offered a critical reading of the political economy of "tone" in electric guitar magazine advertisements: see introduction to Fink, O'Brien, and Wallmark, *The Relentless Pursuit of Tone: Timbre in Popular Music*. For an in-depth discussion of timbral descriptors of a flue organ pipe, see Vincent Rioux and Daniel Västfjäll, "Analyses of Verbal Descriptions of the Sound Quality of a Flue Organ Pipe," *Musicae Scientiae* 6, no. 1 (2002).

[27] See the second part to Leman, *Embodied Music Cognition and Mediation Technology*.

[28] Rafael Ferrer and Tuomas Eerola, "Semantic Structures of Timbre Emerging from Social and Acoustic Descriptions of Music," *Journal of Audio, Speech, and Music Processing* 11(2011).

[29] Paul Théberge, *Any Sound You Can Imagine: Making Music/Consuming Technology* (Middletown, CN: Wesleyan University Press, 1997); Feld *et al.*, "Vocal Anthropology."

[30] "Vocal Anthropology," 324-25. For another helpful account of common categories of timbral description using empirical methods, see Graham Darke, "Assessment of Timbre Using Verbal Attributes" (paper presented at the Conference on Interdisciplinary Musicology, Montreal, Quebec, Canada, 2005).

snare drum tone might be variously described as (1) /dz:::/, (2) *buzzy*, (3) *dry*, (4) "second-line," and (5) "like the drummer for Rebirth Brass Band." I will focus here on the idea of "pure" metaphor advanced by Thomas Porcello, one of Feld's co-authors. In contrast to much of the timbre literature—which views descriptors in primarily *adjectival* terms—Porcello acknowledges the fundamentally metaphorical nature of timbre description: timbre is understood in relation to another domain of a different kind. He defines "pure" metaphor thusly:

> "Pure" metaphor, in which timbral features are invoked not via sonic iconicity but with reference, generally, to other sensorial domains. Examples include *wet*, *dry*, *deep*, *bright*, *round*, each a synaesthetic metaphor. The majority of pure metaphors used to describe musical timbre are synaesthetic to touch and sight, the latter set almost always invoking spatial relationships. Many of the pure metaphors are highly codified, especially among sound engineers who recognize, for example, that a sound described as "too boxy" (a negative timbral trait) can regularly be corrected by dampening all frequencies between 250 and 500 Hz.[31]

Without using the terminology, what Porcello describes here is conceptual metaphor: we understand "bright timbre" in the auditory domain with reference to "brightness" in the visual domain. While highlighting the metaphorical and—importantly—synaesthetic nature of timbre conceptualization, however, Porcello does not explore *why* people describe qualities of timbre using the conventional lexicon they do. Are the metaphors we use to talk about timbre arbitrary? In the context of the studio, after all, the main concern is the clear communication of a desired timbral quality; we could as well call a particular timbral parameter "doorknob" so long as it possesses widely shared referential currency and the communicative efficacy to produce consistent sonic results. So why don't we?

The burden of the preceding discussion implies that "pure" metaphors of timbre are not merely fanciful linguistic designations, but indexes of bodily relationships to sound. The force

---

[31] Feld *et al.*, "Vocal Anthropology," 324.

schemata of *enablement*, *blockage*, *exertion*, and *ultimate limit* are involved in most if not all timbral-corporeal events. These gestalts, moreover, entail some very commonplace metaphorical projections, three of which are of particular relevance to this project (Fig. 3.12): INSTRUMENTS ARE VOICES, SOUND IS MATERIAL, and NOISE IS FRICTION.

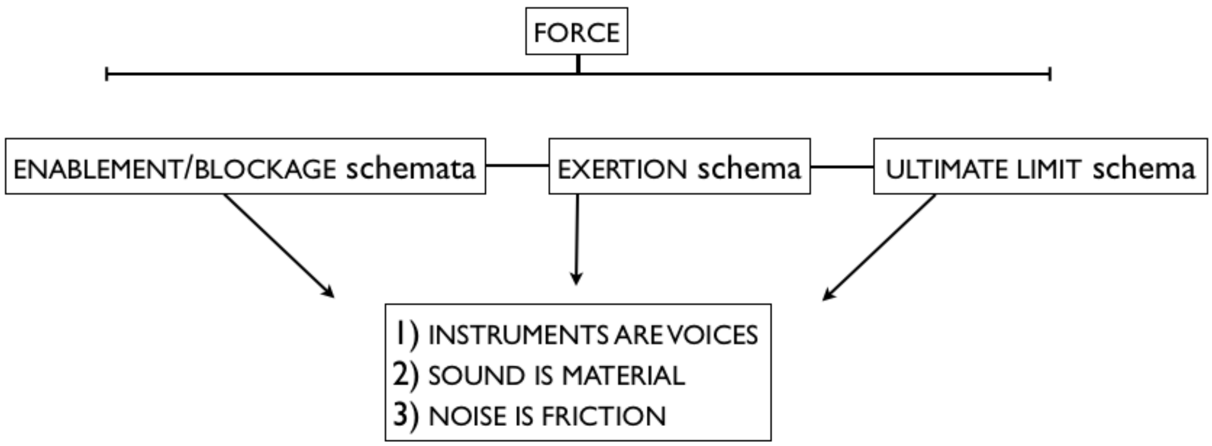


FIGURE 3.12: Image schemata and metaphorical projections of (noisy) timbre

To discuss each of these in turn: As noted by Cox, we often understand aspects of non-vocal timbre in terms of the human and non-human voice, thus the metaphor INSTRUMENTS ARE VOICES.[32] This metaphor manifests itself in many everyday descriptions of instrumental sound, from the *throaty* clarinet to the *cackling* trombone (Tab. 3.2). Considering the ample evidence for a privileged position of the voice (via subvocalization) in timbre perception and cognition, as advanced in Chapter 1, this metaphor is hardly mysterious. Less commonly, moreover, it works the other way around: VOICES ARE INSTRUMENTS describes vocal sound in analogy to instruments (e.g., *horn-like* singing, voices that are *flutey* or *brassy*, etc.).

---

[32] Cox, "The Mimetic Hypothesis and Embodied Musical Meaning," 202.

Next, facilitated by the schematic structures identified previously, we commonly conceptualize timbre as a *material thing*. Thus, with the help of this metaphorical scaffolding, timbre is described as possessing weight, mass, shape, texture, temperature, and every other quality of an object (Tab. 3.2).[33] It is important to stress (with Porcello) that this metaphorical type is fundamentally *synaesthetic*: not only are we understanding one domain in terms of a different domain; SOUND IS MATERIAL enables us to conceptualize one *sense* (hearing) in terms of adjectives from another sense (usually touch and vision). While it may be argued that such metaphors are simply a matter of verbal convention and nothing more, I maintain that SOUND IS MATERIAL demonstrates an ecological understanding of sounds as "sounds of things": we cannot escape the fact that sound-producing interactions involve bodies and objects (each with weight, shape, texture, etc.).

TABLE 3.2: Common timbre metaphors

Metaphors	*Examples*
INSTRUMENTS ARE VOICES	adenoidal, barking, belching, bellowing, biting, braying, breathy, cackling, catarrhal, chesty, chirpy, coughing, croaky, crying, groaning, growling, gruff, guttural, hissing, hoarse, howling, husky, lyrical, nasal, pinched, roaring, screaming, shouting, shrieking, sibilant, snarling, spitting, squealing, throaty, wailing, wheezy, whispery

[33] Many words in the example list are multi-modal (e.g., *dull* can describe a personality type just as a physical quality). However, their primary usage references material things. I also want to point out that I am omitting adjectives describing aspects of light contrast and visual appearance from this example list (e.g., bright/dark, shimmering) to focus instead on metaphors with tactile implications.

SOUND IS MATERIAL	airy, bell-like, blunt, boxy, brassy, brittle, clean, clear, cold, crisp, deep, delicate, dense, diaphanous, dirty, dry, dull, empty, fine-spun, fat, flaccid, floppy, fluffy, fluid, frayed, frosty, full, gauzy, glassy, glittery, gossamer, grainy, grimy, hard, heavy, hollow, honey-coated, icy, light, limpid, liquescent, loose, metallic, mucky, muddy, mushy, pliant, plush, reedy, rippling, round, rubbery, sandy, shallow, silky, silvery, sizzling, slick, smooth, soft, sticky, stringy, thick, thin, tight, tinny, translucent, velvety, viscous, voluminous, warm, wet, wooden, wooly
NOISE IS FRICTION	abrasive, bristly, course, cutting, grating, gravelly, grinding, gritty, harsh, jagged, jarring, penetrating, piercing, prickly, raspy, rough, rugged, scouring, scraping, scratchy, sharp, spiky

The SOUND IS MATERIAL metaphor can be grouped into roughly four sub-types, each with increasing cognitive complexity. First, we commonly *name the source directly*. This is the simplest and most obvious level of description; it can be understood as a form of metonymy where we understand the source in terms of its associated sound. Thus, we identify and describe a timbre that matches our model for a tenor saxophone *as* the "tenor saxophone." At this basic level, conceptualization consists in simply naming the material source that produces it.[34]

Second, SOUND IS MATERIAL can reference the *physical qualities of the source* while not naming the source directly. For instance, the sound of a cymbal might be described as *metallic*, a trumpet as *brassy*, and a bass clarinet as *reedy*.

[34] But we should not be so quick to take the relationship between source and sound for granted. To be sure, the sound of a saxophone *is not* the saxophone itself—it is a sonic index indicating the presence of the saxophone, but it is one level removed from the thing itself. Owing to a lifetime of embodied experience, this relationship is usually completely transparent and not given a second thought, but when the source is perceptually confused with something else—for example, hearing an English horn as a soprano saxophone, or a male's voice as a female's—the metaphorical quality of this level of description is made clear.

Third, we use descriptions that *blend physical and connotative elements of source and sound*. Many (perhaps most) metaphors for timbre grounded in the SOUND IS MATERIAL schema work this way by blending known elements of the physical source with physical and emotional qualities associated with other objects, and with the imagined performer. Thus, we might hear a *light* flute, *warm* voice, *delicate* piano, *heavy* electric guitar, *hollow* bassoon, and *thin, metallic, and bright gold* Dylan record. The quality of "lightness," for example, applies to both flute and flute timbre: the flute is a *light* instrument owing to its physical size, pitch range (high), and cultural connotations (the flute is gendered female), and these qualities are transferred onto the sound. But timbral descriptors are also free to borrow from other objects not directly associated with some aspect of the source. Therefore, we might hear a guitar as *heavy* even though the instrument itself would not be described as such, or a piano as *delicate* despite its bulky, 500-pound frame.

And fourth, the SOUND IS MATERIAL metaphor allows us to conceptualize timbre in terms of *physical qualities of unrelated objects*. Consider the *velvety* string section backing Marvin Gaye, or the *silky* voice of the crooner. Velvet, of course, plays no role in violin design, but the tactile qualities of this material are synaesthetically transposed into sound. We call lush strings "velvety," mapping the feeling of touching things that are soft, furry, or downy onto sounds with no literal connection to soft, furry, or downy things. Though there are acoustic reasons for these metaphorical mappings, cultural connotations play the largest role at this level of metaphor. The crooner has a "silky" voice by virtue of its sexiness and intimacy, conjuring up situations where silk undergarments might well be worn or discarded. It is difficult to imagine the low-spectral centroid, vibrato-rich, relaxed (low *exertion*) tone of Johnny Hartman, for example, conceptualized in terms of negative sensations such as "harshness" or "abrasiveness"; but it is also unlikely that we would metaphorically compare his smooth, frictionless singing to something as unromantic as, say, synthetic motor oil. In short, we hear qualities of timbre through the cultural filter of appraisal, grouping like categories of experience together.

However, it needs to be stressed that metaphorical mappings of timbre, while culturally contingent, are not entirely *culturally constructed*; that is, they rely on certain shared experiential structures of meaning originating in the remarkably consistent physiology of the human body. While my examples so far have been drawn entirely from the English language, schematic structures for timbre—and the metaphorical mappings they produce—appear to be largely consistent across many languages and cultures, indicating basic (though I hesitate from saying "universal") recursive structures of biomechanical and bioacoustic accordance between timbre production and perception, and between experiential patterning and linguistic conceptualization.[35] There are only so many ways that a human can make sense of timbre: as Johnson repeatedly points out, schematic structures *constrain* possibilities for meaning as much as they enable them. We describe timbral-corporeal events as *brittle* and *gravelly*, we do not call them *fishy* or *deciduous*—these terms would simply not make sense as metaphors, since they (as yet) have no coherent schematic structure grounding them. To be sure, more latitude exists in metaphorical conceptualization of other parameters of musical sound: pitch, for instance, is notoriously flexible.[36] But timbre, in its ecological directness and resistance to being segmented, ordered, and rationalized, appears to be held under a good deal more conceptual constraint at this basic level. While there are obviously many other timbre metaphors besides the three presented in this chapter, then, I suspect they are

[35] Support for the hypothesis that timbre semantics often exhibit a high degree of inter-language consistency comes from studies comparing descriptors in different languages, e.g., Greek, German, and Czech: see Asterios Zacharakis, Konstantinos Pastiadis, and Joshua D. Reiss, "An Interlanguage Study of Musical Timbre Semantic Dimensions and Their Acoustic Correlates," *Music Perception* 31, no. 4 (2014). I review in depth the Japanese vocabulary used to talk about *shakuhachi* timbre in other research. Fales also acknowledges the cross-cultural validity of synaesthetic metaphor in timbre description: see Fales, "The Paradox of Timbre," 57.

[36] For example, Lawrence Zbikowksi reviews cultures that conceptualize pitch as "young/old" and "light/heavy": Zbikowski, *Conceptualizing Music: Cognitive Structure, Theory, and Analysis*, 67-68.

much more limited in number than metaphors for musical pitch, rhythm, motion, harmony, and form.

Where things *do* differ tremendously between contexts, however, is in the process of appraisal. We share underlying experiential structures, and these structures limit the way we can conceptualize timbre; but ultimately concepts are read into a specific social/symbolic framework, where the majority of human life is lived. The schema of *exertion*, for example, is a pattern shared by all people. Understanding sound production in terms of increasing and decreasing *exertion*, therefore, is largely reactive (in Huron's sense that it is quick, pre-cultural, self-protective, and automatic). An extremely high exertion timbral-corporeal event, therefore, is only likely to be conceptualized through a limited number of metaphorical mappings (expressing a shared experience of bodily exertion); it is much more likely to be described as *piercing*, for instance, than *delicate*. However, the concept of "piercing-ness" (like "delicate-ness") is subject to contextual appraisal: it could be read as negative (a *bad sound*, e.g., a painful bodily intrusion) or positive (a *good sound*, e.g., a powerful, salient voice in a musical conversation). To illustrate, Paul Théberge points out that the conventional affective polarity of timbre terms is often reversed in popular music settings: *dirty* is often preferred over *clean*, *fat* over *thin*, *wet* over *dry,* etc.[37] There are thus three meaning structures that we should disambiguate: the *bodily* meaning (reactive, image schematic), the *conceptual* meaning (metaphoric), and the broader *appraised* meaning, which forms a composite of all the other meaning structures and is most sensitive to cultural input.

The appraisal dimension of metaphor is exhibited most vividly in common timbre descriptors that are negatively loaded, bringing us to our third metaphorical entailment: NOISE IS FRICTION. With this metaphor, we hear "noisy" qualities of timbre in terms of rubbing, jostling

---

[37] Théberge, *Any Sound You Can Imagine: Making Music/Consuming Technology*, 207-8.

material interaction (see Tab. 3.2 again for examples). In this sense the metaphor overlaps with SOUND IS MATERIAL; however, in addition to textures (*rough*, *course*, *gritty*, etc.), it is based on the experience of forceful engagement involving resistance and friction (*scraping, grating*, etc.). Where SOUND IS MATERIAL is affectively broad, then, NOISE IS FRICTION conceptualizes timbre specifically in terms of haptic experience; it enables us to *hear* "noisy" timbre in reference to *feeling* things (often, as we will see, unpleasant things). It is also related to the INSTRUMENTS ARE VOICES metaphor in drawing a conceptual link between high-exertion, high-arousal conditions of the voice—a timbral-corporeal event that involves friction against a source of material *blockage*—and the noisy timbral qualities to which they are typically correlated (*high-frequency energy, noise,* and *roughness*). Considering these complexities, this metaphor is going to take a bit more unpacking.

Noise and the Tactility of Sound

> The fact that voices and other sounds can impinge on the senses is itself a proof of their corporeal nature. What happens is that, when atoms of voice in greater numbers than usual have begun to squeeze out through the narrow outlet, the doorway of the overcrowded mouth gets scraped. Undoubtedly, if voices and words have this power of causing pain they must consist of corporeal particles.
>
> — Lucretius, *De rerum natura*[38]

As the reflections of Titus Lucretius Caro (99–55 B.C.E.) reveal, the conceptualization of sound in material terms is hardly new. And nowhere is this linkage clearer than for qualities of sound imagined to carry the most immediate, drastic corporeal effects. The passage here is primarily concerned with developing an account of how sound can "impinge on the senses" and "cause pain"; in short, Lucretius is exploring sound's corporeal consequences. In this conceptualization,

[38] Lucretius, *On the Nature of the Universe*, trans. R.E. Latham (London and New York: Penguin, 1951), 108.

sound, and specifically the voice, can only cause corporeal pain if it itself possesses a "corporeal nature" (*natura corporealis*): sound therefore *has to be* a physical, tactile substance ("atoms of voice") in order for it to have any bodily impact.

What is the bodily source of sound's corporeal powers in this Roman account? Following the model introduced earlier, sound here is a function of *force* in different schematic forms: the *blockage* of the "narrow outlet," the *exertion* as the atoms of voice squeeze out through the doorway, "scraping" the mouth with high friction and presumably somewhat smarting energy as they go. Sound can *cause pain* in the listener because its underlying timbral-corporeal event *is painful*. Now, let us further imagine what Lucretius's hypothetical voice would sound like. The intense scraping of corporeal atoms of voice against the surfaces of the mouth would not produce a *smooth*, *dark*, or *gossamer* quality of timbre; instead, it would sound *harsh*, *grating*, *raspy*, and *rough*—in short, "noisy."

But significantly, it is not just noisy sounds that possess corporeal impact, but also *words*. How could words—and specifically for our purposes, metaphors—transmit tactile sensation, like sound? Lucretius, regrettably, does not speculate on the matter. Following up on the implications of this poetic account with insights from modern-day cognitive sciences (including the neuroimaging experiment from Chapter 2), however, sheds additional light on this question. Tactile metaphors for timbre, of which the NOISE IS FRICTION metaphor is the most extreme example, do not just reflect a handy way to describe noxious, possibly painful characteristics of sound. Rather, as Lucretius intuited, they demonstrate a fundamental correspondence between tactile sensation, timbral-corporeal events, and the sounds generated in these forceful interactions.

Hearing and touching seem to share a special link. Eitan and Rothschild (2010) found pervasive tactile conceptualization of all dimensions of music, including timbre.[39] They explain this result by hypothesizing a mechanism whereby low-level sensory input—a *rough* quality of timbre, say—is mapped onto higher-level conceptual domains, an account consistent with image schemata and conceptual metaphor. Moreover, Eitan and Rothschild reveal an underlying evaluation in tactile musical metaphors. As with Théberge, most metaphorical pairs have a *good* and a *bad* side: *soft, smooth, warm, lightweight,* and *blunt* are positively valenced, while their antipodes—*hard, rough, cold, heavy,* and *sharp*—are negatively valenced. The authors conclude: "tactile metaphors for musical sound are neither coincidental nor subjective, but relate systematically to basic qualities of sound."[40]

This insight is echoed in recent neurophysiology research, especially the growing literature on synaesthesia. Neuroscientists suggest that synaesthesia—the phenomenon of cross-sensorial interaction where someone *feels* sounds or *tastes* sights, etc.—is really just an extreme case of "normal" brain functioning: the sensorium is a fluid and promiscuous system, pervasively cross-wired and prone to slippage.[41] For example, in one now-classic study, non-synaesthetic subjects were asked to choose the "correct" spoken name for two shapes, "bouba" and "kiki" (Fig. 3.13).[42] A full 95% of respondents labeled the sharp, pointed figure "kiki" and the rounded, curvy figure "bouba," indicating that timbral aspects of phonemes—and the physical acts re-

---

[39] Zohar Eitan and Inbar Rothschild, "How Music Touches: Musical Parameters and Listeners' Audio-Tactile Metaphorical Mappings," *Psychology of Music* 39, no. 4 (2010).

[40] Ibid., 464.

[41] For example, researchers have determined that sound-color synaesthesia is just an intensification of "normal perception": see Jamie Ward, Brett Huckstep, and Elias Tsakanikos, "Sound-Color Synaesthesia: To What Extent Does it Use Cross-Model Mechanisms Common to Us All?," *Cortex* 42, no. 2 (2006). A helpful general discussion of the implications of synaesthesia for metaphor can be found in Slingerland, *What Science Offers the Humanities: Integrating Body and Culture*, 160-62.

[42] V.S. Ramachandran and Edward Hubbard, "Synaesthesia: A Window into Perception, Thought and Language," *Journal of Consciousness Studies* 8(2001). This effect was originally discovered as early as 1929 by Wolfgang Köhler.

quired to produce vocal sounds—are coupled with different associated shapes and textures. Cross-culturally and even among very young children, the plosive, velar-stopped (or *blocked*) /k/ along with the high-frequency-rich vowel /i/ seem to imply something sharp and jagged, while the bilabial-stopped /b/ and smooth, timbrally dark /o/ point to something softer and rounder.[43]

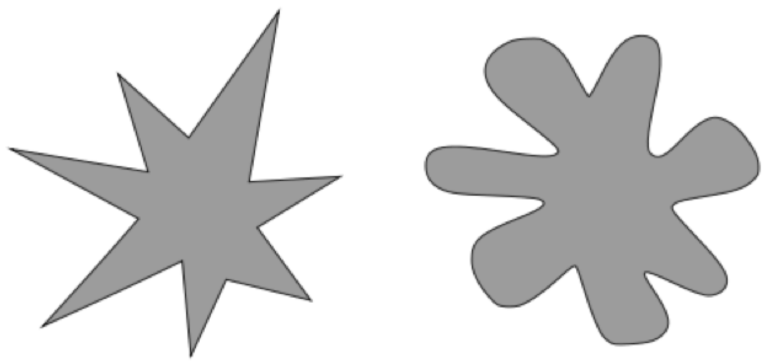


FIGURE 3.13: Bouba/kiki effect
(after Ramachandran & Hubbard 2001)

Though the bouba/kiki effect does not explicitly address tactile sensation, we know that auditory and tactile sensory areas of the brain are closely connected. Numerous studies have shown wide functional overlap when we (non-synaesthetes, that is) *hear* and *touch* things, indicating that these two sensory regions work together when processing both sounds and tactile sensa-

[43] Ramachandran and Hubbard had American college students and Tamil speakers in India perform this selection task. Another study extended this finding to 2.5 year-old children: see D. Maurer, T. Pathman, and C.J. Mondloch, "The Shape of Boubas: Sound-Shape Correspondences in Toddlers and Adults," *Developmental Science* 9, no. 3 (2006). In the 1970s, Manfred Clynes reached a similar conclusion in his seminal study on the affective qualities of different shapes and lines: see Manfred Clynes, *Sentics: The Touch of Emotions* (New York: Doubleday, 1977).

tions.[44] At the neural level, then, motor mimesis is not the only mirroring network involved in listening; we also hear sound through its implied tactile qualities.[45]

Evidence for shared auditory-tactile processing is compelling, but are such couplings also involved in language processing, a key step in conceptualization? We know that speech-events trigger brain activity in areas far afield from the traditionally-defined "language centers," and that action words (e.g., *kick* and *punch*) have motor-mirroring properties.[46] But what about metaphors with sensorimotor implications? Researchers are just now beginning to uncover the role of the somatosensory cortex—the main area involved in processing the sense of touch—in tactile metaphor processing. Using fMRI, Lacey, Stilla and Sathian (2012) presented subjects with metaphorical and literal sentences of the same meaning (e.g., "she had a rough day" and "she had a bad day"), focusing entirely on tactile metaphors.[47] They found activity in areas of the brain specific to tactile sensation when subjects heard the metaphor but not the non-metaphorical sentence, indicating that we process metaphor *literally*; hearing the word "rough," even when not in reference to something with a rough texture, induces activity consistent with feeling something rough. Tactile metaphors, therefore, appear to draw on the mirror properties of the somatosensory cortex.[48]

---

[44] J.J. Foxe *et al.*, "Auditory-Somatosensory Multisensory Processing in Auditory Association Cortex: An fMRI Study," *Journal of Neurophysiology* 88, no. 1 (2002); M. Schürmann *et al.*, "Touch Activates Human Auditory Cortex," *Neuroimage* 30, no. 4 (2006).

[45] In a 2004 behavioral study, Peeva et al. found strong haptic associations between the sensation of loudness and tactile roughness: see D. Peeva *et al.*, "Haptic and Sound Correlations: Pitch, Loudness and Texture" (paper presented at the Eighth International Conference on Information Visualization, London, 2004).

[46] Pulvermüller, "Brain Mechanisms Linking Language and Action."

[47] Simon Lacey, Randall Stilla, and Krish Sathian, "Metaphorically Feeling: Comprehending Textural Metaphors Activates Somatosensory Cortex," *Brain and Language* 120(2012).

[48] This hypothesis was first theorized, to my knowledge, by Vittorio Gallese and George Lakoff, "The Brain's Concepts: The Role of the Sensory-Motor System in Conceptual Knowledge," *Cognitive Neuropsychology* 22, no. 3-4 (2005).

This is a deeply suggestive finding. It implies that sensorimotor metaphors—metaphors of sensation and action—are, crucially, *performative*: they induce resonance consistent with their literal meanings, in an important sense acting out their own implications. Returning to our metaphor NOISE IS FRICTION, then, we can hypothesize that the words "rough and gritty blues singer," for example, may effectively activate in the brain *the feeling of roughness and grittiness.* Tactile metaphors for timbre are not just passive and impressionistic descriptions, but are fundamentally enactive; they allow the conceptualizer to sympathetically generate a sensation consistent with that associated quality of sound.

However, this still does not directly address the framing question of this chapter: Why do we use the words we do to talk about timbre? To briefly retrace our steps: tactile metaphors are commonplace in music conceptualization, and the brain is significantly cross-wired between these modalities; further, words for actions and tactile sensations (even used metaphorically) excite mirror networks in associated sensorimotor areas. But what about the preconceptual processing of timbral qualities conventionally described with the help of tactile—and specifically, unpleasant—metaphors? Presumably *calling* a sound "rough" would involve somatosensory areas, but what about *hearing* a "rough," noisy timbral quality without verbal priming? If conceptualization of noisy timbre relies upon the metaphor NOISE IS FRICTION, as I am suggesting, then it follows that regions associated with somatosensation would be active during perception. We know from studies in ecological acoustics that the tactile sensation of roughness has associated acoustic qualities, particularly strength in high frequency components.[49] Perhaps this psychoacoustic linkage (and others) works its way into musical timbre perception as well; hearing rough timbres may

[49] Steve Guest *et al.*, "Audiotactile Interactions in Roughness Perception," *Experimental Brain Research* 146(2002). It would be very interesting to know if auditory *roughness* is related to tactile *roughness*, but that is a question for a different dissertation.

activate tactile areas even without the explicit use of tactile metaphors. If we innately subvocalize musical timbre, perhaps we also *subsense* it.

Although I could not find anything in the neuroscience literature explicitly addressing this hypothesis, we serendipitously happen to have access to brain imaging data already, from last chapter (Experiment 3). With scaled levels of noise and correlated behavioral information, this dataset seems tailor-made to address this question. Using the brain coordinates from Lacey *et al.*, I found that the precise region of overlap between tactile sensation and tactile metaphor is also involved in processing certain noisy, "tactile" timbres (Fig. 3.14).[50] While this result is not consistent for all instruments, the growled saxophone timbres in particular (Sax 2 and 3)—which, after Voice 3, were given the highest average "noisiness" ratings of all the timbres—drive somatosensory activity consistent with tactile sensation (in the left parietal operculum–1 sub-region, or OP1). Processing the *rough* sound of the growling sax (without verbal priming) thus seems to draw upon the exact region of the brain responsible for both feeling rough things and comprehending the metaphor "rough."

The next step in the analysis is to relate this tactile activity to conscious perception: Do changes in *affective valence* (i.e., preference) and perception of "noisiness" in a timbre correlate with activity in this tactile/metaphor area? This is a critical step in connecting activation to "noisy," negative tactile sensation specifically; documenting activity in tactile areas does not imply the sensation of *roughness per se*—it could as well be *smoothness* or any other touch sensation. Correlating activity with these behavioral variables, however, can tell us how a neural reaction reflects appraisal, that is, if somatosensory activity is linked to a positive (*smooth, silky*) or a negative (*rough, harsh*) tactile percept.

[50] For the specifics of this analysis, see the final section of the results in Appendix 3.

Here again the growled saxophone produced an unambiguous pattern: the perception of "noisiness" (p-noise) is strongly correlated with activity in this same somatosensory area (plus another from the Lacey *et al.* study, OP3) (Fig. 3.15a). This indicates that resonance from tactile sensory areas plays into the appraisal of this timbral quality as "noisy." And lastly, the OP3 tactile processing center is negatively correlated with *affective valence* even when collapsed across all instrument categories (Fig. 3.15b): the more people *dislike* a noisy timbre, the more they show activation there. More research is needed to dig into the neurophysiological details of this linkage, but the preliminary implications are clear enough to resolve the questions driving this chapter. We can provisionally extend Eitan and Rothschild's previous conclusion: tactile metaphors for musical sound are neither coincidental nor subjective, but relate systematically to basic qualities of sound *and to the somatic and affective implications of timbral-corporeal events*.[51]

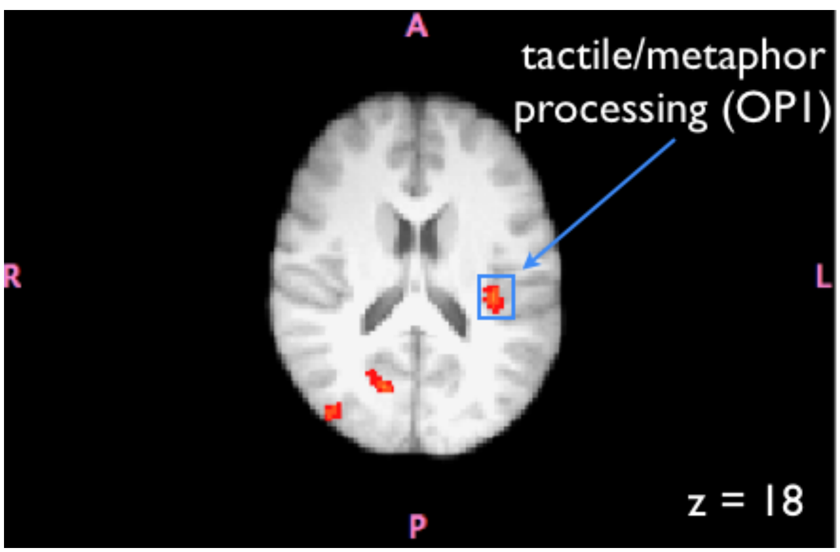


FIGURE 3.14: Growled Saxophone (Sax 3) > Regular Saxophone (Sax 1)

[51] Eitan and Rothschild, "How Music Touches: Musical Parameters and Listeners' Audio-Tactile Metaphorical Mappings," 464.

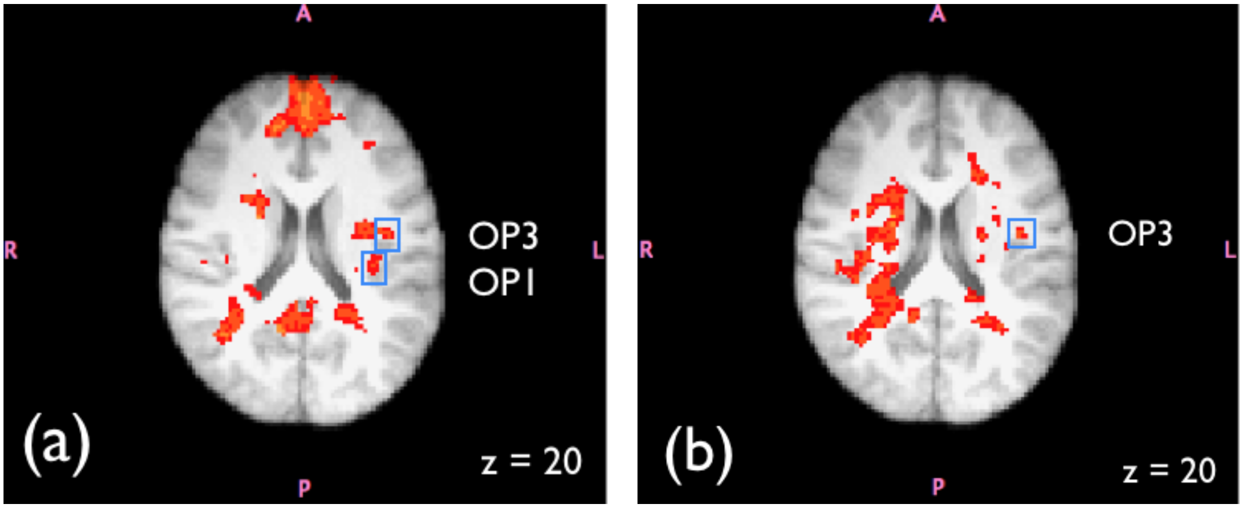


FIGURE 3.15: (a) Saxophone 2 > Saxophone 1 correlated with p-noise; (b) Timbre 2 > Timbre 1 correlated with *negative affect*

As we saw in the last chapter, somatosensory activity was part of the broader mirroring network involved in timbre processing. But these results point to something much more specific—a cross-modal neural link between audition of noisy timbre and tactile sensation. This truly novel result is probably overdetermined. First, the metaphorical mapping of negative, high-friction tactile sensation onto noisy timbre is likely so ingrained as to influence the cognition process; we hear certain timbres as rough, etc., even when not verbally prompted to do so. Maybe "roughness" and other common metaphors for noisy timbre are not simply *post hoc* descriptions, then, but rather constitute an integral, *a priori* aspect of the sensory perception and cognition process itself. Metaphors might well be built into the way we hear timbre.

Another explanation, by contrast, takes us back to the image schematic level. These results—which implicate tactile processing in noisy timbre, and specifically in perception of "noisiness" and negative affect—may actually have little or nothing to do with tactile, high-implied-friction, unpleasant metaphors. Recalling the force schemata reviewed earlier, high-*exertion*, *blocked* timbral-corporeal events often do, quite literally, involve elevated levels of physical friction; the *blockage* schema is premised on materials forcefully interacting in ways that might be conceptualized as "rough," "harsh," and "grating." Perhaps we hear these noisy timbral qualities as "rough" and "harsh" because they literally sonify negatively-valenced, high physical friction, high-exertion acts, and consequently evoke similar sensations in the brain of the listener through auditory-tactile interactions. Seen from this angle, the metaphor NOISE IS FRICTION is really not much of a metaphor at all; perhaps *noise is friction* (or at least its direct sonic correlate).

In the final analysis, these explanations amount to much the same conclusion: metaphors for timbre are performative as much as they are descriptive. This process might appear relatively transparent for most terms (mirroring sensory *smoothness* is presumably positive, for example), but when dealing with particularly potent somatic metaphors, the performativity of timbre descriptors—their ability to influence *perception*, triggering somatosensory and motor reactions—takes on added urgency. The "somatic markings" of metaphor are grounded perhaps in certain shared understandings of physical force: the behavior of the voice under duress, the embodied knowledge that *blockages* have audible consequences, etc. But shared bodily knowledge can be appraised in wildly different ways according to context: once a timbral quality is conventionalized as a *sign*, it becomes polysemous and available for virtually any possible use. Metaphorical projections are not automatically *positive or negative*, therefore—we *construct* these meanings, we actively negotiate them into being.

This means that the way we talk about timbre, like any speech-act, is preeminently social, political, and ethical. NOISE IS FRICTION may start as a negative tactile metaphor in our language (though obviously always prone to transgression, as the history of popular music illustrates). But it is certainly not universally thus; other cultures and languages conceptualize the same acoustic qualities under very different systems of appraisal. That which is here conceived negatively in terms of unpleasant friction, rubbing, and raw materiality, in other words, may well be conceived in a different context as an aural icon for closeness, community, and social interaction (a positive interpretation of the same image schemata). I have proposed that NOISE IS FRICTION is based in certain shared features of human embodiment. But it is not based in universal canons of human meaning.

Conclusion

Musical pitch is eminently communicable. In the Western, equal-tempered system, we can all agree on what constitutes "C-ness" or "G#-ness": the acoustic dimension of frequency has been effectively codified over the centuries (though not without some disruptive burps) into discrete categories, or pitch chroma. Timbre, however, remains undomesticated. Musicians and scholars have long lamented the ill-defined character of timbral description: when talking about sounds, rather than pitches, we are "condemned to the adjective."[52]

This view is wrong. Metaphors for timbre, like all musical metaphors, express certain fundamental image schematic realities of human embodiment. They are grounded by the basic experiential patterns of life: forceful interactions with objects, the feeling of bodily exertion, knowledge of physical limits, etc. As Paul Ricoeur puts it, metaphor "constitutes the primordial reference to the extent that it suggests, reveals, unconceals … the deep structure of reality to which we are related as mortals…"[53] There is certainly room for semiotic ambiguity here: with enough general agreement among an interpretive community, *any* word—even "fishy" or "deciduous"—could consistently signify a particular timbral attribute. The existence of indexical

---

[52] This expression comes from Barthes: see Barthes, *Image, Music, Text*, 184.

[53] Paul Ricoeur, *Hermeneutics and the Human Sciences: Essays on Language, Action and Interpretation*, trans. John B. Thompson (Cambridge: Cambridge University Press, 1981), 240.

signs for timbre does not eliminate the possibility of iconic and symbolic ones. However, the discourse of timbre it is not a linguistic free-for-all either. We use the words we do to describe timbre because they uniquely enable us to connect bodily experiences to coherent mental structures; they help us transform raw perception into meaningful, symbolically ordered thought. Appraisal would be a very different process without metaphorical conceptualization.

But is timbre itself a metaphor? I have argued here that we understand timbre through conceptual metaphor, but we are prepared now, I think, to advance a stronger conclusion: *Timbre is itself metaphorical in nature.* As we have seen, timbre exists where the "acoustic" and the "perceived worlds" converge. The same can be said of all music, of course, but timbre is the primary ingredient for source identification (making aspects of materiality and ecology more perceptually salient here than in other musical domains); it is hard to segment, order, and rationalize; and it is perceived more quickly and directly than other parameters of music. Timbre does considerable though usually invisible work linking perception of the material world to, as Shepherd and Wicke put it, the "logic, structure and inner texture of life."[54] Timbre is thus fundamentally connective; it is defined not primarily by *what it is* (or is not), but by *what it does.*

Sound waves traveling through the air do not alone timbre make; neither do biologically dictated auditory reactions; neither do the panoply of descriptive words we use to talk about them. So what is "timbre" then?

Timbre is not a thing; it is a process. Timbre is what happens in the embodied mind of the listener when sound-generating events are transformed into structures of meaning. We can see this process at play in the brain: sounds produced by "rough" timbral-corporeal events are interpreted according to their rich motor-somatic implications, both inscribing and reinforcing connections with similarly

---

[54] For an exhaustive interpretation of this statement in relation to timbre, see Eidsheim, "Voice as a Technology of Selfhood: Towards an Analysis of Racialized Timbre and Vocal Performance," 243-52.

marked concepts. The experience of timbre is the experience of this translational process. Understood this way, timbre is the name we give to the perceptual bridge between one domain of experience (the sounding world) and another (our own body-mind). The metaphor *is* the mapping process of one domain onto another. Therefore, to refine our earlier definition from Chapter 1, timbre is not just a *verb*, but a *verb as metaphor*. And the metaphor is the living body.[55]

At this juncture, we have indelibly crossed into the realm of the Sign. The living body is part of social discourse; timbral signs, then, may begin with biological distinctions, but end in cultural and historical ones. We will see this plainly in the case studies to come: rooted in the basics of human embodiment and thus prone to seem "natural," timbre as sign is uniquely susceptible to competing, strongly-held, intensely visceral interpretations. Timbral signs can be deployed as powerful, almost subliminal tools in the demarcation of social categories. Designating another's sound "harsh" and "noisy" in contrast to one's own "pure" and "beautiful" sounds conditions a particular somatic response, marking not only the acoustic quality but the social category to which such sound belongs.

Timbre is a metaphor—perhaps *the* musical metaphor—for difference.

---

[55] The terminology here, of course, is Merleau-Ponty's. The idea of the "living," phenomenal body (versus a "corpse") is first mentioned in Merleau-Ponty, *Phenomenology of Perception*, 63.

Chapter 4

The Most Powerful Human Sound Ever Created?
Theorizing the Saxophonic Scream in Free Jazz

> Who, if I screamed, would hear me among the angelic
> orders? And even if one of them
> took me suddenly to his heart: I would die from his
> potent Being. For the Beautiful is nothing but
> the onset of that Terror we can scarcely endure,
> and we are fascinated because it calmly disdains
> to obliterate us. Every angel is terrifying.
>
> — Rainer Maria Rilke, "The First Elegy"[1]

Introduction

Writing about John Coltrane's *Ascension* record in 1967, *Down Beat* magazine critic Martin Williams commented: "The performance soars, and it sings. And it rages, blares, shouts, screams, and shrieks."[2] Here, in somatic (and somatically-marked) terms, Williams dips into a shared metaphorical pool common to all free jazz commentators during the middle of the 1960s. Free jazz saxophone playing is understood in vocal terms—it "sings"—but additionally, and more pointedly for enthralled or anxious listeners, it seems to transcend (or transgress) normal "musical" speech. *It sings but it does not sign.* The vocal-mimetic sound of Coltrane's saxophone seemed

---

[1] Rainer Maria Rilke, *The Duino Elegies*, trans. Leslie Norris and Alan Keele (Rochester, NY: Camden House, 1923/1993), 3.

[2] Quoted in Carl Woideck, ed. *The John Coltrane Companion* (New York: Schirmer, 1998), 46.

to many listeners to occupy a cognitive in-between: vocal but aggressively non-signifying, music-like but also noisily disruptive to conventional musical codes—as the Greeks would say, it is *phone* (sound), not *phone semantike* (sound that signifies linguistically). Coltrane's voice on the saxophone does not just sing to you—it also screams and shrieks at you. But what was he trying to express with these screams and shrieks? What was this a timbral sign *for*?

Reactions to the screaming sounds of jazz's "New Thing" were fiercely divided from the beginning. Some listeners heard it as ecstatic rapture, the "most powerful human sound ever created" whose affective potency resided in its seeming ability to reach beyond the normative signifying practices of music. Others heard the same sound as a righteous cry against racial injustice, the sonic version of a raised fist. Still others heard it as "death-to-all-white-men wails," an interpretation that seemed to harmonize with the race riots in the streets outside (Harlem in 1964, Watts in 1965). The sound of free jazz was famously polarizing, and reactions tended to be more emotional than intellectual. Eschewing dispassionate evaluation, listeners tended towards the extremes—free jazz saxophone playing was either embraced or rejected. Either way, reactions were intensely *visceral*.

Applying the theoretical model sketched in Part I, this chapter sets out to explore the relationship between the timbral landscape of free jazz saxophone playing—particularly its most iconic gesture, what I will call the *saxophonic scream*—and its contentious reception in the volatile mid-1960s. How did the same sound generate such diametrically opposed interpretations among different listeners, leading one side to empathize with this breach of "music" and the other to deny empathy entirely, hearing it instead as pure noise? To address this question, I will apply an embodied cognitive analytical approach to timbre as a "hermeneutic window" through which to theorize the broader affective and social dimensions of the saxophonic scream. Though reception history serves to structure the chapter, then, we will take significant detours to explore the reactions underlying the perception of this uniquely charged timbre. Examining the *act*, *sound*, and *perception* of saxophonic screaming, I suggest that polarized appraisals of this particular timbral sign—its "thick" meanings—are ineluctably bound up in the "thin" contingencies of its reaction. Untangling this interaction is of crucial importance to understanding the reception of the saxophonic scream, since its sonic identity and the various meanings ascribed to it—and, significantly, to the (black) people producing it—are contingent on how (or whether) listeners made sense of its noisy, a-musical, "screaming" quality. Finally, although my analysis focuses broadly on reactions to the saxophonic scream irrespective of specific recorded examples, for the sake of illustration I concentrate on the reception of John Coltrane's "late period" recordings from 1965, particularly *Ascension* and *Meditations*.

Before proceeding, a brief note regarding terminology is in order. Documenting historical reactions to timbre alone poses many perhaps insurmountable challenges, not least of which the fact that timbres are always attached to pitches, rhythms, etc., and also—following an ecological approach—to bodies and things. Moreover, reference to "sound" in jazz, as George Lewis has observed, often means much more than just literal *sound*.[3] In the following pages I try to isolate critical commentary dealing with timbre, though I acknowledge in this bracketing that inevitably the focus does not remain sharp for long. Also, the term "saxophonic scream" is shorthand for altissimo notes with added noise. There are a variety of playing techniques and timbral shades associated with screaming, but I gloss over these distinctions here in favor of a unified perceptual gestalt.

---

[3] "'Sound,' sensibility, personality and intelligence cannot be separated from an improviser's phenomenal (as distinct from formal) definition of music. Notions of personhood are transmitted via sounds, and sounds become signs for deeper levels of meaning beyond pitches and intervals." See George E. Lewis, "Improvised Music Since 1950: Afrological and Eurological Forms," *Black Music Research Journal* 16, no. 1 (1996): 117.

Escaping from Notes to Sounds: John Coltrane's Timbral Turn

> What [Coltrane's playing] showed was that … through changes of timbre, you could play anything from any of the twelve notes […] to noise.
>
> — Steve Reich[4]

> Coltrane studied sound *intensely.* […] The importance of Coltrane was that he pursued it further than anybody; he went to the edge of it. In going to the edge of it, I mean harmonic sound.
>
> — Charles Moore[5]

> [I'm trying to] point out the divine in a musical language that transcends words. I want to speak to their souls.
>
> —John Coltrane[6]

At the time John Coltrane (1926–1967) began to flirt with and ultimately embrace the "New Thing," he was already widely regarded as one of the most influential tenor saxophonists in the jazz world.[7] However, despite the critical and popular praise of his modal and hard-bop innovations from the 1950s, he continually expressed discontent with his own achievements. Creative restlessness, fueled by a newfound spiritual yearning, impelled the saxophonist to explore new modes of expression; but contrary to earlier innovations, which were mainly harmonic, his search

---

[4] Quoted in Ben Ratliff, *Coltrane: The Story of a Sound* (New York: Farrar, Straus and Giroux, 2007), 145.

[5] Quoted in ibid., 154. Italics in original.

[6] Quoted in Iain Anderson, *This is Our Music: Free Jazz, The Sixties, and American Culture* (Philadelphia: University of Pennsylvania Press, 2007), 116.

[7] The life and music of John Coltrane is well rehearsed in numerous sources, so I begin this narrative *in medias res*: for detailed accounts, see Cuthbert Simpkins, *Coltrane: A Biography* (New York: Herndon House, 1975); J. C. Thomas, *Chasin' the Trane* (New York: Da Capo, 1975); Eric Nisenson, *Ascension: John Coltrane and His Quest* (New York: St. Martin's Press, 1993); Woideck, *The John Coltrane Companion*; Lewis Porter, *John Coltrane: His Life and Music* (Ann Arbor: University of Michigan Press, 2000); Ratliff, *Coltrane: The Story of a Sound*.

in the early- to mid-1960s came to revolve primarily around a radical form of timbral experimentation.[8] In short, he wanted to "escape from notes to sounds."[9]

This timbral turn was not original to Coltrane, though he quickly became its most audible proponent. Echoing the aesthetic orientation of the younger "sound-players" of the jazz avant-garde—including saxophonists Ornette Coleman, Eric Dolphy, Albert Ayler, Archie Shepp, Marion Brown, and Pharoah Sanders—he sought to liberate sound itself from the confines of pitch, elevating timbre to a primary expressive position in his improvisatory vocabulary, a shift into what Ekkehard Jost calls "sound improvisation."[10] Marion Brown, alto player on Coltrane's *Ascension* session (1965), explains: "The emphasis was on textures… You can hear it, in the saxophones especially, a reaching for sound and an exploration of the possibilities of sound."[11]

Generating the expanded timbral palette of "sound improvisation" required players to explore a variety of new timbral-corporeal approaches to their instrument, the saxophonic scream foremost among them. Coltrane scholar and transcriber Andrew White dubs this arsenal of extended techniques "effectual saxophonics," which David Ake describes as "instrument-specific gestures irreducible to the page."[12] Such techniques allowed players to explore hitherto uncharted territories of musical sound and affect; they were a crucial component in Coltrane's

---

8 For instance, the so-called "Coltrane matrix" first used in the tune "Giant Steps" (1960). For more on his harmonic innovations, see Karlton Hester, "The Melodic and Polyrhythmic Development of John Coltrane's Spontaneous Composition in a Racist Society" (Ph.D. Diss., CUNY, 1990).

9 This is Albert Ayler's formulation. Quoted in Anderson, *This is Our Music: Free Jazz, The Sixties, and American Culture*, 113.

10 Ekkehard Jost, *Free Jazz* (New York: Da Capo, 1981), 169. He writes: "In solos, there is a gradual emancipation of timbre from pitch that leads to a-melodic structures primarily delineated by changes in color and register" (95).

11 A. B. Spellman. Liner notes to John Coltrane's *Ascension*. Impulse Records AS-95, 2000, CD. Originally released in 1965.

12 David Ake, *Jazz Cultures* (Berkeley and Los Angeles: University of California Press, 2002), 138. Saxophonist/ ethnomusicologist David Borgo lists such effectual saxophonic techniques as follows: extreme vibrato, exaggerated articulations, harmonics, extended range devices, multiphonics, vocalizing effects such a growls, sung tones, smears, kinetic shapes, circular breathing, and recitation tones: see David Borgo, "Between Worlds: The Embodied and Ecstatic Sounds of Jazz," *Open Space* 5(2003).

quest to "speak to the soul."[13] The visceral, seemingly unmediated affective power of the "New Thing" was closely linked to the effectual saxophonics (specifically the scream) that served as its timbral *métier*; as Jost puts it, "[sound-playing] expresses an emotionalism heightened to the ultimate degree."[14] This connection between "emancipation of timbre" and "heightened emotionalism" is ubiquitous in the reception literature: effectual saxophonics such as the scream are commonly characterized as synonymous with affective intensification.[15] But while this particular timbre played a major role in the perceived emotional fervor of the style—for reasons that we will explore later—it was also at the root of its power to polarize listeners. The immediate emotionality of this saxophonic expression led to a bifurcation in appraisal that strained against the definition of "music." By escaping from notes to sounds, I suggest, Coltrane played at the threshold between music and its Other, whether spiritual bliss or just plain noise.

The critical community did not know what to make of Coltrane's new sound. Despite his continued success in *Down Beat* magazine polls, already by 1961 the critics were fiercely divided over the saxophonist's new artistic direction, a polarization that would only intensify by mid-decade.[16] The opening salvo to the so-called "Coltrane controversy" came in November of that year with a review by *Down Beat* editor John Tynan, who famously described Coltrane as "deliberately intent on … pursuing an anarchistic course … that can only but be termed anti-jazz."[17] Fellow *Down Beat* critics Ira Gitler and Pete Welding soon joined in dissent, specifically linking

---

[13] Quoted in Anderson, *This is Our Music: Free Jazz, The Sixties, and American Culture*, 116.

[14] Jost, *Free Jazz*, 95.

[15] For a sampling, see David Such, *Avant-garde Jazz Musicians: Performing "Out There"* (Iowa City: University of Iowa Press, 1993), 17 & 56; David Liebman, "John Coltrane's Meditations Suite: A Study in Symmetry," *Annual Review of Jazz Studies* 8(1996): 177-80; David Borgo, *Sync or Swarm: Improvising Music in a Complex Age* (New York and London: Continuum, 2005), 50; Anderson, *This is Our Music: Free Jazz, The Sixties, and American Culture*, 71 & 115.

[16] Coltrane won the much-vaunted *Down Beat* Critics Poll in 1961 and every year between 1964–66, and the Readers Poll in all but two years from 1960–1966.

[17] John Tynan, "Take Five," *Down Beat*, November 23, 1961.

Coltrane's new "anti-jazz" stance to his new timbral approach. Gitler, who had earlier coined the term "sheets of sound" in reference to Coltrane's style, opined: "Coltrane may be searching for new avenues of expression, but if it is going to take this form of yawps, squawks, and countless repetitive runs, then it should be confined to the woodshed," adding (with a telling *blockage* metaphor), "there are places when his horn actually sounds as if it is in need of repair."[18] Welding likewise took aim at Coltrane's new sound, but was particularly concerned with its affective intensity and rawness:

> … Its gaunt, waspish angularities, its ire-ridden intensity, raw, spontaneous passion, and, in the final analysis, its sputtering inconclusiveness, seems more properly a piece of musical exorcism than anything else, a frenzied sort of soul-baring. It is a torrential and anguished out-pouring, delivered with unmistakable power, conviction and near-demoniac ferocity—and as such is a remarkable human document. But the very intensity of the feelings that prompt it militate against its effectiveness as musical experience. It's the old problem of the artist's total involvement as a man supplanting his artistry, which is based after all, to some greater or lesser degree, in detachment.[19]

Here, in language influenced by the Kantian notion of "disinterested interest," is a case that undermines Coltrane's "intense," "raw," "torrential," and "anguished" sound *qua* musical expression. It is simply *too emotionally direct* to qualify as effective music, Welding contends; the artist's "total involvement" in body and emotion diminishes its status as art. Though this "controversy" came some four years before his definitive turn to free improvisation, we can already see a latent critique embedded in these and other sentiments, one that is taken up and amplified in reception of his late albums: Coltrane fails on the grounds of music. To some listeners, moreover, this criticism was pushed one degree farther: it is not just "failed music," but noise.

---

[18] Ira Gitler, "Double View of Coltrane 'Live'," ibid., April 26, 1962.

[19] Pete Welding, ibid.

The divide between music and noise permeates the literature on late Coltrane and free jazz in general. There are a number of elements to this improvisatory language that confound clear distinction along the "music/noise" spectrum for many listeners, including its lack of intuitive periodic structures, erratic rhythmic pulse, dense textures, sheer volume, and eschewal of standard melodic formulae. But timbre is the sonic parameter that most clearly blurs the threshold between musical sound and noisy incoherence, and this threshold was contested from the beginning. Critic Philip Larkin, for instance, dismissed free jazz saxophonics as "tumults of noise" and "the most appalling noise," but others saw in this same "noise" a form of meaningful expressive purity that transcended standard notions of "music."[20] As biographer Eric Nisenson comments: "These screams, howls, and shrieks were pure sound, pure emotion, the last frontier of music. Coltrane had come to agree [with Albert Ayler] that it was time to go 'beyond notes.'"[21] It is hard to deny that a quality of "beyond-ness" does indeed mark the timbral approach to free jazz saxophonics; the agitated, screaming sound seems to reach past the constraints of the instrument, over the limits of the body, and beyond signification as a musical code. But what does this "beyond" actually signify, and to whom?

Admirers of Coltrane's late style interpreted the saxophonic scream in unmistakably spiritual terms.[22] As biographer J. C. Thomas puts it: "There had to be something else *besides* music there; in reality there was a force beyond music that was communicating with Trane's audience on quite a different, higher level of meaning. Call it Universal Consciousness, Supreme Being,

---

[20] Larkin, *All What Jazz*, 232 & 30.

[21] Nisenson, *Ascension: John Coltrane and His Quest*, 191.

[22] We will be returning to the spiritual implications of noisy timbre throughout. For more, see Borgo, "Between Worlds: The Embodied and Ecstatic Sounds of Jazz."; Franya Berkman, "Appropriating Universality: The Coltranes and 1960s Spirituality," *American Studies* 48, no. 1 (2007); Ratliff, *Coltrane: The Story of a Sound*; Marcel Cobussen, *Thresholds: Rethinking Spirituality Through Music* (Aldershot, UK: Ashgate, 2008); David Ake, *Jazz Matters: Sound, Place, and Time Since Bebop* (Berkeley and Los Angeles: University of California Press, 2010), Ch. 1.

Nature, God."[23] In contrast to critics such as Welding, who saw Coltrane as too personal and subjective, Thomas and other spiritually acquiescent listeners heard it as transcending personal expression altogether. Coltrane's ability to reach beyond the normal bounds of musical communication opened up a path to ecstasy. But significantly (and perhaps even paradoxically), this ecstasy was entirely mediated by the listening body. "Standing outside oneself" (or *ek-stasis*, in the original Greek) was not license for a disembodied, intellectual transcendence, but had to be accomplished through physical toil and pain, extreme sensory overload interpreted by fans as a form of transcendent rupture. In order to achieve the enlightenment beckoning from the performance, the listener's "nervous system [had to be] dissected, overhauled, and reassembled" first.[24] For example, Nisenson compares a Coltrane concert experience to a "mind-boggling religious revelation" where "my body felt exhilaration, transport, even as much as my mind and spirit," and Berendt and Huesmann describe it as "hymnlike, ecstatic music of the intensity of a forty-minute orgasm."[25] Jazz critic Nat Hentoff, in his liner notes to the *Meditations*, makes a similar connection with this extraordinary catalog of conceptual metaphors:

> [Coltrane and Sanders play] as if their insights were of such compelling force that they have to transcend ordinary ways of musical speech and ordinary textures to be able to convey that part of the essence of being they have touched. The emotions are imperious; they cannot be braked or polished into conventional ways of "beauty" or "symmetry." They must explode as felt—in the rawness of palpable, visceral, painful, challenging, scraping, scouring self-discovery.[26]

---

[23] Thomas, *Chasin' the Trane*, 172-3. Italics in original.

[24] Spellman, liner notes to John Coltrane's *Ascension*.

[25] See respectively Nisenson, *Ascension: John Coltrane and His Quest*, xvii; Paul Hegarty, *Noise/Music: A History* (New York: Continuum, 2007), 48.

[26] Nat Hentoff. Liner notes to John Coltrane's *Meditations*. Impulse Records A-9110, 1996, CD. Originally released in 1965.

Sound here is a vehicle for exaltation; the very fact that it pushes beyond normal definitions of "music" is evidence of its cosmic force. As the language makes clear, the timbres of *Meditations* and other late-period albums were conceptualized with extravagant tactile metaphors, specifically NOISE IS FRICTION. They are "palpable" and "painful," they "scour" and "scrape." Noise serves a symbolically haptic spiritual function, scrubbing away the dirt and grime of illusion—or as Coltrane put it, using a classic Zen metaphor, "cleaning the mirror"[27]—to expose the naked Self. And to those who submitted to this sonic "scouring" with ecstatic joy, Coltrane's playing was without equal. Frank Kofsky described *Live at the Village Vanguard Again!* (1966) as "the greatest work of art ever produced in this country," and *Down Beat*'s Bill Mathieu went as far as to praise *Ascension* as "possibly the most powerful human sound ever created."[28] Potent sounds demanded potent (or at the very least, inflated) rhetoric.

The potency of these sounds, however, cut the other way: the same sorts of physical, visceral, extreme reactions marked *negative* critical assessments. While admirers willingly followed Coltrane into the theophanic whirlwind, alienated listeners flinched away in horror. To them, the saxophonic scream was not spiritual revelation—the Other of music that induced psychic ascension through profound bodily experience—but simply sonic violence, the opposite of music, equivalent to an attack on the body. Like the positive accounts, critical responses to this style were physicalized from the beginning. Cultural critic Clive James, for instance, memorably described Coltrane's sound as "full, face-freezing, gut-churning hideosity," adding, "there is not a phrase that asks to be remembered except as a lesion to the inner ear."[29] In a 1965 review of

---

[27] Ibid., 10.

[28] See respectively Ratliff, *Coltrane: The Story of a Sound*, 170-71; Bill Mathieu, "Review of *Ascension*," *Down Beat*, June 23, 1965.

[29] Clive James, *Cultural Amnesia: Necessary Memories from History and the Arts* (New York: Norton, 2007), 191.

Meditations, critic Joe Goldberg wrote, "[I] cannot be scoured or scraped any more. […] I feel only that I am being wildly assaulted, and must defend myself by not listening."[30] The same tactile metaphors employed by Hentoff in his liner notes (to which Goldberg is presumably responding) are retooled here for a different purpose, suggesting that, like fingernails on a chalkboard, the sound cuts deeper than any rational criticism can capture. Rather than "self-discovery," to Goldberg this experience is akin to an "assault," the sound of the saxophone being a source of physical pain that induces fight or flight. These reactions imply physiological vulnerability to sound, not just an intellectual dismissal of it: all listeners sense its power, but while some are swept up into ecstasy, others are confronted with their own susceptibility to undesired aural influence.

Perhaps the most outspokenly vituperative of Coltrane's many critics was the poet-cum-jazz writer Philip Larkin (1922–1985). Larkin was uniquely disturbed by Coltrane's late records, practically to the point of apoplexy, and again, the main problem was timbral. He explains: "It was with Coltrane […] that jazz started to be *ugly on purpose*: his nasty tone would become more and more exacerbated until he was fairly screeching at you like a pair of demoniacally-possessed bagpipes."[31] In an unpublished piece commemorating (celebrating?) Coltrane's death in August 1967, he cantankerously describes the musician's sound as "that reedy, catarrhal tone, sawing backwards and forwards" and "that willful and hideous distortion of tone that offered squeals, squeaks, Bronx cheers and throttled slate-pencil noises for serious consideration."[32] Larkin concludes with a backhanded eulogy:

> Virtually the only compliment one can pay Coltrane is one of stature. If he was boring, he was enormously boring. If he was ugly, he was massively ugly. To

[30] Quoted in Woideck, *The John Coltrane Companion*, 236.

[31] Larkin, *All What Jazz*, 21. Italics in original.

[32] Ibid., 187.

> squeak and jibber for sixteen bars is nothing; Coltrane could do it for sixteen minutes, stunning the listener into a kind of hypnotic state in which he read and reread the sleeve-note and believed, not of course that he was enjoying himself, but that he was hearing something significant. Perhaps he was. Time will tell. I regret Coltrane's death, as I regret the death of any man, but I can't conceal the fact that it leaves in jazz a vast, and blessed silence.[33]

To Larkin, the experience of spiritual transcendence reported by fans was really just a sort of stupefied numbness in the face of revolting noise. They were being tricked, led to hear as *significant* (i.e., something that *signified*) an experience that should rightly have been rejected with disgust.[34] Characterizing this oft-repeated reaction, Ben Ratliff introduces the useful concept of the "monotony of the sublime:"[35] in this formulation, the vast wasteland of the experience sparks in listeners an awareness of the Burkean Sublime, but despite the supposed seriousness and profundity of the experience, its uniform unrelentingness makes for some boring listening. (Its status as "enormously boring" is part of what *makes it* sublime.) To listeners such as Larkin, sometimes ugliness is just that—bad sounds—with nothing profound or "significant" behind it. Thus, while Mathieu heard in *Ascension* "the most powerful human sound ever created," Larkin reviled *Meditations* as "the most astounding piece of ugliness I have ever heard."[36]

Besides the word "noise," "ugliness" appears often in Coltrane criticism. A 1965 *Jazz Journal* review of *John Coltrane Quartet Plays*, for example, begins: "This recording … finds Coltrane well advanced down the path of ugliness which he started to tread in 1961."[37] Remarks about the "ugliness" of Coltrane's sound tap into the same fundamental appraisal as comments about

---

[33] Ibid., 187-88.

[34] As will become clear later, Larkin advocated rejection of Coltrane's late records not because they *failed to signify*, but because they signified something very disturbing to him: black radicalism.

[35] Ratliff borrows this expression from Robert Lowell: see Ratliff, *Coltrane: The Story of a Sound*, 139.

[36] Larkin, *All What Jazz*, 172.

[37] Quoted in Ake, *Jazz Cultures*, 136.

"noise": they imply that it is not really *music* at all, in the process keeping the binarism separating music and noise crisp and neat. But it also keeps the door open for a range of possibly competing appraisals. Revealing the polysemous character of this timbral sign, "noisy" and "ugly" can be positive concepts when their antipodes ("musical," "beautiful," and "symmetrical") are associated with unwanted norms begging for transgression; thus, we see that "ugliness" was taken up as a term of praise among certain advocates of the "New Thing."

In his 1967 short story "The Screamers," Amiri Baraka (*quondam* LeRoi Jones, 1934–2014) describes a late-night jam culminating in a jubilant march down the street accompanied by the sound of the screaming saxophone. He writes:

> The repeated rhythmic figure, a screamed riff, pushed in its insistence past music. It was hatred and frustration, secrecy and despair. It spurted out of the diphthong culture, and reinforced the black cults of emotion. There was no compromise, no dreary sophistication, only the elegance of something that is too ugly to be described, and is diluted only at the agent's peril. All the saxophonists of that world were honkers … the great sounds of our day. Ethnic historians, actors, priests of the unconscious. That stance spread like fire thru the cabarets and joints of the black cities, so that the sound itself became a basis for thought, and the innovators searched for uglier modes.[38]

In this account, the sheer ugliness of the saxophonic scream redeemed it from the "compromise" and "dreary sophistication" of mere music to endow it with a more drastic expressive force, something "past music" and beyond verbal description, even a "basis for thought." As we will note later, Baraka heard the anguished ecstasy of the scream in racial terms ("hatred and frustration"), but for now I want to emphasize one final point before moving on. As both sides of the critical debate over Coltrane amply evince, emotionally polarized responses to his scream set up a complex interplay between visceral reactions and culturally-situated appraisals. It would seem

[38] Imamu Amiri Baraka, *The LeRoi Jones/Amiri Baraka Reader* (New York: Basic, 2000), 174.

that both positive and negative valuations of these particular timbral gestures were processed at a gut level, yet how different listeners made sense of the experience was largely determined by their willingness to either engage and empathize with this acutely embodied sound (as a new kind of thought; as blissful, physical transport), or reject it (as thoughtless and nasty; as noisy, physical pain).

We will pick up this hermeneutical thread later in the chapter. But before going further, I want to turn to a representative musical example from the *Meditations* album. As saxophonist David Liebman has described, noisy timbre is the primary expressive ingredient in this late masterpiece, the main carrier of its sensory impact: "The color spectrum was … an assault upon the senses, especially to those who were used to the more traditional quartet sound."[39] Since it was a lighting rod for many of the critical comments just reviewed, next we will take a closer analytical look at *Meditations*—namely the effectual saxophonics on its opening track, "The Father and the Son and the Holy Ghost"—in order to explore how Coltrane's "scraping and scouring" sounds work in musical context.

"The Father and the Son and the Holy Ghost"

The opening cut on *Meditations*, "The Father and the Son and the Holy Ghost" begins with Coltrane and Sanders playing an incantatory drone (E♭5 and A♭4) with the use of multiphonics in the altissimo range (see Table 4.1 for a complete list of timbral happenings).[40] The effect of this dry, hoarse, and sputtering timbral gesture is distinctly vocal, with yodel-like cracks between the

[39] Liebman, "John Coltrane's Meditations Suite: A Study in Symmetry," 177.

[40] For more, see ibid.

two pitches and a palpable sense of strain. However, though vocally mimetic, this opening is entirely alien to standard modes of vocal and instrumental expression; in his liner notes, Hentoff likens this figure to glossolalia, or speaking with the "gift of tongues."[41] Already in the first few seconds of the album the listener is immersed in a wash of saxophonics with a distinctly transgressive vocal-mimetic flavor.

TABLE 4.1: Major timbral-corporeal events

Time	*Event*	*Techniques/Timbre*
0–0:47	Drone (Coltrane/Sanders)	Altissimo multiphonics
0:47–1:56	Head motive (Coltrane)	Exaggerated vibrato
1:56–2:06	Drone (Coltrane/Sanders)	Altissimo multiphonics
2:06–6:30	Coltrane solo	
2:06–	*Variations on head motive*	
2:30–	*Screaming (F#5)*	Altissimo with added noise elements
2:33–	*Low register*	Overblowing to produce lower harmonics
2:56–	*Mixed register, screams (B♭5)*	Screams alternating with rapid figuration across the full range of the saxophone
3:43–	*Duet: screaming & low*	Overblowing, sustained screaming
4:00–	*Duet: screaming & middle*	Overtones, multiphonics on fast 3-note figure
4:20–	*Duet: screaming & low*	Call-response repetition of related patterns in extreme high and low registers
4:47–	*Rapid mid-register w/screams*	Overtone-rich, occasionally growled middle register punctuated with screams
5:39–	*Timbral anomaly by extraction*	Spectral energy jumps between overtones of same fingered pitch
6:20–	*Screamed head (C6–E6)*	Highest altissimo playing of the solo
6:30–	*Cool down, Sanders begins*	
6:30–10:28	Sanders solo	
6:30–	*Purring and hissing*	Rolled /r/ while playing
7:17–	*Shrieking*	Inharmonic screaming, biting reed and vocalizing

[41] Hentoff, liner notes to John Coltrane's *Meditations*. In the Pentecostal context, glossolalia is only "gibberish" if the listener does not understand; to insiders who hear the signal in the noise, such as other members of the religious community, it *signifies*.

9:04–	*Coltrane joins*	
10:28–	Drone (Coltrane/Sanders)	Altissimo multiphonics
10:52–	Head motive (Coltrane)	Exaggerated vibrato
11:22–end	Drone (Coltrane/Sanders)	Altissimo multiphonics

At 0:47, Coltrane breaks away to state the melody while Sanders continues with the drone. Contrary to free jazz's reputation for "difficulty," the head motive is entirely diatonic, even child-like: beginning in D♭-major, Coltrane spins out a repetitive four-note melodic fragment in a range of different key centers connected to each other by ascending fourth or descending minor third (Fig. 4.1). His middle-register tone, though unmarked by the noisy extended techniques of the opening, carries an exaggerated vibrato. As David Borgo points out, extreme vibrato to Coltrane and other avant-garde saxophonists was part of the sonic arsenal for the expression of ecstatic states, a timbral sign redeemed from earlier "un-hip" usages and fashioned into a signifier of spirituality.[42] Seen from this perspective, by oscillating the pitch so dramatically Coltrane adds a sense of otherworldly timbral dynamism to what might otherwise be a trite sequence of phrases. At 1:56 he rejoins Sanders—fiercer now—for another restatement of the opening incantation.

[42] Borgo, "Between Worlds: The Embodied and Ecstatic Sounds of Jazz."

FIGURE 4.1: Head motive

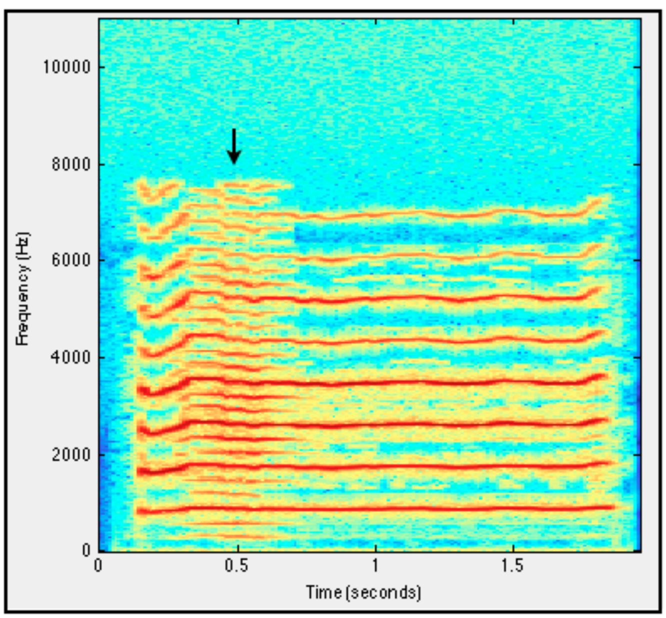
Beginning at around 2:06 and lasting for the next four and a half minutes, Coltrane's maelstrom of a solo is structured entirely around contrasting extended techniques in contrasting registers of the instrument. These can be roughly divided into three recursive patterns (from low-register to high): (1) low-note overblown playing, which drives energy into the lower harmonics; (2) rapid figuration in the middle-register, particularly fleet three-note patterns with added overtones, growls, and multiphonics; and (3) sustained altissimo screaming. In ever-shifting combinations, these three registral/timbral topics make up the primary vocabulary of the solo.

After playing variations on the head motive, at 2:30 Coltrane fires off the first unambiguous scream of the record, a short-lived high F# that quickly tumbles down to the lowest notes of the instrument. Staying at the bottom of the saxophone (2:33), Coltrane overblows to produce the first few harmonics, adding a bristling, coruscating sheen to his sound. From here, he moves into a mixed register passage (2:56) exploiting the full range of the instrument, and follows with the first "duet" of the solo. During his late period, Coltrane was fond of improvising duets with himself at the registral extremes of the instrument—beginning at 3:43, we hear heavily overblown bombs in the basement of the saxophone (A♭/B♭2) punctuated by screams. Playing in the low and middle register, it should be noted, Coltrane typically sticks to hard-articulated, relatively quick figuration, but in the altissimo register, screams are sustained. This is doubtless due to the mechanics of altissimo technique—since notes that are not "supposed" to be played on the instrument require upper harmonics via vocal support, they are not as dexterous as the normal register. Phenomenologically speaking, this quirk of extended technique brings the saxophonic scream even closer to its vocal correlate by providing momentary escapes from the mechanical finger-work of the faster passages. If the middle- and lower-register patterns are digital (i.e., dominated by rapid finger movement), the altissimo sections showcase the saxophone as analogue of the human voice.

After a series of duets between Coltrane's three registral/timbral personae and an extended mid-register passage characterized by cascading streams of distorted, overtone-saturated sounds, at 5:39 we arrive at a section that explicitly foregrounds the acoustic principles intrinsic to effectual saxophonics. Of course, the whole solo relies on these principles, but here the veil is pulled back, so to speak, to sonify the timbral-corporeal *act* of moving between overtones of the same fingered pitch. Cornelia Fales describes timbral manipulation of this sort, which is common to Tuvan throat singing and didgeridoo playing, as "anomaly by extraction." She explains: "a subset of the harmonics comprising a tone's timbre is amplified to a point of sufficiently greater intensity than surrounding harmonics, that they stand out in relief from the remainder of the tone's timbre."[43] By channeling energy into different harmonics of the same fundamental pitch, a

[43] Fales, "The Paradox of Timbre," 67.

musician can create a timbral kaleidoscope where the "true" pitch remains perceptually ambiguous. To illustrate, Fig. 4.2 shows a spectrograph of an overtone-rich scream (indicated with the arrow) transitioning into a normal altissimo B♭5 (933 Hz). The fingering remains the same here, but a shift in energy during the scream—executed with the aid of the larynx, it should be added—fills the spectrum with additional harmonic (and weaker inharmonic) overtones both above and below the sounded fundamental.[44] The harmonic regularity of this series acoustically reveals the "true" pitch—what we may have heard as the fundamental (933 Hz) is really the third harmonic of E♭4 (311 Hz).[45] This moment of perceptual sleight of hand gives way to a clear and steady altissimo fundamental, but the "extracted" harmonics, like ghostly afterimages, remain in memory to add an element of phenomenological disbelief to what we are hearing.



[44] This phenomenon, known as harmonic distortion, will be explored in relation to electric guitar distortion in the next chapter.

[45] I would be hard pressed to call this timbre entirely *harmonic*, however, owing to the salience of inharmonic energy. Although most of the spectral energy is concentrated in harmonically-related bands—311, 622, 933, 1244 Hz, etc.—there is quite a bit of fuzziness in between.

FIGURE 4.2: Timbral anomaly by extraction[46]

This solo passage exemplifies "timbral anomaly by extraction" in a range of different contexts. We hear Coltrane searching for different overtones within one fingered pitch; at times, a note will morph suddenly up or down an octave, fifth, or third, indicating Coltrane's bodily struggle to usher energy into different discreet spectral bands. Often, when the energy smears between such bands, this technique imparts a feeling of being perceptually untethered to the rational grounding of fixed scales. Sonifying this physically arduous timbral-corporeal search, the passage is a fitting corollary of Coltrane's (and the listener's) spiritual journey. As Borgo explains: "exploring the upper partials can … greatly expand a musician's range and the searching and frequently unpredictable quality of these extended harmonic forays has been interpreted by many as conveying a sense of spiritual yearning."[47] Following this exciting section, the "yearning" intensifies as the saxophone is catapulted into the highest register of the solo for a climactic, screamed statement of the head motive before Coltrane gradually cools off and passes the baton to Sanders.

Though timbrally abrasive, Coltrane's solo is tame compared to the playing of his bandmate, tenor saxophonist Pharoah Sanders (1940–). What Sanders's solo lacks in registral/timbral diversity and motivic richness, it more than makes up for in sheer audacity—even Coltrane supporters have described Sanders's playing as "more noise than music."[48] The solo begins with a purred, wheezy timbre likely produced by forming a hard, rolled /r/ sound in the back of the

---

[46] Signal downloaded from the University of New South Wales Acoustics Laboratory; spectrograph generated with PsySound3 for MATLAB.

[47] Borgo, "Between Worlds: The Embodied and Ecstatic Sounds of Jazz."

[48] Nisenson, *Ascension: John Coltrane and His Quest*, 177.

throat while blowing into the instrument. It is temporally dynamic, with rapid amplitude fluctuations and an unsteady tendency to squeak and squeal. This particular technique is one of the central timbral ingredients of the solo.

Beginning at 7:17, we are confronted with another of Sanders's ingredients, "ultrahigh altissimo shrieking."[49] Sanders has a saxophonic scream, like Coltrane's, but of a different subspecies, much higher and choked with inharmonic energy, burying the pitch in a ferocious squall of noise. To play some of the highest of his shrieks, Sanders bites the reed, sometimes with considerable force bordering on (and even exceeding) the *ultimate limit* of the instrument. According to widely circulated stories, Sanders occasionally bit his reed so hard during performances that it shattered in the mouthpiece, forcing him to spit out a mouthful of bloody wood shards and replace it before he could resume.[50] Sanders was also known to scream directly into his instrument.

Critic A. B. Spellman describes his first experience of Sanders in a *Down Beat* concert review from 1965:

> He went on for minute after minute in a register that I didn't know the tenor had… Those special effects that most tenor men use only in moments of high orgiastic excitement are the basic premises of his presentation. His use of overtones, including a cultivated squeak that parallels his line, is constantly startling. He plays way above the upper register; long slurred lines and squeaky monosyllabic staccatos…[51]

This particularly startling effect came to be Sanders's timbral signature, his primary artistic voice. But what sort of a voice is this? If Coltrane's sound mimics a screaming human voice, Sanders's is a shriek of a different, non-human sort. As a colleague once put it, Sanders's playing is "dying animal music" at its most disturbing, representing perhaps the most strident timbral gesture in

[49] This is Liebman's term: see Liebman, "John Coltrane's Meditations Suite: A Study in Symmetry," 177.

[50] I thank New York-based saxophonist Nolan Lem for this piece of oral history.

[51] Quoted in Hentoff, liner notes to John Coltrane's *Meditations*.

the free jazz repertory.[52] Its stridency, moreover, is compounded by sublime implacability ("[going on] for minute after minute…"); Sanders's tendency to sustain his screams for long intervals undoubtedly contributed to critics feeling "under assault." To be sure, Sanders *demands* your attention—he does not ask politely for it.

At the peak of this sonic onslaught, Coltrane returns (9:04) and the two overlap for about a minute of additional soloing using the vocabulary developed in their respective individual solos. The piece closes symmetrically by restating the opening drone and head motive passages, though this time with greater intensity. After a final flare-up of multiphonic droning, the voices of both saxophones gradually fade.

I have described one representative late Coltrane recording in detail to show that the saxophonic scream, as defined in this chapter, is really quite a varied and diverse timbral-corporeal phenomenon, by no means a nondescript slab of noise (though acoustically noisy it certainly is). But I have not yet gone into any depth exploring what exactly constitutes the saxophonic scream. In the next section, we turn to this timbre *per se*.

Saxophonic Screaming

The timbre of the scream is not unique to free jazz saxophone playing, of course. As Baraka readily points out, the "ecstatic" sonic mode in black music, religious ritual, and other cultural practices can be traced back to the days of slavery and before, serving an important role in Afro-

---

[52] This description can be attributed to Lindsay Oura in conversation with the author, New York, 1999.

diasporic "low-down spirituality" (in Hakim Bey's formulation).[53] Neither were Coltrane and company the first saxophonists to experiment with this registration and timbre: screaming has a long history as an R&B "freak effect" perfected by Illinois Jacquet, Big Jay McNeely, and others in the 1940s and 50s.[54] However, the free jazz cohort were the first to place this electrifying timbre at the center of their aesthetic agenda: it is not mere sound effect or priapic display, but an essential expressive ingredient. It is the sonic signature of the genre.

The saxophonic scream in free jazz has been interpreted as spiritual ecstasy, a rejection of Western instrumental heritage, the "primal scream" method of psychoanalysis, and radical racial politics.[55] Underlying these interpretations is one commonality: *the saxophonic scream is tantamount to its vocal equivalent*. Practitioners explicitly drew this link: Archie Shepp, for example, remarked that effectual saxophonics "gets a quality of … male and female voices."[56] As a "superexpressive voice,"[57] the high register and distorted timbral components of the saxophonic scream directly mimic the vocal scream and its accompanying bodily strain. The "scream" designation, then, is not only a felicitous INSTRUMENTS ARE VOICES metaphor; it is impossible to disembody the saxophonic scream by separating it from its vocal analog. The sound is heard *as if* it were an actual screaming voice.

---

[53] For more, see respectively Michael Spence Washington, "Beautiful Nightmare: Coltrane, Jazz, and American Culture" (Ph.D. Diss., Harvard University, 2001), 149 & 248; Hakim Bey, "The Utopian Blues," in *Sounding Off!: Music as Subversion/Resistance/Revolution*, ed. Ron Sakolsky and Fred Wei-han Ho (Brooklyn: Autonomedia, 1995). McClary and Walser offer further explanation of this cultural phenomenon, particularly the "ring shout"; see Susan McClary and Robert Walser, "Theorizing the Body in African-American Music," *Black Music Research Journal* 14, no. 1 (1994): 77.

[54] See Nisenson, *Ascension: John Coltrane and His Quest*, 188. This lineage of saxophonic screaming, of course, continues to the present (e.g., Clarence Clemons of Bruce Springsteen's E Street Band).

[55] For more on the various interpretations of the scream, see ibid., 188-90; Borgo, *Sync or Swarm: Improvising Music in a Complex Age*, 47.

[56] Quoted in Simpkins, *Coltrane: A Biography*, 188.

[57] This term, to recall, comes from Juslin and Laukka: see Chapter 1.

Theoretical approaches to the vocal scream itself, then, may help shed light on its saxophonic equivalent. One of the earliest modern writers to ruminate on the aesthetics and psychology of the scream in a musical context was Richard Wagner, who employed the effect first in *Der fliegende Holländer* (1843) then regularly throughout his career. To Wagner, screaming was "the fundamental element of all human appeal to the hearing," an utterance of Schopenhauer's Universal Will that was primordial, emotionally direct, and unmediated by musical (read: cosmopolitan, Jewish) convention.[58] He explains: "Without any reasoning go-between we understand the cry for help, the wail, the shout of joy, and straightaway answer it in its own tongue."[59] The scream, in short, transcends musical expression. Characterizing Wagner's thinking on the topic, Philip Friedheim writes: "[Screams] are extroverted and elementary, indeed primitive. At these moments, natural energy manifests itself not in word and sentence, but in pure sound that allows the emotion to engage the listener's attention directly."[60] The same thing seems to be true of the saxophonic version. (Spellman, incidentally, compared the intensity of *Ascension* to Wagner in his liner notes.) The reception history of free jazz saxophonics reproduces with remarkable fidelity Wagner's premise: screaming is a reflection of "pure sound, pure emotion," an existential wallop of potent affect and bodily arousal that breaks through the rationalizing code of "music."

What sort of a timbre is capable of conveying this heightened immediacy? And how is it that this sound came to be heard as a "scream"? At this point, I want to shift our focus back to the early, reactive stages of ASPECS to consider the saxophonic scream as *act*, *sound*, and *perception*. The intensely polarized reception history chronicled in this chapter, while clearly discursive,

---

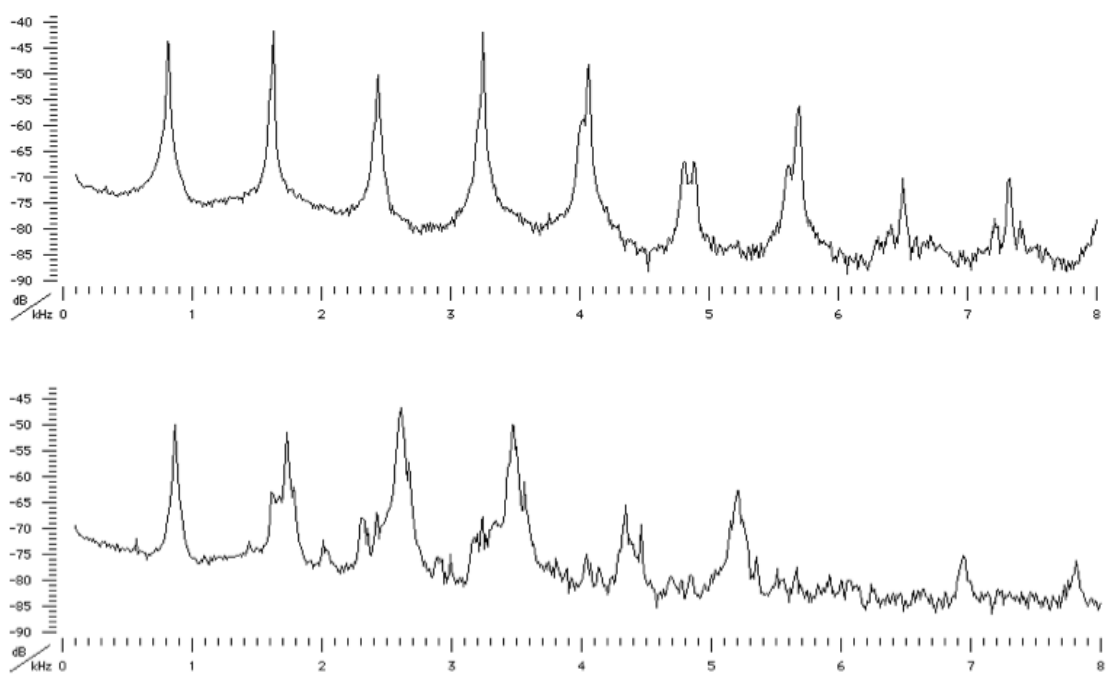
[58] Quoted in Philip Friedheim, "Wagner and the Aesthetics of the Scream," *19th-Century Music* 7, no. 1 (1983): 67.

[59] Ibid.

[60] Ibid., 68.

is constrained by basic properties of sounds and bodies. These properties work their way up to affect how the saxophonic scream is interpreted *qua* sign.

Let us begin with sound. Consistent with the vocal "sound of arousal" discussed in Chapter 1, saxophone screaming pushes more energy into higher frequency components, introduces new harmonic and inharmonic overtones, flattens the spectrum, and exacerbates auditory roughness.[61] To demonstrate this, Fig. 4.3 compares spectra of a standard altissimo note with its screamed equivalent: note the smooth, reasonably clean harmonic peaks in the first timbre that are flattened out and muddied in the screamed version, especially in the higher frequency range.[62]



[61] For a detailed acoustical account, see Vincent Gibiat and Michèle Castellengo, "Period Doubling Occurences in Wind Instruments Musical Performance," *Acustica* 86(2000).

[62] Of course, screaming is not a binary, on/off technique, and the lower spectrum here is a fairly mild version of this effect (though still, I would argue, phenomenologically a saxophonic scream). A more intense version would have a flatter, more inharmonic spectrum still.

FIGURE 4.3: Standard altissimo (top) and screamed altissimo (bottom)
(from University of New South Wales Music Acoustics Laboratory)

As with its vocal equivalent, the playing technique required to produce this effect on the saxophone typically involves a relatively high degree of physical exertion. Altissimo notes require the player to sound the right pitch with his or her vocal tract: ascending higher into the altissimo range, therefore, requires the same vocal demands as upper-register singing.[63] Figure 4.4 shows the relationship between vocal tract tuning and sounded pitch:[64] in the normal range, the two are unrelated, but as the player ascends into the altissimo range vocal tuning begins to play a vital role. The screaming saxophone is not just experienced as such by the listener—it requires bodily exertion and vocal strain on the part of the player as well.

---

[63] Players have long been aware of the connection between altissimo performance and the voice. In his 1989 saxophone manual, for instance, David Liebman dedicates a whole chapter to larynx control. See David Liebman, *Developing a Personal Saxophone Sound* (Medfield, MA: Dorn, 1989). As Nolan Lem tells me, even most advanced players tire quickly when performing these sorts of effectual saxophonics. Email communication with Nolan Lem, May 7, 2010.

[64] Jer-Ming Chen, John Smith, and Joe Wolfe, "Experienced Saxophonists Learn to Tune Their Vocal Tracts," *Science* 319(2008).

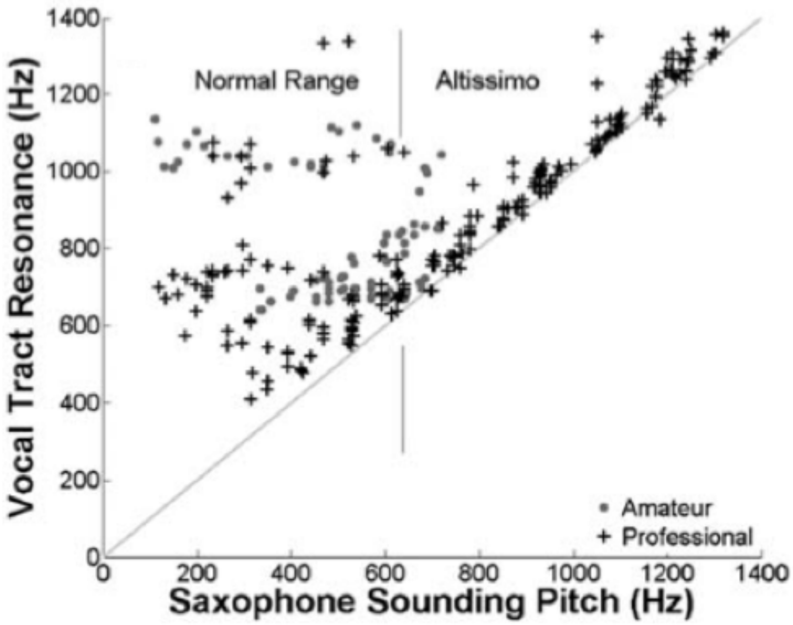


FIGURE 4.4: Role of vocal tract in saxophone altissimo
(from Chen, Smith, & Wolfe, 2008)

The saxophonic scream relies on vocal techniques to produce a sound marked with the timbral indices of overexertion also typical to the voice. But simply noting this correlation does little to address its contentious history. What accounts for the uncommonly polarized tenor of Coltrane reception, the reason listeners tend to hear it in diametrically opposed terms (as either ecstasy or painful noise) despite shared grounding in the voice?

Saxophonist/ethnomusicologist David Borgo has written persuasively on the link between effectual saxophonics and the perception of ecstatic, altered states of consciousness.[65] Though often interpreted as an icon for spiritual search by pushing beyond the expected bounds of register and timbre, he argues, "there is nothing inherent in the nature of extended saxophone tech-

[65] See David Borgo, "Reverence for Uncertainty: Chaos, Order, and the Dynamics of Musical Free Improvisation" (Ph.D. Diss., UCLA, 1999); "Between Worlds: The Embodied and Ecstatic Sounds of Jazz."; *Sync or Swarm: Improvising Music in a Complex Age*.

niques that triggers ecstatic states in musicians and listeners."[66] While I am in substantial agreement, the empirical studies of this dissertation suggest Borgo might be selling effectual saxophonics just a little bit short. As found in the experiments outlined in Chapter 2—which only look at growling and multiphonics in the comfortable middle register, let alone altissimo—it appears that extended, "noisy" saxophone techniques *do* drive up activity in regions of the brain that play a major role in strong affective reactions to music, ecstasy included.[67] I am not reductively arguing that this sound universally "triggers ecstasy"—appraisal is too contextual for such a specific response—but it does, it seems, make substantial limbic, motor, and (per Chapter 3) somatosensory demands on the listener via the neurophysiology of act–sound coupling. Such extremity of body and sound, moreover, demands an equally extreme appraisal, hence the tendency to default to one of two strong positions in the reception literature. One would be hard pressed, indeed, to find a jazz listener with an indifferent view of the saxophonic scream. While not necessarily always "ecstatic" in reception, I contend that effectual saxophonics are at the same time not *arbitrarily* associated with ecstasy or pain either—listeners feel strongly about them one way or another owing in large part to the dynamics of embodied timbre cognition.

Exploring these dynamics, psychologists Mark Gridley and Robert Hoff have co-authored a number of quirky and fascinating papers addressing the cognitive psychology of Coltrane criticism.[68] They posit a role for the mirror neuron system in critics' misattribution of emotions—hearing "ecstatic" sounds as "angry"—and the data from Experiment 3 appear to support this

---

[66] Quoted in Ake, *Jazz Cultures*, 141-2.

[67] As we recall from Experiment 3, noisy saxophone timbre is linked to activity in amygdala, which is one of the primary structures identified by Becker in deep listening and musical trance: see Becker, *Deep Listeners: Music, Emotion, and Trancing*.

[68] For example, see Mark C. Gridley and Robert Hoff, "Who's Actually Angry, John Coltrane or His Critics?," *Psychology Journal* 4, no. 4 (2007); "Are Music Perceptions Biased by Priming Effects of Journalism?," *Psychology Journal* 7, no. 2 (2010).

basic idea.[69] Noise elements in the saxophone set off a cascade of neural activations in a wide swatch of regions, including mirror neurons. To illustrate this visually, Fig. 4.5 compares the same view of the brain in two conditions, (a) listening to a growled, noisy saxophone timbre, and (b) listening to regular saxophone tone. The difference is stark. It is clear that processing a noisy saxophone timbre is much more cognitively demanding than its non-noisy counterpart, with activity in a variety of sensorimotor and limbic areas not active in the other condition.[70] Furthermore, activity in tactile (and tactile metaphor) processing areas might help to explain why noisy saxophone timbres were heard in terms of the NOISE IS FRICTION metaphor (e.g., "scouring and scraping"). It thus appears that noise in a saxophone timbre is always already somatically marked, and the marking here is specifically negative.[71] In the self-protective manner of all reactions, that is, the sound seems to be saying: "pay attention—something intense and effortful is going on here!" The listener has no choice but to respond: to listen to the saxophonic scream is in some way to internally scream along with it. (Or, as Wagner might put it, "straightaway answer it in its own tongue.")[72]

---

[69] "Do Mirror Neurons Explain Misattribution of Emotions in Music?," *Perceptual & Motor Skills* 102, no. 2 (2006).

[70] These include supplementary motor area (SMA), pre/primary motor cortex (PMC), primary somatosensory cortex (PSC), and limbic areas (amygdala [A] and hippocampus [H]). Note the intensification of activation in the primary auditory cortex (PAC) as well.

[71] Recall once again that the saxophone sounds used for these images were not even screams, but growls and multiphonics in the middle register; presumably an altissimo scream would have produced even more activity in these and possibly other regions.

[72] The mimetic properties of the saxophonic scream, moreover, are evident in the slippage between instrumental and vocal screaming common in the reception history. On many documented occasions, "internal screaming"—motorically mirroring this potent timbral-corporeal event—escaped the level of covert mimesis to translate into overt screaming, both by performers and audiences. It was not uncommon for Coltrane and Sanders to put down their horns and scream directly into their microphones; similarly, audiences during Coltrane's late period often broke into spontaneous screams, either in appreciation or while running out the door with hands over their ears. Bassist Art Davis explains: "I would be there and people would just be shouting, like you go to church, a holy roller church or something like that… John had that sprit—he was after the spiritual thing. You could hear people screaming and this was something, despite the critics who tried to put him down." Drummer Billy Hart paints a similar picture, recalling, "I had never heard that many people screaming." Furthermore, saxophones were not the only things screaming during the recording of *Ascension*: as Brown recounts, "We did two takes, and they both had that kind of

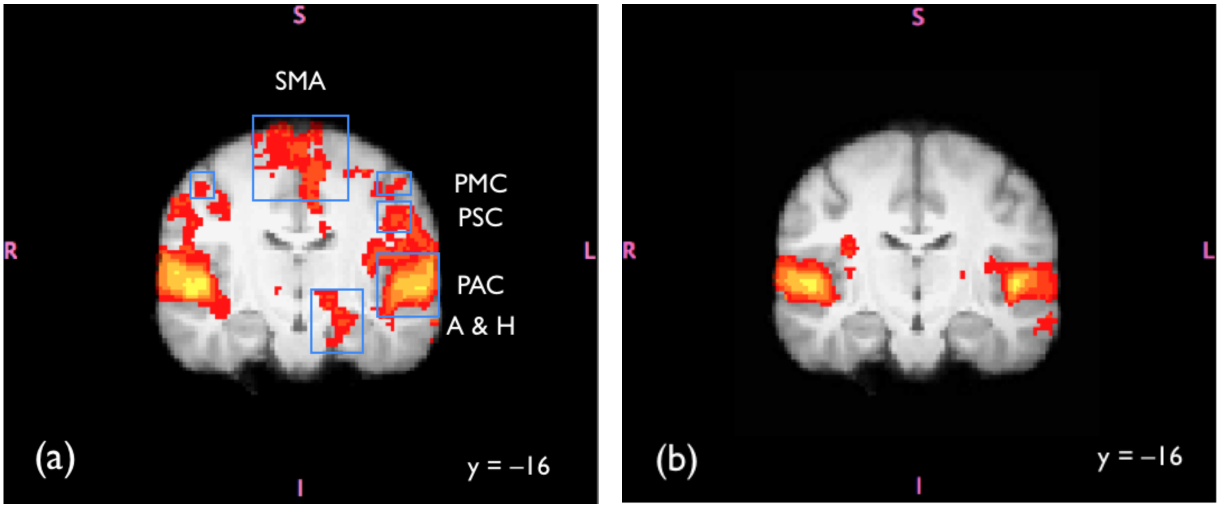


FIGURE 4.5: (a) Growled saxophone; (b) Regular saxophone[73]

How might we read these empirical results into the reception history of saxophonic screaming? Phenomenology can be helpful in theorizing how mechanisms of perception contribute to emergent aesthetic experience; the key here seems to be the concept of *possession*. As Merleau-Ponty explains, "crude sounds of instruments, as presented to us in natural perception, are insufficient to provide the musical sense of music," adding that in order for raw noises to be heard as music, one must "*possess* the articulatory and acoustic style as one of the modulations, one of the possible uses of my body."[74] We are all, as just observed, psychophysiologically subject to the drastic corporeality of sound—the saxophonic scream illustrates this dramatically—but whether or not we choose to "possess" what we are hearing, to appraise it positively in spite of its

---

thing in them that makes people scream. The people who were in the studio *were* screaming. I don't know how the engineers kept the screams out of the record." Accounts like these illustrate the point at which motor mimesis manifests itself as shared action—the saxophonic scream is transmogrified into its referent, the actual scream. For more, see Nisenson, *Ascension: John Coltrane and His Quest*, 207; Washington, "Beautiful Nightmare: Coltrane, Jazz, and American Culture," 284-85; Simpkins, *Coltrane: A Biography*, 206-10; De Sayles Grey, "John Coltrane and the 'Avant-Garde' Movement in Jazz History" (Ph.D. Diss., University of Pittsburgh, 1986), 113; Spellman,

[73] Abbreviations of brain regions include: SMA (supplementary motor area), PMC (pre/primary motor cortices), PSC (primary somatosensory cortex), PAC (primary auditory cortex), and A & H (amygdala and hippocampus).

[74] Merleau-Ponty, *Phenomenology of Perception*, 208 & 10. Italics mine.

negative somatic markings, to *know with our bodies* that it is good, is a matter of considerable dissention. Thomas Clifton frames this particular perceptual polarity in terms of the opposing experiences of "closing in" or "opening toward":

> A raw sound event is experienced as such when an involuntary proximity of sounds is felt as closing in—a condition we wish to be free of by fleeing from it, putting it at a distance. This can be done literally or figuratively; even ignoring a sound event is fleeing from it, in a sense. On the other hand, music is experienced as such when a voluntary proximity of sounds is felt as opening toward: a condition *toward* which we use our freedom to affect the closure between ourselves and music.[75]

This conceptualization wields considerable explanatory power when applied to the saxophonic scream. For fans, the scream was heard as "opening toward" rapturous extra-musical possibilities: in a sense, taking bodily possession of the scream—feeling it as "music," or something "beyond" music—allowed listeners to empathize with its implied anguish or ecstasy. As Clifton notes, "What makes sound remain merely sound is the absence of any bodily complicity with it."[76] Here the body is fully complicit and in league with the screaming timbre (and the screaming musician producing it). Consistent with these imaging results, moreover, I would hypothesize that *possession* to acquiescent listeners is strongly correlated with activity in larynx control centers of the brain, indicating the intercorporeal co-experience—at the subvocal level, at least—of the scream.[77]

But "bodily complicity" is not the exclusive province of musical sounds to which we "open up," as Clifton proposes. Figure 4.5 demonstrates that bodily complicity, defined as broad

---

[75] Clifton, *Music as Heard: A Study in Applied Phenomenology*, 279. Italics in original.

[76] Ibid.

[77] To recall from Chapter 2 (Experiment 3), positive valence—i.e., liking a timbre—correlates with activity in the Rolandic operculum, the major larynx area. A similar result was found by Koelsch et al. (2006) cited in the first two chapters.

sensorimotor activity, appears to be involved in the perception of a-musical, noisy, "closing-in" sounds as well. Cross-fertilizing these empirical findings with the phenomenological speculations of Merleau-Ponty and Clifton, I want to suggest that the compulsion of certain listeners to "flee" from the noisy timbres of free jazz saxophonics—to "defend themselves by not listening," as Goldberg put it—demonstrates a different kind of bodily complicity. When a sound "closes in"—a phenomenological prerequisite for the experience of "noise"—listeners feel they have no choice but to disengage from the noise; but this is only because they are *already involuntarily complicit* with it. Mirror neurons in the motor cortex are firing, but this sensorimotor engagement is not freely elected. The sounds "close in" because they are neurophysiologically coercive; the listener does not like what the sound of the scream is *making them do*. "Disinterested listening" is simply not an option—one is forced to take sides.

Side-taking demands explanations. One does not tend to feel strongly about something without a narrative justifying one's appraisal, and indeed, justificatory story-telling is thick on the ground in Coltrane reception. We have already explored one story about the saxophone scream, a spiritual interpretation that sees it as a sign for ecstatic transcendence of the self. But what of other, less uplifting interpretations for the scream's overwhelming somatic intensity?

Gender and Race in Coltrane Reception

> It is a basic fact of life (psychologically and physiologically) that any one thing repeated for too long a time without variation becomes boring and/or dulling. It is the artist's job to be sensitive to the fine line to where a continued effect is building interest but if carried any further will lose its interest … In the great bulk of Coltrane's work we get a good deal of filigree or decoration … but very little "meat"…
>
> — Don Ellis, review of *The John Coltrane Quintet Plays* (1965)

> [Ellis's review] shows the white's utter contempt for black creativity. The white's ability to condescendingly dismiss black music as "filigree and decoration." […] The white's mind still, even in '66, seeks to castrate the black's music and write it off in the form of European-based technical criticism and control—white control … The feeling of this music is more important to me than the technical matters; a feeling that you, Mr. Ellis, have insulted, thereby declaring yourself as another of my many white enemies. And for that, along with your ideals and artifacts from ancient history, you must die.
>
> — Charles Moore, in response to Ellis's review

> I hope the *rational* reader will take the trouble to listen to the record in question … In your letter you imply you are mainly interested in "feelings." Well, there is one large difference between us then—I am interested in good ART.
>
> — Don Ellis, in reply to Moore's response[78]

To say that the rhetoric surrounding John Coltrane reception was "heated" would be a vast understatement: in the last two years of his life, it took on an incendiary fervor typically reserved for sectarian strife or religious fanaticism, even to the point of death threats. Witness the above exchange, published in *Jazz* magazine in 1965–66. Here, Ellis—a white trumpeter—criticizes Coltrane on the familiar grounds that it is not good "music" (or "ART," capitalization and all); that it is boring; that "*rational*" critique of its shortcomings is better than emotional ardor and "feelings."[79] Enter black trumpeter Charles Moore, whose response is not so much a defense of Coltrane as a repudiation of the reviewer: Ellis is blinded by a narrow, ethnocentric understanding of "music"; he aims to "castrate" black music; he is more concerned with European notions of technique than "feelings." Some well-rehearsed lines of criticism are repeated here, only with added vehemence. As Ratliff elegantly observes about the escalating polarization of Coltrane reception, "What was taking shape here was an ugly circle of irritation, based on reductive white-

[78] The whole exchange is quoted in Ratliff, *Coltrane: The Story of a Sound*, 163-65. Italics in original.

[79] Sentiments such as these echo Welding and Larkin, as previously discussed.

listener notions and reductive black notions of the white notions and reductive white notions of the black-listener notions."[80] By 1966 Coltrane criticism had reached an impasse; it had morphed into all-out combat, with competing tribal allegiances threatening to drown out the music. This was no longer the "Coltrane controversy"—this was war. Music, of course, was not the only thing at stake.

The Ellis-Moore exchange was particularly harsh: most writers and listeners, presumably, did not consider Coltrane's sound something worth killing for.[81] Nevertheless, their fight accurately reflects the racial anxieties, political preoccupations, competing aesthetic agendas, and gender dynamics common to the discourse surrounding Coltrane's "language-transcending" music. Violently contrasting views about race and gender were projected onto timbre with uncommonly furious zeal, I will argue, because of timbre's very amorphousness as a sign.

Race has been the stock-in-trade of Coltrane reception and scholarship—often wrongheadedly, as we will see—but gender remains undertheorized. Let us begin, then, by gendering the most common INSTRUMENTS ARE VOICES metaphors in the reception of the saxophonic scream, namely the familiar terms "scream," "shriek," "cry," and "wail." Words like these are common coin for listeners and performers in this style, but "screaming" is not a gender-neutral act. It is impossible to dissociate these vocal expressions from their millennia-old connotations in the Western world: "screaming," "shrieking," "crying," and "wailing" are heavily gendered terms, associated since the Greeks with uncontained female vocality and hysteria.[82] As Anne Car-

---

[80] Ratliff, *Coltrane: The Story of a Sound*, 165.

[81] The curious reader will be relieved to know that Don Ellis died at his home in North Hollywood in 1978—of natural causes.

[82] The most comprehensive account of the dangers of female vocality in classical philosophy, to my knowledge, can be found in Part II of Cavarero, *For More Than One Voice: Toward a Philosophy of Vocal Expression*. For an illuminating reading of "noisy" female sound in the context of free jazz improvisation, see Julie Dawn Smith, "Perverse Hysterics:

son explains, women in classical literature are “a species given to disorderly and uncontrolled outflow of sound—to shrieking, wailing, […] screams of pain or of pleasure and eruptions of raw emotion in general.”[83] High-pitched voices, especially those marked by the timbral stress of screaming, are the sounds *women* make, and although (male) free jazz saxophonists readily used these words to describe their own sounds, such designations tap into an implicitly gendered discourse of power.

Noise as symbol is operative in the determination of whether such sounds represented the masculine overriding of earlier connotations, or simply (to again quote Carson) an “ideological association of female sound with monstrosity, disorder and death.”[84] As cultural historian Peter Bailey points out, noise “sexes” men while “unsexing” women,[85] a linkage that is clear in critical and popular characterizations of Coltrane as, in Ratliff’s words, the “primal masculine machine, emitting the ‘cry of jazz.’”[86] Coltrane recuperates the female scream into an icon of power and masculinity.

Gendered readings of Coltrane’s sound figure into the Ellis-Moore exchange as well. Ellis, in language that is coded but obvious, disregards Coltrane’s playing as “filigree or decoration,” in contrast to the real “meat” of good art; he then goes on to observe that Coltrane rarely plays “positive strong statements.”[87] Simply put, Coltrane’s music is fluff, an evaluation with long-held

---

The Noisy Cri of Les Diaboliques,” in *Big Ears: Listening for Gender in Jazz Studies*, ed. Nichole T. Rustin and Sherrie Tucker (Durham, NC: Duke University Press, 2008).

[83] Anne Carson, *Glass, Irony and God* (New York: New Directions, 1995), 126.

[84] Ibid., 121.

[85] Peter Bailey, “Breaking the Sound Barrier: A Historian Listens to Noise,” *Body and Society* 2, no. 2 (1996): 63.

[86] Ratliff, *Coltrane: The Story of a Sound*, 166.

[87] Ibid., 163.

connections to the tastes and talents of musicking women—"pretty but trivial."[88] Indeed, the masculine in music (in music historiography, read: German music) is meaty, profound, and strong, while the effeminate is light, decorative, and frivolous. Moore picks up on the rather transparent gender coding of this critique ("[he] seeks to castrate black music"), reaffirming free jazz's (and his own) virility in response. The gender meanings of the saxophonic scream were held under constant, vigilant contestation during this period, and responses were polarized along the same lines as other appraisals.

Additionally, the very fact that the *body* enters so prominently into the foreground of Coltrane's late music is itself loaded with multivalent gender connotations, matters further vexed by the color of that body. The musicking body—especially displays of "excessive" carnality in performance—has long been connected with femininity and effeminacy.[89] Such considerations played an important role in the way listeners interpreted Coltrane and other performers in concert, where audience members were confronted, often in close proximity, with the sheer, unbridled physical intensity of black, male performers "screaming" through their instruments. As Michael Spence Washington explains, in the 1950s Coltrane played with great bodily restraint, an essential quality to appearing "dignified" (and also masculine) in concert.[90] His free jazz period, however, was marked by complete kinetic abandon. Pictures and film of the saxophonist from this period betray the unbridled physicality of his performances: one can see the contorted face, bulging veins, perspiration, and puffed-out neck consistent with extreme exertion (Fig. 4.6).

---

[88] Of course, Ellis likely did not find Coltrane "pretty" either. The characterization here comes from McClary, *Feminine Endings: Music, Gender, and Sexuality*, 19.

[89] Writes McClary: "The charge that musicians or devotees of music are 'effeminate' goes back as far as recorded documentation about music, and music's association with the body (in dance or for sensuous pleasure) and with subjectivity has led to its being relegated in many historical periods to what was understood as a 'feminine' realm." See ibid., 17.

[90] Washington, "Beautiful Nightmare: Coltrane, Jazz, and American Culture," 284-85.

FIGURE 4.6: John Coltrane in concert, 1965
(photo by Charles Shabacon, from Hentoff, 1996)

Nisenson vividly speaks to Coltrane's bodily performance style in his description of a mid-1960s concert:

> [Coltrane was] playing roiling arpeggios alternating with ribbons of intense lyricism often accentuated by saxophone cries and wails. Playing at this level for even a few minutes would have exhausted most musicians, in terms of both imagination and sheer physical taxation. But Coltrane kept up at this level for what seemed hours… Coltrane was eventually almost completely bent over forward, his face flushed, and at one point saliva poured out of the side of his mouth. He seemed to be not in this world, and I, as well as most of the audience, I am certain, felt that we had long left it far behind, too. At this point I looked at my companion with concern; if this man didn't stop, I thought, surely his heart would give out. Here was a performance where one could no longer objectively judge aesthetics; the feelings it engendered were closer to the awe one felt for a volcano or a mind-boggling religious revelation.[91]

[91] Nisenson, *Ascension: John Coltrane and His Quest*, xvii.

Coltrane's uncontrolled, even slightly grotesque physicality made him appear to be "not in this world," a reaction consistent with the transcendent interpretation of the scream. But it is easy to see how the same experience could lead observers to default to a different set of presumptions: since it makes him relinquish control of his body to the gendered sounds coming out of his horn, Coltrane's screaming has the power to feminize him. Coltrane was aware of this precarious balance between transcendent, masculine power and female hysteria, sometimes to the point of his own embarrassment. Following one physically uncontained performance, for example, Coltrane reportedly fled the venue in tears after his friend Benny Golson witnessed him in such an ecstatic state.[92] Such uncontrolled sounds and bodily movements were considered a threat to masculinity, a risk to which Coltrane felt vulnerable.

The gender connotations of the scream are further complicated when the issue of race is brought into the discussion. Beyond any doubt, it is this interpretive lens that has played the most formative role in the reception of free jazz: Coltrane's was not just a screaming body, but a screaming *black* body. Baraka, for instance, dubs the vocal elements of the free jazz saxophone "quintessentially African," and Frank Kofsky writes of the "growling, raspy" sound quality and "eerie shrieks," concluding that these are the "quintessence of Negro vocal patterns."[93] Nina Eidsheim has convincingly shown that "the black voice" is a fantasy, though a very real (and dangerous) one: there is nothing racially immanent about vocal timbre, only differing pedagogies,

---

[92] Washington, "Beautiful Nightmare: Coltrane, Jazz, and American Culture," 285.

[93] See respectively Anderson, *This is Our Music: Free Jazz, The Sixties, and American Culture*, 110; Frank Kofsky, *Black Nationalism and the Revolution in Music* (New York: Pathfinder, 1970), 134.

perceptions, and ideologies.[94] The "quintessentially black voice" is a cultural construct, and a particularly slippery one at that, not an innate biological or acoustical reality. Nevertheless, the saxophonic scream, in its connection to screaming black voices, carried strong racial associations that show up again and again in critical and sympathetic accounts by both black and white writers of the 1960s and 70s.

This association was, of course, hardly new to the "New Thing." Sociologist Jon Cruz points out that black singing was once heard as "noise" to white listeners in the days before abolition, and whites have long linked unruly vocality to blackness.[95] While white unease over black sound is a topic too big to adequately tackle here, suffice it so say that it is not surprising that many critics interpreted free jazz in explicitly racialized terms, as the sound of Black Nationalism. In 1965, a year of race riots and escalating civil rights tension, the image of a black man screaming (even through an instrument originally designed for white European military bands) was particularly formidable, and many critics who viewed the movement positively still interpreted these sounds through the lens of racial struggle.[96]

This interpretation has entered into conventional wisdom during the intervening decades: it was proffered in Ken Burns widely-viewed documentary *Jazz* (2000), and it has found its way into a number of jazz history textbooks.[97] The connection, however, is tenuous at best. Mark

---

[94] Eidsheim, "Voice as a Technology of Selfhood: Towards an Analysis of Racialized Timbre and Vocal Performance." For more on the fantasy of a radically Other black music, see Ronald Radano, *Lying up a Nation: Race and Black Music* (Chicago: University of Chicago Press, 2003).

[95] See Chapters 2 and 3 in Jon Cruz, *Culture on the Margins: The Black Spiritual and the Rise of American Cultural Interpretation* (Princeton, NJ: Princeton University Press, 1999).

[96] Interestingly, this interpretation also influenced reception outside the American context. In his cultural history of free jazz in France, Eric Drott quotes critics calling it "the concrete cry of the ghetto," "a political instrument," and "a pure product of the black American's anger." See Eric Drott, "Free Jazz and the French Critic," *Journal of the American Musicological Society* 61, no. 3 (2008): 545.

[97] For example, Brain Harker's textbook *Jazz: An American Journey*, has this to say: "As the civil rights movement advanced and white southern reactionaries dug in… the music exploded in a metaphorical cry of impatience and frus-

Gridley has exhaustively documented that the linkage of free jazz and civil rights is largely spurious, grounded in misunderstood cause-and-effect relationships, unrepresentative sampling, and misconceptions of hearing.[98] Coltrane and most of his colleagues—though supportive of mainstream civil rights and open-minded about the Black Power movement—never fully embraced the politics of Black Nationalism, and even expressed frustration at the phalanx of journalists who uncritically circulated this connection.[99] The major exception here was saxophonist Archie Shepp, but his was the minority position. Coltrane explained his creative goals in unmistakably spiritual terms: "what music is to me—it's just another way of saying this is a big, beautiful universe we live in, that's been given to us, and here's an example of just how magnificent and encompassing it is;"[100] he also commented, "My music is the spiritual expression of what I am, my faith, my knowledge, my being."[101] In a similarly apolitical vein, Marion Brown indicated, "When I play my music I'm not playing anything else at all. I'm not putting down anything that you could express in words. I don't play about religion, or the Universe, or Love, or Hate, or Soul."[102]

It could be argued that musicians were eliding their radical affiliations in the white press, but it seems clear that the Black Nationalist interpretation of the saxophonic scream (and free jazz in general) was projected onto this potent timbral sign with very little input from the musi-

---

tration, producing yet another species in the evolution of jazz styles: free jazz." See Brian Harker, *Jazz: An American Journey* (Upper Saddle River, NJ: Pearson, 2005).

[98] Mark C. Gridley, "Misconceptions in Linking Free Jazz with the Civil Rights Movement," *College Music Symposium* 47(2007). This article is a good source for primary and secondary commentary on this topic, and most of the quotations from this paragraph are found here as well.

[99] Coltrane dedicated a song ("Alabama") to Martin Luther King, Jr., and performed benefit concerts in support of the movement, for instance. He also adopted certain African topics in his music (e.g., the album *Africa/Brass*), and performed with Nigerian percussionist Babatunde Olatunji. Thank you to David Ake for pointing these out.

[100] Don DeMichael, "John Coltrane and Eric Dolphy Answer Their Critics," *Down Beat*, April 12, 1962.

[101] Quoted in Borgo, "Between Worlds: The Embodied and Ecstatic Sounds of Jazz."

[102] Quoted in Gridley, "Misconceptions in Linking Free Jazz with the Civil Rights Movement."

cians creating it. As Coltrane's long-time producer Bob Thiele put it: "…for the literary fraternity, the music of Coltrane and others… really represented black militancy. Most of the musicians, including Coltrane, really weren't thinking the way their militant brothers were. I mean, LeRoi Jones [Amiri Baraka] could feel the music was militant, but Coltrane didn't feel that it was."[103]

So where did this particular widely circulated appraisal come from, if not from the musicians themselves? Most influential here was Amiri Baraka, who repeatedly drew the link between free jazz and Black Nationalism, serving as a reliable polemicist for the avant-garde's radical racial interpretation throughout the mid- to late 1960s. In "The Screamers," to recall, the sound of the saxophone was "hatred and frustration," and the protagonist's "horn spat enraged sociologies."[104] Baraka heard in the effectual saxophonics of free jazz a fitting corollary to the civil rights struggles transpiring in American streets: the sound was "too ugly to be described," but only because ugliness, hatred, and frustration was the truth of the African-American experience.[105] To Baraka, the screaming saxophones on *Ascension* were the sonic equivalent of a race riot, and John Coltrane was "Malcolm X in New Super Bop Fire."[106]

Baraka was not alone in making this connection. Writer Henry Dumas (1934–1968), in his mid-1960s short fiction "Will the Circle Be Unbroken?," tells the story of a saxophonist (or "afro-hornist") whose "lethal vibrations" have the power to kill unsuspecting white listeners.[107] This story takes "The Screamers" one step further: going beyond sonic icons for rage, Dumas

---

[103] Quoted in Ted Fox, *In the Groove: The People Behind the Music* (New York: St. Martin's Press, 1986), 196.

[104] Baraka, *The LeRoi Jones/Amiri Baraka Reader*, 175.

[105] Gridley points out that the logical error of mistaking correlation for causation is rampant in the free jazz/black militancy connection. Some of Baraka's writings are a fine example of this.

[106] Baraka, *The LeRoi Jones/Amiri Baraka Reader*, 271. For more on Baraka's role in the reception history of Coltrane, see Washington, "Beautiful Nightmare: Coltrane, Jazz, and American Culture," 298.

[107] Henry Dumas, *Echo Tree: The Collected Short Fiction of Henry Dumas* (New York: Coffee House Press, 2003), 108.

weaponizes the timbre of the saxophone against white bodies incapable of absorbing its "truer vibrations."[108] In line with the empirical findings reported previously, Dumas intuited the somatic vehemence of noisy timbre, literalizing it along racial lines. In addition to certain black authors and journalists, moreover, some musicians—particularly the outspoken Archie Shepp—did express an affinity with the ideals of various Black Nationalist organizations. A number of interviews appeared in which Shepp connected the sounds of free jazz to the more militant factions of the Civil Rights Movement, including one in which he commented: "Some of [Coltrane's] solos have exactly the rage that was being expressed in the streets, by the Muslims and the Panthers, and many people thought Trane's music was very angry."[109] In sum, it was not white anxiety (or titillation) alone that linked the sound of the saxophonic scream to Black Nationalism.

But to be sure, voices among the white literati did their part to connect free jazz to political/racial freedom movements. Most influential among this cohort to take an "ethno-sympathetic" position was the Marxist historian Frank Kofsky (1935–1997), who believed free jazz represented "the ghetto's vote of 'no confidence' in Western civilization and the American Dream."[110] Kofsky is perhaps best known today for an extended 1966 interview with Coltrane that has proved a singular document in the literature.[111] In this oft-cited interview, the historian pressed Coltrane to own up to the racial and political valence of his music with a series of leading

---

[108] Ibid., 110.

[109] Quoted in Ashley Kahn, *The House that Trane Built: The Story of Impulse Records* (New York: Norton, 2006), 131.

[110] Kofsky, *Black Nationalism and the Revolution in Music*, 134. The concept of "ethno-sympathy" comes from Jon Cruz.

[111] A full transcript of the interview can be found in *John Coltrane and the Jazz Revolution of the 1960s* (New York: Pathfinder, 1998), 425.

questions designed, in Ake's assessment, to elicit an "angry black man" response from the gentle saxophonist.[112] To briefly sample from their dialogue:

> *Kofsky*: Some musicians have said that there's a relationship between some of Malcolm [X]'s ideas and the music… Do you think there's anything in that? […]
>
> *Coltrane*: I feel that it expresses the whole thing… The whole of the human experience at the particular time that it is being expressed.
>
> *Kofsky*: What do you think of the phrase "the new black music"?
>
> *Coltrane*: Well… I don't know. Phrases… They don't mean much to me. […]
>
> *Kofsky*: The people who use that phrase argue that jazz's particularly related to the black unity … that's why I asked you about your reaction to Malcolm.
>
> *Coltrane*: Well I think it's up to the individual where you can… call it what you may, for any reason you may.
>
> *Kofsky*: But it does seem to be a fact that most of the, changes in the music, the innovations have come from black musicians.
>
> *Coltrane*: Yeah well that is … how it is.
>
> *Kofsky*: […] Have you ever seen that the racial composition of the audience seems to determine how the people respond?
>
> *Coltrane*: Well, sometimes "yes" and sometimes "no."
>
> *Kofsky*: Any examples?
>
> *Coltrane*: Well… it's hard to say, man, you know sometimes people like it or don't like it no matter what color they are.
>
> *Kofsky*: You don't have any preferences yourself about what kind of an audience you play for?
>
> *Coltrane*: Well to me… it doesn't matter.

As clearly evidenced in this cringe-worthy exchange, Kofsky was trying to extract a political confession from the besieged saxophonist, despite the overwhelming prevalence of "cosmic mysticism" tropes in Coltrane's album and song titles, imagery, and published commentary.[113] To

---

[112] Ake, *Jazz Cultures*, 140. To be clear, Coltrane's reputation as an "angry tenor" dates to the late 1950s, though his timbre was not the focal point of this earlier perception.

[113] Kofsky himself uses this term: see Kofsky, *Black Nationalism and the Revolution in Music*, 187.

writers like Kofsky, the sonic qualities of free jazz—particularly its timbral attributes—were simply too black sounding *not* to be related to racial politics. As for Baraka and Dumas, Kofsky could not hear the highly aroused sounds of Coltrane's saxophone without hearing them as an angry black man screaming.

An avowed leftist, Kofsky considered himself an ally of the Black Power movement; the animus he heard in the music of Coltrane was not something to be feared as much as celebrated. Less sympathetic perspectives, however, predominated—fearful condemnation was a relatively common stance among white critics and listeners. As might be expected, the most egregious examples of white anxiety in the Coltrane critical literature come from the Tory Philip Larkin. Repudiating the new direction of jazz in the mid-1960s was not just a point of intellectual disagreement for Larkin, and the perceived racial overtones of the "New Thing" contributed significantly to the intransigence of his position. Lamenting the radicalization of jazz in 1968, he explains: "I soon saw how quickly jazz was passing from mystification… to outrage. […] From using music to entertain white people, the Negro had moved to hating him with it."[114] In his essay commemorating Coltrane's death, he writes: "[Coltrane] did not want to entertain his audiences: he wanted to lecture them, even to annoy them. His ten-minute solos, in which he lashes himself up to dervish-like heights of hysteria, are the musical equivalent of Mr. Stokely Carmichael."[115] Again, despite Coltrane's well documented resistance to being associated with political or racial causes—a resistance of which Larkin, an influential and well-informed jazz critic, would likely have been aware—his playing was heard in terms of hysteria and black rage. He also repeatedly mocked Shepp for his political sympathies, writing: "Anyone who thinks that an Archie ('America's done

[114] Larkin, *All What Jazz*, 24.

[115] Ibid., 187.

me a lot of wrong') Shepp recording is anything but two fingers extended from a bunched fist at him personally cannot have much appreciation for what he is hearing."[116]

Intriguingly, these comments and others reveal an underlying point of convergence between celebratory and anxious accounts. To Baraka and Kofsky, Coltrane was the jazz world's own Malcolm X, and to Larkin he was a jazz Stokely Carmichael: that his sounds reflected black anguish was not under dispute. Baraka heard a positively empowered black man calling out at white injustice, and Dumas a sonic whitey-slayer—while Larkin, not surprisingly, flinched away from "death-to-all-white-men wails."[117] It may appear that musical discourse is always far removed from "thin" fight-or-flight evolutionary triggers, but here is a potent counterexample: through this timbral effect, free jazz saxophonists tapped into the somatic markings of threat, pushing their listeners to embody its sense of (racialized/spiritualized) menace. The screaming saxophone said in "pure sound" what Moore and Dumas had said in words—*you must die*.

Conclusions

In this chapter I have gone into depth exploring the reception of a specific timbral gesture, the saxophonic scream, in mid-1960s free jazz. As we have seen, "the most powerful human sound ever created" functioned as a Rorschach test to listeners in this turbulent decade, and arguably still does today.[118] On both sides of the highly polarized divide, extended timbral-corporeal engagements on the saxophone seem to push "beyond music," conveying emotionally intense forms of embodied knowledge with an unmediated swiftness that catch listeners off guard. I have sug-

---

[116] Ibid., 24.

[117] Ibid., 179. There is more than a hint of dry humor in Larkin's writing; it is difficult, therefore, to take him entirely at his word. His sincerity in profoundly disliking free jazz, however, is impossible to dispute.

[118] I owe this Rorschach metaphor to Fales: see Fales, "The Paradox of Timbre," 78.

gested in this chapter that there are certain psychological explanations for this initial reaction: the neural demands of noisy saxophone timbre, specifically in mirror neuron and limbic areas, are substantial indeed, owing in no small measure to our innate capacity to motor-mimetically empathize with the sounds of others. The saxophonic scream in the hands of Coltrane and his "sound-player" cohort confronts the threshold between music and its Other by manipulating high-arousal, high-intensity, negative-affect (even physical threat) reactions.

Yet the deeply passionate battle over timbral interpretation evidenced in the reception history obviously betokens much more than a mere reflex; the phenomenological and social complexity of its "beyond-ness" is a matter no brain scan could possibly reveal. The reaction, then, was shared, but the appraisal was contingent upon how the individual listener (or community of listeners) chose to hear this ugly, noisy, scraping and scouring, gendered, angry, ecstatic timbre. "Escaping from notes to sounds" was thus also, in a crucial sense, escaping from the normative bindings of musical signification to… *what exactly?* In the absence of conventional interpretive clues, listeners had to decide: what do these noisy sounds mean?

For they have to mean *something*. The somatic demands of the saxophonic scream are too great for the gesture to be entirely without purpose—we feel the strong affective pull of the sound, we know on a primal level what it signals. Yet despite its unambiguous bodily meanings, by placing noisy timbres at the center of the listener's phenomenal field, Coltrane conspicuously withheld the rational, codified, signifying aspects of musical sound, the "musical sense of music" (to again quote Merleau-Ponty). We have bracketed timbre from the other elements of music for the sake of analysis, but at this point we must dissolve the chapter's framing *epoché*: although timbre, I have argued, does much of the affective work in free jazz, the "musical sense of music" was disturbed by the absence of containing tonal forms and rhythmic periodicity as well (not to mention the extreme volume of its presentation). Not only was the timbral content noisy; the frame was cattywumpus. Taken in aggregate, this creates a certain cognitive dissonance: we understand it to be meaningful on a bodily level, but not when appraised according to the rubric of "music" (or as Ellis would have it, "ART"). Coltrane's playing was "gobbledegook" and "musical nonsense" to Tynan, an "exercise in gigantic absurdity" to Larkin.[119]

The pregnant space between meaning and meaninglessness, knowing and unknowing, is a symbolically fraught one, to say the least. As David Lidov has noted, aporias of musical meaning can create a *tabula rasa* to be readily inscribed with the strongest sorts of appraisals: "when clues to interpretation are withheld, monotony [noise!], if not rejected, appears transcendental and spiritual."[120] And also perhaps, in its very ambiguity, *terrifying* (to echo Rilke). To be sure, the terror evinced in both positive and negative accounts is both somatically- and symbolically marked. Fales accounts for timbre's semiotic "edgework" in the baldest of terms: "many [timbre-based] formulations of an essentially epistemological duality end in metaphysics: the ultimate unknown is God."[121] What is more terrifying—and, possibly, Beautiful—than that?

In the end, it is precisely this amorphousness of meaning that I want to emphasize. As the multifarious reception of Coltrane's screaming saxophone illustrates, timbre is uniquely malleable as a sign—especially in contexts where the "noise!" impression is reinforced by other musical and social factors—and this very malleability makes it susceptible to a range of fiercely contested symbolic projections. This is especially the case with acoustically noisy timbre, which seems to doubly withhold clues to (musical) interpretation. In a way that often evades conscious awareness, qualities of noisy sound are painted with the desires, ideals, and fears of the listening individual

---

119 Tynan, "Take Five."; Larkin, *All What Jazz*, 21.

120 David Lidov, *Is Language a Music? Writings on Musical Form and Signification* (Bloomington, IN: Indiana University Press, 2005), 157. Brackets in original.

121 Fales, "The Paradox of Timbre," 90. The term "edgework" comes from Chris Jenks.

and group. Fales trenchantly comments to this point: “timbre is free to operate with little direct scrutiny by a listener, creating effects that are intense but also hazy in definition, difficult to articulate, and freely attributable to other features of the musical context,” adding, “listeners make sense of their experience in whatever expressive currency is around.”[122] In the case of mid-1960s free jazz, the main expressive currency proved to be ecstatic spirituality and racial/gender politics. But the semiotics of timbre is slippery indeed, and the expressive currency is constantly shifting according to context.

In the next chapter, we shift to another case study of embodied timbre cognition in action, one featuring a different panoply of “noisy” sounds, though—*mutatis mutandis*—similar manners of embodiment. As we will see in our reading of extreme heavy metal, the embodied perceptual and acoustic boundary separating music and noise can serve, when transgressed, not only as a sonic icon of spiritual ecstasy—it can also work as a simulacrum of sacred violence.

[122] Ibid., 91 & 78.

Chapter 5

The Sound of Evil: Timbre, Body, and Sacred Violence in Death Metal

> It's guitar tuned down so low that only dogs can hear it. It's songs about the Devil, revenge from the grave, death by garden tools and other tales from the dark side delivered, for the most part, in a Linda Blair/Exorcist-like Satanic growl.
>
> — Marc Shapiro[1]

> Heavy metal rock is nothing more than a bunch of noise; it is not music, it's distortion—and that is precisely why its adherents find it appealing.
>
> — Lester Bangs[2]

Introduction

Death metal music is easy to despise. To many, this gore-obsessed genre—the pathologically aggressive outcast of the popular music family—is simply a "fundamentalist" caricature of everything hyper-masculine, risibly puerile, and utterly ridiculous about heavy metal.[3] Since the inception of the style as an offshoot of thrash metal in the mid-eighties, death metal has provoked a range of highly emotional reactions, from fear to disgust, rage, and ridicule.[4] To many rock crit-

---

[1] Quoted in Ronald Bogue, "Violence in Three Shades of Metal: Death, Doom and Black," in *Deleuze and Music*, ed. Ian Buchanan and Marcel Swiboda (Edinburgh: Edinburgh University Press, 2004), 106-07.

[2] Quoted in Weinstein, *Heavy Metal: A Cultural Sociology*, 2.

[3] The "fundamentalist" designation comes from ibid., 48.

[4] As we recall from Experiment 2 (Ch. 2), brief clips of heavy metal were far and away the *least liked* genre, with an average "affective valence" score of just 16 (on a scale of 0 to 100 from "strongly like" to "strongly dislike"). The polarization of its perception appears to have a long history: heavy metal has always been a controversial genre, even in

ics, it is normal metal pushed overboard by a juvenile fixation on being "extreme" to the detriment of good taste and listenability; it is silly schlock calculated to shock.

Yet the fact that most people perceive it to be—in Bangs's assessment of '70s heavy metal—"a bunch of noise" is exactly what empowers fans, transforming the derision of the mainstream into the passion of a subculture. The noisiness and extremity that so alienate and horrify critics *are* precisely what fans find attractive and powerful about the music. But how do we locate this elusive "extreme" that so perplexes and repels mainstream audiences? Beyond the obvious offensiveness of lyrics and imagery, why is death metal music so repugnant to so many people? To answer this question, I propose in this chapter that we turn to the timbres that define the style, and to their bodily implications. As *Terrorizer* magazine editor Nick Terry points out, "it's the musical ferocity that marks out death metal […], not the lyrics, attitude or imagery. The extremism of death metal […] comes down to inhuman vocals, grinding guitars and vicious blast beats, not to any political or religious ideology."[5] Death metal has primarily been studied as a sociological phenomenon rather than as a coherent musical system, and this has led to a general critical deafness to the issue of sound.[6] In this chapter, I attempt to fill this critical lacuna. As

---

the era of its greatest commercial success. Robert Walser cites a 1989 survey showing that while 10% of the sample liked the style, a full 19% had actual antipathy toward it. Presumably, opinions toward today's extreme metal subgenres of death, grindcore, black, doom, etc., are more polarizing than this earlier "heavy metal" generic identification, a term that, significantly, applied to artists like Van Halen, Judas Priest, and even Bon Jovi. Although certain observations from Walser and other scholars of heavy metal apply in our case, I take care not to conflate death metal with its broader parent style, for the two differ markedly. For more, see Walser, *Running with the Devil: Power, Gender, and Madness in Heavy Metal Music*, xi. For sociological and cultural-theoretical perspectives on the polarizing tendencies of the genre, see Bryson, "'Anything But Heavy Metal': Symbolic Exclusion and Musical Dislikes."; Michelle Phillipov, *Death Metal and Music Criticism: Analysis at the Limits* (Plymouth, UK: Lexington, 2012).

[5] Quoted in Albert Mudrian, *Choosing Death: The Improbable History of Death Metal and Grindcore* (Los Angeles: Feral House, 2004), 23.

[6] The major books on the genre include works by the sociologists Natalie Purcell, *Death Metal Music: The Passion and Politics of a Subculture* (Jefferson, NC: McFarland, 2003); Keith Kahn-Harris, *Extreme Metal: Music and Culture on the Edge* (Oxford and New York: Berg, 2007); Weinstein, *Heavy Metal: A Cultural Sociology*. Recent work by cultural theorist Michelle Phillipov explores the (anti-)politics of death metal, the trope of "extreme," and its challenge to musical critique: see Phillipov, *Death Metal and Music Criticism: Analysis at the Limits*. All of these publications are rich in social/cultural insight but provide little musicological detail.

Terry suggests, central to the genre's identity—and its power to both fascinate and repel—are its timbral qualities, specifically the "extreme" sounds coaxed from electric guitars and the human vocal tract. More than any other single parameter, noisy timbre acts in complex and sometimes surprising ways to articulate a range of meanings in this often-mocked genre.

Surveying the ethnographic literature on death metal, both scholarly and vernacular, reveals a specialized emic vocabulary of timbral metaphors that serve to define the style. In addition to the SOUND IS MATERIAL metaphor "heavy," which features in the conceptualization of all subgenres under the heavy metal umbrella, sociologist Natalie Purcell stresses the importance of "brutal" sounds in the death metal scene.[7] Additionally, scene members value "evil" timbres: documentary filmmaker Sam Dunn, for example, defines the entire genre as "the sound of evil."[8] But what exactly constitutes this heavy, brutal, and evil sound? I suggest that the "brutal" sound in death metal is largely attributable to its high levels of acoustic noise, "bad sounds" associated with a variety of negative appraisals (including chaos, suffering, and death). Crucially, the production of noise in the context of "brutal" and "evil" death metal timbres transcends aesthetic representation to strike at a drastic, experiential awareness of the material limits it embodies: distorted, overloaded guitar and vocal sounds are the audible manifestations of imminent bodily or mechanical breakdown (the *ultimate limit* schema). Extremity of timbre is thus the sounding presence of an embodied liminal state. Drawing on the theoretical work of Jacques Attali, René

[7] Purcell, *Death Metal Music: The Passion and Politics of a Subculture*, 16. See also Kahn-Harris, *Extreme Metal: Music and Culture on the Edge*, 53. The value of "brutality" in death metal is too widespread to fully enumerate here. It is the ultimate marker of timbral power and identity in the genre; there is even a German website ("Brutal Sounds TV") dedicated to the tone.

[8] Sam Dunn, "Metal: A Headbanger's Journey." Warner Home Videos, 2005. Although the terminology of "brutality" and "evilness" are remarkably pervasive in the scene, even most in-group explanations of what constitutes "that evil sound" focus on elements of melody and modality (tritones and the prevalence of the Locrian and Phrygian modes) rather than *sound per se.*

Girard, and others, I claim that this physically punishing liminality serves an essential, sacrificial function in the highly ritualized space of the music.[9]

Death metal timbres do not just represent (and celebrate) evil; rather, they embody evil and chaos in order to demonstrate mastery over them. As sociologist Keith Kahn-Harris has observed, the dialectic between chaos and control is central to the genre, with timbre playing an essential role.[10] Death metal musicians create a timbral environment of extreme violence only to bring it under meticulous order. Although these sounds relate to listeners first through their drastic immediacy, manipulating evolutionary mechanisms in the process, they also serve an entirely symbolical function in the ritual of simulated sacred violence, playing the role of sacrificial victim: ugly sounds, chaotic and abject, are summoned for the sake of being harnessed and mastered. This symbolic display is performed through an interconnected web of timbral "extremes," from warped and sludgy guitar sounds to the animalistic growl of the vocalist. In order for the timbral threat to serve a cathartically satisfying function in the sacrificial ritual of the music, moreover, its sense of danger and horror must be absolute. "The sound of evil," which is often taken literally as the proud signifier of a morbid and nihilistic subculture, is actually a deliberate manifestation of disordered noise for the purpose of its ultimate subsumption into order.

The chapter unfolds in two acts: first, I assess what these timbres are, then I move on to analyze what they do in their native setting. Our inquiry into the timbral landscape of death metal begins at the broad level of record production before venturing into an analysis of the

---

[9] Deathmetal.org, for instance, describes the genre as "a ritual with a sense of overwhelming evil power, regality and clarity."

[10] Walser identifies the central poles of heavy metal as freedom/control, but this schema is not apropos to the central dynamic of death metal. Weinstein is the first to articulate the theme of chaos as a fundamental trope, and Kahn-Harris puts this concept into dialog with its opposite, control, as a defining characteristic of extreme metal. For more, see Walser, *Running with the Devil: Power, Gender, and Madness in Heavy Metal Music*, 54; Weinstein, *Heavy Metal: A Cultural Sociology*, 48; Kahn-Harris, *Extreme Metal: Music and Culture on the Edge*, Ch. 2.

"brutal" guitar tone (specifically the components of volume, distortion, EQ, and detuning). After finishing with the guitar, I assess the vocal technique of the "death growl," a timbre that arguably demarcates the genre more than any other single sonic quality. With the "evil" timbral attributes of the genre enumerated and dissected, I will finish with an analysis of how these sounds function as a *sign*, teleologically, psychologically, and ritually.

Florida Sound / Swedish Sound

In the early years of genre formation (c. 1986–1993), two distinctive timbral palettes developed simultaneously in two of the biggest death metal scenes, Tampa, Florida and Stockholm, Sweden.[11] Tampa was home to some of the pioneering acts of the late eighties, including Death, Obituary, Morbid Angel, and Deicide, and while opportunities for live performance in this medium-sized city were relatively scant, the scene cohered around mutual respect and a friendly recording studio that came to define Florida death metal, Morrisound. So ubiquitous and influential would the productions of this one studio become that by the early nineties, bands from around the world were traveling to Florida to record at the death metal capital, and up-and-coming American bands like Buffalo's Cannibal Corpse permanently relocated to the Tampa area.

At the center of the "Florida sound" was the producer Scott Burns.[12] Burns brought a high level of sonic clarity to the fledgling movement, with a crisp, clean production style hitherto

---

[11] In addition, New York death metal (exemplified by bands like Suffocation and Dying Fetus) has developed a parallel third stylistic stream characterized by heavier bass, slam-oriented rhythms, and a more mosh-friendly sensibility. For more, see Purcell, *Death Metal Music: The Passion and Politics of a Subculture*, 19.

[12] Burns served as producer on albums by Cannibal Corpse, Death, Deicide, Obituary, Sepultura, and Suffocation, among many others.

rare in death metal recordings.[13] He also employed a significant amount of digital compression, which went a long way toward evening out and restraining the sonic onslaught of his blaringly loud, blazingly fast artists. Studio compression did not detract from the brutality quotient of Burns's productions, however, since a hallmark of his style can be observed in the "forwardness" of his spatial approach to sound.[14] Sound on a Burns recording is aggressively present and thick, a fitting corollary to the timbral brutality, volume, and speed of the bands he documented. Burns recalls the difficulties of keeping such a forward approach to the "sound box" consistent with his other goal, clarity: "Perhaps recording fast heavy bands is more technically challenging because every member wants everything up front and audible."[15] Successfully negotiating between contradictory impulses—keeping the brutal, "up front and audible" quality that courts the overload essential to the genre, while also ensuring a transparent, clean, and "produced" sound, which keeps the symbolic violence under sonic control—Burns's production style was essential to the popularity of Florida death metal in the late eighties and early nineties.

It also came to be an albatross around his neck. When the scene overheated in the mid-nineties and a glut of bands rushed to record at Morrisound, the perceived sameness of the "Florida sound" drew criticism. Burns responded: "I'm not a miracle worker. I probably did eighty death metal records, but you've got drums, bass, guitar, they're all playing double bass, they're all detuned and they're all barking, so how many ways can you make it sound different?"[16] As we shall see, balancing the dual tendencies of chaos and control can be a precarious undertaking for musicians, audiences, and producers alike. As Kahn-Harris notes, even under representative, ex-

---

[13] Kahn-Harris stresses the "produced" quality of Burns's recordings: see Kahn-Harris, *Extreme Metal: Music and Culture on the Edge*, 103.

[14] The term comes from Allan Moore, *Rock: The Primary Text* (Surrey: Ashgate, 2001).

[15] Interview with Scott Burns, "Voices from the Dark Side," <http://www.voicesfromthedarkside.de/ interviews/scottburns.htm>. Accessed March 8, 2011.

[16] Quoted in Mudrian, *Choosing Death: The Improbable History of Death Metal and Grindcore*, 206.

emplary conditions, extreme genres like death metal continually "[teeter] on the edge of formless noise."[17] If a producer or a band pushes the distortion level, tempo, or volume an iota too far, it can easily devolve into amorphous cacophony. On the other hand, if too much technical control is exerted over the music, if it sounds *too* produced, then it ceases to be generically identifiable—it's not death metal. The producer in this genre is thus constrained by the exacting poles of chaos and control, and there is not a lot of wiggle room for "sounding different" in the narrow space between the two.

Scott Burns's model was not the only one, however, despite the inherent limitations of the genre. Coeval with Morrisound's ascent, Tomas Skogsberg, an engineer at Stockholm's Sunlight Studios, managed to nudge the sound a little bit closer to chaos while avoiding disintegration into random noise.[18] The Swedish production style was much rougher than the Florida counterpart, with a DIY ethos that reflects the producer's punk background.[19] Skogsberg's recordings were also heavy in the mid-range, a fact that sets the Swedish model apart from the conventional wisdom (to be discussed) regarding heavy metal EQ preferences. Skogsberg remarked: "Some of the guys call me the king of midrange. I love midrange, that's my frequency."[20] Aside from these production elements, however, perhaps the most identifiable single attribute of the "Swedish sound" is the much-vaunted brutality of the guitar sound.

Working with the bands Entombed and Dismember, Skogsberg's productions featured a level of guitar distortion not heard in U.S. recordings, a grisliness exacerbated by radically detuned guitars. This eminently dark and brutal sound is exemplified in Entombed's full-length

---

[17] Kahn-Harris, *Extreme Metal: Music and Culture on the Edge*, 5.

[18] For accounts of the "Swedish sound," see ibid., 32 & 106.; and Mudrian, *Choosing Death: The Improbable History of Death Metal and Grindcore*, 98-99.

[19] *Choosing Death: The Improbable History of Death Metal and Grindcore*.

[20] Ibid.

debut, *Left Hand Path* (1990), a seminal record that features a "chainsaw" guitar tone highly influential in the death metal scene (as we will explore later).[21] Bandleader Nicke Andersson explains: "It was our guitarist, Leffe, who actually came up with that sound from the Boss Heavy Metal pedal. You have the midrange on full. I think you have everything on full. He bought the pedal and just cranked it."[22] This setup is confirmed by Skogsberg: "we used the distortion pedal DS-1, the orange one. I think I used it for everything in the beginning; for the vocal, for the hi-hat—everything."[23] Using a telling NOISE IS FRICTION metaphor—guitar distortion as motorized saw blade—listeners made sense of this guitar timbre in terms of the physical "brutality" it evoked. One record reviewer described *Left Hand Path* (positively) as "so disgustingly rough and raw it sounds like death metal gone through a disastrous abortion."[24]

It is a matter of no small irony that one of the most important guitar sounds in the genre was the creation of a Swedish teenager who simply cranked the levels on his cheap digital distortion pedal up to the proverbial eleven. For a genre obsessed with meticulously crafted (and often expensive) guitar brutality, this haphazard approach is all too reminiscent of death metal's scrappy predecessor, punk rock. In the booklet to *Left Hand Path*, the band explains the willy-nilly creation of this sound: "We have Peavey amps, we have a distortion pedal and a Heavy Metal pedal and everything on 10… it's not hard really. The amp is actually very little, only 40 watts!"[25] Significant here—besides the maximalist knob-twiddling approach—is the relative pu-

---

[21] Using chainsaw metaphors to describe this guitar sound is now a death metal cliché. For instance, Purcell describes it as "the aural equivalent of a chainsaw through flesh"; she later goes on to mention the "saw-like" guitar sound: see Purcell, *Death Metal Music: The Passion and Politics of a Subculture*, 22 & 58.

[22] Mudrian, *Choosing Death: The Improbable History of Death Metal and Grindcore*, 99.

[23] Ibid.

[24] Reviewer on <metalarchives.com>. Accessed February 17, 2011.

[25] The pedals referenced here are the classic, orange BOSS DS-1 Distortion and the BOSS HM-2 Heavy Metal (now discontinued).

niness of the transistor amplifier. When flooded with massive amounts of gain from the two distortion pedals, the amp was pushed to its limits; there are numerous moments on *Left Hand Path* where the amplifier simply gives up, chopping (or better, chainsawing) off large amounts of the distorted signal flowing through its transistorized signal path.[26] There is a direct exploration of material boundaries contained within Entombed's influential "chainsaw" tone: mechanically (an amp driven over the edge of its reproducing capabilities), acoustically (a level of inharmonic distortion that veers dangerously close to white noise), and materially (strings detuned to a loose and floppy fourth below standard tuning). Players in recent years have mimicked this quintessential tone by using large, powerful amps with cascading pre-amp gain, as well as custom-designed seven- (or eight-) string guitars for the gravelly low notes, but—as is common in the history of the electric guitar—the sound was originally born from a spirit of boundary-testing, pragmatic experimentation, and poverty (it is fair to assume that cheap amps and effects were not a purely aesthetic choice).[27]

Production plays a major role in delineating the sonic characteristics of death metal, sometimes even defining what counts and what doesn't. (Picture, for instance, a death metal band recorded by Phil Spector.) However, as popular music genres go, extreme metal is decidedly *not* a producer's craft, due to the aesthetic and practical constraints discussed above. Production is often downplayed in styles with a live concert emphasis, a feature that, as we shall see, applies here. As evinced in the "Swedish sound," it is a single element of production—the achievement of distinctive timbres, specifically of the guitar and the vocals—that sonically defines death metal. We

---

[26] For a good example of this effect, listen to "Drowned," especially 0:15. Guitar aficionados carefully separate the harsh "unmusical" sound of solid-state clipping from the smooth onset of harmonic distortion when valve amplifiers are simply overdriven.

[27] For a comprehensive and theoretically sophisticated history of the electric guitar, see Steve Waksman, *Instruments of Desire: The Electric Guitar and the Shaping of Musical Experience* (Cambridge, MA: Harvard University Press, 1999).

now turn to this element, beginning with a closer assessment of the components that make up the "brutal" guitar sound.

Volume, Distortion, EQ, and (De-)Tuning

It is an article of faith in the extreme heavy metal world that the music, both in concert and on record, *must* be experienced at very high volume levels. More than any other style of music, metal is preeminently, unapologetically *loud*: so closely tied is its "loudestness" to the identity of the genre that listening at a low, safe level would be completely missing the point.[28] The sounds of death metal, moreover, transmit a feeling of excessive loudness irrespective of the actual decibel level. As Walser and Theodore Gracyk point out, overdrive—which has traditionally been achieved through high volume levels—is needed to saturate the signal chain in order to manifest the complex, distorted timbre central to the music.[29] Overdrive alters the overtone spectrum through harmonic distortion, and the guitar sound, even when reproduced at low volume levels, is thus shot through with the timbral traces of loudness. In a very real sense, the brutal sound of the death metal guitar is impossible without the right level of volume, whether real or simulated (in the form of cascading gain stages).

Physical volume also serves an imperative phenomenological function in the music, overwhelming listeners, drawing them into the complex nexus of the sound, and explicitly appealing to them at the site of the body. Like trance or dubstep, death metal is at its best when it is *felt*, not just heard. As Deleuze theorizes, music has the power to "deeply traverse our bodies, and [put]

---

[28] This term comes from Robert Duncan: see Walser, *Running with the Devil: Power, Gender, and Madness in Heavy Metal Music*, 45.

[29] Ibid.; Theodore Gracyk, *Rhythm and Noise: An Aesthetics of Rock* (Raleigh-Durham: Duke University Press, 1996), 110 & 12.

an ear in our belly, in our lungs"; nowhere is this more true than in extreme metal genres, where loudness at the threshold of pain is required to break down the barriers separating inside (Self) and outside (Other).[30] There is a psychophysiological basis for this: music above approximately 90 dB, especially in the "heavy" frequency range of 100–300 Hz, evokes both acoustic and vibrotactile sensation in the vestibular system, an essential control center for our sense of balance, movement, and physical orientation.[31] Beyond mere reflexive response, this potent reaction is weighted with symbolic meanings: it is a forceful articulation of power. Deena Weinstein explains: "The essential sonic element in heavy metal is power, expressed as sheer volume. Loudness is meant to overwhelm, to sweep the listener into the sound, and then to lend the listener the sense of power that the sound provides."[32] As I shall demonstrate later, loudness as an expression of power and a vital link to the body (via schemas of *force*) are fundamental to the ritual potential of the music that fans find so captivating. Combined with other manifestations of embodied power, like distortion, volume is crucial in the transmission of death metal's deepest meanings.

Besides loudness, which is more a matter of (psycho)acoustics than music *per se*, no single signifier is as closely associated with death metal and heavy metal more generally than an extremely distorted electric guitar. Along with the indecipherable, bestial growl of death vocals, distortion—and massive amounts of it—transmits genre identity like nothing else in the music, materializing the noise, volume, and violence so essential to the style.[33] Distortion is essentially a

---

[30] Quoted in Bogue, "Violence in Three Shades of Metal: Death, Doom and Black," 114. For more, see Walser, *Running with the Devil: Power, Gender, and Madness in Heavy Metal Music*, 45; Weinstein, *Heavy Metal: A Cultural Sociology*, 25 & 214; Kieran James, "From 'The Undead Will Feast' to 'The Time to Kill is Now': Frankfurt School and Freudian Perspectives on Death-Metal," *Musicology Australia* 31(2009): 33.

[31] Neil P. McAngus Todd and Frederick W. Cody, "Vestibular Responses to Loud Dance Music: A Physiological Basis of the 'Rock and Roll Threshold'?," *Journal of the Acoustical Society of America* 107, no. 1 (2000).

[32] Weinstein, *Heavy Metal: A Cultural Sociology*, 23.

[33] Although a diverse array of gear is used by death metal guitarists to get their characteristic distortion, the Swedish sound is generally associated with the BOSS Heavy Metal HM-2 pedal while the Florida sound favors the BOSS

form of clipping that results when an input signal exceeds the voltage threshold of the amp. Figure 5.1 demonstrates how this works: the first diagram (a) shows a clean, unclipped signal; in the second (b), the energy of the signal exceeds the threshold, resulting in a distorted, "hard-clipped" waveform resembling a square-wave (c). In a literal, electro-acoustic sense, distortion is the sonification of uncontained power: the incoming guitar signal is too great for the amp to replicate without the introduction of nonlinear, noisy timbral components.

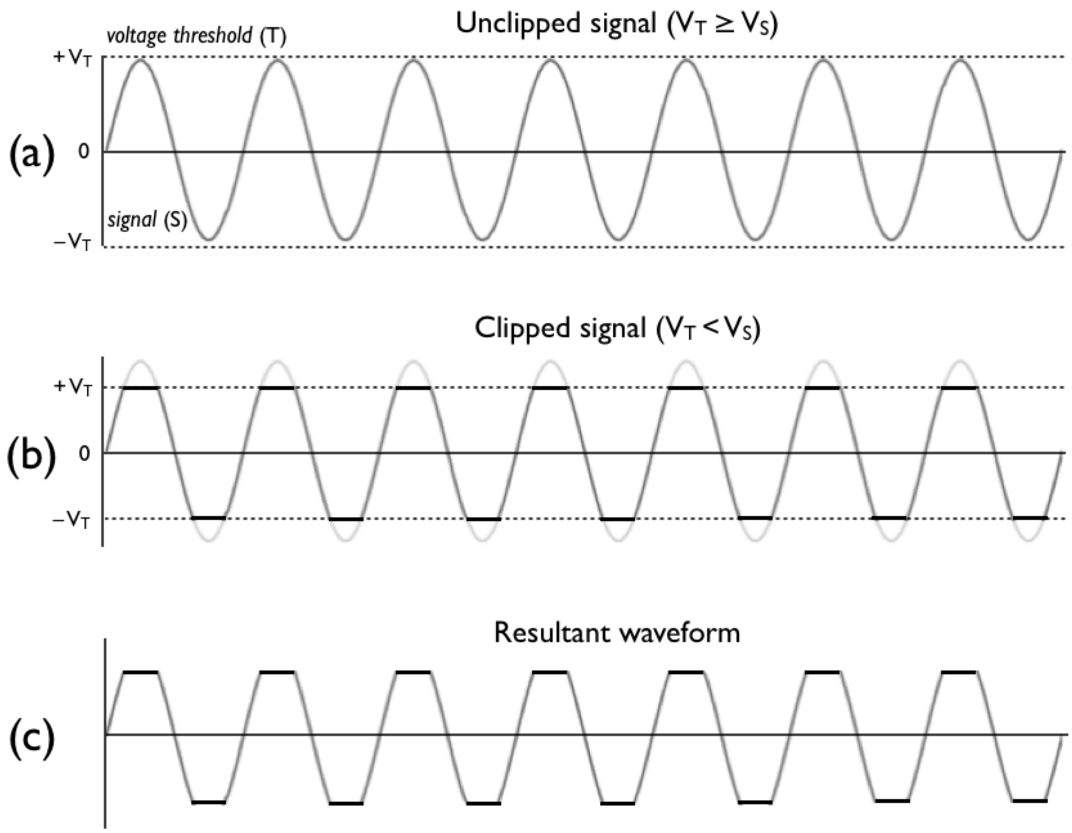


FIGURE 5.1: Diagram of guitar clipping

The electro-acoustic phenomenon above resembles in uncanny ways the timbral-corporeal schema of *blockage* discussed in Chapter 3: the voltage threshold of the amp works to

Metal Zone MT-2. Many other groups eschew pedals entirely, using amp distortion only in the creation of their brutal guitar tone.

"block" the incoming signal, and distortion is the sonic manifestation of this figurative friction. However, appraisal of this basic experiential structure has shifted considerably over the years. Although it began as a timbral sign of impending amp damage or failure, the electromechanical *blockage* of a clean, unadulterated guitar signal quickly became a marker of power, and a dialectical flip from "noise" to "music" was underway.[34] Distortion signifies power not because it overrides this earlier connotation of failure, but rather because it conducts and controls our recognition of imminent breakdown in order to sonically materialize a sense of bodily risk and danger; it plays at the *ultimate limit*, teetering on the verge of collapse yet always somehow managing to make it through.[35] Distortion, moreover, is scalar: the more distortion present in a guitar sound, the more acutely manifest the danger. (In electrical terms, clipping grows more pronounced the greater the differential between the signal and the voltage threshold.) An extreme genre like death metal, which traffics in transgressive risk and violence, is thus an ideal arena for high levels of distortion. Harris Berger and Cornelia Fales have shown through spectrographic analysis that the "heaviness" of metal guitar distortion has quantifiably increased over time in a teleological drive of extremity that will be addressed in various guises throughout the chapter.[36]

Assessing the iconicity and materiality of this particular timbre, specifically its relationship to machines and to bodies, it is natural that distortion would come to play such a genre-defining role in metal music. Death metal embodies a range of different sonic "extremes," and distortion is a palpable result of mechanical extremity. As Walser explains, distortion communicates "ex-

---

[34] Walser, *Running with the Devil: Power, Gender, and Madness in Heavy Metal Music*, 41.

[35] In addition to this daredevil risk of failure, distortion communicates power through its electro-mechanical link with amp *saturation*. By increasing sustain, saturation—of which distortion is an audible byproduct—flattens out the dynamic and articulatory range of the instrument, thus giving the player more leeway for mistakes. Pushing the amp to the edge of failure thus actually has the inverse effect of making the player *less* likely to fail.

[36] Berger and Fales, "'Heaviness' in the Perception of Heavy Metal Guitar Timbres: The Match of Perceptual and Acoustic Features over Time."

treme power and intense expression by overflowing its channels and materializing the exceptional effort that produces it."[37] Distortion is aural testimony to the determined breaching of the guitar amp's material limits, a sign of over-exertion that carries with it the risk of failure. In its dangerous testing of technological boundaries and the *ultimate limits* of the material, distortion thus also flirts with silence, the sonic equivalent to death.

But we do not merely respond empathically to the painful sound of an overtaxed amplifier fighting to do its job. Read through the lens of ASPECS, distortion should also be understood from the perspective of the body, particularly our own internal sound generators, our vocal chords. When voices are pushed beyond their natural limits distortion is the result, and for the same basic reason as the amplifier: the energy being channeled through the corporeal sound-producer is greater than the actual physical capacity of the vocal apparatus to transmit the acoustic signal completely and without added transients or inharmonic tones. By overdriving its material structure, the voice is manifesting the risk of its own (painful) crash.

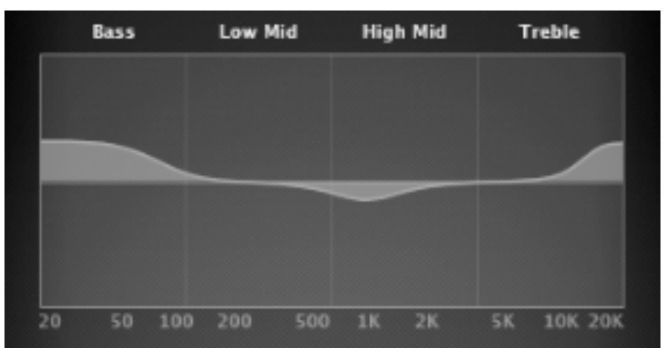
Physical exertion is universally exciting, even if not (as found in Experiments 1 and 2) necessarily positively valenced. As guitarist Marc Ribot reminds us, guitar distortion is tied to this "desire to strain": "broken voices… eroded by screaming… [signify] surpassing the capabilities" of the body.[38] As with the saxophonic scream in free jazz, then, distortion is heard as a psychoacoustic correlate to the vocal "sound of arousal" (screaming, growling, etc.). Distortion, a somatically-marked alarm signal that conveys exertion and a threat to material boundaries, whether of the guitar amp or of the body, thus contains within it a dangerously alluring physical obstacle, where violent collapse—reaching the *ultimate limit* schema—is always a concrete possibility.

---

[37] Walser, *Running with the Devil: Power, Gender, and Madness in Heavy Metal Music*, 42.

[38] Marc Ribot, "Earplugs," in *Arcana: Musicians on Music*, ed. John Zorn (New York: Granary Books, 2000), 233.

Also related to such dangerous, schematically-grounded boundary states is the control of EQ. Equalization is a major component of the "brutal" guitar sound. The most characteristic heavy metal tone (and the best documented in the trade literature), despite Skogsberg's affection for mid-range, is a "mid-scoop" EQ, wherein lows and highs on amps, pedals, and mixing boards are accentuated and mids are diminished (Fig. 5.2).[39] This polarization of the sonic spectrum reflects the emphasis on timbral extremes that is fundamental to the genre: with chiming highs and chest-thumping lows (and not much between), a scooped guitar sound concretizes the binarized borders of the tonal spectrum. The scooped sound enhances a frequency split already implicit in heavy metal guitar practice: as Walser first pointed out, distortion disproportionately increases the energy of higher partials, while the ubiquitous power chord (perfect fifth) produces difference tones below the fundamental pitches of the vibrating strings.[40] Working in concert, mid-scoop EQ, heavy distortion, and power chords all serve to push guitar tone to the spectral extremes.



[39] For discussions of the "mid-scoop" in metal music, see Walser, *Running with the Devil: Power, Gender, and Madness in Heavy Metal Music*, 43; Harris Berger, "Death Metal Tonality and the Act of Listening," *Popular Music* 18, no. 2 (1999): 165 & footnote 2. Additional reference can be found in Dave Hunter's exhaustive book on effects processors: David Hunter, *Guitar Effects Pedals: The Practice Handbook* (New York: Backbeat, 2004), 28. This sonic model also plays an important role in hip-hop and black music more generally, and can be subsumed under Olly Wilson's "heterogeneous sound ideal."

[40] Walser, *Running with the Devil: Power, Gender, and Madness in Heavy Metal Music*, 43.

FIGURE 5.2: Mid-scoop EQ[41]

For all the references to mid-scoop EQ in the literature, however, the practicalities are actually considerably more complex, and scene members widely dispute the "myth of scooped mids."[42] While it is difficult to deny the primacy of extremities in the death metal sonic spectrum, the common perception of a mid-scoop EQ may actually be the result of an aural illusion: with grossly detuned guitars and other sonic signifiers of "heaviness," including thick textures and low frequencies in the bass and drums, intense distortion (which increases the power of the upper harmonics), as well as the propensity for the lead guitar to occupy the screaming upper register during solos, the ear is tricked into focusing on the poles of low and high while ignoring the middle.[43] Certainly the *perception* of scooped-mids EQ, regardless of the actual settings involved in its creation, is paramount to the experience of timbral extremity in the genre.

As soon as we turn to the fundamental pitch of a vibrating guitar string, though, the death metal drive towards frequency extremes is unambiguous. It is an absolute convention in metal to detune the guitar down one or more half-steps. This technique has been a staple of death metal guitar playing since its earliest days, and detuning has followed an ineluctable teleological drive towards lower and lower registers ever since the genre's birth, a matter we will discuss in greater detail later. In the search for the most "extreme" sonic utterance possible, perhaps this gradual sinking is inevitable: a band can always distinguish itself as occupying the outer limits of the style if the guitarists push the tuning one notch lower in pursuit of an even deeper level of brutality. The chronological trend here is remarkable: seminal Florida band Death's debut LP, *Scream*

---

[41] This visual EQ plot is the "mid reduce" preset in Apple's GarageBand.

[42] Commenter "SPL," on <guitartricks.com>. Accessed 10 February 2011.

[43] For more on the musical metonymy of "heavy," see Bogue, "Violence in Three Shades of Metal: Death, Doom and Black," 100.

Bloody Gore (1987), uses a by-now standard D-tuning, in which every string on the guitar is dropped a whole step (E-A-D-G-B-E becomes D-G-C-F-A-D); in the last fifteen years, however, it is not uncommon for groups to tune their instruments to C2 or lower. Cannibal Corpse's *Kill* (2006), for example, was recorded in A♭-tuning, which meant lowering each string a full *minor sixth* from standard pitch.[44] On a regular 6-string guitar, this makes for increasing loss of tension, especially in the bass; players have to develop specific techniques for dealing with the flaccid setup, keeping the tone brutal while also exercising the formidable technical control necessary to execute punchy, fast riffs. They also need to play on the thickest string gauges available.[45]

Despite precautions and special playing techniques, at a certain point in the progressive downward slide, the strings literally become unplayable; there is a limit to how low one can detune an electric guitar. As Paradise Lost vocalist Nick Holmes puts it: "After a point, I mean, how heavy or downtuned can it get—until the guitar strings are literally hanging off the neck?"[46] While in Chapter 3 we discussed the *ultimate limit* image schema as a function of the upper bounds of material possibility, in the case of death metal we might conceptualize this schema (which finds most obvious expression in the metaphor "heavy") in the opposite terms: the death metal genre plays at the *lower* bounds of the material, including low pitches and loose-tension guitar strings (and vibrating vocal cords). Figure 5.3 maps this reorganization. However, in terms of the *exertion* schema, flirting with the *ultimate* (heavy, brutal) *limit* obviously does not imply extreme

[44] See comments on the "Lowest Tuning" thread on <ultimate-guitar.com>. Accessed 6 March 2011.

[45] Kahn-Harris, *Extreme Metal: Music and Culture on the Edge*, 32.

[46] Quoted in Mudrian, *Choosing Death: The Improbable History of Death Metal and Grindcore*, 222.

lassitude, but the contrary: it takes physical work to play such lower-threshold-tempting sounds.

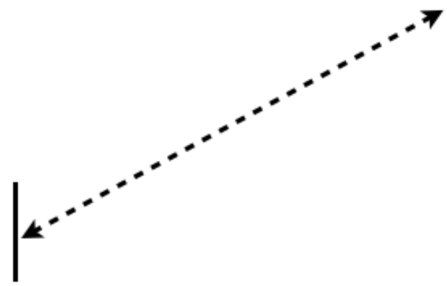


FIGURE 5.3: *Ultimate limit* schema for death metal

Although lower pitches effectively symbolize the "heaviness" of the genre, detuning in death metal is preeminently a matter of timbre. First, auditory roughness is much greater at lower intervals: in the C0 through C2 range, even a perfect fifth can exhibit considerable "beating."[47] In addition, loose strings sound a particular way irrespective of pitch, a matter corroborated by the fact that groups with 7- or 8-string guitars *also* commonly detune them. If easily reaching low notes without the physical inconvenience of limp strings was really the issue, why push it even lower on an added-string guitar, which results in the same string slackness as before? The Swedish progressive death metal band Meshuggah, for example, routinely uses 8-string guitars tuned down to F1 or D#1. For one song—"Spasm," from the album *Nothing* (2002)—they drop their tuning a perfect fifth to B0 (Fig. 5.4). This is lower than a standard electric bass, but pitch is only an ancillary issue: detuning in death metal is about *both* fundamental frequency *and* the audible timbral-corporeal consequences of encroaching *ultimate limits*. The "flappy" sound quality of a profoundly downtuned guitar is an important quality of the timbre—it is sonic evidence that the peripheries of the instrument are being violated in a transgressive and "brutal"

[47] For more on the acoustics of roughness in the lower frequency range, see Vassilakis, "Perceptual and Physical Properties of Amplitude Fluctuation and their Musical Significance."

manner. (Writers on the official Meshuggah forum celebrate the audacious boundary crossing of "Spasm," commenting that the guitar is "totally flapped out.")[48] This "flappy" metaphor is also a quantifiable attribute of the vibrational properties of thick strings themselves. Thick strings, regardless of pitch, produce more inharmonicity and vibrate in a less predictable manner than thin, high-tension strings.[49] Before the signal reaches the amp, therefore, it is already marked with timbral "brutality."

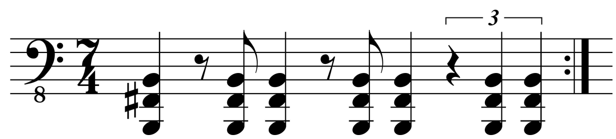


FIGURE 5.4: "Spasm" guitar riff

Finally, standard guitar speakers are not designed to cleanly reproduce such low frequencies. As a result—as in the infamous "Swedish sound"—the speaker cones are subjected to audible strain.[50] Like excessive volume, thick distortion, and polarized EQ, detuning, in its flappiness and inharmonicity, amplifier stress, and menacing low register, helps death metal guitarists perform the timbral "edgework" necessary for the sonic violence fundamental to this highly ritualized genre.[51] We will interpret the ritual dimensions of this guitar tone later, but at this point I want to address the other defining timbral quality of the genre: the death growl.

---

[48] Commenter "Nem," on "Official Meshuggah Forum" <http://www.tandjent.com/meshforum/archive.html>. Accessed 8 March 2011.

[49] See Chris Forster, *Musical Mathematics: On the Art and Science of Acoustic Instruments* (San Francisco Chronicle, 2010), Chapter 4.

[50] Swedish death metal guitarists from the early 1990s liked to play on relatively small, cheap transistor amps (Peavey being the brand of choice).

[51] For more on Jenks's idea of "edgework," see Kahn-Harris, *Extreme Metal: Music and Culture on the Edge*, 29.

Growls, Grunts, and Gutturals: Death Vocals in Theory and in Practice

For many people, the sound of a heavily distorted electric guitar has long been leached of the power to transgress. Gradually metabolized into the bloodstream of popular music, metal-like distortion can be found in a wide range of music today; it is featured in the background to more than one Chevy truck commercial, and plays a prominent role in both *The Daily Show* and FOX News intros and transitions. In short, this particular timbral sign has been tamed and naturalized to serve a variety of functions in the popular music economy.

Not so death metal vocals. Of all the sonic signifiers of the genre, perhaps none is as unrelentingly foreign and off-putting to the outsider as the "death growl."[52] As Walser points out, "alienating noise and exclusivity" has always appealed to metal fans as a point of sub-cultural pride.[53] In death metal, the vocals, more than everything else, serve as a Shibboleth demarcating the borders of a noise-identified community: only the strong can withstand the pure ugliness and brutality of this *sui generis* sound. As Dan Lilker of the band Brutal Truth explains, "[death metal] has to be something that's ugly, where your average person would say, 'the music's okay, but when the vocals come in I just don't get it.'"[54] In this section, I venture to "get" the beast-like bellow of the death growl as a technology of the voice, as a discursive strategy, and as a somatic marker for the communication of physical danger.

Before digging into these issues, however, I need to address one additional facet of death metal vocality that will help frame our inquiry: death vocals are uniquely non-verbal. As Keiran James notes, the guttural style converts semantic meaning into instrumental sound, doing vio-

---

[52] This style is also referred to as death vocals, extreme vocals, harsh singing, cookie monster singing, and a smattering of other terms. For the present purposes, I will stick with "death growl" and "death vocals."

[53] Walser, *Running with the Devil: Power, Gender, and Madness in Heavy Metal Music*, xviii.

[54] Dan Lilker, in "Death Metal: A Documentary." Grimoire, 2004.

lence to language in the process.[55] Bogue follows up this idea: "The most important function of the vocals is to provide a broadly affective, percussive reinforcement of accents and phrases, to fuse vocal noises with the instrumental sounds and create semi-human, semi-machine, blocks of sound."[56] The intelligibility of the lyrics is almost completely obliterated in the process, although death metal CDs almost invariably include lyrics in the booklet.

Interestingly, it is the *lyrics* (as well as the imagery) of death metal songs that have prompted the most outrage from religious groups and the Parents Music Resource Center (PMRC) over the years, even though the words are impossible to hear in the majority of death metal songs. Opening a 1997 senate hearing on the social impact of media violence, for example, Connecticut Democrat Joseph Lieberman referenced the "vile work of the death metal band, Cannibal Corpse… which recorded one song describing the rape of a woman with a knife and another describing the act of masturbating with a dead woman's head." He went on to note: "To me, [Cannibal Corpse's music] is the equivalent of yelling 'fire' in a crowded theater when there's no fire."[57] This raises an important question: If one cannot discern these vile words, then how can the vocal sounds be said to "describe" anything? It is clear that critics are responding not only to the imagery and the lyric sheets of death metal albums; they are also responding to the threatening, "brutal" quality of the sounds themselves, stripped of semantic value.[58] As indicated by Lieberman, critics heard the music of Cannibal Corpse and others as a dog whistle that

[55] James, "From 'The Undead Will Feast' to 'The Time to Kill is Now': Frankfurt School and Freudian Perspectives on Death-Metal," 34.

[56] Bogue, "Violence in Three Shades of Metal: Death, Doom and Black," 107.

[57] Senate Committee on Governmental Affairs, *Music Violence: How Does it Affect Our Chidren?*, 105th Congress-395, 1997.

[58] This assumes, of course, that critics are *bothering to listen*, which Bob Dole notably did not do in 1995 before publically condemning Cannibal Corpse.

induces violent behaviors.[59] The threat is arguably amplified precisely because the vocals are so aggressively non-verbal. Even in the perceptual absence of explicitly violent words, listeners can *hear* violence being done to language and to the singing body (and to the bodies ostensibly represented in the lyrics). One does not have to know that a given song is about torture and mutilation—in a visceral way, one already knows.

In death metal, sound precedes semantics. Vocalists pursue harsh, ugly, painful, scary, abject, violent sounds in response to the unremitting brutality of the guitar riffs and lightning fast drumming, and the words are selected to mirror the dark affective state of the music. As singer Karl Sanders of the band Nile indicates: "The lyrics have to keep up with the music. The music is almost horrifying itself and you can't have lyrics that don't live up to the music in some way."[60] Erik Sayenga of Dying Fetus amplifies this point: "death metal is for the music, not the lyrics… [the lyrics just] cover the aggression that's heard in the music."[61] Language to these musicians is thus a *response* to the sound of the band, not the central organizing principle of the music. The privileged position of sound in this approach can even lead singers to eschew words entirely. Obituary's John Tardy explains: "I won't hesitate to make up a sound instead of an actual word if that is what it takes to make a song feel right. To me it is the sound of my voice that is what it is all about and not the meaning of what I am saying."[62] Dying Fetus's John Gallagher feels similarly: "Personally, I feel that the music and the sound of the vocals are the most important part of

---

[59] Behavioral psychology research on this topic has not revealed any link between listening to heavy metal and the induction of anger: see Gowensmith and Bloom, "The Effects of Heavy Metal Music on Arousal and Anger."

[60] Purcell, *Death Metal Music: The Passion and Politics of a Subculture*, 131.

[61] Ibid., 129.

[62] Quoted in James, "From 'The Undead Will Feast' to 'The Time to Kill is Now': Frankfurt School and Freudian Perspectives on Death-Metal," 34.

death metal. I really don't care what someone is yelling into a mic as long as it sounds good."[63] Many vocalists view the death growl as a natural correlate to the music, as something not just well suited to the style, but necessary. Dan Nelson of Malignancy explains: "the vocals, the grunting—it's the way that the music *has to be*."[64]

"Death vocals" is an umbrella term that covers a panoply of related techniques, including growls, grunts, pig squeals, screams, and deep gutturals.[65] Loosely put, the term describes an exaggerated style of "dist" singing, thick with raspy vocal distortion and a tense, overstrained quality. In order to properly generate the death growl, melodies are virtually impossible and, as previously mentioned, words are difficult to completely annunciate (this is especially true for gutturals, the lowest of the death vocal techniques).[66] As with guitar EQ, singers tend to conceptualize the range of their tonal options in three categories: highs, mids, and lows. While some singers traverse the entire range of options in the course of their songs, many vocalists in the genre stake a claim to one primary register. For example, George "Corpsegrinder" Fisher of Cannibal Corpse typically sings in a low-mid register; Chuck Schuldiner of Death sang in the middle register (with occasional blood-curdling screams); and Frank Mullen of New York-based Suffocation is a low (guttural) singer.

In addition to registral differences between the various techniques, each style has a unique timbral makeup and associated technology of the voice. The grunt, for example, is known for its "gurgly" sound quality, requiring the singer to hold a small amount of saliva in the back of his

[63] Quoted in Purcell, *Death Metal Music: The Passion and Politics of a Subculture*, 129.

[64] Dan Nelson, in "Death Metal: A Documentary." Italics mine.

[65] The following discussion draws on observations from a number of YouTube death metal vocal tutorials.

[66] As Chuck Stelzner has demonstrated in his spectral analysis of the growl, however, a fundamental pitch is always present in the growl, despite the high levels of noise. For more see Chuck Stelzner, "Death Metal/Throat Vocal Analysis," Unpublished paper.

(and it is nearly always a male) throat to bubble away as the vocal cords vibrate; the pig squeal focuses on the vowel /i/; gutturals can have many different sound components, including a "burpy" quality.[67] What unites all of these sounds, however, is a common vocal technique that, contrary to the wild, seemingly untamed quality of the sound, requires careful control, practice, and training to master.[68] Many vocalists warm up before performing with a vocal fry, the sound produced by humming then modulating the voice into its deepest register to generate a rough, gravelly buzz. After warm up, death vocals are produced in a variety of ways, though many singers employ some level of overtone singing and, in the case of deep gutturals, the false vocal cords technique (the same method used by Tibetan monks to generate the incredibly deep drones used in ceremonies and meditation).

When executed correctly, death vocals have minimal negative impact on the body. Performed poorly, growling can easily result in soreness or even painful injury. YouTube tutorials stress the importance of staying well hydrated when singing in order to avoid vocal damage. Additionally, "Corpsegrinder" Fisher suggests that the sound has to come "from deep within"; while growling is possible in the throat, if one does not sing from the diaphragm, stamina can only be maintained for a few minutes before pain and fatigue set in.[69] One of the main tendencies for beginning growlers is to tense up and tighten the throat in order to squeeze out the desired sounds. This is another potentially harmful mistake: vocalists are coached to keep their vocal cords as loose and floppy as possible, despite the natural inclination to link the distorted, high-energy sound quality of the growl to a tighter throat. For false vocal cord singing, which involves the vestibular folds rather than the "true" vocal folds, this is especially essential. Intriguingly, then, the

---

[67] Dan Lilker, in "Death Metal: A Documentary."

[68] Vocal training facilities such as Los Angeles's Extreme Vocals specialize in teaching these techniques.

[69] George Fisher Interview on "Metal Hammer," available on YouTube. Accessed 15 March 2011.

same equalization schema and detuning strategies used in death metal guitar are operative in the vocal technique as well, and as with detuning, it serves a similar function: to push the voice into an extreme and inhuman sounding sonic space. The death metal voice is denaturalized, pressed to do things that are outside the range of its ordinary applications. Learning to sing death vocals is learning how to reshape "musical" vocal timbre into something that evokes primal horror. Even human language is transmogrified into formless noise.

Purcell points out that death vocals should be "beast-like," and should "give you the impression that you're hearing a demon."[70] This inhuman quality is a critical affective ingredient of the sound, carrying with it a long and storied history in Western culture. According to music historian Olivia Bloechl, forms of vocalization that transgress the norms—subhuman, animal sounds, otherworldly timbres, and lewd, inappropriate singing—have for hundreds of years been associated with demonic possession.[71] Accounts of seventeenth-century exorcisms, not to mention cinematic depictions of possession as exemplified by Linda Blair in *The Exorcist*, attest to the mixture of "hideous noise" and song exhibited by individuals under demonic possession: in such instances, pure timbre and semantic meaning were decoupled, with words associated with song (the human part of the possessed) and timbre associated with the meaningless noise of the Devil.[72] This dichotomy is also present in Gary Tomlinson's assessment of "heightened vocalization" when he suggests that sounds like this convey meaning in the manner of a binary code, commu-

[70] Purcell, *Death Metal Music: The Passion and Politics of a Subculture*, 11. Nick Terry calls the vocals in death metal "inhuman": see Mudrian, *Choosing Death: The Improbable History of Death Metal and Grindcore*, 23.

[71] Olivia Bloechl, *Native American Song at the Frontiers of Early Modern Music* (Cambridge: Cambridge University Press, 2008), 55-74.

[72] Ibid., 74.

nicating either "comforting, communal sameness or discomfiting otherness."[73] In death metal, "discomfiting otherness" is precisely the point, and the heightened form of the death growl conveys this dangerous, demonic difference by means of a set of vocal technologies rarely explored in Western society. By breaching the boundaries of normal, acceptable vocal timbre and eliminating the stabilizing tethers of semantic signification in the process, death vocals traffic in the uncanny, summoning a chimerical, grotesque order of meaning linked in the Western imagination with the barbaric, monstrous, and devilish.

As explored in Part I, cultural associations aside this uniquely denatured and noisy style of singing has an affective immediacy conditioned as much by psychobiology as history (or, put differently, conditioned by the "deep history" of our genome). Whether produced on an instrument or with the voice, the sound of the growl is not just an acoustic *sign* of aggression; it is a preconscious trigger that enacts, via act–sound coupling, a specific physiological response consistent with high arousal states. Before this timbre is inscribed with the cultural meanings outlined above, therefore, its affective and evolutionary valence is gleaned by the most primitive parts of the brain.

The guitar and the voice in death metal share certain perceptually noisy features—particularly their elevated levels of high-frequency energy, inharmonic noise, and roughness—that activate subcortical responses consistent with high-arousal negative emotions, specifically (as found in Experiment 3) in the amygdala. Death metal musicians' embrace of extreme volume, speed, and these particular timbral qualities thus demonstrates a dynamic manipulation of listeners' ingrained biological mechanisms. Prior to the symbolic ordering of these sounds, which will be the topic of the rest of this chapter, meaning is ordered around listeners' innate arousal: as

[73] Gary Tomlinson, *The Singing of the New World: Indigenous Voice in the Era of European Contact* (Cambridge: Cambridge University Press, 2009), 194.

corroborated by Experiment 2, we are "tricked" into hearing the high-tension timbres of the guitar and the voice as instantiations of actual physical danger, when in fact they are carefully controlled and regulated. This element of *simulation* is crucial to the symbolic function of timbre in the socio-ritual space of the music.

Brutal sounds simulate high arousal states by forcing the listener, on some level, to mimetically engage through subvocalization and other forms of embodied enaction. Perceiving these timbral qualities, then, is never simply a passive, receptive act, but rather always involves a level of tacit imitation. As ethnographic research on this topic suggests, listeners are often acutely aware of how the brutal sounds of extreme metal affect them, and reactions can be highly physical. In interviews, I found that many listeners are conscious and articulate about why they dislike the genre.[74] One person explained: "[extreme metal vocals] makes you feel uncomfortable, like the singer's in distress." Another reacted with a wincing gesture: "It hurts my ears… it's the textures. When he just starts yelling, it hurts me." Other people explained their reactions in similar terms, indicating that the brutal sound of this genre "gets to me," "makes me uncomfortable," "sounds forced," and "gives me an uneasy feeling." These common reactions illustrate in stark terms the mimetic properties of these timbral qualities. Instead of reflecting a detached, aesthetic critique, death vocals can actually "hurt" to hear, and a large part of this can be attributed to the mechanisms of embodied cognition sketched in Part I.

Death metal singers are keenly aware of the visceral power of the sounds they elicit from their vocal tracts. As Bobby Bray of The Locust eloquently explains, these vocal techniques "elicit an evolutionary alertness which can be translated as 'PAY ATTENTION.' Perhaps it sounds like someone in extreme agony or someone dying, truly one of the first means of verbal communica-

[74] All quotes were transcribed, anonymously, from interviews conducted in association with Experiment 3, reported in Chapter 2.

tion we must have had."[75] The expressive efficacy of the death growl is closer to a self-protective reaction than a conditioned aesthetic judgment, just as intensely loud sounds have been shown to substantially raise adrenaline levels irrespective of their conscious interpretation as sounds.[76] The drastic impact of this unearthly timbre lies in its corporeal powers, its ability to radically tear the listener out of "music-listening" mode, forcing him or her to actually *feel* the horror.

But horror is a moving target. Like guitar detuning, death vocals over the last twenty-five years have gradually polarized in the relentless pursuit of (the most brutal) tone. For a set of vocal techniques intended, since the earliest days of the genre, to communicate the outermost regions of sublime suffering through embodied expressions of pain, the attainment of an even higher level of "extreme" might seem an impossibility, or perhaps simply an absurdity. How can such an explicitly liminal timbre, one that exists on the absolute borders of the body, possibly be pushed any further? A limit is, by definition, a *limit*, not simply a comfortable stopping point on the scale of possibility. Yet despite this incongruity, the teleological drive referenced earlier is just as operative with vocals as with guitars, and it is toward this complex impulse that we now turn.

Timbral Teleology

"Extreme" is not just a sound; it is also a philosophical position, an audible ideology. Written into its code is an unspoken claim about the nature of musical "ends." Since every wall begs to be torn down, every line overstepped, an extreme is an invitation for defiance, an opportunity to subvert the boundary and establish a new colony in the unmapped hinterlands of the sonic wilderness. The end, it turns out, never actually materializes; instead, it is inevitably postponed

---

[75] Stelzner, "Death Metal/Throat Vocal Analysis," 25.

[76] Weinstein, *Heavy Metal: A Cultural Sociology*, 214.

whenever a new liminal space is colonized. In search of a timbral Manifest Destiny, the frontier keeps moving.

Musicians and fans understand this peculiar logic. As Dunn puts it, "there's an ongoing battle in heavy metal to be more evil than the band that came before you."[77] This linear conception of ever-increasing evilness implies an end point. As King Fowley of Deceased has astutely observed, "as it became more 'extreme,' bands found any way they could to be more 'over the top,' be it lower tuning, deeper vocals, more extreme artwork, etc." He concludes: "There's only so low you can tune or so fast you can play a blast beat."[78] There are the limits of style and aesthetic sensibilities, which are quite malleable, and then there are the limits of the human body and our machines (though this dichotomy will be confounded later). Both invite transgression, but the material limits of the body are unsurpassable—if these boundaries are breached, you bleed. The question of musical ends thus differs profoundly from bodily ends: on one end, you get death metal; on the other, you reach the ultimate in excess: death.[79]

"Realistically, you can't progress and expand on something that was meant to be a conclusion," explains Albert Mudrian.[80] Yet somehow, the death metal teleological drive has pushed "extreme" into further and further reaches of sonic density, loudness, speed, and human ability. The ever-lower guitar tunings of the last twenty-five years were noted earlier, but this same drive applies to other parameters of the music as well: death vocals, which parallel the guitar in key ways, have also been pushed into progressively more extreme regions during the lifespan of the

---

[77] Dunn, "Metal: A Headbanger's Journey."

[78] Purcell, *Death Metal Music: The Passion and Politics of a Subculture*, 74.

[79] As Derrida explains, "the excess of life is death." Transposed into music, according to Pattison, "rock and romanticism share an aesthetic appreciation of death as the ultimate form of excess." See Georges Bataille, *Erotism: Death and Sensuality*, trans. Mary Dalwood (San Francisco: City Lights, 1986), 42; Jack Harrell, "The Poetics of Destruction: Death Metal Rock," *Popular Music and Society* 18, no. 1 (1994): 97.

[80] Mudrian, *Choosing Death: The Improbable History of Death Metal and Grindcore*, 222.

genre. The band Death's aforementioned 1987 debut, *Scream Bloody Gore*, for instance, features a mid-range growl punctuated by screams. This style, revolutionary in its day, was adopted by other seminal Florida and Swedish groups like Morbid Angel, Deicide, Entombed, and Dismembered. While heavily distorted, the vocals for these bands are raw and human—one can even discern the lyrics. But by as early as 1992, singers in the scene were beginning to experiment with deeper guttural singing; Cannibal Corpse's first vocalist Chris Barnes, for example, barked the entire *Tomb of the Mutilated* album in a thick, husky, and impenetrable growl. By the middle of the nineties, extremely deep gutturals were *de rigueur* for many groups, including Suffocation, Dying Fetus, and Nile. Combined with lower guitar tunings and faster tempi, the boundaries of "extreme" kept pushing outward, a process has continued to the present.

The most extreme sub-genre of death metal today is the fittingly titled "brutal death metal" scene, as exemplified by the bands Disfiguring the Goddess and Guttural Secrete. Every parameter of this style reflects extremity, from the low gutturals and pile-driving, detuned guitar, to the unpredictable song structures and blazingly fast "blast beats." Guttural Secrete's 2006 album *Reek of Pubescent Despoilment*, moreover, is mastered at ear-damagingly loud levels; one has to turn down one's sound system, in fact, to keep from potentially blowing speakers (or eardrums) upon pressing "play." Nick Terry describes death metal as an "arms race to produce the fastest, heaviest, most brutal music on the planet."[81] In this particular sub-genre, bodily ends are definitively reached: it is difficult to imagine anything faster, heavier, and more brutal. Brutal death metal, therefore, begs the question: Have we finally entered the endgame?

It is true that bodily, material limits have been reached in the death metal scene, but this in itself does not augur the end to this teleological pattern, at least not yet. In the arms race of

[81] Ibid., 23.

extreme, our human bodies are aided and abetted by our technologies. Every step of the way, we create tools to expand the limits of possibility, and formerly unsurpassable boundaries are, with the purchase of new gear, suddenly available to us. Thus emerges the paradox: if a timbre is *not* the audible expression of an extreme physical *act*, despite the fact that is *sounds* like it, is it still "extreme"? Is it cheating to "fake" extreme?

In Chapter 2 we explored these same questions, and the answer—using both behavioral and imaging methods—seems clear: the "sound of arousal" matters more than the real thing. (Recall that distorted guitar produced significantly higher *bodily exertion* ratings than a clean guitar signal.) This finding is particularly relevant in our discussion of death metal timbre. The sonic world of death metal is, as we have already seen, a charnel house of mirrors involving the pervasive use of aural illusions. The amp sounds like it is about to blow, but is actually so powerful that failure is a virtual impossibility. The detuned strings flop on the neck, but they were specifically optimized to do so. The register of some guitar playing sounds impossibly low, but in fact is performed on an added-string guitar. The death growl sounds like the singer is about to tear his larynx, but is actually a precisely controlled vocal technique that requires training and practice. (And further, this vocal technique implies loudness, but is actually comparatively quiet and relies on intense amplification.) The "blast beat" drumming in brutal death metal sounds inhumanly fast and precise, and it is—drum machines are routinely used to supplement the acoustic drums. In sum, the sonic vocabulary of death metal *sounds* at the borders, the frayed space at the margins of the body and our machines. In reality, however, part of this effect is pure trickery; as Ribot slyly reveals, the violence is entirely simulated.[82]

---

[82] Ribot, "Earplugs," 233.

In the next section, we will explore in greater theoretical depth this duplicity in the context of the genre's status as ritual, but for now, suffice to say that mechanized extremity is a different ontological beast from the material extremes of the body. It could be argued that the two are blurred in perception; indeed (such a position goes), who cares that the voice is carefully controlled as long as it *sounds* tortured? I keep emphasizing the word "sound" here in reference to ASPECS because, with the help of technology, what we *perceive* in death metal can be vastly different from the reality of the physical *acts* underlying its production. The very disjuncture between *acts* and *sounds* in this formulation reveal the fundamental illusion of the style, even if our perceptual systems lie. These timbres sound in*tense,* although their physical production requires *looseness* (of guitar strings and vocal cords); they are conceptualized as "brutal" and "evil," even if they do not involve any real physical violence or risk; they convey the idea and sensation of torture, although no real torture is involved in their creation. As we shall see, this is not necessarily a sign of creative failure. In the next section, I investigate how these evil timbres work in their ritual context, employing the theoretical perspectives of Jacques Attali and René Girard to examine what they *do* for participants. As I will argue, this trickery is essential to the sacrificial drama of the music.

Death Metal as Ritual Sonic Sacrifice

Timbral brutality in death metal always takes the form of noise. From sheer loudness to "inhuman" vocal and guitar distortion, both of which imply a forceful breaching of boundaries, noise is the operative category of experience when assessing death metal timbre, the anti-code that goes up against the normal, acceptable code of "music." A large part of this identification with noise, as certain writers have observed, is social in nature: scene members value the harsh sound quality precisely because the population at large has such antipathy toward it. As Kahn-Harris points out, extreme metal involves the "aestheticization of sounds that many people find repellent."[83] To be sure, these sounds are aestheticized *because* they are repellent, and valuing them gives the subculture a powerful form of cohesion and distinction. But noise is not courted purely to define the border between in-group and outsiders; it also serves a crucial function within the music itself.

The poles of chaos and control are central to death metal, and noise is a central element of chaos. Even to fans, the music should be "harsh on the ears":[84] aspects of death metal are indeed *meant* to be heard as noise—as King Fowley points out, "metal ain't meant to be pretty."[85] These elements, what Walser identifies as the "deliberately transgressive, noisy, violent" side of heavy metal, are not given totally free reign, however.[86] Chaotic elements ("noise") are summoned only to be restrained and brought under meticulous control ("music"). Disorder and order are thus simultaneously present, and scene members hear it dialectically as both noise and music at the same time. Returning to the Lester Bangs epigraph: "heavy-metal rock is nothing more than a bunch of noise; it is not music, it's distortion—and that is precisely why its adherents find it appealing."[87] To Bangs (echoing Larkin's assessment of free jazz), metal is unredeemable noise, sonic disorder. To adherents, however, it is *ordered* noise, a dialectical relationship between the binary poles of chaos and control. As musicians readily point out, listeners have to "get beyond the 'noise' reaction" in order to really understand death metal, an observation that presup-

---

[83] Kahn-Harris, *Extreme Metal: Music and Culture on the Edge*, 5.

[84] Purcell, *Death Metal Music: The Passion and Politics of a Subculture*, 12-13.

[85] Ibid., 188.

[86] Walser, *Running with the Devil: Power, Gender, and Madness in Heavy Metal Music*, 14.

[87] Weinstein, *Heavy Metal: A Cultural Sociology*, 2.

poses two fundamental levels of meaning.[88] On the surface level, death metal can be quite noisy; beneath this seeming chaos is a deeper layer of order.

Noise and chaos are essential to the death metal timbral palette, but how precisely are they restrained and channeled into order? As we have seen, technological "controls" (big amps, thick strings, etc.) are the norm. This principle applies to larger systems of musical organization as well: order and control are expressed musically through a number of stylistic conventions. Death metal is structured around tight, unison guitar riffs, a dense framework where little deviation is possible. In the typical song this framework can be very complex, featuring sudden and dislocating shifts in time signature and tempo.[89] This makes individualistic guitar solos problematic. Compared to other metal genres, melodic improvisation—perhaps the paradigmatic musical signifier of freedom—is relatively rare in death metal.[90] When guitar solos do occur, moreover, they are generally brief and used as transitional elements. Also, while the guitar tone contains a high level of distortion, techniques that represent an *actual* loss of control over sound (like feedback) are virtually never used.[91] Instead, one of death metal's most characteristic guitar techniques is palm-muting, a rhythmic method that involves strumming the strings while simultaneously covering them with the palm of the right hand. The strings are not allowed to vibrate freely, and the resulting sound is a highly distorted, but dull rhythmic thud. Perhaps more than anything else, palm-muting—a technique that represents a player's fierce repression of the vibra-

---

[88] "Death Metal: A Documentary."

[89] For more, see Harrell, "The Poetics of Destruction: Death Metal Rock," 92; Bogue, "Violence in Three Shades of Metal: Death, Doom and Black," 104.

[90] Kahn-Harris, *Extreme Metal: Music and Culture on the Edge*, 34. This aspect of death metal differs markedly from the soloistic virtuosity of mid-1980s heavy metal discussed in Walser's book.

[91] Ibid.

tory energy of heavy, stiff guitar strings—illustrates death metal's musical orientation toward control.

How are we to understand this dialectic chaos and control, the subsuming of noise into a higher order?[92] Drawing from a range of anthropological theorists, Jacques Attali in his influential theoretical tract *Noise* (1985) contends that noise, in both its acoustic disorder and in its relationship to actual physical pain through high volume levels, is a symbolic stand-in for ritualized murder.[93] Turning noise into music thus asserts control over this primal violence; it is the channelization of destructive chaos into socially productive modes of organization. He writes: "Noise is a weapon and music, primordially, is the formation, domestication, and ritualization of that weapon as a simulacrum of ritualized murder."[94] This ritual function defines the first period in Attali's tripartite division of musical history, *sacrificing*. Sound here is put to the service of replicating the same ritual and social codes that expel collective violence through the form of a scapegoat and blood sacrifice, a theory developed by René Girard in his seminal *Violence and the Sacred* (1977).[95] The relationship between noise and music, then, articulates the same cultural patterns,

[92] There are many potentially fruitful theoretical approaches to this question. For instance, death metal could be read in the context of Antonin Artaud's "Theater of Cruelty," which postulates a drastic, "extreme" relationship between audience and performer. There is also an element of the Sublime to the death metal genre, particularly the Burkean "delightful horror" concept (Kant's "negative pleasure"). These philosophers contend that terror and pain are always the cause of the Sublime.

[93] Jacques Attali, *Noise: The Political Economy of Music*, trans. Brian Massumi (Minneapolis: University of Minnesota Press, 1985).

[94] Ibid., 24.

[95] René Girard, *Violence and the Sacred*, trans. Patrick Gregory (Baltimore: Johns Hopkins University Press, 1977).

anxieties, and attitudes towards violence as *religion*, and noise is therefore a presumptively sacred phenomenon fundamental to the ordering of collective identity and social stability.[96]

The death metal genre seems tailor-made for Attali's thesis.[97] Inherent in the structure of the guitar and vocal timbres, I contend, is a sacrificial dynamic between the chaotic violence of the noise elements and the systems in place to bring these timbres under control. Noise is a stand-in for the sacrificial victim; it is something deliberately ugly, for onto it is projected the violence that is ultimately to be subsumed into order. It is thus, in cultural anthropologist Victor Turner's words, a blend of "lowliness and sacredness," low because it is ultimately rejected, and sacred because it purifies the violence that would otherwise infect the whole group.[98] Death metal musicians cultivate the chaotic and impure violence of noise in order to channel it into a socially productive ritual. These timbres are invested both with the noisy corporeal violence of their production (even if imaginary), and with the projected violence of the listener, a dynamic that eschews semiosis. They *have to be* associated with abjection, with liminal zones, and with physical pain, therefore, in order to attract and sponge up our negative feelings and violence, which is their ultimate purpose as victim in the sacrificial drama. As with a real sacrificial victim, once the sound is invested with various forms of violence, it is ready to be purified (i.e., sacrificed, controlled).

---

[96] Phillip Chidester profitably relates death metal rhetoric to the four stages of Kenneth Burke's symbolic human drama consisting of (1) order, (2) victimage, (3) purification, and (4) transcendence. For more see Phillip Chidester, "Through the Blood: Death Metal and the Rhetoric of Song as a Transcendent Discursive/Presentational Form" (Ph.D Diss., University of Kansas, 2004). These stages correspond closely with ethnographer Arnold van Gennep's classic trifecta of (1) separation, (2) margin [transformed by Victor Turner into "liminal" stage], and (3) aggregation. All of these ideas deserve to be worked through more thoroughly, which, unfortunately, falls outside the scope of the chapter.

[97] Attali writes: "The hypotheses of noise as murder and music as sacrifice are not easy to accept. They imply that music *functions* like sacrifice; that listening to noise is a little like being killed; that listening to music is to attend a ritual murder, with all the danger, guilt, but also reassurance that goes along with that…" (p. 28). Vis-à-vis death metal, perhaps this hypothesis *is* easy to accept.

[98] Victor Turner, *The Ritual Process: Structure and Anti-Structure* (London: Routledge, 1969), 96.

This dynamic plays itself out both in the timbres themselves and in the ritual of performance, a matter I will discuss in greater detail in the next section.

Death metal timbres, as noted previously, are not, practically speaking, as tortured and physically liminal as they let on. Despite the apparent horror, the loudness and the distortion—the *noise*—is actually very carefully controlled. Its connotation of pain is a simulation brought about by the contingencies of human perception that we hear extreme timbres—the "sound of arousal"—as a reflection of extreme material states, both of the body and of our technologies. This, of course, is a false perception: amps in death metal rarely blow, guitar strings almost never fall off, and vocalists rarely spit blood as a result of their inhuman laryngitic contortions.[99] These timbres sound at the limits, but they rarely actually are. Ribot sees this as a form of inauthenticity, a trick that diminishes the power of volume and distortion as sacrificial tools in the concert's ritualized "theater of pain."[100] But what Ribot fails to acknowledge is the fact that trickery is actually a fundamental element of sacrifice. As Girard indicates, sacrifice is an act of symbolic *substitution*: internal group violence is projected onto the victim, who is purged in lieu of the destabilizing murder that is bound to occur if such rage is left unchecked. He explains: "The function of ritual is to 'purify' violence; that is, to 'trick' violence into spending itself on victims whose death will provoke no reprisals."[101] In timbral terms, these sounds "trick" us into imbuing them with actual physical violence—into hearing them as the sonic incarnate of brutality and evilness—but in reality, they are safe. These sounds have sacrificial power because they *resemble* the audible af-

---

[99] It should be acknowledged that the level of physical risk is higher in this genre than most. Singers and drummers have permanently injured themselves as a result of these techniques. These are the exception rather than the rule, however; for the most part, singers, guitarists, and drummers are able to maintain a consistent, repeatable level of brutality night after night. For more on injuries associated with making this music, see Mudrian, *Choosing Death: The Improbable History of Death Metal and Grindcore*, 262.

[100] Ribot, "Earplugs," 236.

[101] Girard, *Violence and the Sacred*, 36.

ter-effects of material violence; in Attali's terminology, they are a simulacrum of ritual murder. Yet they are not implicated in any significant *real* violence—the ritual is entirely symbolic, and the sacrifice (like all sacrifices) requires an element of sleight-of-hand.

This does not mean, however, that substitution is continually a fixed relationship in the genre, that timbres with a certain level of faked violence will always continue to signify in the same way and provide the same sacrificial use-value for participants. As we have seen, the timbral balance in death metal is hardly stable: since the genre's birth, groups have had to continually up the ante of "extreme" on a teleological treadmill of increasing brutality simply to stay in the same phenomenological place. Arguably Chuck Schuldiner (of Death) sounded just as "extreme" in 1987 as Cameron Argon (of brutal death metal band Disfiguring the Goddess) does in 2014. The sacrificial scope of the genre has shifted, and new liminal timbres have been created to fill the void. In other words, the power of "extreme" as a timbral simulacrum of sacrifice has lost its potency over time, requiring sounds with a higher and higher brutality quotient just to get the same effect. In the context of Girard's theory of sacred violence, this is a quintessential "sacrificial crisis."

Girard postulates that "impure" violence must always be kept separate from "pure" violence; if the two intermingle and the difference is collapsed, a "crisis of distinctions" could result, throwing the whole hidden mechanism of the sacrificial dynamic into sudden, irrevocable doubt.[102] Sameness and equivalency thus undermine the whole cultural endeavor of sacrifice: that which is Other, abject, and impure—the "noise" in a given system—is polluting, and thus

[102] Ibid., 39-49. He writes: "As long as purity and impurity remain distinct, even the worst pollution can be washed away; but once they are allowed to mingle, purification is no longer possible" (p. 38).

must be separated and contained.[103] If this category is normalized and incorporated into the whole, if difference is erased, then the ritual logic of the sacrifice is invalidated and violence will overspill its banks. Applied to the teleological drive of death metal, we can see that, since the beginning, extreme timbres like grisly, detuned guitar distortion and death growls have been defined by their very extremity and Otherness, by the fact that they exist on the borders of the body and our machines. Mastering (i.e., sacrificing) such noisy chaos in a ritual of performance is the *raison d'être* of the genre: noise is summoned only to be dominated. But the meaning of "extreme" is bound to shift as what were once liminal timbres become the norm of the genre; without a clear distinction between chaos (noise) and control (techniques to channel and order this noise), the symbolic system of violence loses its meaning and collapses.[104] Like an overused medication, sounds can incrementally lose their power to signify extremity.

Standardization is the enemy of extremity.[105] When Chuck Schuldiner was the only person singing with the gravelly "death" tone, it was appropriately liminal; when dozens of groups are all singing the same way, the sound loses sacrificial power. The abjectness of timbre necessary for purgative sacrificial uses was thus rendered normal, and the distinctions between "noise" and "music" were dissolved. A symbolic distinction, we should note—though grounded in embodied knowledge of timbral-corporeal events—this dissolution of the threshold separating "music" from "noise" precipitated a sacrificial crisis that led musicians to define a new order of noise. In re-

---

[103] This notion is taken up by Mary Douglas in her classic work on symbolic categories, dirt, and taboo: see Mary Douglas, *Purity and Danger: An Analysis of the Concepts of Pollution and Taboo* (London and New York: Routledge Classics, 2002).

[104] One might argue that the spate of murders and church burnings that came out of the 1990s Norwegian black metal scene represents precisely this sort of sacrificial crisis: transgressing the tightly structured sacrificial dynamic of their death metal counterparts, this scene sought to unleash noise entirely. This is a suggestive link between the cultural history of black metal and the genre's aesthetic conceptualizations, particularly as it pertains to timbre.

[105] It is also a major reason why the genre is so aggressively "underground." As Incantation's John McEntee puts it, death metal needs to stay a sub-cultural phenomenon, where "there's no limits to how extreme it can be." See Mudrian, *Choosing Death: The Improbable History of Death Metal and Grindcore*, 269.

sponse, bodies and machines were pushed into ever more marginal regions to recapture the energy and danger of the original timbre. Girard explains: "The more critical the situation [the crisis of violence], the more 'precious' the sacrificial victim must be."[106] In death metal terms, greater sacrificial use-value is pursued through increasingly brutal timbres, those packed with a greater level of abjection—lower gutturals and guitars, thicker and more viscous distortion, "flappier" strings, etc. The search for more "precious" sacrificial noise is the engine driving this teleological momentum.

When the precarious dynamic is upset, when formerly extreme sounds are neutralized, then a more sublime level of noise is needed to establish a new sacrificial equilibrium. Attali speaks precisely to this point: "Noise only produces order if it can concentrate a new sacrificial crisis at a singular point, in a *catastrophe*, in order to transcend the old violence and recreate a system of differences on another level of organization."[107] "Catastrophe" is a telling word here; it signals the *ultimate limit* schema, the furthest reaches of material possibility. For the last twenty-five years, catastrophic reorientations of timbral power were available to musicians looking for a more "over-the-top" sound. This has led us to the proverbial brink: human bodies are becoming inadequate to the task, and machines are used as a substitute.[108] Nonetheless, it is still the fundamental ground of the *body* on which the symbolic sacrifice takes place, even if it appears in the form of the body's prosthetic extension, technology. Any catastrophe, therefore, is a crisis of the corporeal, both in terms of the expressive power of the timbres (as embodied extremity) and as an

---

[106] Girard, *Violence and the Sacred*, 18.

[107] Attali, *Noise: The Political Economy of Music*, 34. Italics in original.

[108] Although this is not timbral, the use of drum machines in brutal death metal exemplifies this slide into greater mechanization and, I would argue, is a harbinger for timbral manipulations that are bound to follow.

effect of the ritual experience for audiences and musicians. This experience calls for closer examination.

Mutilation by Proxy: Real and Imaginary Bodies at the Death Metal Concert

At this point, I want to turn from a theoretical understanding of sacrificial violence to look at the way these sounds function in the actual experience of the death metal concert. The performing body, so fundamental to the ordering of meaning in these liminal timbres, is also an essential aspect of the live music spectacle, serving a multivalent function in a ritual that contains at its core a sonic simulacrum of sacred violence.[109]

Purcell, Weinstein, Robin Sylvan, and others have noted the centrality of live concerts to the death metal scene, demonstrating the fundamentally sacred aspect of these events, a connection also articulated by fans.[110] One scene member, for instance, describes the concert experience as "like a religious experience… you're really connected."[111] Drawing on observations like this, Sylvan likens the concert to a form of ritualized unveiling: "Ideal metal concerts can be described as hierophanies in which something sacred is revealed."[112] Since violence, indeed, is a crucial definer of the sacred—Girard goes as far as to call it "the heart and secret soul of the sacred"—and

---

[109] For an interesting Bakhtinian account of the role of the body in death metal, see Karl Beckwith, "'A Human Being This Once Resembled': Bodily Transgression in Extreme Metal Music," Unpublished paper(2006). Beckwith interprets the lyrical landscape of bodily violence to represent the "grotesque" body, which is brought under the control of the "classical" body (technical mastery, etc). Additional theoretical speculation on the body in death metal can be found in Purcell, *Death Metal Music: The Passion and Politics of a Subculture*, 175.

[110] See *Death Metal Music: The Passion and Politics of a Subculture*, 192; Weinstein, *Heavy Metal: A Cultural Sociology*, 213; Robin Sylvan, *Tracing the Spirit: The Religious Dimensions of Popular Music* (New York: NYU Press, 2002), 163. Weinstein introduces this topic with explicit recourse to Durkheim's notion of the "sacred of transgression" (p. 39), while Harrell calls the death metal concert a "sacred festival": Harrell, "The Poetics of Destruction: Death Metal Rock," 95.

[111] Sylvan, *Tracing the Spirit: The Religious Dimensions of Popular Music*, 166.

[112] Ibid., 168.

death metal is a “purifying” form of sonic violence, I argue that the death metal concert ultimately functions as “the awesome machinery of ritual” in which collective violence is pooled then sacrificially expunged.[113] To analyze this function, we will begin with a discussion of the musicians on the stage, then move on to the fans in the audience, focusing on the role that both the literal and the symbolic body play in the sacred violence of the concert experience.

As the people who push their bodies and gear to violent extremes in front of an audience (or at least seem to), death metal musicians might appear as the sacrificial victims of the ritual-concert. The opposite, however, could also be true: with the power to produce annihilating levels of volume, musicians occupy the role of sacrificer (with the audience being the victim). This dual identity has been noted by writers: Attali, for example, refers to the musician as “sacrificed sacrificer.”[114] Observing, too, that musicians presumably experience louder volume levels than audiences, Ribot comments: “in this illusion, the musician is both sacrificial victim and magical protector who filters the dangerous volume levels through his/her body (literally standing between amp and audience) to protect the audience.”[115] In this scenario, the musicians suffer for the audience, and the audience’s pain is justified by the musicians’ sacrifice. (There is a strong element of self-mortification to this: the musicians are wielding the weapon of sound against their own bodies first and foremost.)

Bodily violence appears in three different aspects in the typical death metal performance. First, violence is present, as we have seen, in the timbres themselves. Although they do not usually involve real pain to the performers, they *communicate* physical pain via the mechanisms of embodied timbre cognition. Second, themes of bodily destruction are the lyrical stock-in-trade of the

---

[113] Girard, *Violence and the Sacred*, 31 & 19.

[114] Attali, *Noise: The Political Economy of Music*, 30.

[115] Ribot, “Earplugs,” 236.

genre;[116] the words provide a specific annunciation of the inchoate bodily violence heard in the timbres.[117] And third, the extreme loudness that characterizes the death metal concert serves as a literal form of violence in its potentially damaging effects on the ear. Though these three forms of violence combine in ambiguous, contextually defined ways, bodily threat is omnipresent in the death metal concert, both literally and symbolically, and this sense of physical danger and risk is essential to its power as a sacrificial ritual. Occasionally these energies overwhelm the bounds of the purely symbolic, and real blood is spilt: Morbid Angel guitarist Trey Azagthoth, who aspires to "spiritualism in music" and "being ritualistic" in concert, for instance, has been known to cut himself on stage and display his bleeding arm for the audience (as shown in Fig. 5.5).[118] Most of the time, however, bodily mortification is either entirely symbolic (the lyrics) or represented in a form that involves no real harm to the musicians ("faked" timbres and earplugs effectively eliminate the risk to musicians' bodies).[119] But fleshly mortification is not the exclusive right of musicians in the death metal concert; it is also projected out onto the audience, who, in Phillip Chide-

---

[116] Illustrating this perhaps obvious point are lyrics from a randomly selected and entirely representative Cannibal Corpse track, "Mummified in Barbed Wire" (from the 1996 album *Vile*): "Wrap the wire around the neck / pull it tight to suffocate / his mouth is spewing steaming blood / hang the body with a cord / gag his mouth with shards of steel / he doesn't know the pain he'll feel / interwoven deadly spikes / tightening cables lacerate / assorted gashes open wide / ruptured veins are gushing blood / scraps of skin are flying off / the body of the helpless prey."

[117] Although lyrics are seldom clearly decipherable, audiences often know the words to songs by their favorite artists, so the connection is made between undefined sonic violence in the form of signifiers of bodily suffering, and the specific linguistic annunciations of violence against the body contained in the song lyrics. As Kahn-Harris points out, death metal lyrics express an "almost obsessive desire to explore that which is dangerous, that which is scary, that which points towards obliteration, formlessness; a delight in exploring the body in its ways of being cut up, destroyed, and mutilated" (in Dunn, "Metal: A Headbanger's Journey."). However, it is often difficult to determine just *whose* body is being mutilated in the typical song-text. Chidester, in his study of death metal rhetoric, concludes that most songs are structured with a good deal of ambivalence in this regard: listeners are guided to identify sometimes with the victim, sometimes with the victimizer, and other times with both at once. (See Chidester, "Through the Blood: Death Metal and the Rhetoric of Song as a Transcendent Discursive/Presentational Form," 174.) The multidirectionality of violent impulses makes for a greater flexibility of sacrificial use-value in each lyrical and timbral trope. What is always present, however, are themes of bodily violence that treat the corporeal as, in Kahn-Harris's formulation, "ludicrous and revolting but endlessly fascinating" (Kahn-Harris, *Extreme Metal: Music and Culture on the Edge*, 36.).

[118] Mudrian, *Choosing Death: The Improbable History of Death Metal and Grindcore*, 215.

[119] Ribot, "Earplugs," 236.

ster's words, can choose to undergo "mortification by proxy" through the same means as the musicians (timbre, lyrics, and volume).[120] The bodies of the audience members are thus substituted into the sacrificial drama.

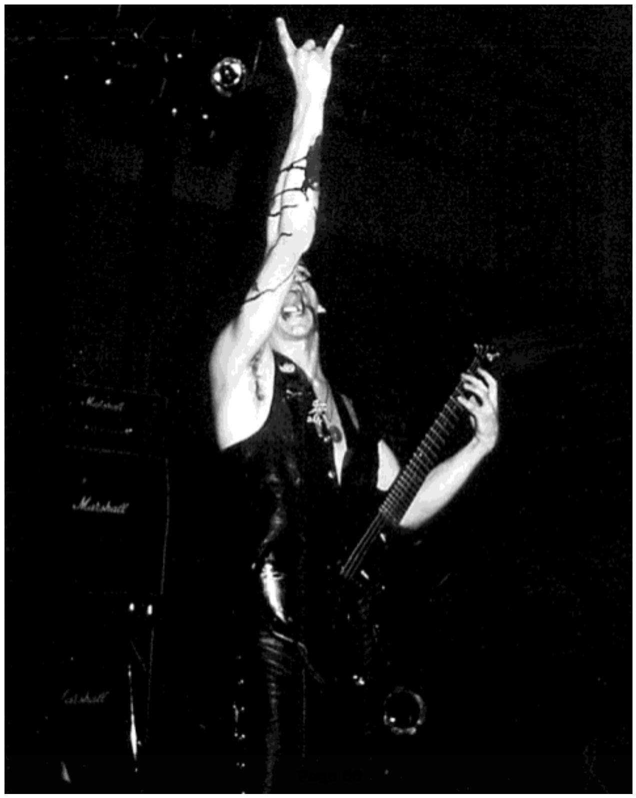


FIGURE 5.5: Morbid Angel's Trey Azagthoth bleeding in concert
(Photo by Frank White, from Purcell 2003, p. 60)

It is striking how body-centered accounts of death metal concerts from audience members can be. Fans talk about the "intensity" and "energy" of concerts as a major part of the appeal, relating the live music experience explicitly to ecstatic modes of being.[121] As one fan put it, the

[120] Chidester, "Through the Blood: Death Metal and the Rhetoric of Song as a Transcendent Discursive/Presentational Form," 41.

[121] See Kahn-Harris, *Extreme Metal: Music and Culture on the Edge*, 52; Weinstein, *Heavy Metal: A Cultural Sociology*, 213; Sylvan, *Tracing the Spirit: The Religious Dimensions of Popular Music*, 159. Georges Bataille revealingly tells us that "ecstasy begins where horror is sloughed off": see Bataille, *Erotism: Death and Sensuality*, 68.

experience "stirs something inside of you, it's like fucking adrenaline you know what I mean?"[122] Audiences at death metal concerts often respond to the violence of the music with bodily violence of their own; more than any other genre, it is known for the ferocity of its head-banging, stage diving, and mosh pits.[123] These aggressive forms of audience participation have raised the ire of critics who point to this as evidence of the danger of the music. However, the violence of these activities is highly circumscribed by ethical codes: like the timbral dynamic of the style itself, chaos is never given free reign. When someone falls in the mosh pit, they are helped to their feet; when someone jumps off the stage, they are typically caught. One scene member explains that moshing is "mainly a release of aggression, tension, stress. Trying to get as wild as you can without killing yourself."[124] It is notable that this form of physical violence, like the sounds, is concerned with approaching the *ultimate limit*, but not overstepping it. It is a carefully mannered sort of brutality, a controlled chaos that never blurs the distinction between "pure" (that which expels tension in a circumscribed way and builds group cohesion) and "impure" violence (the unchecked, uncontrolled variety). If an individual transgresses against this code and swerves into the latter form of violence during a show, they are expelled from the group.[125]

Like many rituals, the death metal concert involves an element of physical endurance. Bodies are forced to withstand brutal sounds, extreme loudness, and the physical agitation of the mosh pit; they are forced to undergo, like the musicians, a form of symbolic mutilation. This is another major component of the question of "brutality" and "evilness" in death metal: the more

---

[122] Kahn-Harris, *Extreme Metal: Music and Culture on the Edge*, 52. This "adrenaline rush" aspect is confirmed by accounts in Sylvan, *Tracing the Spirit: The Religious Dimensions of Popular Music*, 167.

[123] For more, see Kahn-Harris, *Extreme Metal: Music and Culture on the Edge*, 44.

[124] Sylvan, *Tracing the Spirit: The Religious Dimensions of Popular Music*, 160.

[125] James, "From 'The Undead Will Feast' to 'The Time to Kill is Now': Frankfurt School and Freudian Perspectives on Death-Metal," 35.

brutal and violent the experience, the more powerful the individual feels for having survived something that the average person would shrink from in horror.[126] As one scene member puts it: "The real over-the-top either overboard Satanic overboard distortion vocals or overboard eeriness to the music—that's when my medication was being filled, as being 'yeah, this is good.'"[127] To fans like this, extremity (the "over-the-top," "overboard" quality) is required for the music to work as a medication, for it to serve a purgative function in the ritual-concert. Again, physical endurance is key.

The ritual is fundamentally geared towards its survivors, those who walk away from witnessing death (metal) with a feeling of reassurance and power. This is, to most fans, the ultimate goal of the ritual: as a "here and now" religion, what death metal does for people is of crucial importance in a cultural assessment of the genre.[128] We can safely bet that nobody in the mosh pit is consciously aware of the sacred dynamic that underlies the ritual; indeed, as Girard points out, "celebrants do not and must not comprehend the true role of the sacrificial act."[129] So what is the (conscious) point of all this? What role does the death metal ritual, and all the sounds that make it work, play in the lives of ordinary scene members?

Conclusion

In this chapter I have argued that the "brutal" timbral qualities of death metal are used to symbolically enact the sacred violence essential to the ritual of the music. Girard and Attali stress that

---

[126] For more on this component of brutality, see Purcell, *Death Metal Music: The Passion and Politics of a Subculture*, 186; Sylvan, *Tracing the Spirit: The Religious Dimensions of Popular Music*, 169.

[127] *Tracing the Spirit: The Religious Dimensions of Popular Music*, 180.

[128] The term here is from Thomas Bossius. For more, see Kahn-Harris, *Extreme Metal: Music and Culture on the Edge*, 10.

[129] Girard, *Violence and the Sacred*, 7.

sacrifice is ultimately aimed at curbing the violent energies of a group, preventing contagious cycles of murder by rechanneling aggression toward a symbolic victim. Sacrifice thus protects a community from its own violence.[130] On an individual level, however, the significance and meaning of sacrifice is more diffuse; controlling group behavior through ritual involves one set of social dynamics, but individuals function under a different set of psychological drives. If sacrifice purges harmful violence from a group, then on an individual level the ritual can be said to serve a cathartic function, allowing people to pour their rage and negativity into the receptacle of the sacrificial victim.[131] This is why the victim-object needs to be *bad* (or in this instance, a "bad sound"): it has to be able to attract these violent energies to itself. The "evil" sounds of the genre, therefore, must be the way they are in order to serve as an effective conduit for violence. Timbral "evilness" and "brutality" are necessary prerequisites for ritual purification and release.

The ritual of the death metal concert is a high-volume exercise in cathexis and catharsis—violence is extracted (in the form of loudness, timbral brutality, and moderated physical aggression); accumulated (around the ugly, noisy object of sacrificial focus); then purged (controlled).[132] A successful ritual-concert, like an efficacious sacrifice, curbs and rechannels individuals' violent energies into a feeling of physical energy, ecstasy, and release. Participants' aggression has been exorcized—it is no longer a threat to the group and to themselves. As one fan explains: "I would either go and kill people or I would kill myself or both… because I was so angry. I found a release when I went to concerts. You know, I went in there angry. I did my moshing. I

---

[130] Ibid., 8.

[131] Ibid., 99.

[132] This maps well onto James's tripartite scheme: acknowledge – control – rechannel. For more, see James, "From 'The Undead Will Feast' to 'The Time to Kill is Now': Frankfurt School and Freudian Perspectives on Death-Metal," 35.

did my stage-diving. I did my screaming… When I walked out of that club, and I was straight, I felt great."[133]

This liberating function has been noted by writers and fans, but catharsis is not the only role the music plays in people's emotional lives. Kieran James, in his Freudian interpretation of the genre, calls it "controlled group therapy," arguing that it is a safe, regulated place for individuals to explore their repressed death drive.[134] Minus the Freudian terminology, Berger follows this line of interpretation when he writes that the music is "not merely intended to evoke simple cathartic release, but to help listeners explore their emotional life actively, to examine the feelings that hold them back and to overcome these."[135] In contrast to the conventional understanding of the genre as promulgating anger, violence, and nihilism, death metal helps fans to acknowledge and confront their fears, anxieties, and frustrations. But drudging such dark emotions to the surface requires a potent medicine. Paradoxically, it is only through extreme sonic representations of death, torture, decay, and dismemberment that the ultimately positive and life-affirming psychological effects of the ritual can be realized.

How do such disturbing and violent themes (and timbres) function to produce a vital feeling of ecstasy? According to Georges Bataille, sacrifice involves the ritual transformation, through violence, of separate, particular, discontinuous being (life) into the limitless, infinite, and sacred continuity of wholeness (death).[136] He writes: "It is the common business of sacrifice to bring life and death into harmony, to give death the upsurge of life, life the momentousness and the vertigo

---

[133] Sylvan, *Tracing the Spirit: The Religious Dimensions of Popular Music*, 167.

[134] James, "From 'The Undead Will Feast' to 'The Time to Kill is Now': Frankfurt School and Freudian Perspectives on Death-Metal," 35.

[135] Berger, "Death Metal Tonality and the Act of Listening," 173.

[136] Bataille, *Erotism: Death and Sensuality*, 90. To Bataille, sacrificial violence is not just invested with the vitality of life, but with erotic power.

of death opening on to the unknown. Here life is mingled with death, but simultaneously death is a sign of life, a way into the infinite."[137] This speculation sheds considerable light on our topic. The "brutal" sounds of death metal, in their embodiment of pain, chaos, and death, are actually invitations for the listener to confront, examine, accept, and overcome their own finiteness and isolated particularity. In the sacrificial context, death is the ultimate liberation from discontinuity, and the presence of this eternal wholeness inspires awe and ecstasy. As Bataille succinctly puts it: "consciousness of the void… throws us into exaltation."[138] The music also gives listeners a feeling of power in the face of the unknown, a dark force all the more mysterious in a society that routinely sanitizes, denies, and ignores its presence. Weinstein concludes: "Heavy metal's insistence on bringing chaos to awareness is a complex affirmation of power, of the power of the forces of disorder, of the power to confront those forces in the imagination, and of the power to transcend those forces in art."[139]

Chaos plays a major role in death metal, but not in the way critics might believe. In sum, the "evil" sounds discussed throughout this chapter—excessive volume, guitar distortion, the death growl, and other sonic icons of physical extremity—are calculated precisely *to be* ugly and repellant. Their very noisiness and association with corporeal pain and material limits (albeit largely illusory) makes them a fitting simulacrum for the victim of sacred violence unleashed by the ritual-concert. The brutality of these sounds is never allowed to truly run free; sonic violence is harnessed and purified for the benefit of the audience, who leave the experience refreshed, connected, and vivified. After analyzing the nature and symbolic power of the "evil" sounds of death metal, it is clear that timbre actively shapes, mediates, and orders the ritual dimensions of

---

[137] Ibid., 91.

[138] Ibid., 69.

[139] Weinstein, *Heavy Metal: A Cultural Sociology*, 38.

this oft-misunderstood genre. Furthermore, in making the dark and the dangerous "utterly, stunningly *present*," inviting listeners to find "stark, even visceral beauty" in terror, death metal timbres perform powerful psychological work, helping fans to cope with the aggression, negativity, and ominous unknowns in their lives.[140] Deleuze and Guattari claim that some music "necessarily gives us a taste for death… Music has a thirst for destruction, every kind of destruction, extinction, breakage, dislocation." They go on, however, to conclude, "music is never tragic, music is joy."[141] The "evil" timbres of this extreme genre indeed give us a "taste for death"; they allow us to suffer "mortification by proxy," to vicariously experience the pain and brutality they embody. But sacrifice undoes all of this. Death ultimately leads toward the illumination of its opposite.

---

[140] Chidester, "Through the Blood: Death Metal and the Rhetoric of Song as a Transcendent Discursive/Presentational Form," 343. Italics in original.

[141] Quoted in Bogue, "Violence in Three Shades of Metal: Death, Doom and Black," 113.

Epilogue

The 'other' need not be a person: it could be music.

— Thomas Clifton[1]

It is commonplace to the point of banality: you are scanning the dial on the car radio, quickly appraising your musical options.[2] Even before hearing the "tune," you already *know* what you're hearing—a twang, a growl; a choir, an accordion; a belter, a crooner—each betokening a different genre and a different target audience.[3] You continue scanning: not right, not me, not in the mood, not good; too harsh, too smooth, too old, too young. Goldilocks at 60 mph, you finally find—in a split-second—your "just right" (or at least acceptable). You begin to sing along.

Scanning the dial is an everyday exercise in *musical empathy*, one mediated largely by timbre. In this dissertation, we have examined the affective agency of timbre by exploring its connection to embodiment; we have traced its perceptual/cognitive arch through the ASPECS model; we have explored its role in two musical contexts. But we have not squarely addressed its social

---

[1] Clifton, *Music as Heard: A Study in Applied Phenomenology*, 285.

[2] This example is not original to me: I thank Anne Dhu McLucas for the memorable classroom illustration (at the University of Oregon in 2006), and Mitchell Morris for the same trope (in his keynote address for the 2011 Echo conference at UCLA).

[3] Listeners are generally quite adept at linking consistent, stereotyped impressions of musical genres to their associated social groups (e.g., the association of heavy metal with young, white, alienated, lower-middle class males). See Peter J. Rentfrow and Samuel D. Gosling, "The Do Re Mi's of Everyday Life: The Structure and Personality Correlates of Music Preferences," *Journal of Personality and Social Psychology* 84, no. 6 (2003); "The Content and Validity of Music-Genre Stereotypes among College Students," *Psychology of Music* 35, no. 2 (2007); Peter J. Rentfrow, Jennifer A. McDonald, and Julian A. Oldmeadow, "You Are What You Listen To: Young People's Stereotypes about Music Fans," *Group Processes & Intergroup Relations* 12, no. 3 (2009).

valence, particularly its link to empathy, despite the recurrence of this theme throughout the project. This is clearly too big a task for this short epilogue. By way of conclusion, then, I want to present some telegraphic thoughts and questions that will guide my research program in the next years. As this example suggests, musical ordering—based to a significant degree on rapid perceptual differences in timbre—is closely connected to social ordering. Understanding timbre's role in this dynamic is thus crucial, I maintain, for future research in the sociology, ethics, and social psychology of music.

According to philosopher Theodor Lipps, empathy (*Einfühlung*) is a form of radical self–other exchange, a way to "feel into" the other by psychologically projecting oneself into his or her body.[4] I use the term "musical empathy" in this sense: as embodied simulation of musical acts.[5] By this definition, this dissertation is very much about timbre-mediated empathy, about timbre's uniquely "subterranean" (Fales 2002) ability to facilitate empathic resonance, both with sound itself and with its associated bodies. This raises a number of questions: Can musical empathy—and more pointedly, motor empathy toward timbre-producing actions—exist independent from appraisal of specific human bodies?[6] How do we disentangle *timbre* from *timbred* (or *timbre-ing*) *people*?

Clifton, of course, is not the only writer to suggest that musical sound can work as an ersatz human actor. Marc Leman amplifies this point: "music can be conceived as a virtual social

---

[4] Theodor Lipps, *Ästhetik* (Berlin: B. G. Teubner, 1907).

[5] This usage of aesthetic empathy is discussed in relation to mirror neurons in David Freedberg and Vittorio Gallese, "Motion, Emotion and Empathy in Aesthetic Experience," *Trends in Cognitive Sciences* 11, no. 5 (2007).

[6] It is suggestive, as the mirror neuron literature indicates, that individuals who exhibit the most motor empathy with human-made action sounds are also the most *empathic*, as measured by scores on the Interpersonal Reactivity Index (IRI), a common psychometric for evaluating interpersonal intelligence and empathy. This result has been replicated in numerous studies: see Gazzola, Aziz-Zadeh, and Keysers, "Empathy and the Somatotopic Auditory Mirror System in Humans."; Jonas T. Kaplan and Marco Iacoboni, "Getting a Grip on Other Minds: Mirror Neurons, Intention Understanding, and Cognitive Empathy," *Social Neuroscience* 1, no. 3-4 (2006).

agent, and … listening to music can be seen as a socializing activity in the sense that it may train the listener's self in social attuning and empathic relationships."[7] As I hope to have shown throughout the dissertation, musicking cannot be severed perceptually from the musicking bodies behind it: music can function as a "virtual social agent"—and thus an object of empathy—precisely because of the way it sonifies the expressive bodily movements of human agents.

This observation is intuitive for a melody, perhaps less so for timbre. So let's sharpen our focus: What role does timbre *specifically* play in making empathic evaluations of music? Is the appraisal factor in timbre perception related to the appraisal of people?[8] To explore these questions, I ran one final experiment (see Appendix 4d): participants rated 400-millisecond "blips" of popular music recordings (same as discussed in the Introduction) according to two scales—how much do you identify with this *music*, and how much do you identify with the *people* who like this music?[9] Comparing results between the two conditions, I found that participants' appraisals of timbral snapshots were roughly synonymous with their social evaluations. This is probably unsurprising: liking a certain excerpt (or its representative genre) corresponds with a feeling of "belonging" to its associated community (hip-hop fans, metal-heads, punks, club kids, etc.). Suggestively, however, musical and social empathy appear to be largely exclusionary: participants identified with people who like the music they themselves like, but *not* with people who don't.[10] When we con-

---

[7] Leman, *Embodied Music Cognition and Mediation Technology*, 126.

[8] There is strong evidence in the affirmative. People form consistent snap judgments about others' personalities in as little time as it takes to hear a novel voice saying the word "hello," roughly the same duration (400 ms) as this study: see Phil McAleer, Alexander Todorov, and Pascal Belin, "How Do You Say 'Hello'? Personality Impressions from Brief Novel Voices," *PLOS One* (2014). I also might point to the uncanny resemblance of timbral appraisal to the "dual-process" theory of moral judgment developed by psychologist Joshua Green: see Joshua Green and J. Haidt, "How (and Where) Does Moral Judgment Work?," *Trends in Cognitive Sciences* 6, no. 12 (2002). Many thanks to Mark Johnson for alerting me to this connection.

[9] Unlike the experiment reported in Appendix 4c, here I used a Likert-modeled, 5-point ordinal scale (from "strongly do not identify" to "strongly identify," with a neutral "no opinion" in between).

[10] Positive correlation between musical identification and social identification reaches statistical significance when comparing like-like categories (e.g., asking subjects who identify strongly with EDM how they feel about EDM fans).

sider that a split-second is enough time to trigger this musical/social sorting, it is clear that differences in timbre are perceptually isomorphic to social difference. That twang or growl is enough to signal "me" or "not me."

I have argued in this dissertation that timbre is best conceptualized as a quality of bodily movement (a source's behavior) rather than an intrinsic quality (a source's identity). But of course, it is both-and, not either-or. Because all non-electronic timbres are materially determined, voices and instruments have timbres specific to them alone. Timbre is performative, as Eidsheim has suggested of the voice, but only to a limited extent: a cello cannot be made to sound like a trumpet; a clarinet can "sing" only metaphorically; an "old-sounding" voice cannot magically sound young again (without extensive surgical intervention). With the exception of electronic sound, timbre is immutably bound (and chained?) to the material; it reflects the presence of others by sonically manifesting their physical particularity. Like a fingerprint, this makes it a potent marker of individual and collective identity: our perceptual systems evolved to connect sounds with things (and people) as quickly and transparently as possible, and timbre does heavy lifting to enable this seamless equation of sound and source, though always—even if seemingly "natural"—within a specific cultural context. "Perceiving sound objects," writes Stephen Handel, "is much like perceiving faces."[11]

If timbre is an aural face, then how do we interpret situations were musical empathy is irrevocably thrown into question on the battlefield of timbral appraisal, cases such as free jazz

---

Correlation (positive or negative) between musical and social identification does *not* reach significance when comparing un-like categories (e.g., EDM fans' feelings about country listeners).

[11] Handel, "Timbre Perception and Auditory Object Identification," 458.

saxophonics and extreme heavy metal?[12] How do we make sense of "noisy timbre"? Anthropologist David Howes explains: "within each sensory field … sensations deemed relatively unpleasant or dangerous will be linked to 'unpleasant,' 'dangerous' social groups."[13] In certain conditions, negative timbral appraisals—like the contextual reading of a certain "bad sound" as "noise"—are not merely a matter of "personal distaste" but of "social ordering." Although noise can be viewed independently as entirely acoustic or entirely epistemological, in many instances of audition there is a good deal of slippage between the two. Acoustic qualities that register as "noise" can quickly shift from a purely auditory perception to judgments of the individuals or even the kind of people who produce such disagreeable sound.[14] Just as bodies are gendered, raced, and classed, they are timbred, and negative appraisals often serve as a referendum on the character, worth, sophistication, morality, civility, virtue, and intelligence of the sonic offenders, among many other qualities. Thus, as we have seen, dismissive critics heard an angry black man in the saxophonic scream, and pathological violence in extreme heavy metal, though it bears repeating that these appraisals reveal more about the *listener* (and his or her milieu) than the "sounds themselves" and the specific bodies they sonify. A breakdown of empathy for musical timbre—hearing some "humanly organized sound" as disorganized noise—can also be an ethical failure toward the people whose bodies stand behind it, a failure to deem meaningful the timbres

[12] I coin the expression "aural face" in reference to the ethical philosophy of Emmanuel Levinas. The classic account of ethical demand in the form of the Other's "face" can be found in Emmanuel Levinas, *Totality and Infinity: An Essay on Exteriority*, trans. Alphonso Lingis (Pittsburgh: Duquesne University Press, 1969).

[13] David Howes, ed. *Empire of the Senses: The Sensual Culture Reader* (Oxford and New York: Berg, 2005), 10.

[14] A chilling example of this can be found in a common psychological test for evaluating the possibility of mental illness in young children: *not* showing an aversive reaction to white noise as a toddler has been found to correlate with psychopathy as an adult. See Jennifer Kahn, "Can You Call a 9-Year-Old a Psychopath?," *New York Times Magazine*, 11 May, 2012.

of the Other. "Aesthetic judgments," notes Simon Frith, "are tangled up with ethical judgments."[15]

In the final analysis, of course, the problem of timbre remains, obstinate and mercurial, oblivious to the epistemic and discursive thresholds it ruptures. I wrap up this dissertation, therefore, with a feeling less of consolidation than splintering, less of conclusion than commencement. It will take us a while yet to sort through all the "problems" of timbre, both empirically and theoretically. In the meantime, scan that radio dial. You might like the faces you meet.

[15] Frith, *Performing Rites: On the Value of Popular Music*, 70.

Glossary

Introduction: Statistical Significance *vs*. Musicological Relevance

Before humanities scholars crack the Appendices, a brief word of clarification is in order, both about the specific statistical concepts employed therein and more generally about the epistemological questions such quantitative techniques might raise within the context of a humanistic project like this one. We should begin by acknowledging one important fact: empirical research rests on statistical verification and the assumptions and conventions underlying it. This is one of the major philosophical and methodological divides separating the sciences from the humanities; the researcher in empirical fields must subject his or her data to a gauntlet of statistical procedures in order to make sure there is a signal in the noise, to search for perhaps invisible relationships between variables, and to ensure the results amount to a reasonably high probability of accuracy. I hesitate from saying "high probability of being *true*," since this involves a much heavier philosophical burden of proof than I think most scientists are prepared to shoulder when advancing their conclusions. As Huron memorably put it, "science probes, it doesn't prove." An empirical knowledge claim, then, is not a statement of ultimate Truth, even if backed up by a ream of data to "prove" it, but rather a preliminary explanation for a given phenomenon always subject to revision or, if supplanted by an even better model, wholesale rejection. Statistical inference plays a central role in defining the terms of such claims.

Numerous statistical procedures dot the Appendices, but two in particular are most important to grasp. These two concepts help drive the analyses presented here in ways that, I hope, augment and enrich the narrative account of Chapter 2.[1] First, as a measure of correlation between two variables, I use standard *Pearson's product-moment correlation* coefficients, notated by the letter "*r*." Correlation, we should remind ourselves, is not *causation*, but it often does reflect a linear relationship between phenomena, and *r*-values are a way to quantify the strength of this relationship. The scale for this statistic is intuitive: an *r*-value of −1 is a perfect *negative* correlation, meaning that the two variables under observation move in exact inverse proportions, while an *r*-value of +1 is a perfect *positive* correlation. A random data set where the variables bear no relation to each other will yield a value of 0. The significance of a correlation coefficient will vary depending on sample size, context, and convention, but generally speaking, a correlation of, for example, $r = .86$ would be considered strong, while one of $r = -.13$ would be weak, though still loosely correlated.

It is logical enough that correlations increase in strength approaching an *r*-value of negative or positive 1. But how do we determine if the configuration of data showing a correlation is not the result of pure, dumb chance, even if the *r*-value is high in absolute terms? (Or conversely, how do we know that a correlation with a low *r*-value *does not* reflect a systematic, linear relationship between variables, even though the relationship might not be a strong one?) To answer this, researchers commonly employ a statistical method called Null Hypothesis Significance Testing (NHST).[2] The test—a ubiquitous measure of "statistical significance" in virtually all empirical

---

[1] A great primer to experimental design and data analysis in music research can be found in W. Luke Windsor, "Data Collection, Experimental Design, and Statistics in Music Research," in *Empirical Musicology: Aims, Methods, Prospects*, ed. Eric F. Clarke and Nicholas Cook (Oxford: Oxford University Press, 2004). For more information on statistical procedures, see the helpful resources Steve Miller, *Experimental Design and Statistics* (London: Routledge, 1974); Colin Robson, *Experiment, Design and Statistics in Psychology* (Harmondsworth, U.K.: Penguin Education, 1973); John L. Phillips, *How to Think About Statistics*, 6 ed. (New York: Henry Holt, 1999).

[2] It should be noted that the use of *p*-values to determine "significance" of a Pearson's *r* is a somewhat controversial practice. NHST is premised upon certain assumptions regarding a parametric distribution of data; however, correlations make no such assumption. Thus, *r*-values cannot truly be given an accurate *p*-value since the distribution may not necessarily conform to the assumptions of parametric statistics. In the proceeding Appendices I *will* be giving

fields—asks this: What is the probability that a pattern in your data is really just the result of random chance? To do this, NHST compares the "null hypothesis"—that your result is the work of chance—to your actual data in order to produce a probability, or *p*-value, indicating the likelihood of a false-positive result. Put differently, the *p*-value gives your odds of being *wrong*: the lower the value, the better the probability that the data reflect a "real" pattern (the "alternative hypothesis," in NHST terminology). The convention in many scientific fields is to set the threshold of statistical significance at $p<.05$, meaning that there is a less than five percent chance of a false-positive conclusion. In this instance, the null hypothesis should be rejected, and the alternative hypothesis retained. When the probability of error exceeds the five percent mark, however, the risk of a false-positive mounts; the prudent course of action in such a circumstance is to reject the alternative hypothesis. Your results, in this case, are null: there is too great a chance that they do not reflect anything other than chance.

NHST is not universally accepted among scientists, nor does it claim that *p*-values represent the chance that the alternative hypothesis is *not true*; but whether $p<.05$, $p<.01$, or whatever the level specifying statistical significance, NHST does set a binary threshold, one that is entirely arbitrary and deemed significant through consensus among the scientific community, not through any innate quality. So what does this mean for the study of music? Huron rightly points out that what constitutes significance for one field, like cognitive psychology, might not be appropriate to a field like musicology, which is both more data-poor and "low-stakes" (in terms of generating hypotheses with relatively minor consequences of false-positive claims).[3] Should we toss out results that fail to meet the *p*-threshold as "insignificant" simply because there is a greater

significance levels for *r*-values, but I do so advisedly—for better or worse, such a practice has become a convention. I thank Roger Kendall for pointing out this important distinction.

[3] Huron, "The New Empiricism: Systematic Musicology in a Postmodern Age."

than five percent chance that the results are due to chance? Surely something looser and more contingent would suit musicological purposes better than a hidebound threshold for statistical significance. The question we should be asking, as far as quantification of musical experience is concerned, is what constitutes *musicological relevance*. This is not to abandon the *p*-value entirely, as NHST constitutes an accepted method for verifying results (not to mention a near-universally acknowledged marker of meaning in empirical discourses). Rather, it requires some flexibility in the context of musicological argumentation. Since we are not generally in the business of "proving" a given hypothesis anyway, and since the stakes of musicological claims rarely hold lives and major resources in the lurch, an empirically-informed musicology would do best to treat statistical significance less as a binary threshold and more as a sliding scale of possible interpretations. If $p<0.0001$, then a very strong case could be made that the results are relevant (a case, incidentally, that would be intelligible to empiricists as well); but $p<.10$ or even $p<.25$ could still be quite relevant to the sorts of questions we ask. After all, a one in ten chance of being *wrong*, some wag might argue, is still a nine in ten chance of being *right.* That ain't bad odds.

Flippancy aside, we should acknowledge the very real epistemological boogieman lurking these pages, and stop now to look under the bed and point it out. Empirical methods, of which NHST forms an essential part, proffers a materialist, quantifiable view of reality, one that—to some observers—might appear incommensurable with the interpretative methods and theoretical frameworks that form the basis of the contemporary Anglo-American humanities, musicology included, even in the era of "new materialism" scholarship. The thought that one might be able to verify to *x* degree of probability that, for instance, Schubert was gay—a controversial "new musicological" assertion from the 1990s—is a patent absurdity, since this constitutes a qualitative, interpretive claim drawn from historical sources and musical documents, not an empirical claim subject to experimental manipulation, falsifiability, statistical analysis, and so on. This is undoubtedly true for many of the questions asked under the aegis of musicology: trying to determine *p*-values for musicological knowledge, even with some highly inventive data coding, would in many (most?) cases constitute a simple category error.

In the Appendices, I employ *p*-values (and a traditional threshold of .05) to drive home the statistical power of certain results, but I also want to make it clear that results lacking $p<.05$ validation are not, conversely, necessarily *insignificant* to the sorts of interpretations and analyses advanced throughout the rest of the dissertation. Therefore, at certain—though rare—junctures in the analysis, I will point out results that I think constitute "musicological relevance," even when they do not strictly meet this threshold, by setting a parallel line of "relevance" at $p<.10$. Although just as arbitrary as .05 and probably not publishable in any major scientific journal, this line of significance—a one in ten chance of a false-positive result—is suitable to certain musicological purposes and, I think, worthy of reporting rather than rejecting offhand as "insignificant."

In addition to the inclusion of these two statistical operations, one brief note regarding scientific prose is in order. In the Appendices, you will notice my adoption of certain passive constructions that vary in tone from the body text of the dissertation. This is a writerly convention of the sciences, and I use it here—advisedly—for a specific reason. Although it might just look like bad writing to some in the humanitities (and indeed, in some case *it is* just bad writing), science writing is passive for two reasons: the first, in removing the agency of the researcher, is a strategy to appear objective (e.g., "it was observed that variable x…"). I do not abide by this first usage. But the second reason is to hedge claims, adding a necessary level of uncertainty to the statement of conclusions (e.g., "this appears to indicate…," "this might suggest…," etc.). Considering that most empirical claims are based on NHST, there is indeed *always* some probability that a result is pure chance. Passive hedging, in this instance, underscores this uncertainty.

Neuroimaging Primer

Before turning to the images in Appendix 3 (and Chapters 2–4), a brief primer in how to read neuroimaging data is in order. Imaging is mediated by a set of conventions and statistical assumptions, not to mention aesthetics; understanding in broad strokes how the process works is thus crucial to making sense of the results. Without getting too technical, fMRI works by dividing the brain into very small cubes, or *voxels*, then measuring hemodynamic blood-flow change (BOLD signal change) within each voxel.[4] Once all of the individual subject data is collected, averages are taken for the group, corresponding to the mean activation of voxel *x* in response to experimental variable *y*. The statistical principles introduced above—including NHST and correlation—apply here as well; rather than listing *p*- or *r*-values in huge tables or graphs (as the brain consists of thousands of individual voxels), however, neuroimaging studies tend to display this information in visual form, i.e., as images of the brain with colorful clusters of activated voxels representing activity. (Keep in mind that these brain-shaped graphs are group-level abstractions representing different levels of statistical significance, not anatomical pictures of any one individual's brain.) The brighter the color, from dark red to light yellow, the lower the *p*-value of the associated cluster of voxels, meaning the higher the probability that activity in that specific region is related to the stimuli, not to chance. This procedure makes imaging highly sensitive to the *threshold* at which the researcher sets the level of significance (at the level of both individual voxels and clusters): for example, at $p<.01$, an image might return no results at all, but if the threshold is lowered to $p<.02$, it might light up with patterns of significant activation. All of this is to say that threshold levels matter profoundly, both for the researcher analyzing the data and for the reader

[4] For more details, see Scott A. Huettel, Allen W. Song, and Gregory McCarthy, *Functional Magnetic Resonance Imaging*, 2 ed. (Sunderland, MA: Sinauer, 2008).

interpreting the results. In this experiment we used both Z=2.3 ($p<.01$) and Z=1.7 ($p<.05$) thresholds, depending on whether or not the higher threshold produced any results. While both levels reach "statistical significance" as defined earlier, the 4% difference in the probability of a false-positive error is not trivial and should be kept in mind when reviewing the images.

Brain images are usually presented in the form of *contrasts*; an image shows clusters of voxels that are activate in one experimental condition, but not in another. Contrasts are denoted with a "greater-than" sign (>) to show the direction of a particular relationship: for example, Gtr. 3 > Gtr. 1 would show areas that are active when subjects hear Gtr. 3 that are *not* active when they hear Gtr. 1. And since the brain is a three-dimensional object, moreover, contrasts typically appear in three views: the *sagittal* plane visualizes the brain from the side (*x* axis); the *coronal* plane looks at it front-to-back (*y* axis); and the *axial* plane looks at it from top-to-bottom (*z* axis). In Experiment 3 [Appendix 3], we created the following contrast maps:

(1) *Each of the 12 timbres > baseline (silence)*. Baseline contrasts are the most simple form of measuring brain activation, since you are only seeing an on-off, binary relationship (in this case, the difference between hearing each timbre and hearing silence). Baseline contrasts can tell us what regions are active during a particular stimulus, but not how this activation compares to other stimuli.

(2) *Inter-instrument contrasts* [Gtr 3>1, Gtr 2>1, Gtr 3>1, etc.]. To compare two different conditions, we created contrasts between the three "noisiness" levels of each sound generator.

(3) *Inter-timbre contrasts* [all condition 3 > all condition 1, etc.]. Collapsing sound generator categories allowed us to compare levels of "noisiness" irrespective of the specific instrument.

(4) *All timbres (combined) > baseline*. This contrast shows areas of the brain that are active when listening to *all* of the stimuli in sum.

To summarize, brain images are statistical maps showing degrees of probability that the BOLD signal change in a specific voxel or cluster of voxels is the result of experimental manipulations and not chance. But it is also possible to extract numerical values for BOLD signals directly. Beyond images proper, Experiment 3 uses two different species of correlational analyses: *single-voxel* and *voxel-by-voxel* correlations. In the first approach, we extracted data (*beta estimates*) from individual voxels with the highest local values per cluster, then correlated these numbers with other data (e.g., activity in other regions, behavioral data, etc.). In the second approach, we added behavioral data into the computer model to perform a voxel-by-voxel correlation at the level of the whole brain through a statistical procedure called *multivariate regression*. For more details, see Appendix 3.

Glossary of Empirical Concepts

Analysis of Variance (ANOVA): A statistical test that evaluates how much of the variance within subject responses is *systematic*, i.e., changes resulting from an experimental manipulation, *vs.* how much is *unsystematic*, i.e., changes resulting from random fluctuation. The test produces a ratio between these two values labeled F (see *F-ratio* below). Besides number of subjects (N), degrees of freedom and number of variables are contributing factors in ANOVA. A *repeated measures* ANOVA measures variance in data for the same subject group under different conditions.

Correlation: (or "Pearson's product-moment coefficient") A measurement of the linear relationship between two variables or sets of data, most commonly indicated with the statistic r (see R below).

Covariance: A measurement of how much two random variables change together.

Cronbach's α: A measurement of the internal consistency of a data set (given as a coefficient) commonly used to evaluate the reliability of psychometric tests. Higher values (approaching 1) indicate greater consistency of responses.

***F*-ratio**: A common statistical ratio between systematic and unsystematic variance (see *ANOVA* above). The higher the *F*-ratio, the more of the total variation can be accounted for by the experimental manipulation. All *F*-ratios have a corresponding *p*-value.

Greenhouse-Geisser correction: Corrects data against violations of sphericity (see *sphericity*).

M: (or "mean") A group average.

Multicollinearity: A condition in multiple regression analysis in which two or more independent variables are so closely correlated as to undermine the regression.

Multiple Regression: A statistical approach for determining the relationship between a dependent variable and multiple independent variables. The R^2 value shows how much of a given effect can be explained by one independent variable (compared to the others), allowing the researcher to determine relative weightings for multiple variables.

N: The total number of subjects or data points within a given sample.

Null Hypothesis Significance Testing (NHST): A statistical way of testing whether a result is attributable to an independent variable or is likely the result of chance, given in a *p*-value (see *P-value* below).

SD: (or "standard deviation") A statistic showing variation from the mean. Low values indicate data points tend to be close to the mean; high values indicate greater spread of data point values from the mean.

Sphericity: A statistical assumption for repeated measures ANOVA often determined with *Mauchly's test*. Sphericity is indicated with the Greek letter ε.

***P*-value**: A statistic giving the probability of achieving a certain result assuming that the null hypothesis is true (i.e., the probability of a false-positive result). The conventional threshold for *p*-values is under 5% (or $p<.05$).

Post hoc testing (Tukey's HSD): Methods for analyzing data for patterns after primary analysis (such as ANOVA) is complete. Tukey's HSD (honest significant difference) is a common post hoc test for differences between means typically used in conjunction with ANOVA.

R: Standard correlation coefficient ranging from −1 (perfect negative correlation) to +1 (perfect positive correlation), with $r=0$ indicating no linear relationship between two sets of data. Assuming normal distribution, *p*-values can be calculated from *r*.

T-test: Test comparing means for two groups to detect whether there is a statistically significant difference between them. The test statistic for this procedure is labeled *T*.

Wilcoxon sign-rank test: Similar to T-test, but for non-parametric statistics (i.e., no assumption of normal distribution is required). Compares means for two groups to detect significant difference.

Z-score: Standard score measuring relationship to the group mean in a normally distributed dataset. Z-scores have a corresponding *p*-value (the higher the Z, the lower the *p*).

Appendix 1

Experiment 1: Perceptions of Embodiment and Affect in Isolated Timbres

N.B.: A full discussion of the context, central questions, and key conclusions of all experiments appear in the main text. This appendix is intended to serve as a parallel text to the analysis presented in the main body of the chapter, providing the experimental procedures and data analysis that drive the larger interpretive claims of the dissertation. For this reason, I will proceed in the appendices straight to explanations of method, experimental procedure, and results without the usual extensive introduction, discussion, and conclusions common to empirical publications. Certain explanations common to more than one appendix and to the Chapter 2 body text are replicated so that each appendix can stand alone.

Hypotheses

An increase in the tension and amplitude of human and non-human mammal vocal production generates an increase in high frequency energy, inharmonicity, roughness, and other "nonlinearities" in the acoustic output (Blumstein, Bryant & Kaye, 2012; Hauser, 2000; Fitch, Neubauer & Herzel, 2002; Tsai et al., 2010). Some maintain that the expressiveness of musical timbre is related to iconic similarities to vocal expression (Spencer, 1857/1951; Juslin and Laukka, 2003; Eidsheim, 2008). It is hypothesized, then, that these particular timbral qualities, whether vocal or instrumental, may convey a sense of both heightened bodily exertion and negative affect based in evolutionary psychology. It is further hypothesized here that the perception of "noisiness" (or, *p-noise*) in musical signals is mediated by the relationship between the acoustic qualities above (or, *a-noise*) and their psychobiological significance. To explore these hypotheses, short recordings of isolated instrument and vocal timbres with varying acoustic profiles were presented to experimental subjects for a variety of ratings tasks.

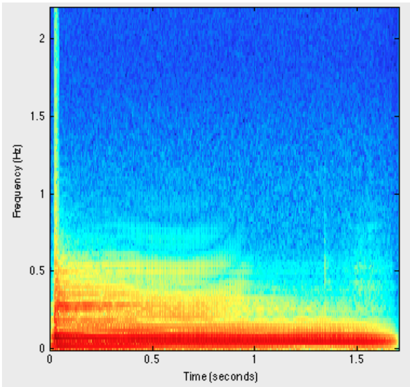
Stimuli

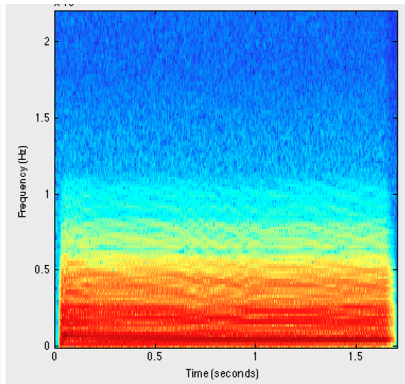
Twelve signals of instrument and vocal timbres were created: 3 electric guitar, 3 tenor saxophone, 3 *shakuhachi* flute, and 3 female vocals. Within each category of sound producer, the three different signals were scaled according to increasing levels of acoustic noise (i.e., a-noise) defined with the musicians' assistance according to the specificities of each source: (1) the low-noise condition, (2) medium-noise condition, and (3) high-noise condition. For the electric guitar, condition 1 corresponded to a "clean" tone (as in jazz), condition 2 to moderate distortion (as in rock and Chicago blues), and condition 3 to heavy distortion (as in heavy metal and punk); for the saxophone, condition 1 corresponded to a regular tone (as in traditional jazz), 2 to slight growl (as in R&B and rock), and 3 to a harsh growl (as in free jazz); for the shakuhachi, condition 1 corresponded to a regular tone (*zennon*), 2 to slight air turbulence (*sasane*), and 3 to heavy air turbulence (*muraiki*); and for the voice, condition 1 corresponded to a sung vowel "open-O" /ɔ/ (no vibrato), 2 to a sung open back unrounded vowel /a/ with additional breathiness, and 3 to a sung close front unrounded vowel /i/ with throat tension causing additional raspiness. The instruments and specific techniques selected reflect the primary timbral features discussed in the dissertation. They were also selected to represent a range of different material means of production, including electro-acoustic (electric guitar), reed (saxophone), air-stream (shakuhachi), and vocal production.

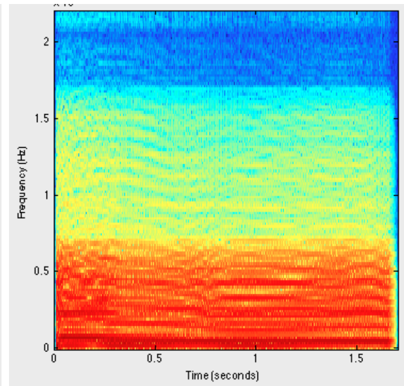
The pitch for all signals was B-flat3 (233 Hz). Signals were recorded specifically for this experiment using Apple Logic software, a Neumann TLM-103 microphone, and a MOTU 896 interface at a sampling rate of 44.1 kHz. Signals were 2 seconds in length, with attack and cutoff unaltered for maximum ecological validity. Additional signal-specific details are as follows:

> Electric Guitar: Guitar signals were recorded with a 2000 Fender Deluxe Stratocaster (made in Mexico), performed by the author. A Dunlop medium thickness pick and D'Addario nickel-plated steel string (.10) were used. The amplifier was a Vox AC30C2 30-watt 2x12 combo, with all potentiometers set to "5" for balanced equalization. No additional modifications were made to the amplifier. This clean signal was edited in Apple GarageBand for the addition of digital distortion. GarageBand guitar presets were selected due to the program's ubiquity and easy reproducibility. For condition 1, the preset "Classic Blues" was selected; condition 2 used "Seattle Sound"; and condition 3 used "Heavy Riffing."
>
> Tenor Saxophone: Saxophone signals were recorded with a Selmer Super Balance Action tenor saxophone manufactured in 1951, performed by a professional jazz saxophonist. A Vandoren V16 ebonite mouthpiece and Vandoren ZZ #3 reed were used. For condition 2, a moderate growl was produced by performing a vocal growl in the back of the throat, similar to a hard "R" in French. For condition 3, an intense growl was produced by both vocally growling and humming an unspecified pitch while blowing.
>
> Shakuhachi: Shakuhachi signals were recorded with a 1.8 *ji-nashi* shakuhachi flute made by John Kaizan Neptune in Kyoto in the late 1990s or early 2000s, performed by the author. Condition 1 was a regular, unadorned *zennon* tone on the note *RO* (D4). In condition 2, the tongue partially perturbed the airstream to produce additional breathiness (*sasane* technique, or "bamboo grass blowing"). And in condition 3, the tongue fully perturbed the airstream to produce a pitch-occluding admixture of noise (*muraiki* technique, or "thrashing breath"). To conform the signals to the same frequency as the other instruments, the Audacity "Change Pitch" function was used to lower the signals from D4 to B-flat3.
>
> Vocals: Vocal signals were recorded with a female singer with formal musical training and professional performance background in opera and avant-garde concert music.

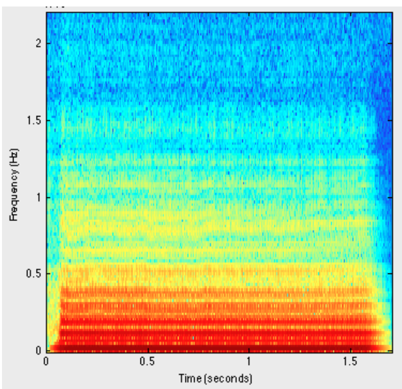
In addition, two "test" signals were included to demarcate the outer boundaries of "noisiness": a 2 s clip of pink noise generated with Audacity software, including 10 ms amplitude ramps, for the exemplar of "noise" (i.e., high spectral centroid, entirely inharmonic, high auditory roughness, etc.); and a 2 s sine wave generated with Audacity software, also including 10 ms amplitude ramps, for the exemplar of the opposite of noise (i.e., low spectral centroid, no inharmonicity, no roughness, etc.). These two signals were excluded from the ANOVA test detailed later due to asymmetry of study design. Loudness was equalized manually for all signals.

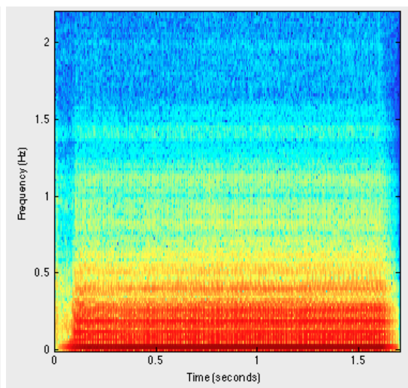


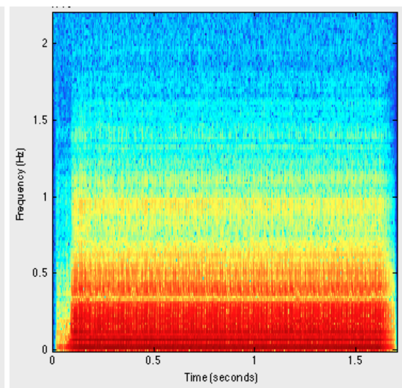




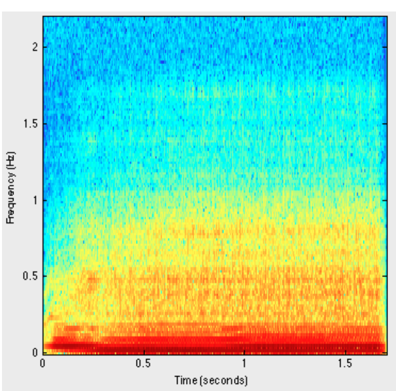
Electric Guitar: 1 2 3

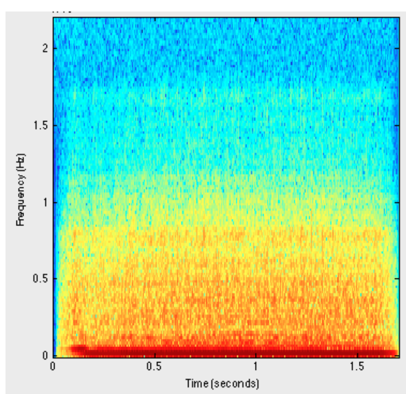


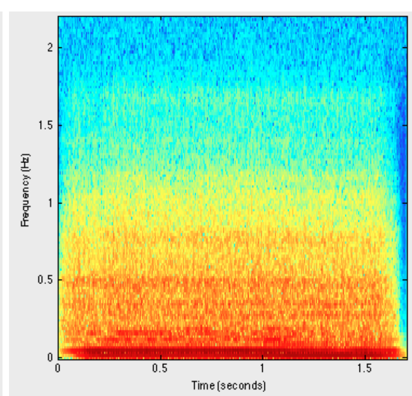




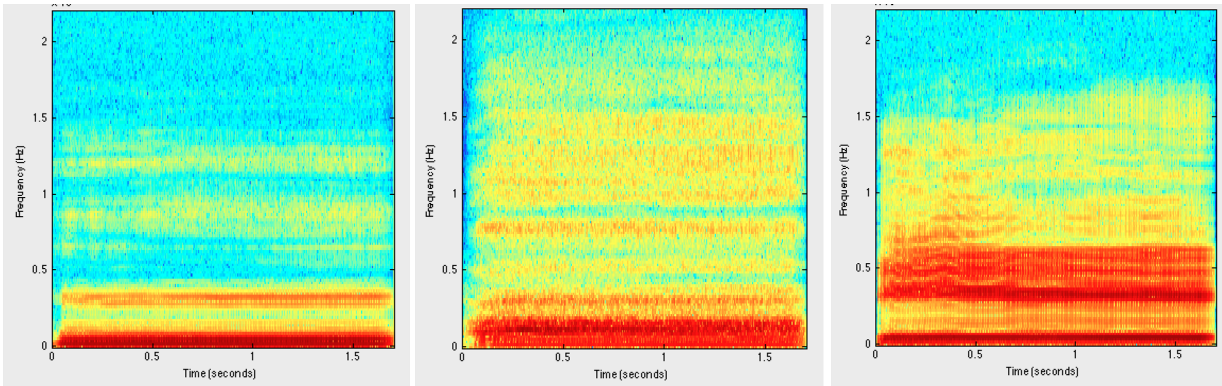
Saxophone: 1 2 3



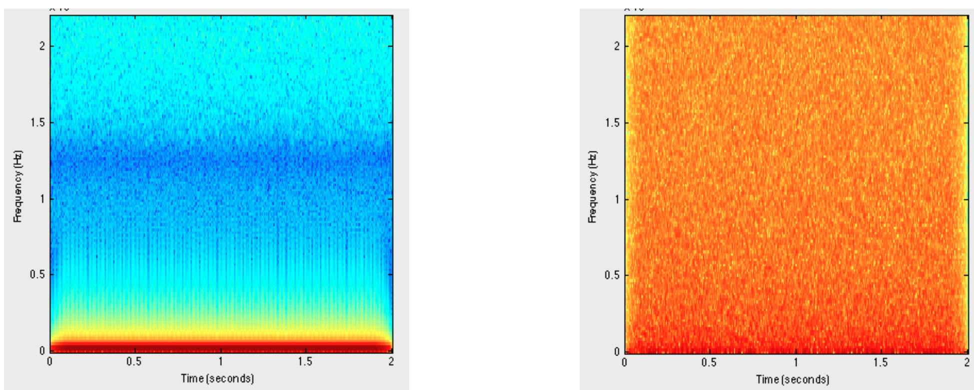




Shakuhachi: 1 2 3



Voice: 1 2 3



Test Signals: Sine wave (233 Hz) White Noise

FIGURE A1.1: Spectrographs of experimental stimuli

For the sake of comparison, Figure A1.1 shows spectrographs of all 14 signals. In all instrument families, Timbre 1 generally contains more of its energy in the fundamental and first few partials (this is especially true for saxophone and voice). In Timbre 2 and 3, however, two general changes can be observed: first, more of the total energy creeps into the higher registers, indicating an increase in spectral centroid. The guitar signals illustrate this most starkly, but the pattern obtains for all signals, even when not made transparent in spectrographic form. And second, the bands of energy representing harmonic partials thicken and grow ever fuzzier in Timbres 2 and 3, perhaps best illustrated by the saxophone. Here, what were discrete partials in

Timbre 1 morph into a turgid band of energy in Timbre 3. In contrast to the twelve natural timbres, the two test signals show the artificial, polar extremes of the a-noise scale. A sine wave contains all of its energy in the fundamental (at least in theory), and thus has the lowest spectral centroid possible and no inharmonicity (and no harmonicity, in fairness, since there is no energy in the harmonic partials either). Such a sound is exceedingly rare in nature. At the other end of the scale, white noise—or more precisely, *pink* noise—spreads its energy across the entire frequency range in random oscillations (Kosko, 2006). Its spectral centroid is high, and the spectrum is almost perfectly flat.

Subjects

Thirty-six participants were recruited from the UCLA community. The study group consisted of 22 females and 14 males between the ages of 18 and 31 (age $M = 20.3$, $SD = 2.07$). The subjects self-reported their number of years of formal musical training ($M = 6.83$ years, $SD = 5.93$); all subjects were non-music majors with musical backgrounds ranging from no formal experience (7 subjects) to 20 years (1 subject). Every subject who participated in this experiment also participated in a related experiment (Experiment 2). The order of the two experiments was randomized. Each subject was paid $20 in cash after completion of both experiments.

Procedure

Experimental control was achieved through Music Experiment Development System (MEDS 2002-B-1) software (Kendall, 1990-2002). Subjects listened to the mono signals through high-quality headphones at a comfortable listening level, approximately 65 dB SPL, in a quiet room with minimal visual distractions.

The experiment consisted of four separate ratings and evaluation tasks that were presented in a predetermined order:

(1) First, as a measurement of the "embodiment" dimension of the signals, subjects were asked to rate the perceived *bodily exertion* required to produce each timbre with the use of a numbered horizontal rating scale (0-100) with bipolar labels (*low exertion-high exertion*) consistent with the semantic differential paradigm (Osgood, Suci, & Tannenbaum, 1957). In order to familiarize themselves with the software, the experiment began with a "practice" run, with subjects receiving the following instruction:

> This experiment investigates the perception of energy and emotion in the timbre (or tone color) of different musical sounds. It is not a test of musical aptitude or ability.
>
> You will begin by hearing a short series of vocal and instrumental tones. Your task will be to rate the level of bodily exertion implied by these sounds on a scale of `low` to `high.` After you hear each tone, drag the cursor to the appropriate location on the scale and press OK. There will be 14 tones total. Please try to use the entire scale in your judgments. This first run is practice only to familiarize yourself with the range of sounds and the rating scale. Press OK to begin.

Presentation of the 14 individual signals was randomized. Following this practice run, subjects were given the following prompt:

> We are now ready for the main experiment, which is the same as the trial run you just completed. You will be asked to rate each tone on a scale of `low exertion` to `high exertion,` corresponding to the level of physical exertion implied by each sound. There will be 14 tones total. Please try to use the entire scale in your judgments. Press OK to begin.

Subjects were allowed to ask questions throughout the procedure. As before, presentation of the signals was randomized. Listeners were allowed one audition of each signal, and were required to press the "OK" button on the screen before going on to the next signal.

(2) Next, subjects were asked to evaluate the *affective valence* of each signal on a numbered horizontal rating scale (0-100) with bipolar labels (*dislike-like*). *Affective valence* has been shown to equate to musical *preference* (Eerola, Ferrer, & Alluri, 2012). The procedure was the same as that outlined above, but without a practice run. Subjects were prompted with the following instruction:

> Next, you're going to tell us how much you like or dislike each of these tones. Please rate the same sounds on a scale of `dislike` to `like.` Click OK to begin.

(3) Next, subjects were asked to categorize the perceived emotional content of each signal by selecting an adjective from a list that best describes the "*emotion conveyed*" by each sound. The list consisted of the five primary emotions determined by Juslin (2001) to be successfully communicable by performers, and used by Krumhansl (2010) in a similar experiment: (1) Happiness, (2) Sadness, (3) Anger, (4) Fear, and (5) Tenderness. Subjects could only select one word from the list. The following text was presented to each subject:

> Next, we'd like to know how you would categorize the emotion conveyed by these tones. Please choose one of the following five emotions that best characterizes each sound: happiness, sadness, anger, fear, tenderness. Click OK to begin.

(4) Finally, subjects were asked to rate the "brightness" and "noisiness" of the signals using a VAME-modified semantic differential paradigm (Kendall & Carterette, 1993). A numbered horizontal scale (0-100) with bipolar labels was used (*not bright-bright, not noisy-noisy*), and subjects were asked to make both evaluations after a single audition of each signal. The prompt text read:

> Last, we'd like to know how you would rate the tones in terms of two qualities: brightness and noisiness. After hearing each sound, please rate it on a scale of `not bright` to `bright,` and `not noisy` to `noisy.` Click OK to begin.

The duration of the full experiment was approximately 5-10 minutes.

Analysis and Results

Acoustic Data Analysis

Since the analysis of timbre acoustics can only be as good as the researcher's acoustic measurements, the specific details of extraction are not of trivial importance. For this reason, spectral components of the signals were computationally extracted using a range of software applications, including MIRtoolbox1.4 (Lartillot & Toiviainen, 2007), PsySound3 (Cabrera, 1999), and Timbre Toolbox (Peeters, Giordano, Susini, Misdariis, & McAdams, 2011), all in the MATLAB environment. MEDS was used for additional spectral centroid calculations. The aim of extracting the same or similar acoustic descriptors from a variety of different software was to enable the comparison of acoustic data to perceptual ratings, a task that has not, to my knowledge, been systematically carried out. Data that did not correlate significantly to any of the perceptual ratings were discarded in favor of extraction tools that related in a meaningful way to the behavioral data. (For example, the roughness and spectral centroid calculations in PsySound3 and spectral features in Timbre Toolbox did not seem to bear any relevance to the perceptual data; they have therefore been omitted from this analysis.) In an attempt to isolate spectral characteristics associated with the behavioral variables measured, the analysis considered only spectral and not temporal or spectrotemporal timbral features.

Measurements of all acoustic parameters using MIRtoolbox were derived from a 12-order FFT with 50% overlap (50ms per frame), producing approximately 40–45 frames for each signal at a sampling rate of 44.1 kHz. The resulting calculations were based on an average across all frames. Similar extraction parameters were employed in MEDS, with the addition of a Bartlett window. Transients were cut (above a "unit-less centroid" value of 20), and the resulting spectral centroid calculations were an average across all remaining frames.

First, correlations between timbral qualities were assessed (see Table 1.1; Pearson's r entries in bold indicate statistical significance at $p<.05$). Five acoustic dimensions of interest were generated based on high inter-acoustic r-values, including *spectral centroid* (SC), *inharmonicity* (I), *spectral flatness* (SF), *zero-cross rate* (ZC), and *roughness* (R):

(1) *Spectral centroid* (SC), or the fulcrum point of energy distribution across a spectrum, commonly experienced as "brightness," "nasality," or "sharpness." Spectral centroid has been shown to be perhaps the most perceptually salient feature of timbre (see sources listed in Ch. 1); it has also been implicated in the conveyance of escalating musical tension (Huron, 2006). The importance of this particular quality impelled me to measure it with two separate approaches: MIRtoolbox, which uses a cut-off threshold of 1500 Hz to measure the percentage of energy above this threshold; and MEDS, which captures the full average of the spectrum without any thresholding. Additionally, MEDS allows for a "unit-less" centroid measurement by dividing the spectral centroid by the fundamental frequency (Kendall & Carterette, 1996). This long-time average centroid value can be acquired through the following equations (EQ 1.1 and 1.2, respectively), where t is time, n is partial index, N is the total number of partials, f_n is the frequency of partial n, a_n is the amplitude of partial n, E is the total energy of the spectrum, and $F0$ is the fundamental frequency:

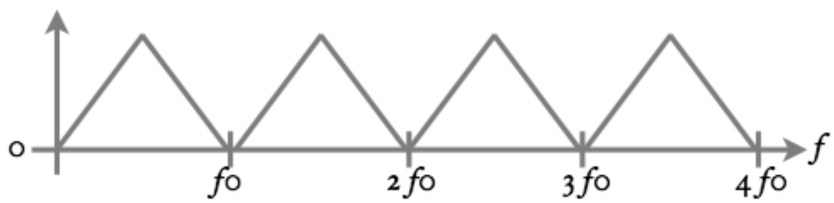
$$centroidMIR(t) = \frac{\left(\frac{\sum_{n=1}^{N} f_n a_n}{\sum_{n=1}^{N} a_n} \right) \geq 1.5kHz}{E} \quad \text{(EQ 1.1)}$$

$$centroidMEDS(t) = \frac{\left(\frac{\sum_{n=1}^{N} f_n a_n}{\sum_{n=1}^{N} a_n} \right)}{F0} \quad (1.2)$$

(2) *Inharmonicity* (I), or energy outside the harmonic series of a given fundamental, is often said to be experienced as "noisiness" (Peeters et al., 2011). Such a "noisiness" measure (as extracted us-

ing MIRtoolbox, as with all remaining timbral variables) consists of a ratio of the sum of all energy contained in the ideal harmonic series over the total energy of the spectrum, where *a* is the amplitude, *f0* is the fundamental, *2f0…* are harmonic partials, and *E* is the total energy, as in (EQ 2.1) with triangular weighting (2.2):[1]

$$inharmonicity = \frac{\sum\left(a_{f0}, a_{2f0}, a_{3f0,}...\right)}{E} \quad \text{(EQ 2.1)}$$



(2.2)

(3) *Spectral flatness* (SF), another common measurement of perceived "noisiness," indicates the relative smoothness or spikiness of a signal. It is Wiener entropy of the spectrum given as a simple ratio between geometric and arithmetic means (EQ 3), where *x(n)* represents the amplitude of frequency domain of frame number *n*:

$$Flatness = \frac{\sqrt[N]{\prod_{n=0}^{N-1} x(n)}}{\left(\frac{\sum_{n=0}^{N-1} x(n)}{N}\right)}$$

(EQ 3)

(4) *Zero-cross rate* (ZC), another simple indicator of noisiness, counts the number of times the signal crosses the *x*-axis per time unit. It is given as follows (EQ 4), where *sign* function is 1 for positive arguments and 0 for negative arguments and *x(n)* is the time-domain signal for frame *t*:

---

[1] My gratitude to Olivier Lartillot, the programmer of MIRtoolbox, for assistance with equations, and for the graphic.

$$Z_t = \frac{1}{2}\sum_{n=1}^{N}\left|sign(x[n]) - sign(x[n-1])\right| \qquad \text{(EQ 4)}$$

(5) *Roughness* (R), or "sensory dissonance," results from amplitude fluctuation (beating) when a pair of sinusoids are close in frequency. It can be calculated as a frequency ratio of the two sinusoids with additional weighting (EQ 5), where *E* is energy, *s* is frequency separation, and *f* is frequency (Sethares, 1998; Vassilakis, 2001 & 2005; Vassilakis & Kendall, 2010):[2]

$$R = E^{-3.5s(f_2 - f_1)} - E^{-5.75s(f_2 - f_1)} \qquad \text{(EQ 5)}$$

TABLE A1.1: Matrix of inter-acoustic correlations

	SC (MIR)	(MED)	I	SF	ZC	R
SC (MIR)		**.61**	**.57**	.37	.46	.33
SC (MEDS)	**.61**		**.70**	**.94**	**.93**	**.89**
Inharmonicity	**.57**	**.70**		**.68**	.50	**.53**
S Flatness	.37	**.94**	**.68**		**.90**	**.94**
Zero-Cross	.46	**.93**	.50	**.90**		**.95**
Roughness	.33	**.89**	**.53**	**.94**	**.95**	

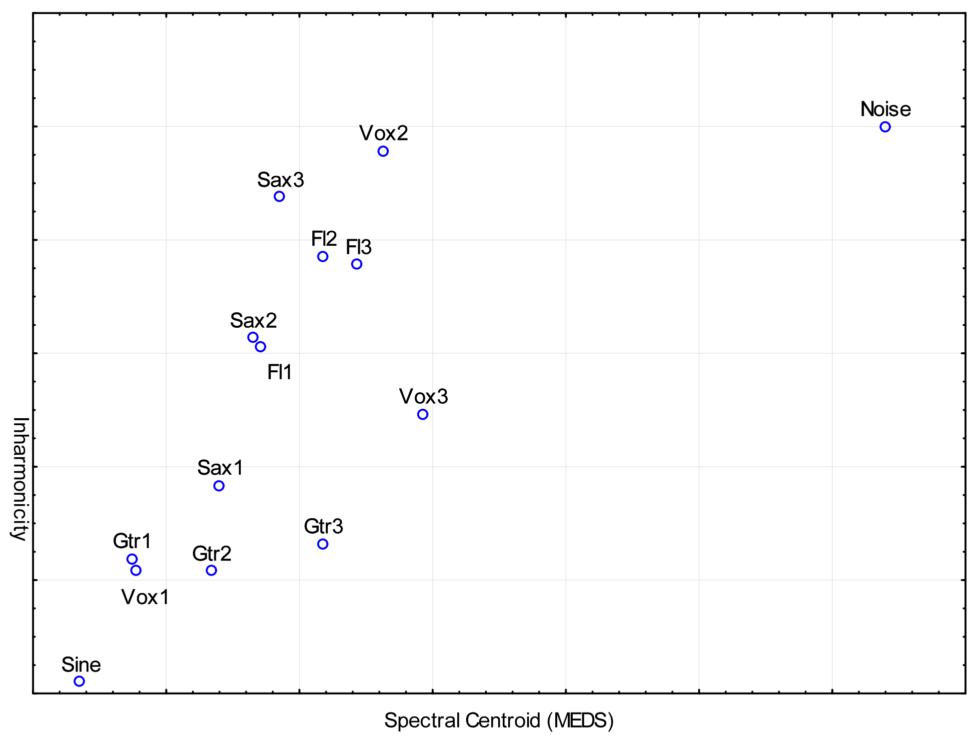
(N = 14)

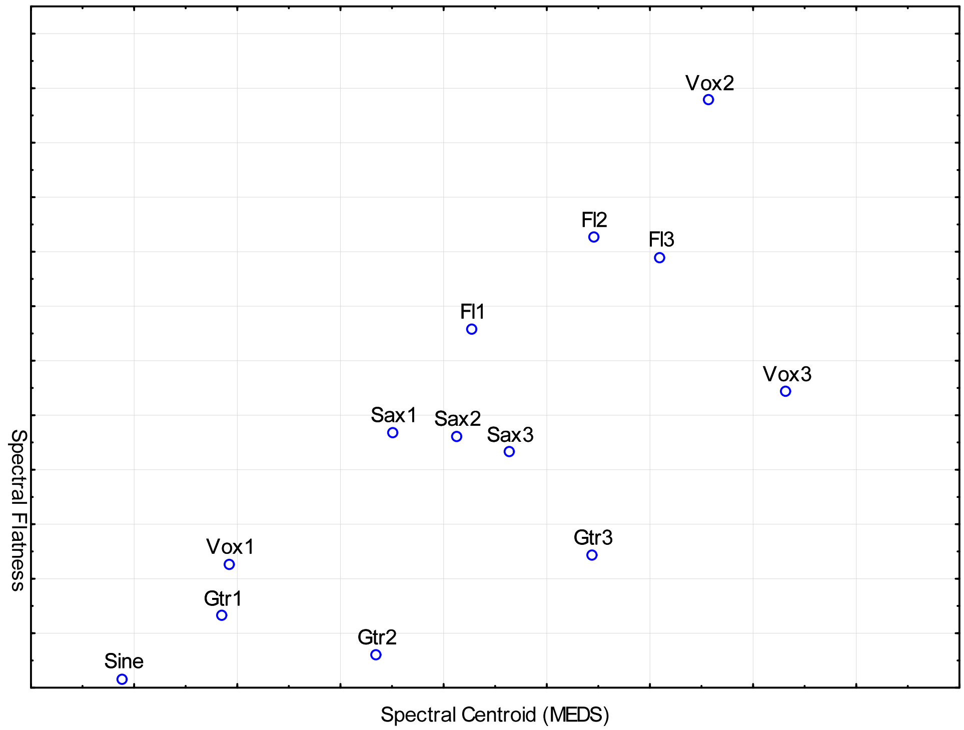
As indicated in Table A1.1, the five acoustic variables that form of focus of this study are all positively correlated, and many of the correlations are substantial. To illustrate, spectral centroid as calculated in MEDS relates closely to all the other four variables, including inharmonicity (r = .70), flatness (r = .94), zero-cross (r = .93), and roughness (r = .89), as plotted in Fig. A1.2. (Interestingly, the lowest correlation of this category, although still significant, is with spectral

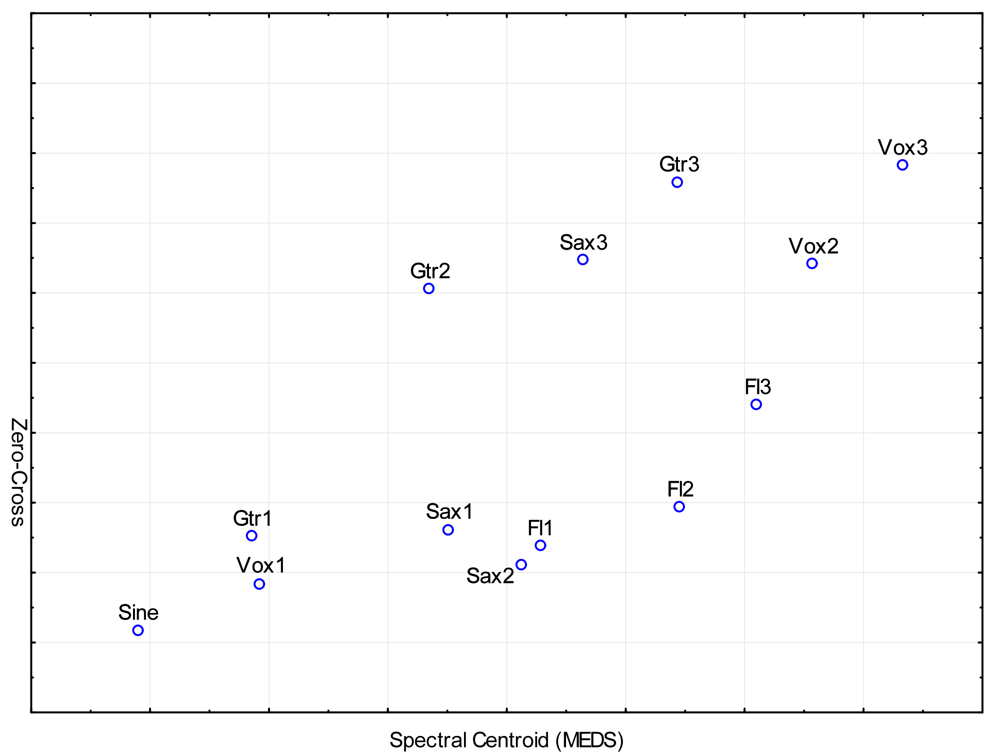
[2] *Mirroughness* measurements were calculated with the addition of the "Vassilakis" function, which adds the weightings shown here. Frequency separation can be expressed as: $s = \frac{0.24}{0.0207(f_1) + 18.96}$

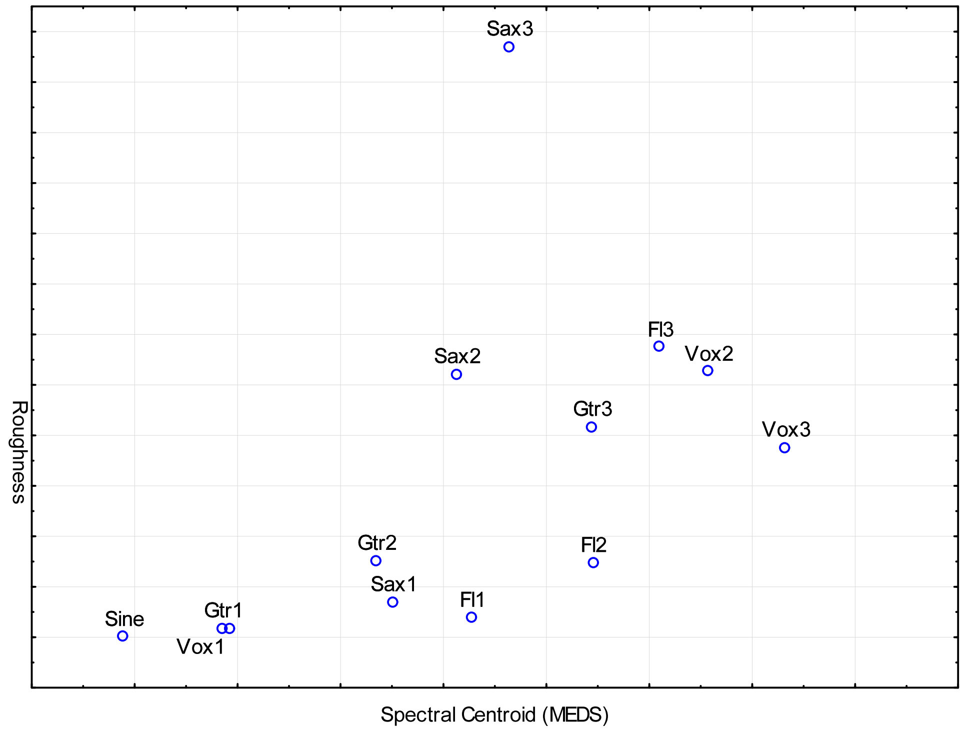
centroid calculated by the MIRtoolbox, r = .61. This hints at the variability of acoustic measurements between different analytical tools.) Taken together, correlation analysis indicates that all five of these acoustic parameters (six counting the dual SC measurements) tend to move together in a positive relationship, with SC predominating. Psychoacoustically speaking, then, brightness, noisiness, and sensory dissonance often go together.

It should also be mentioned that, in general, the acoustic parameters outlined here increase from the first timbral condition ("low-noise") to the third ("high-noise"). For instance, in Fig. A1.2 we can see that Timbre 1 (Gtr1, Vox1, etc.) is fairly low in both spectral centroid and inharmonicity, with Timbre 2 and 3 gaining in these qualities. The sine and noise signals define the outer edges of the "noisiness space" within which the rest of the signals are arrayed.




Vox2
Fl2
Fl3
Fl1
Vox3
Sax1
Sax2
Sax3
Gtr3
Vox1
Gtr1
Gtr2
Sine
Spectral Flatness
Spectral Centroid (MEDS)


Vox3
Gtr3
Sax3
Vox2
Gtr2
Fl3
Fl2
Gtr1
Sax1
Fl1
Vox1
Sax2
Sine
Zero-Cross
Spectral Centroid (MEDS)



***Note**: noise signal omitted from some plots because of its extremely high values. Scale not given because units of measurement are of comparative (not absolute) value only.*

FIGURE A1.2: Plots of spectral centroid (MEDS) against inharmonicity, flatness, zero-cross, and roughness

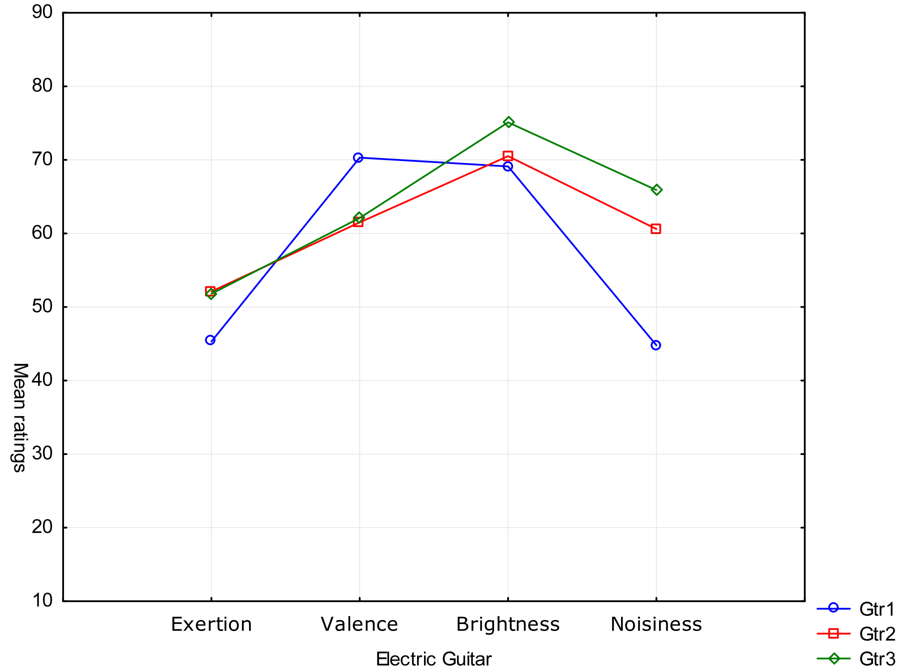
Perceptual Data Analysis

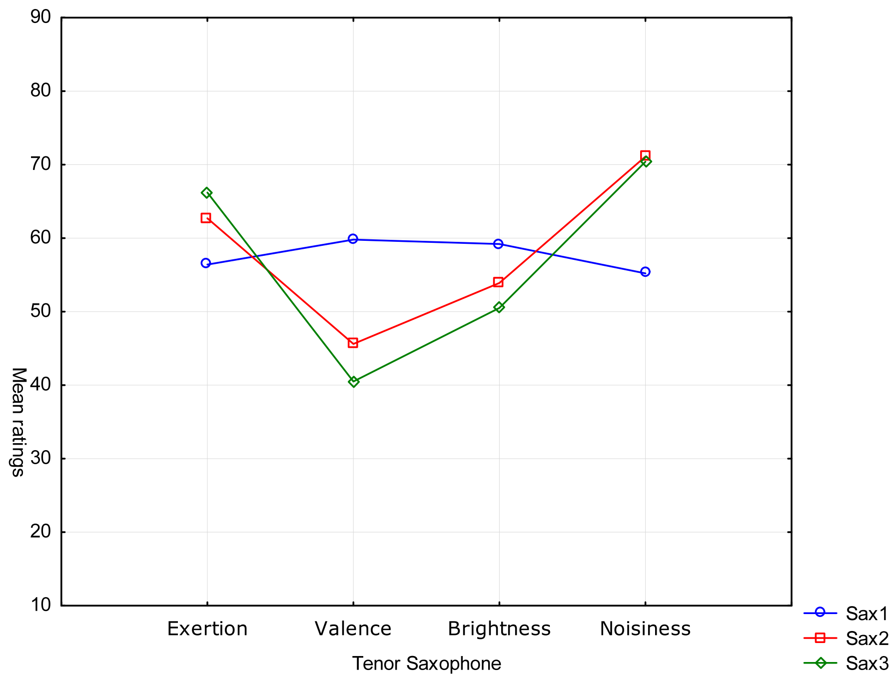
None of the subjects reported any difficulty in performing the experimental task, and acceptable inter-subject consistency in the ratings was observed (Cronbach α for all scales = .74).[3] The consistency of responses indicates fairly uniform associations between the continuous perceptual qualities rated—*bodily exertion*, *affective valence*, *emotion conveyed*, "brightness," and "noisiness"—and the acoustic stimuli presented.

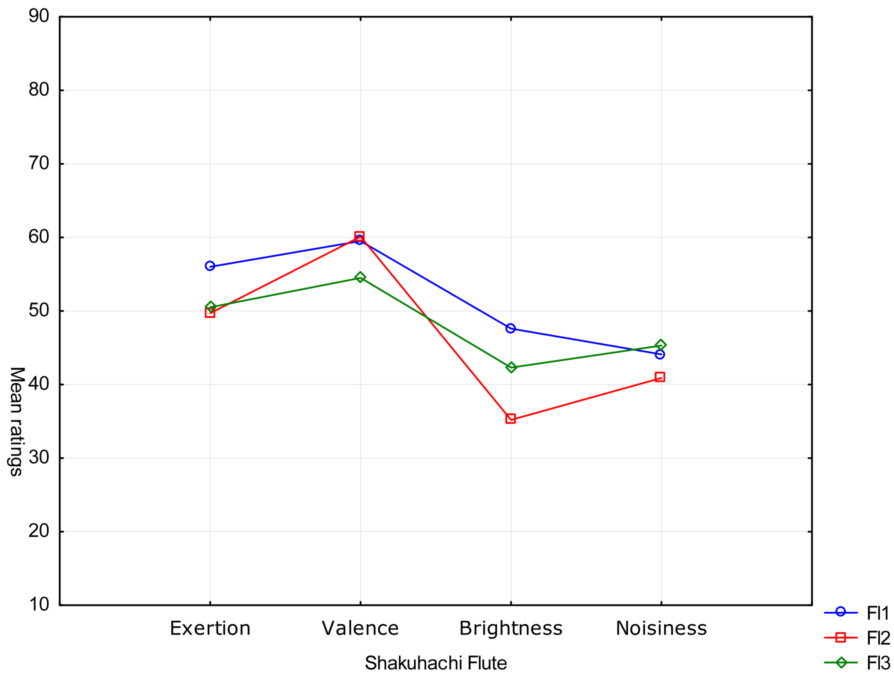
[3] All statistical analyses were carried out using STATISTICA and R software.

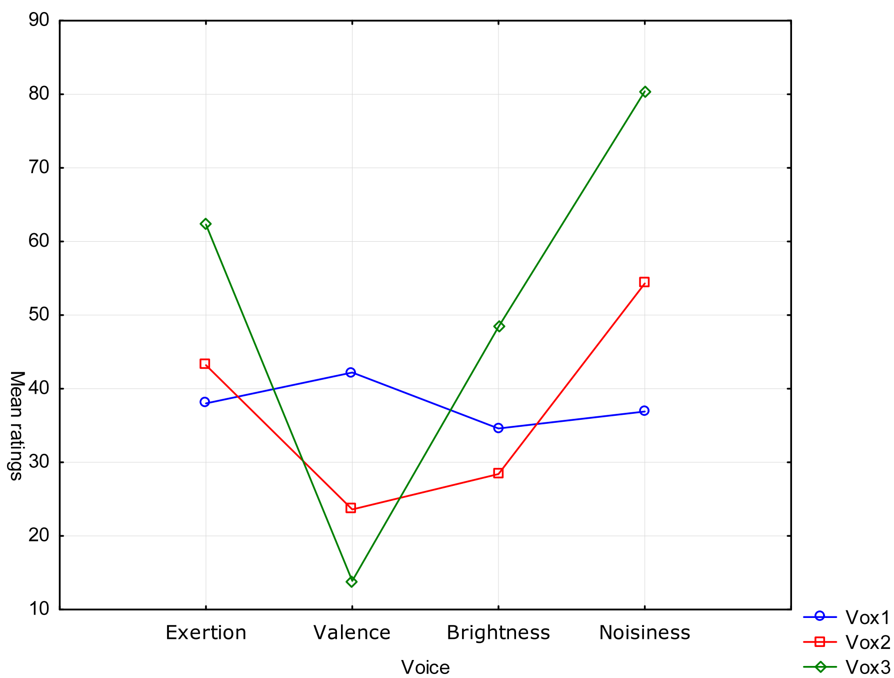
Next, in order to verify and evaluate differences between the ratings, a three-way repeated measures analysis of variance (ANOVA) was performed. The ANOVA employed a 4x4x3 design, with four categorical levels for both Instrument (Guitar, Saxophone, Shakuhachi, Voice) and Perceptual Quality factors (Bodily Exertion, Affective Valence, Brightness, and Noisiness), and three ordinal levels for the Timbre factor (low noise, medium noise, high noise). All factors yielded significant differences of within-subjects mean ratings: for the main effect of the Instrument factor, $F(3, 105) = 36.51$, $p<.0001$; Perceptual Quality, $F(3, 105) = 3.54$, $p<.017$; and Timbre, $F(2, 70) = 7.15$, $p<.0015$. Interactions between factors also demonstrated a very high level of significance: Instrument/Timbre, for instance, yielded $F(6, 210) = 11.5$, $p<.0001$; Perceptual Quality/Timbre, $F(6, 210) = 29.13$, $p<.0001$; Perceptual Quality/Instrument, $F(9, 315) = 14.95$, $p<.0001$; and a three-way interaction of Instrument/Timbre/Perceptual Quality gave $F(18, 630) = 7.25$, $p<.0001$. Following the ANOVA, a post hoc test (Tukey's HSD) was performed, confirming the significance of results on all pair-wise comparisons of means relating to the hypothesis. Sphericity of ratings was assessed using Mauchly's test, which indicated symmetry of variance for all factors ($p<.05$) except Perceptual Quality, $\varepsilon = .51$. To remedy this, a Greenhouse-Geisser correction was performed on the Perceptual Quality factor for an adjusted p-value of .02. Taken together, these results indicate a strong relationship between timbral qualities, instrument, and perceptual experiences including embodiment and affect.

After determining the reliability of the data, means of all subject ratings were extracted (14 signals x 4 interval ratings tasks = 56 total), a full summary of which can be found in Figure A1.3 below.


90
80
70
60
50
40
30
20
10
Mean ratings
Exertion
Valence
Brightness
Noisiness
Electric Guitar
Gtr1
Gtr2
Gtr3


90
80
70
60
50
40
30
20
10
Mean ratings
Exertion
Valence
Brightness
Noisiness
Tenor Saxophone
Sax1
Sax2
Sax3


90
80
70
60
50
40
30
20
10
Mean ratings
Exertion
Valence
Brightness
Noisiness
Shakuhachi Flute
Fl1
Fl2
Fl3


90
80
70
60
50
40
30
20
10
Mean ratings
Exertion
Valence
Brightness
Noisiness
Voice
Vox1
Vox2
Vox3

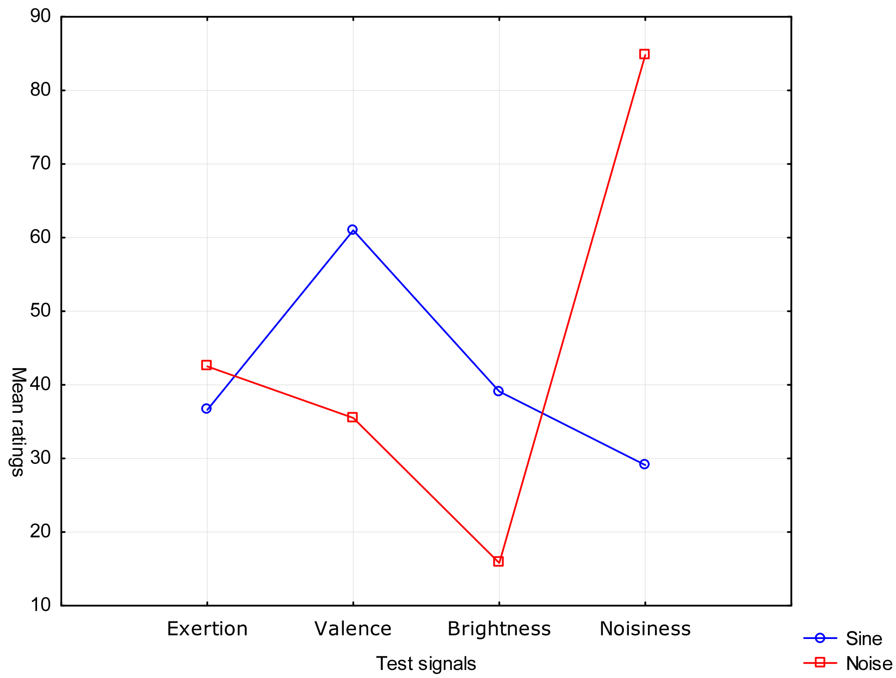


FIGURE A1.3: Mean ratings for all signals (by instrument) in four perceptual conditions

To briefly identify the salient features of each instrument group, Electric Guitar signals exhibited the general pattern hypothesized: signal 1 (the "low-noise" timbre) had the lowest *bodily exertion* and p-noise means, and the highest *affective valence* rating of the set, while the distorted signals (2 and 3) showed higher *exertion*, "brightness," and p-noise (i.e., perception of "noisiness"), in addition to lower *valence.* (No significant difference in ratings between 2 and 3 were found in the first two perceptual categories.) Interestingly, higher levels of *exertion* were heard in the distorted signals, even though the physical means of producing guitar distortion do not correlate with increased physical energy on the part of the player. It could be argued that the acoustic qualities under observation (see guitar signals in Fig. A1.2) somehow "trick" the listener into hearing increased bodily exertion, although no more work is required to produce distorted electric guitar than clean tones, and in fact, the same original sample was used for all three timbres with the addition of different levels of digital distortion. (This result plays a role in the analysis of simulated timbral violence in "extreme" heavy metal in Chapter 5.)

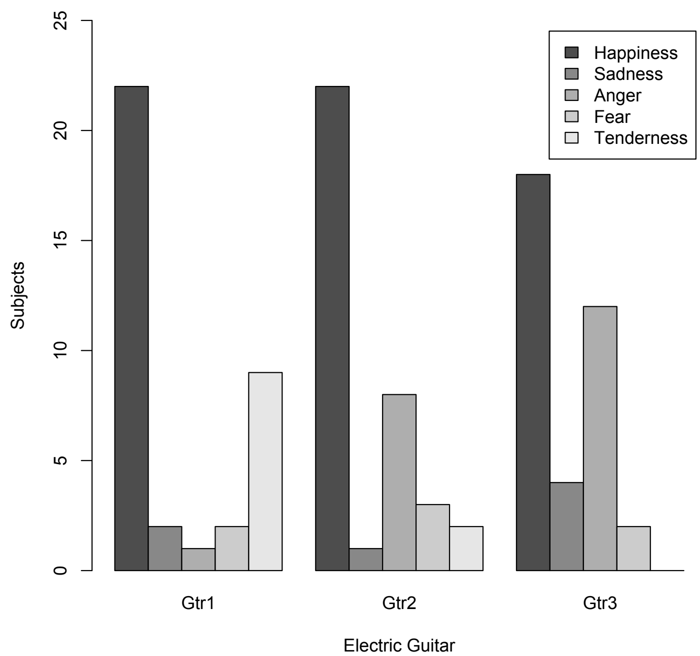
Saxophone signals exhibit a similar pattern, though the direction and degree of change in means between conditions is very different. With this instrument, *exertion* ratings follow the hypothesized arrangement, as does *valence* and p-noise: the low-noise timbre (1) sounds the least physically effortful and the least noisy, and is most liked. However, "brightness" ratings are the inverse relationship from what might be expected if one were to define "brightness" purely in terms of spectral centroid: signal 1 was rated as brightest, with 2 and 3 behind (as shown in Fig. A1.1, spectral centroid increases with each timbre). This seems to indicate that, with this instrument at least, "brightness" is negatively related to both *exertion* and p-noise. This cursory conclusion will be discussed in more depth here and in Appendix 2.

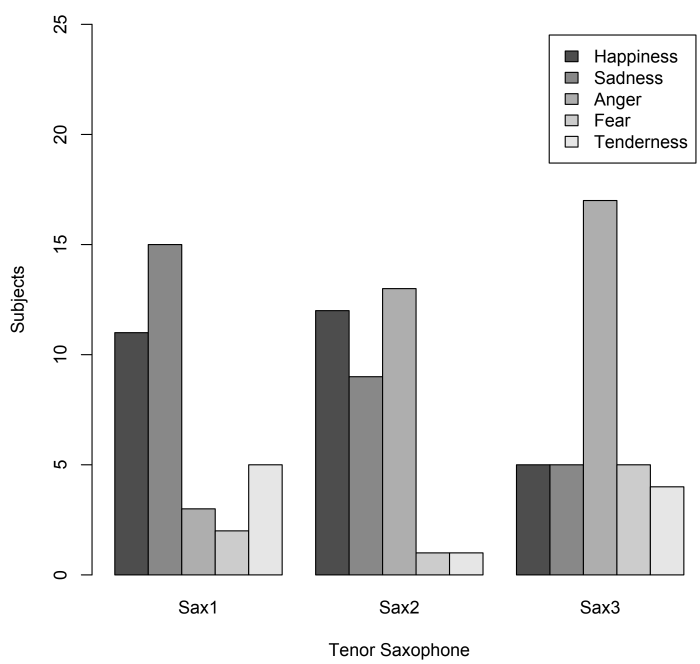
Shakuhachi signals exhibit a high degree of ambiguity in all perceptual categories. There is a discernable pattern between conditions, but the data do not reveal any significant difference between the three timbres: mean ratings tend to cluster closely together and hover close to 50. The much greater ambiguity of responses to this instrument might indicate a higher level of unfamiliarity with the source. Unlike the guitar and saxophone—and especially the voice—the material means of shakuhachi sound production are probably not well understood to study participants. Without a clear idea of how sound is generated on the instrument and the specific techniques involved—techniques that, incidentally, *do* involve higher degrees of physical effort as we move through the three timbre levels—subjects were not able to make these evaluations consistently.

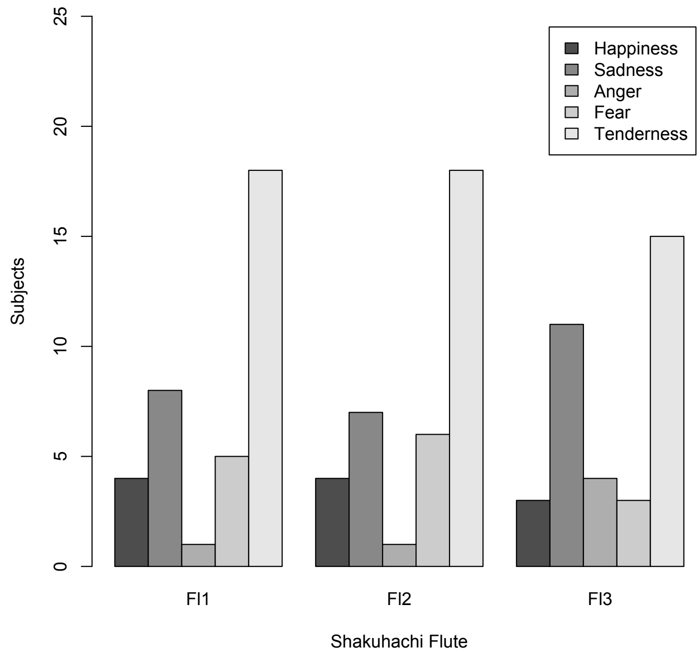
The voice is obviously the most physically intuitive to all of us, and intriguingly, it generated both the greatest range of responses and the most dramatic shifts in mean scores between conditions. The pattern here follows the Guitar and Saxophone, only the differences are much more pronounced. For instance, where the Saxophone *exertion* means range from the mid-50s (Timbre 1) to the mid-60s (Timbre 3), the voice covers the mid-30s to the mid-60s. This is true for all perceptual conditions. Additionally, it should be noted that the "high-noise" signal for the voice (Timbre 3) is both the least liked (M = 13.9) and the most noisy (M = 80.4) of the natural signals. It would seem that appraisals of this particular timbre are more potent and polarized than any of the instruments. If, as it was hypothesized, human vocality provides us with a privileged interpretive mechanism for understanding the affective dimensions of musical timbre more generally, then this result would not surprise.

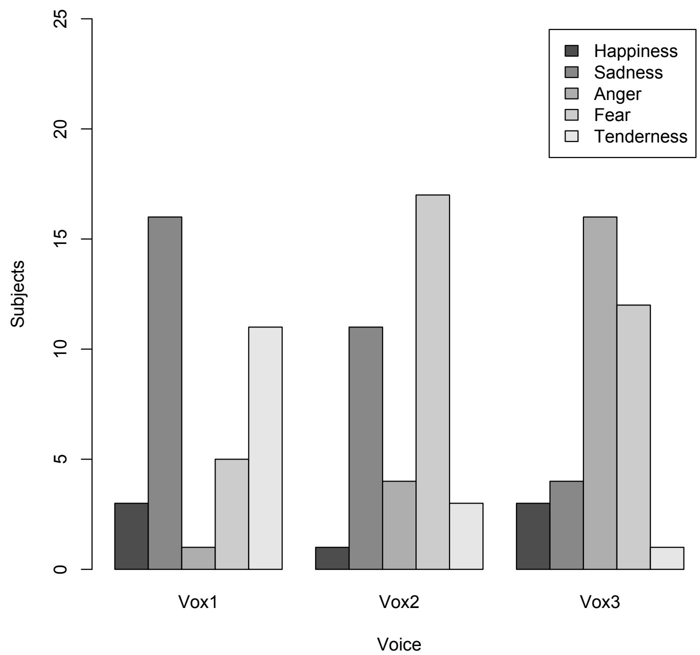
Ratings for test signals indicate a similar pattern. Like the Saxophone, "brightness" ratings differ from what might be expected if we look at the signals entirely from the physical frame of reference, with the Sine showing considerably higher "brightness" ratings than Noise. Predictably, subjects gave higher *valence* ratings to the Sine, and judged Noise to be noisier by a margin of some 50 points. Although both signals were computer-generated, Noise was given a higher *exertion* score than Sine, possibly (like the Electric Guitar) as a result of the presence of certain acoustic qualities that are often correlated in human experience with higher levels of physical exertion.

Next, turning very briefly to the nominal "emotion conveyed" data, frequency tables were developed and responses compared. As shown in the Fig. A1.4 histograms, the Guitar was thought to convey "happiness" to a greater degree than the other instruments, though "anger" increased between Timbre 1 and Timbre 3. Saxophone signals saw a similar rise in "anger" evaluations, with "happiness" and "sadness" (which predominate in Timbre 1) declining through the other two signals. Shakuhachi is the only instrument of the set in which "tenderness" predominates in all three Timbres. Voice signals generated three very different "emotion conveyed" profiles: signal 1 led with "sadness" and "tenderness," signal 2 "sadness" and "fear," and signal 3 "anger" and "fear." The Sine test signal generated a polarized response (almost exclusively "sadness" and "tenderness"), while Noise was the most mixed of the group, with a fairly flat distribution across the five emotion categories.






Subjects
0
5
10
15
20
25
Happiness
Sadness
Anger
Fear
Tenderness
Fl1
Fl2
Fl3
Shakuhachi Flute


Subjects
0
5
10
15
20
25
Happiness
Sadness
Anger
Fear
Tenderness
Vox1
Vox2
Vox3
Voice

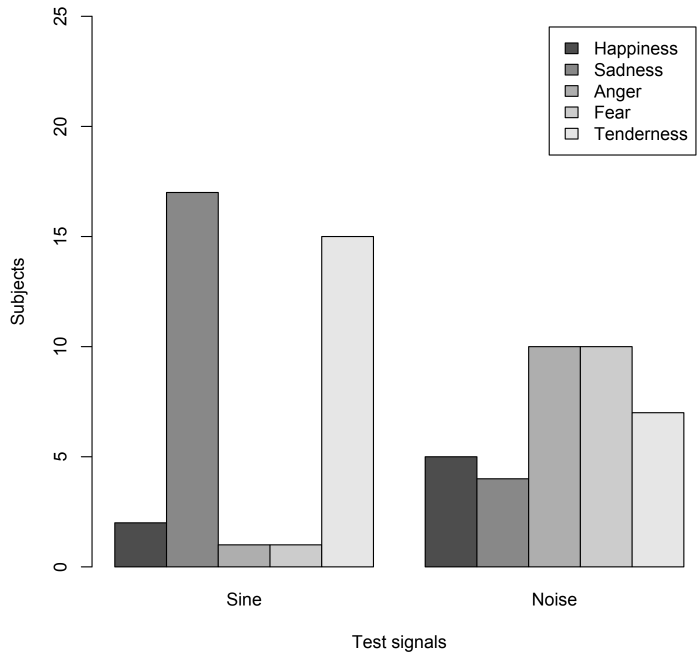


FIGURE A1.4: Histograms of "emotion conveyed" ratings for each signal (by instrument)

After reviewing descriptive statistics for the data set and confirming significance of results, the first task was to explore how the different perceptual qualities interrelate. Table A1.2 shows all correlations between the behavioral categories. First, we can see that *bodily exertion* correlates most strongly with the semantic dimension of "noisiness" (.55) and the "emotion conveyed" factor of *anger* (.68). (Observe as well that *anger* and *fear* ratings increase as signals grow noisier for guitar, saxophone, and voice.) *Affective valence* yields more significant correlations: this quality appears to relate to both semantic dimensions, and to two "emotion conveyed" dimensions. "Brighter" signals tend to elicit more positive valuations (.52), while "noisy" signals are negatively perceived (−.55). Further—and unsurprisingly—positive valence ratings are related to the impression that a signal is meant to convey *happiness* (.56), while low valence responses are related to impressions of *fear* ($r = -.82$, one of the strongest negative correlations of the analysis). The inverse relationship between affective valence and the two semantic qualities ("brightness" and "noisiness") was a novel finding. Researchers have positively correlated impressions of "brightness" with high spectral centroid (Schubert & Wolfe, 2006), and a large ratio of high to low frequency has been shown to negatively correlate with affective valence in both musical and non-musical signals (Eerola et al., 2012; Kumar, von Kriegstein, Friston, & Griffiths, 2012; McDermott, 2012). It might be expected, then, that "brightness" would be negatively valenced, but the opposite appears to be the case.

TABLE A1.2: Matrix of inter-scale correlations

Dimensions:	*Embodiment*	*Affect*	*Semantic*		*Emotion Conveyed*				
	E	V	B	N	H	S	A	F	T
Exertion		−.17	.42	**.55**	.14	−.39	**.68**	−.09	−.40
Valence	−.17		.52	**−.55**	**.56**	.01	−.51	**−.82**	.41
Brightness	.42	.52		.03	**.83**	−.38	.18	**−.59**	−.39
Noisiness	**.55**	**−.55**	.03		.15	**−.59**	**.87**	.34	**−.74**
Happiness	.14	**.56**	**.83**	.15		**−.56**	.10	−.50	−.43
Sadness	−.39	.01	−.38	**−.59**	**−.56**		−.50	−.11	.37
Anger	**.68**	−.51	.18	**.87**	.10	−.50		.16	**−.74**
Fear	−.09	**−.82**	**−.59**	.34	−.50	−.11	.16		−.19
Tenderness	−.40	.41	−.39	**−.74**	−.43	.37	**−.74**	−.19	

(N = 36)

Figs. A1.5 and A1.6 graph the inverse relationships documented above. As indicated in the first scatterplot (Fig. A1.5), p-noise ratings increase along with *bodily exertion* ratings, while decreasing the more subjects report "liking" the timbral quality in question. Both *r*-values here reach significance at $p<.04$. When we look at the semantic variables vis-à-vis *valence* (Fig. A1.6), moreover, we can see that "brightness" ratings increase with liking even as "noisiness" declines. There is much more to be said about these data in Chapter 2.

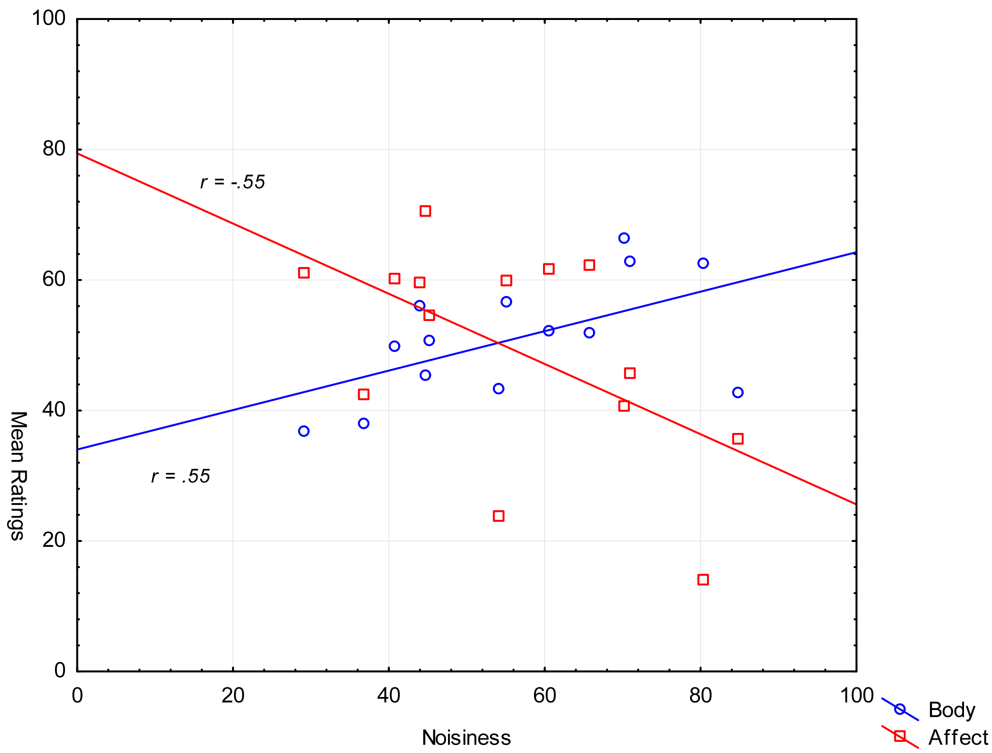


FIGURE A1.5: Mean ratings of p-noise (*x*-axis) related to perceptions of *bodily exertion* and *affective valence* (*y*-axis)

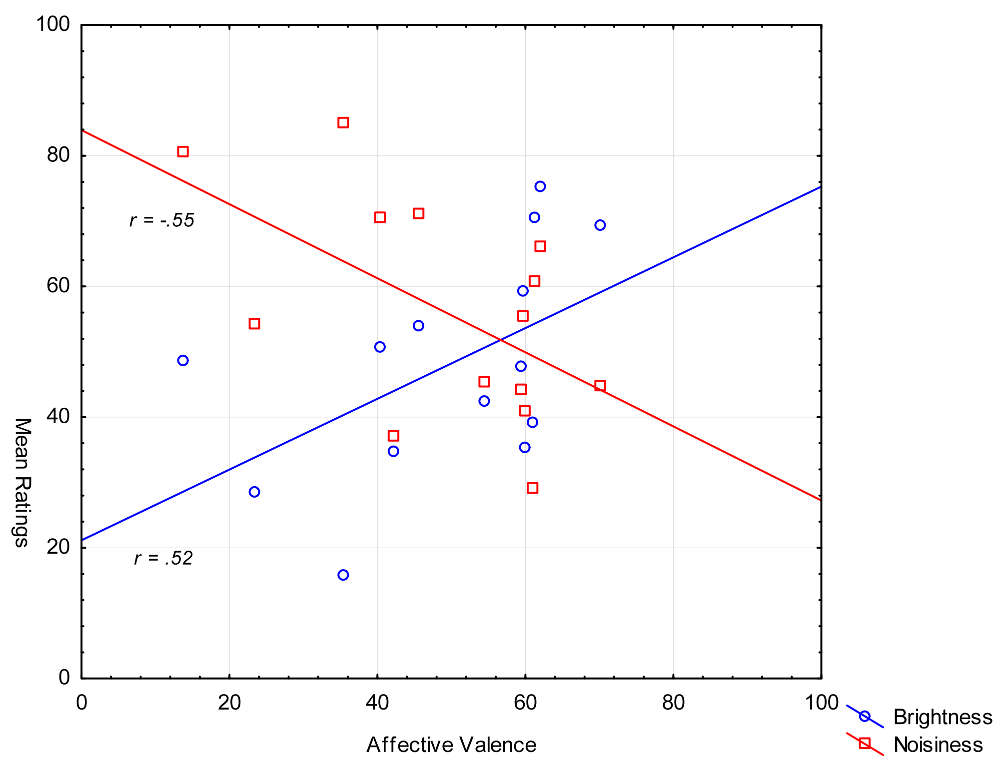


FIGURE A1.6: Mean ratings of *affective valence* (*x*-axis) related to "brightness" and p-noise (*y*-axis)

After assessing both inter-acoustic and inter-perceptual rating correlations, it remained to place both bodies of data into conversation in order to explore which acoustic parameters relate to the perceptual variables observed. Table A1.3 shows the results of correlation analysis between the perceptual and the acoustic variables of the study. To highlight a few results: all six acoustic parameters reached significance on at least one of the perceptual measures. Spectral centroid (MIR) had the greatest number of strong correlations, with the same parameter computed in MEDS following close behind, though for different qualities. (The two measurements differ primarily in robustness of correlations, not direction, though "brightness" is an exception. This is somewhat ironic, considering that spectral centroid is often believed to be the acoustic correlate *par excellence* of "brightness." The p-noise measure found both calculations to be in close agreement.) Since the acoustic qualities are strongly covariant, for most of the perceptual variables, correlations with the acoustic data are either entirely negative or entirely positive (discounting values that fall at or near randomness, $r = 0$). For instance, *valence* is negatively correlated with these six parameters across the board, while p-noise is positively correlated with all six, many of which at strong levels. Concomitantly, we can see that all acoustic parameters are positively correlated with the negatively-valenced emotions of *anger* and *fear*. We can thus conclude that the perception of "noisiness" (p-noise) in a given isolated instrument signal appears to comprise an emergent composite of all these qualities, perhaps among other variables not explored here. In order to assess the relative weightings of each acoustic parameter in perceptual judgments, a multiple regression analysis was carried out; however, this attempt revealed erratic results due to multicollinearity of the acoustic variables. Failure to produce a coherent regression model based on these acoustic qualities was further evidence for their tight covariance.

TABLE A1.3: Correlations between mean perceptual ratings and selected acoustic qualities

	Brightness		*Noisiness*		*Sensory Dissonance*	
	SC (MIR)	(MED)	I	SF	ZC	R
Exertion	**.71**	.02	.27	−.18	−.10	−.16
Valence	−.33	−.50	−.50	−.38	−.40	−.31
Brightness	.07	**−.54**	**−.58**	**−.68**	−.44	**−.54**
Noisiness	**.76**	**.68**	.44	.49	**.69**	**.58**
Happiness	.03	−.25	−.47	−.31	−.07	−.16
Sadness	**−.56**	−.36	−.19	−.18	−.42	−.27
Anger	**.68**	.38	.28	.14	.41	.30
Fear	.24	**.56**	**.60**	.49	.45	.33
Tenderness	−.37	−.17	0	.03	−.27	−.11

Brief Conclusion

This study both confirmed certain *a priori* hypotheses and contradicted others, a full account of which is given in Chapters 1 and 2. The single most important take-away, however, confirmed the larger theoretical insight advanced in the dissertation: *embodiment of musical timbre is positively related to perceptions of noise and negatively related to affective valence.* It was hypothesized that low-noise timbres would elicit lower *bodily exertion* ratings, while very noisy timbres would sound out the effortful physical circumstances of their production, resulting in higher *exertion* ratings.

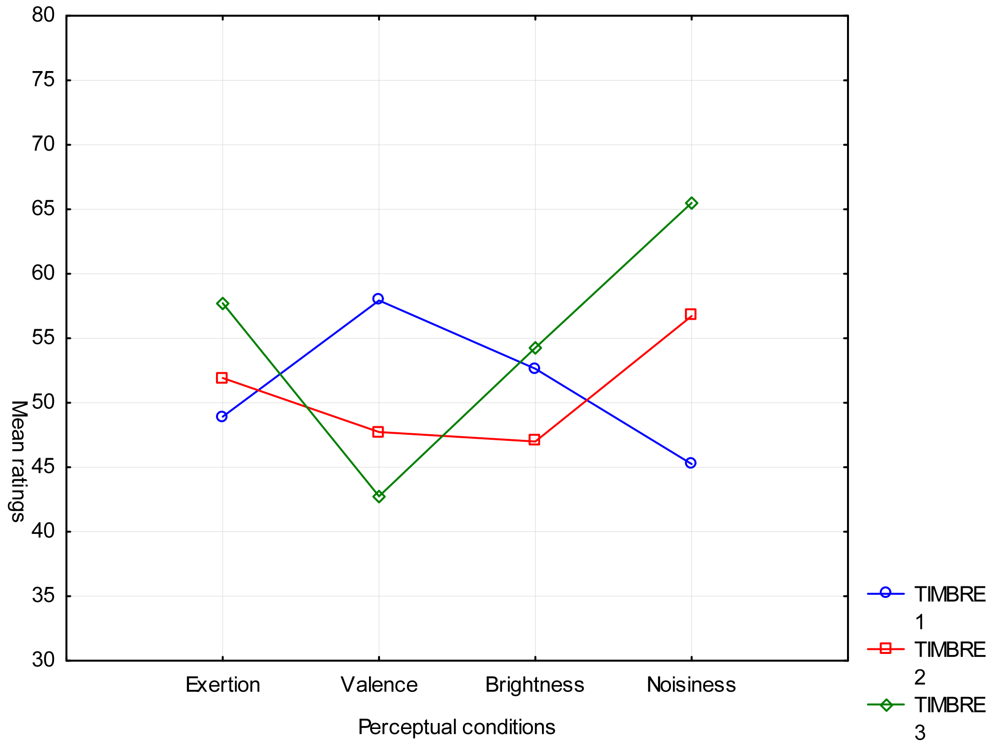


FIGURE A1.7: Summary graph of mean ratings for three timbres (low-, medium-, high-noise) in four perceptual conditions

As exhibited in Fig. A1.7, this was indeed the case: Timbre 1 (the low-noise condition averaged across the four Instrument categories) received the lowest mean *exertion* rating, while Timbre 3 received the highest. (Recall from the ANOVA results that differences in mean Timbre scores constitute statistical significance, $F(2, 70) = 7.15$, $p<.0015$, even though there was considerable variability in individual scores.) In the next perceptual quality condition, *affective valence,* we can see the order reversing: subjects tend to find Timbre 1 to be the most positively valenced, while Timbre 3 received lowest mean ratings. To summarize, then, noisy signals appear to indicate a greater amount of bodily effort going into their production, and they are more negatively viewed than less noisy signals. The "brightness" variable here is quite ambiguous: there is no clear spread between the timbre types, and the clustering right around 50—or pure chance—indicates a good deal of perceptual ambiguity. P-noise, however, is quite clear, with a spread of over 20 points between Timbre 1 and Timbre 3 means, with Timbre 2 right about in the exact middle. As hypothesized, timbres of isolated instruments that are scaled in ecologically valid ways (without manipulating acoustic parameters systematically) into ordinal levels of a-noise are perceived as such, despite varying means of production and vastly different acoustic profiles. As shown in Table A1.3, however, there are acoustic commonalities that tend to unite p-noise, namely high spectral centroid (as measured by either software program), zero-cross, and roughness, all of which correlate with p-noise at $p<.05$. And furthermore, the other two acoustic variables of interest—inharmonicity and spectral flatness—also demonstrate moderate correlations with p-noise.

This study deals only with isolated musical timbres, which are admittedly quite rare in the world of real music. Do these same conclusions hold for the perception of timbre in actual music? In Experiment 2, I endeavor to address this question through a parallel experimental paradigm examining brief recordings of popular music songs.

Appendix 2

Experiment 2: Perceptions of Embodiment and Affect in Brief Popular Music Excerpts

Hypotheses

Timbre research has largely devoted itself to the study of isolated tones. However, timbre is most often experienced not in monophonic settings but in the context of ensemble music. Regrettably, polyphonic timbre (Alluri, 2010; Alluri, 2012)—the "global sound" of music (Aucouturier, 2006) comprising a rich and complex timbral mix of different sound producers—has received a good deal of scholarly neglect, and to my knowledge, no studies have systematically compared the psychoacoustics of monophonic to polyphonic timbre. This study attempts such a project by employing the same experimental paradigm in both modalities. I ask in this study whether the same patterns of embodied listening, affect, semantic association, and emotion perception investigated in Experiment 1 apply to settings of polyphonic timbre as well. To that end, the experiment combines approaches from both timbre and rapid music cognition research, an experimental model recently popularized by Malcolm Gladwell's bestseller *Blink* (2005), as well as many psychoacoustic studies (Schellenberg, 1999; Gjerdingen, 2008; Krumhansl, 2010). Isolating the polyphonic timbre of a range of popular music examples through the use of extremely rapid excerpts (400 ms), it is hypothesized that the same mechanisms of embodied music cognition undergirding perception of monophonic timbres—namely, that increased bodily exertion is perceptible in "noisy" timbres and translated via systems of adaptive psychobiology into negatively valenced affect—are also active when listening to polyphonic timbre.

Stimuli

Eighteen brief excerpts of popular music recordings were created using Audacity software. Excerpts were selected to represent six genres of contemporary popular music: (1) Rock, (2) Electronic Dance Music, or EDM, (3) Hip-Hop/Rap, (4) Pop, (5) Heavy Metal, and (6) Country. Each genre was thus represented by three excerpts. Following the precedent of Schellenberg (1999), Gjerdingen and Perrott (2008), and Krumhansl (2010), excerpts were 400 ms in duration—the "medium-length" condition in these studies—with 10 ms amplitude ramps on both ends of the signal. Loudness was equalized manually.

Excerpts were chosen based on the criterion of "genre representativeness": they were selected to include basic acoustic elements of the "timbral environments" (Ferrer, 2011) with which their representative genres are commonly associated. For example, rock and heavy metal excerpts included prominent distorted electric guitar and vocals; EDM included heavy bass and harsh synthesizer timbres associated with the sub-genre of dubstep; rap included sample-based beat production and rhythmic, "rapped" vocals; pop included a highly polished production style, vocals, and synthesizers; and country included common signifiers of "twang," including vocals, fiddle, and slide guitar (Neal, forthcoming). In order to isolate the timbral components of different singing styles without semantic confounds, excerpts with vocals were selected to exclude any identifiable words. Sources for the stimuli were recorded from the mid-1950s to the present: again, the criterion of "genre representativeness" was applied in determining year of recording. Rock excerpts, for example, were all from iconic songs from the "classic" era (late 1960s to early

1970s) due to high genre identification levels achieved in pilot testing (as opposed to contemporary rock, which exhibited ambiguity of identification in pilot tests); country excerpts originate in source recordings from the 1950s and 60s for the same reason (contemporary country, which is more stylistically mongrel than its mid-century predecessors, yielded highly ambiguous identification results in pilot tests). The process of selecting excerpts was admittedly highly subjective and non-scientific, drawing upon the author's listening history, knowledge of popular genres as a performer, and musicological training. To quantify "genre representativeness"—i.e., to determine if the signals really were *heard* as the genre they were intended to represent—twenty-eight subjects performed a genre selection task with the 18 signals post hoc (Appendix 4a). Excerpt sources are listed in Table A2.1.

Excerpts were brief enough that no melodic and harmonic elements, structural implications, or rhythmic patterns were discernable. A vague sense of tempo (e.g., fast or slow) could be observed in certain excerpts, but no other details were perceptible besides timbre, texture, and artifacts of the recording that might provide stylistic cues and time-period of release. Such "thin slices" (Krumhansl, 2010) were desired in order to isolate the elements of polyphonic timbre under experimental observation.

TABLE A2.1: Excerpted songs ordered by genre

Rock

1. Jimi Hendrix, "Purple Haze"
2. Led Zeppelin, "Whole Lotta Love"
3. Rolling Stones, "Tumbling Dice"

EDM

1. Porter Robinson, "The Seconds"
2. Kaskade (feat. Skylar Grey), "Room for Happiness"
3. Destroid, "Raise Your Fist"

Hip-Hop

1. Snoop Dogg, "Gin and Juice"
2. Public Enemy, "Bring the Noise"
3. Tribe Called Quest, "God Lives Through"

Pop

1. Cody Simpson, "iYiYi"
2. Lady Gaga, "Poker Face"
3. Carly Rae Jepsen, "Tiny Little Bows"

Metal

1. Slayer, "Raining Blood"
2. Cannibal Corpse, "Mummified in Barb Wire"
3. Slipknot, "Wait and Bleed"

Country

1. Hank Williams, "Your Cheatin' Heart"
2. Merle Haggard, "Okie from Muskogee"
3. Lefty Frizzell, "Always Late (With Your Kisses)"

Subjects

The group of participants was the same as Experiment 1. However, one subject was omitted as an outlier due to a Euclidean distance of over 300 in single linkage comparisons with the rest of the subject group. (This participant only responded with "100" or "0," accordingly skewing group means.) Besides this adjustment, the participants were as follows:

Thirty-five volunteers were recruited from the UCLA community. The study group consisted of 22 females and 13 males between the ages of 18 and 31 (age M = 20.1, SD = 2.06). The subjects self-reported their number of years of formal musical training (M = 6.79 years, SD = 5.93); all subjects were non-music majors with musical backgrounds ranging from no formal experience (7 subjects) to 20 years (1 subject). Every subject who participated in this experiment also participated in a related experiment (Experiment 1). The order of the two experiments was randomized. Each subject was paid $20 in cash after completion of both experiments.

Procedure

The procedure was the same as Experiment 1, with a few minor alterations to account for the different stimuli:

Experimental control was achieved through Music Experiment Development System (MEDS 2002-B-1) software (Kendall, 1990-2002). Subjects listened to the mono signals through high-quality headphones at a comfortable listening level, approximately 65 dB SPL, in a quiet room with minimal visual distractions.

The experiment consisted of four separate ratings and evaluation tasks that were presented in a predetermined order:

(1) First, as a measurement of the "embodiment" dimension of the signals, subjects were asked to rate the perceived *bodily exertion* required to produce each polyphonic timbre with the use of a numbered horizontal rating scale (0-100) with bipolar labels (*low exertion-high exertion*) consistent with the semantic differential paradigm. In order to hear the full range of the stimuli before being asked to evaluate them, and to accustom themselves with the extreme brevity of the excerpts, the experiment began with a "practice" run. Presentation of the 18 individual signals was randomized. Following this practice run, subjects proceeded to the experiment. Subjects were allowed to ask questions throughout the procedure. As before, presentation of the signals was randomized. Listeners were allowed one audition of each signal, and were required to press the "OK" button on the screen before going on to the next signal.

(2) Next, subjects were asked to evaluate the *affective valence* of each signal on a numbered horizontal rating scale (0-100) with bipolar labels (*dislike-like*). *Affective valence* has been shown to equate to musical *preference* (Eerola, Ferrer & Alluri, 2012). The procedure was the same as that outlined above, but without a practice run.

(3) Next, subjects were asked to categorize the perceived emotional content of each signal by selecting an adjective from a list that best describes the *"emotion conveyed"* by each sound. The list consisted of the five primary emotions determined by Juslin, (2001) to be successfully communicable by performers, and used by Krumhansl (2010) in a similar experiment: (1) Happiness, (2) Sadness, (3) Anger, (4) Fear, and (5) Tenderness. Subjects could only select one word from the list.

(4) Finally, subjects were asked to rate the "brightness" and "noisiness" of the signals using a VAME-modified semantic differential paradigm (Kendall & Carterette, 1993). A numbered horizontal scale (0-100) with bipolar labels was used (*not bright-bright, not noisy-noisy*), and subjects were asked to make both evaluations after a single audition of each signal.

The full script for the experiment was as follows:

> "This experiment investigates the perception of energy and emotion in very brief excerpts of popular music recordings. It is not a test of musical aptitude or ability. Before beginning, we will listen to all the excerpts to familiarize you with them. Remember, they are VERY short, so listen closely and pay attention! There will be 18 excerpts total. Disregard the scale for now and press OK between each music clip to advance to the next. Press OK to begin.
>
> *[Practice run]*
>
> We are now ready for the main experiment. You will be asked to rate each excerpt on a scale of `low exertion` to `high exertion,` corresponding to the level of physical exertion implied by the music. There will be 18 excerpts total. Please try to use the entire scale in your judgments. Press OK to begin.
>
> *[1. Bodily exertion rating task]*
>
> Next, you're going to tell us how much you like or dislike each of these excerpts. Please rate the same clips of music on a scale of `dislike` to `like.` Click OK to begin.
>
> *[2. Affective valence rating task]*

Next, we'd like to know how you would categorize the emotion conveyed by these excerpts. Please choose one of the following five emotions that best characterizes each clip of music: happiness, sadness, anger, fear, tenderness. Click OK to begin.

[3. "Emotion conveyed" selection task]

Last, we'd like to know how you would rate the excerpts in terms of two qualities: brightness and noisiness. After hearing each clip, please rate it on a scale of `not bright` to `bright,` and `not noisy` to `noisy.` Click OK to begin.

[4. Brightness and Noisiness rating tasks]

The duration of the full experiment was approximately 5-10 minutes.

Analysis and Results

Acoustic Data Analysis

Spectral centroid (SC), *inharmonicity* (I), *spectral flatness* (SF), *zero-cross rate* (ZC), and *roughness* (R) estimates were extracted from the 18 signals using MIRtoolbox1.4 for MATLAB (Lartillot, 2007). Additionally, MEDS was used to calculate spectral centroid (Kendall, 1990-2002). Measurements of all acoustic parameters using MIRtoolbox were derived from a 12-order FFT with 50% overlap (50 ms per frame), producing approximately 8–10 frames for each signal at a sampling rate of 44.1 kHz. The resulting calculations were based on an average across all frames. In MEDS extraction of spectral centroid, I opted for greater temporal specificity: 10-order FFT with 50% overlap (23 ms per frame), producing 34-35 frames with a Bartlett window, also at a sampling rate of 44.1 kHz. The greater resolution of SC (MEDS) calculations should be kept in mind in the proceeding discussion. This difference might account for the higher correlative value of the MEDS measurements over MIRtoolbox on virtually all the variables (see Table 2.7).

First, correlations between timbral qualities were assessed (see Table 2.2; Pearson's *r* entries in bold indicate statistical significance at $p<.05$). The same five acoustic dimensions of interest from Experiment 1 were evaluated, including:

(1) *Spectral centroid*, or the fulcrum point of energy distribution across a spectrum, commonly experienced as "brightness," "nasality," or "sharpness." Spectral centroid has been shown to be perhaps the most perceptually salient feature of timbre; it has also been implicated in the conveyance of escalating musical tension (Huron, 2006). The importance of this particular quality impelled me to measure it with two separate approaches: MIRtoolbox, which uses a cut-off threshold of 1500 Hz to create a ratio of energy above versus below this threshold, and MEDS, which captures the full average of the spectrum without any thresholding (indicated below as SC (MIR) and SC (MEDS)). See Appendix 1 for equations of this and all other acoustic parameters.

(2) *Inharmonicity*, or energy outside the harmonic series of a given fundamental, is often said to be experienced as "noisiness" (Peeters et al., 2011). Such a "noisiness" measure (as extracted using MIRtoolbox, as with all remaining timbral variables) is given as a ratio between the total acoustic energy the total inharmonic energy.

(3) *Spectral flatness*, another common measurement of perceived "noisiness," indicates the relative smoothness or spikiness of a signal. (An oboe, for example, is spiky owing to clearly defined harmonic partials, while white noise is flat since it has, at least in theory, equal energy across the entire frequency spectrum.) It is Wiener entropy of the spectrum given as a simple ratio between geometric and arithmetic means.

(4) *Zero-cross rate*, another simple indicator of noisiness, counts the number of times the signal crosses the *x*-axis per time unit.

(5) *Roughness*, or "sensory dissonance," results from amplitude fluctuation (beating) when a pair of sinusoids are close in frequency. It can be calculated as a frequency ratio of the two sinusoids with additional weighting (Sethares, 1998; Vassilakis, 2001):

As shown in Table A2.2, spectral centroid correlates moderately to strongly with all other acoustical parameters except roughness. In addition, it is interesting to note that the values given by the two SC extraction programs are in far closer agreement here than in Experiment 1. These correlations indicate that among these signals the psychoacoustic quality of "brightness" tracks with all three measurements of *noise* (I, SF, and ZC).[1] To illustrate the robustness of this correlation, Fig. A2.1 plots SC against ZC. As might be expected, the heavy metal excerpts have high SC and ZC values, as do two of the EDM excerpts (incidentally, the ones that scored highest for

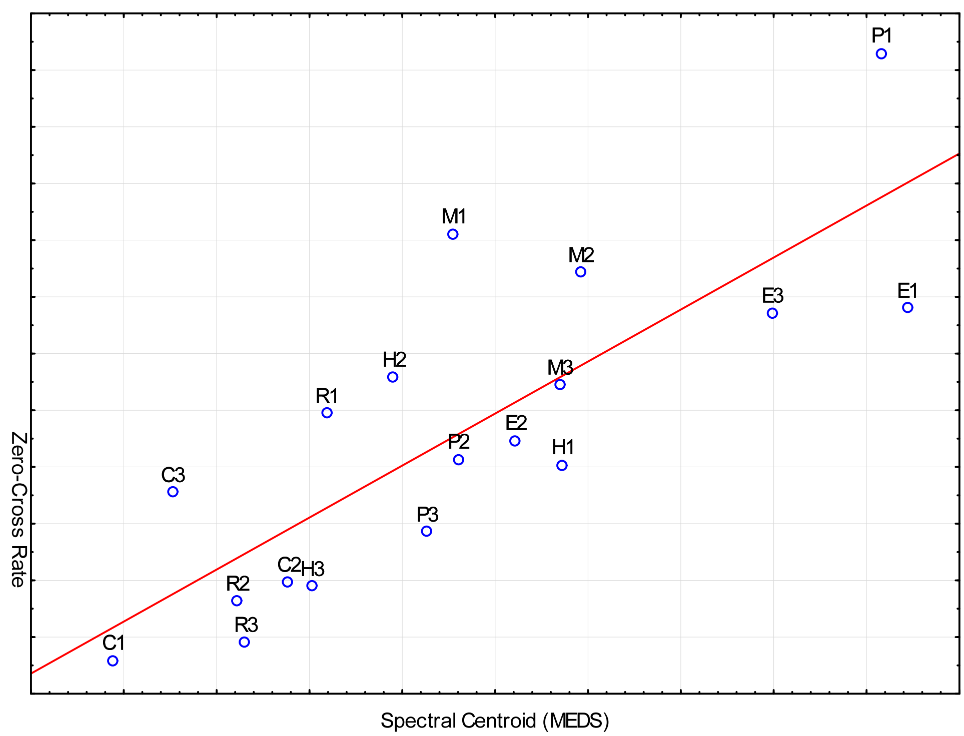
[1] It should be noted that the *inharmonicity* extraction tool in the MIRtoolbox is designed for monophonic signals with only one fundamental frequency. For this reason, results in this category are close to meaningless. This will be borne out in Table 2.6, which shows no strong correlations between *inharmonicity* calculations and any behavioral variable.

genre identification, as shown in Table A2.2). As a genre, Country excerpts have the lowest spectral centroid and zero-cross rates.

TABLE A2.2: Matrix of inter-acoustic correlations

	SC (MIR)	(MED)	I	SF	ZC	R
SC (MIR)		**.73**	**.51**	**.53**	**.71**	.04
SC (MED)	**.73**		**.48**	**.75**	**.80**	−.06
Inharmonicity	**.51**	**.48**		.47	.46	.12
S Flatness	**.53**	**.75**	.47		.44	−.22
Zero-Cross	**.71**	**.80**	.46	.44		.30
Roughness	.04	−.06	.12	−.12	.30	

(N = 18)



***Note**: R = Rock, E = EDM, H = Hip-Hop, P = Pop, M = Metal, and C = Country.*

FIGURE A2.1: Plot of spectral centroid (MEDS) against zero-cross rate

Perceptual Data Analysis

None of the subjects reported any difficulty in performing the experimental task, and acceptable inter-subject consistency in the ratings was observed (Cronbach α for all scales = .69).[2] The consistency of responses indicates fairly uniform associations between the continuous perceptual qualities rated—*bodily exertion*, *affective valence*, *emotion conveyed*, "brightness," and p-noise—and the acoustic stimuli presented. However, consistency of responses varied considerably from variable to variable. Table A2.3 shows complete Cronbach α results for the four "perceptual quality" factors and the six genres (*italicized*). The greatest degree of consistency in ratings can be found in the "noisiness" (p-noise) scale (.91), indicating a fairly uniform understanding of how this semantic dimension maps onto the experimental signals. *Valence*, on the other hand, showed no consistency whatsoever, reflecting the diversity of musical likes and dislikes of the participant group. The wide range of responses to this measure (and correspondingly high *SD*) should be kept in mind going forward.

TABLE A2.3: Cronbach α for each variable

Variable	*Cronbach* α (all scales *M* = .69)
Exertion	.74
Valence	.04
Brightness	.59
Noisiness	.91
Rock	*.67*
EDM	*.73*
Hip-Hop	*.51*
Pop	*.47*
Metal	*.71*
Country	*.69*

(N = 35)

[2] All statistical analyses were carried out using a combination of R and STATISTICA.

To verify statistical significance of differences in means, a three-way repeated measures ANOVA was performed. The analysis revealed significant differences of relevant within-factor mean ratings. The Perceptual Quality factor main effect (4 levels) was significant [$F(3, 102) = 4.28$, $p=.007$], and Genre (6 levels) was highly significant [$F(5\ 170) = 19.88$, $p<.00001$]. Moreover, interactions of these two factors (Perceptual Quality and Genre) proved highly significant as well [$F(15, 510) = 72.18$, $p<.00001$]. As expected, differences between the three excerpts (3 levels) within each genre were not significant [$F(2, 68) = 0.65$, $p=0.52$]. With differences between genres but not between excerpts, this result provided additional support to the genre representativeness of the excerpts selected. ANOVA results were subjected to a post hoc test (Tukey's HSD) confirming significance of all relevant comparisons of means, and Mauchly's test was performed to verify sphericity. Following confirmatory testing, means of all ratings were calculated (18 signals x 4 interval rating tasks = 72), and frequency tables were developed from the nominal "emotion conveyed" responses. Since the ANOVA demonstrated insignificant differences between within-genre excerpts, these data were tabulated and collapsed into the six genre categories (presented in Fig. A2.2).

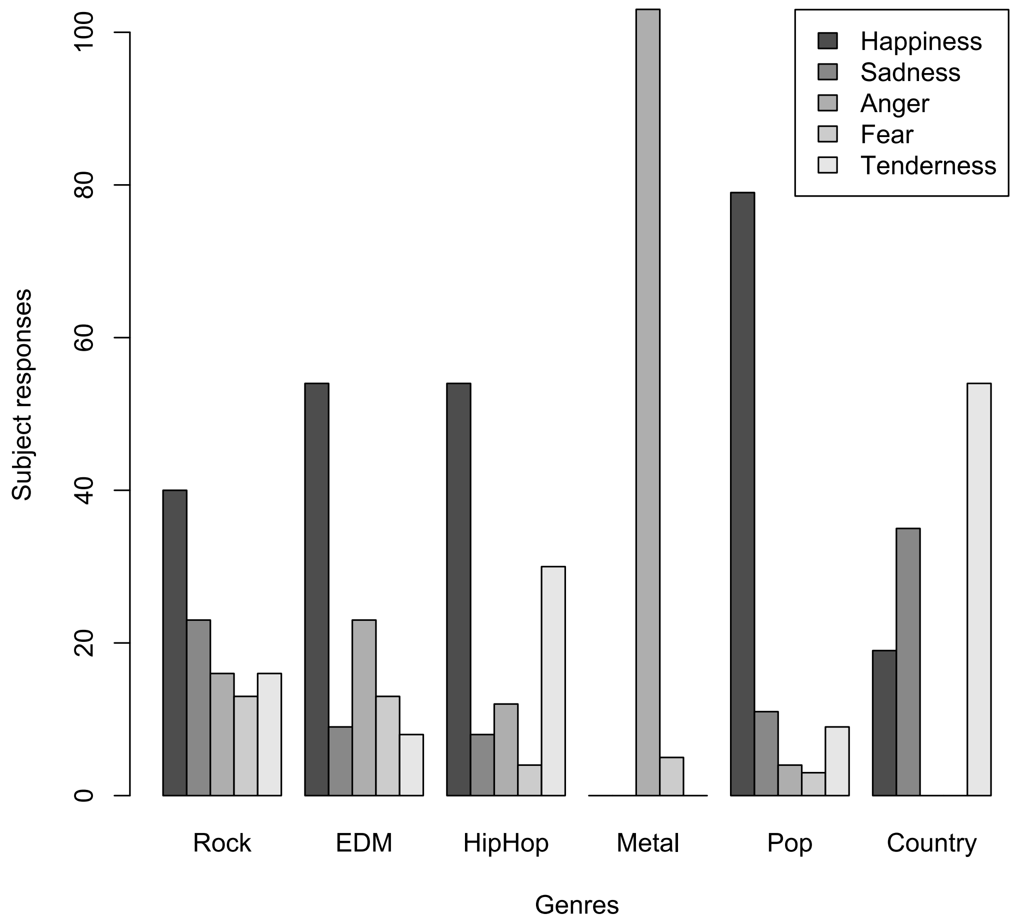


FIGURE A2.2: Histogram of "emotion conveyed" responses by genre

Next, correlations between perceptual rating means and "emotion conveyed" scores were generated. Table A2.4 is a matrix of all such correlations divided among embodiment, affect, semantic, and "emotion conveyed" dimensions. We can see that many significant correlations obtain between the behavioral variables. To present a few notable findings (a full discussion can be found in Chapter 2), *bodily exertion* is negatively correlated with both *affective valence* (r = −.67) and "brightness" (−.65), and also unsurprisingly with the more subdued emotions of *sadness* (−.61) and *tenderness* (−.86). It is also strongly correlated with p-noise—at r = .97, almost a perfect correlation—and with *anger* (.87) (Fig. A2.3). From this first quality we can already begin to see two binary categories emerging: *positively valenced* affective evaluations, semantic qualities, and emotions are grouped together, while *negatively valenced* categories of the same are likewise strongly correlated.

TABLE A2.4: Matrix of inter-scale correlations

Dimensions:	*Embodiment*	*Affect*	*Semantic*		*Emotion Conveyed*				
	E	V	B	N	H	S	A	F	T
Exertion		**−.67**	**−.65**	**.97**	−.34	**−.61**	**.87**	.35	**−.86**
Valence	**−.67**		**.76**	**−.70**	**.70**	.34	**−.85**	−.03	.36
Brightness	**−.65**	**.76**		**−.74**	**.71**	**.47**	**−.89**	−.16	−.38
Noisiness	**.97**	**−.70**	**−.74**		−.37	**−.69**	**.90**	.38	**−.85**
Happiness	−.34	**.70**	**.71**	−.37		−.17	**−.64**	−.13	−.07
Sadness	**−.61**	.34	**.47**	**−.69**	−.17		**−.58**	−.14	**.68**
Anger	**.87**	**−.85**	**−.89**	**.90**	**−.64**	**−.58**		.17	**−.62**
Fear	.35	−.03	−.16	.38	−.13	−.14	.17		**−.50**
Tenderness	**−.86**	.36	−.38	**−.85**	−.07	**.68**	**−.62**	**−.50**	

(N = 35)

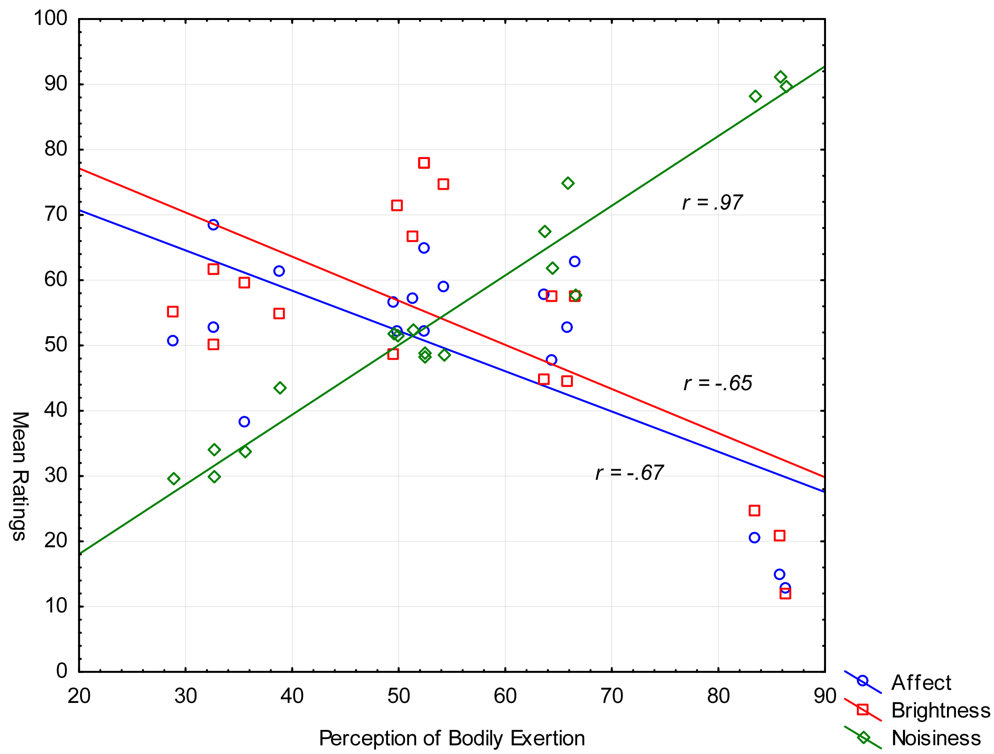


FIGURE A2.3: *Bodily exertion* correlated with *affective valence*, "brightness," and p-noise

This binary configuration is clear from the *valence* results as shown in Table A2.5: "brightness," *happiness*, and *tenderness* are all positively related to "liking," while *exertion*, p-noise, and *anger* are related to "disliking." *Fear*, another negative emotion, does not correlate with *valence*, but it does, albeit moderately, with *exertion* (.35) and p-noise (.38), as well as exhibiting a statistically significant (p<.05) negative correlation with *tenderness* (–.50). For this reason it is included in the "negative" group. Intriguingly, *sadness* is positively correlated with *valence*, as well as the other "positive" qualities of "brightness" and *tenderness*, indicating the complexity of this emotion in musical representation. ("Sad music" may be very well liked, even though the emotion conveyed is ostensibly unpleasant to actually experience.) This challenges the usual bipolar labeling of *happy*–*sad* employed in many studies, since both are positively valenced. The relationship between the two valenced categories is demonstrated perhaps most vividly in the *anger* correlations, all of which (with the exception of *fear*) are statistically significant. The ramifications of this relationship between positive and negative perceptual categories are discussed in depth in Chapter 2.

TABLE A2.5: Behavioral variables arranged into two valence factors, correlated with *affective valence* ratings

	Positive	*Negative*
	Valence (1)	Exertion (−.67)
	Brightness (.76)	Noisiness (−.70)
	Happiness (.70)	Anger (−.85)
	Sadness (.34)	Fear (−.03)
	Tenderness (.36)	

Briefly, it should also be noted that the results above resonate in interesting ways with the "emotion conveyed" responses for each genre (Fig. A2.2). It is clear from the histogram, for example, that heavy metal is the outlier genre in terms of polarity of response: a full 95% of all responses indicate the predominant emotion of *anger*. It is no shock to see, then, that metal is far and away the most *embodied* genre (in terms of perceived physical exertion required to perform it), with a mean rating of 85.2; the least *liked* genre (M = 16); the least "bright" genre (M = 19.1); and the most "noisy" (M = 89.6). (For a summary graph comparing where the other genres fall along these scales, see Fig. A2.4). These data support the sociological and musicological literature on heavy metal (see Weinstein, 1991; Walser, 1993; and Bryson, 1996). Also of note, Pop received the most *happiness* responses (73%), with very little of any other single emotion represented, and also the highest "brightness" score (M = 74.5). Country, on the other hand, had the highest *sadness* and *tenderness* responses (32% and 50%, respectively) and the lowest *exertion* (M = 32.4) and p-noise (M = 31) ratings.

Following inter-acoustic and inter-perceptual rating correlations, another correlation analysis was performed to assess the relationship between acoustic parameters and behavioral data (Table A2.6). There are fewer of the "slam dunk" correlations here than can be found in Table A2.4, but certain patterns are worthy of note. First, generally speaking there is a consistency of direction between correlations across the acoustic parameters and the positively/negatively valenced perceptual categories. Thus, *valence* is negative across the board on all a-noise parameters, as is "brightness" and *sadness*. It is clear, moreover, that the perception of "brightness" in these 400 ms excerpts differs markedly from psychoacoustic "brightness," defined as energy in the high frequency range of the spectrum. There are a number of possible reasons for this. For one, "brightness" in most timbre studies is systematically isolated and manipulated, whereas here it is in constant interaction with the other acoustic variables. Further, it may be the result of complex semantic, cultural, and even synaesthetic associations. Heavy metal, for example, was rated very low in "brightness" despite relatively high spectral centroid measurements, likely because of the associations of metal music with darkness and evil and the penchant of adherents to wear black. Conversely, Pop—the "brightest" genre—might be labeled as such for its emotional connotations (*happiness*), commercial appeal, cheerful, "bubblegum" quality, and so on. In short, there is no discernable relationship between "brightness" ratings and spectral centroid, and in fact they are negatively correlated. As mentioned previously, inharmonicity in MIRtoolbox is designed for monophonic signals only, so the numbers here are somewhat meaningless, but another a-noise measure, ZC, does show correlations with *exertion*, p-noise, *anger*, and *fear* (and, following our binary scheme from earlier, concomitant negative correlations with *valence,* "brightness," *sadness*, and *tenderness*), some of which at relatively strong levels of association.

TABLE A2.6: Correlations between mean perceptual ratings and selected acoustic qualities

	SC (MIR)	(MED)	I	SF	ZC	R
Exertion	.25	.42	−.01	.24	**.51**	**.65**
Valence	−.11	−.12	−.27	.11	−.38	**−.76**
Brightness	−.15	−.19	−.29	.07	−.33	**−.57**
Noisiness	.33	**.51**	.09	.27	**.58**	**.63**
Happiness	.07	.32	−.09	.39	.02	**−.64**
Sadness	**−.51**	**−.69**	−.22	−.40	**−.61**	−.12
Anger	.21	.28	.18	.05	**.47**	**.71**
Fear	.40	.36	−.22	.04	.41	.13
Tenderness	−.32	**−.66**	.05	−.40	**−.66**	−.38

(N = 35)

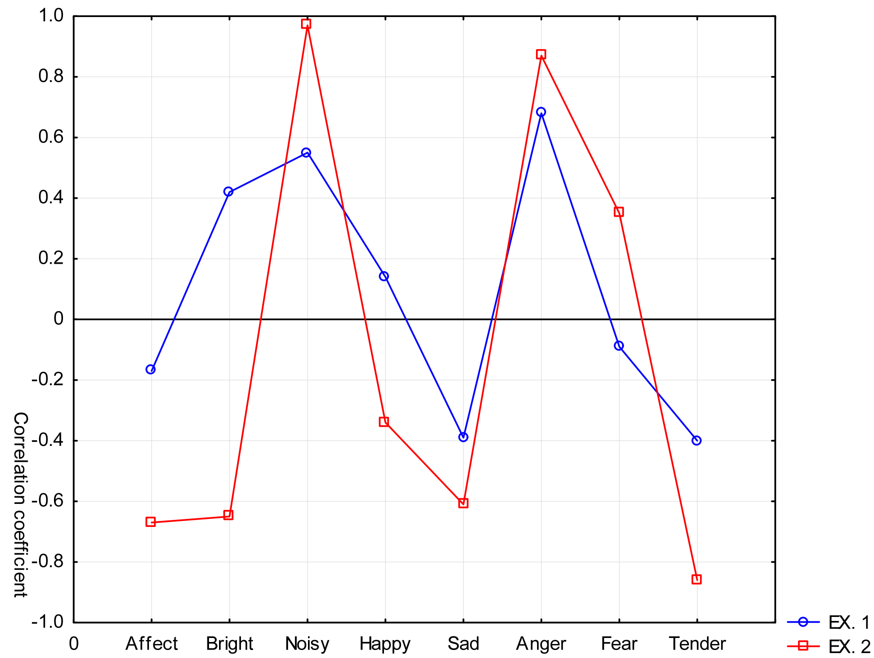
But the most substantial acoustic factor in these perceptual evaluations is roughness. Again, this parameter follows the binary pattern that is now becoming familiar: negative correlations with "positive" categories and positive correlations with "negative" categories. A step-wise multiple regression analysis, as summarized as standardized beta-coefficients in Table A2.7, bears out the significance of R in the perceptual evaluations below, with behavioral variables that fail to meet statistical significance ($p<.05$) omitted. As indicated, a high proportion of the rating means for *exertion*, *valence*, p-noise, *anger*, and *tenderness* could be explained by the six acoustic parameters. For instance, 80% of variance in *exertion* scores could be explained by these acoustic qualities alone, with relative weightings given below, R predominant among them, along with SC (MED) [$R^2 = .80$, $F(6, 11) = 7.53$, $p<.002$]. *Valence* (64% of variance explained) is similarly driven by R (more specifically, it is driven by *low* R numbers); p-noise (80% variance explained) is determined predominantly by SC (MED) and R; *Anger* (63% explained) by R and, to a lesser extent, SC (MED); and *Tenderness* (90% explained) by low SC (MED) and R values. In this regression analysis, R is the only acoustic parameter to reach significance in all perceptual categories.

TABLE A2.7: Standardized beta coefficients (β) as a result of multiple regression on acoustic dimensions

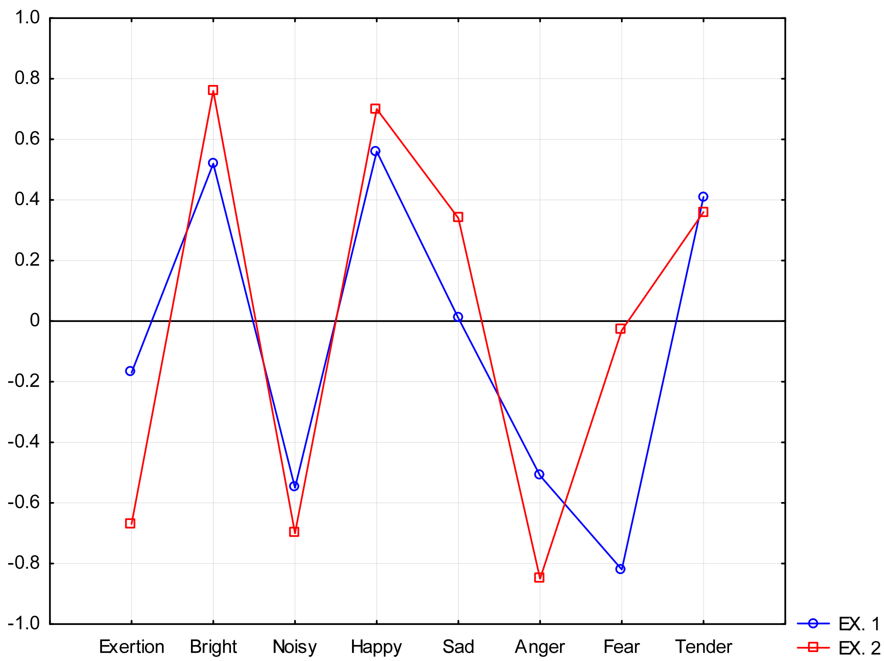
	Exertion $R^2 = .80$	Affect $R^2 = .64$	Noisiness $R^2 = .80$	Anger $R^2 = .63$	Tenderness $R^2 = .90$
SC (MIR)	−.09	.14	−.07	−.10	.29
SC (MEDS)	.69	−.24	**.85**	.50	**−.90**
Inharmonicity	**−.44**	−.20	−.33	−.06	**.55**
S Flatness	.21	.19	.07	−.06	−.17
Zero-Cross	−.12	−.07	−.16	−.03	−.17
Roughness	**.83**	**−.69**	**.80**	**.75**	**−.51**

Comparison with Experiment 1 and Brief Conclusion

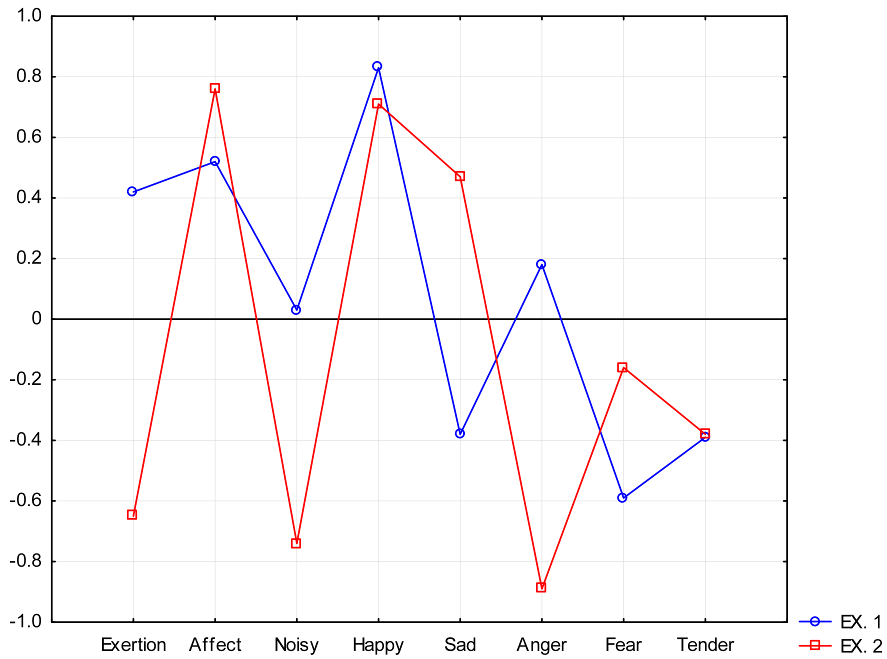
These results extend the major findings of Experiment 1, indicating that polyphonic timbre has much the same properties of perception in regard to questions of embodiment, affective valence, and certain semantic and emotional qualities as monophonic timbre. Moreover, the correlations in this study are much stronger than Experiment 1. Thus, where in Experiment 1 *exertion* correlated with p-noise at $r = .55$, here the association is $r = .97$; where *valence* correlated moderately to *happiness* in Ex. 1 ($r = .56$), here it is a robust $r = .70$. There are exceptions to this: for examples, "brightness" correlated positively to *exertion* in Experiment 1, while correlating negatively here, but the general pattern holds. Figure A2.4 compares correlation coefficients between all perceptual conditions in both Experiment 1 (Ex. 1) and Experiment 2 (Ex. 2).



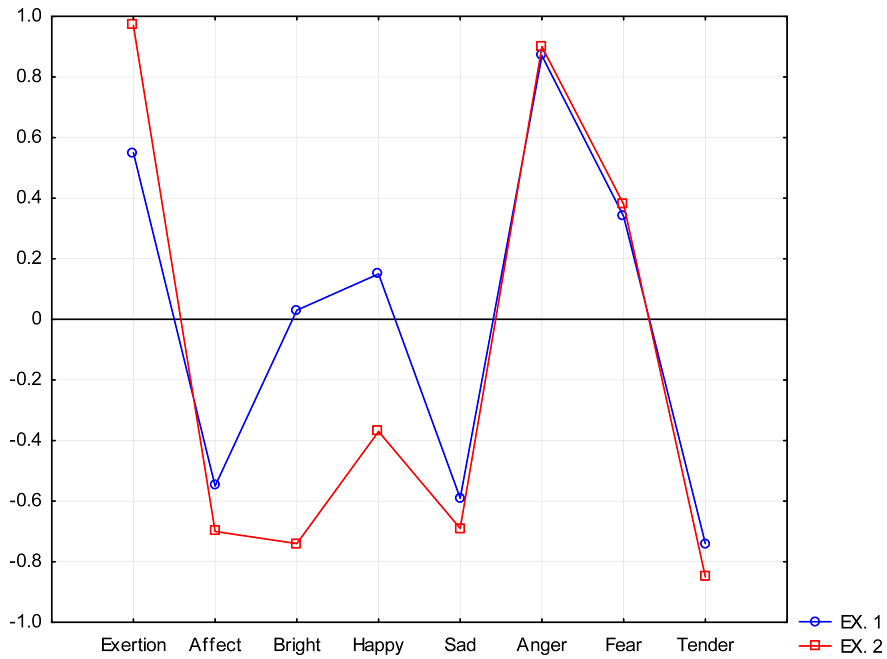
Bodily exertion



Affective valence



Brightness



Noisiness (p-noise)

FIGURE A2.4: Comparison of correlations between Experiments 1 and 2

Figure A2.5 shows a summary of the six genres in the four perceptual quality conditions. As hypothesized, the genres with the highest levels of both a- and p-noise (Metal, Rock, and EDM) tended to be heard as imbued with greater bodily exertion than the other three genres (Pop, Hip-Hop, Country, in descending order of *exertion* means). Further, as in Experiment 1, signals with higher mean *exertion* scores tended to show lower *valence* numbers than signals with lower *exertion* scores. This can be seen plainly here: the three "high exertion" genres listed above—all of which received mean ratings of over 60—exhibit a decline in *valence*, while the genres with lower exertion scores (c. 30–50) demonstrate an increase in valence. Moreover, the slope of the change between these two conditions seems to be related to the distance in the *exertion* condition between the actual mean and some hypothetical "ideal" level of exertion (around $M = 55$). For example, Metal, with a mean rating in the 80s, falls precipitously in the *valence* condition even as Rock and EDM, with means in the low 60s, decline only moderately, with similar inverse patterns at play in the lower *exertion* genres. Like in Experiment 1, "brightness" revealed a good deal of ambiguity here, though scores for the two extremes of the spectrum—Pop and Metal—may be related to the semantic and cultural associations discussed earlier. Illustrating the strength of correlations rooted in categorizations of positive and negative qualities, however, we can see that the perceived "noisiness" (or p-noise) condition brings us back to essentially where we started: the lowest *exertion* genre (Country) is also the lowest p-noise genre, the highest (Metal) is back to being the highest, and the other genres sort themselves into the same order as the *exertion* condition. In fact, the extreme variation in the Metal ratings between conditions exemplifies the correlations discussed earlier—*exertion* and p-noise are close to the same (85.2 and 89.6, respectively), as are *valence* and "brightness" ratings (16 and 19.1). The shape of this relationship between perceptual variables and signals resembles Fig. A1.5 in Appendix 1 in striking ways, just with greater spread, higher correlations, and lower standard deviation: high *exertion* is related to high p-noise and low *affect*, with “brightness” playing an indeterminate role.

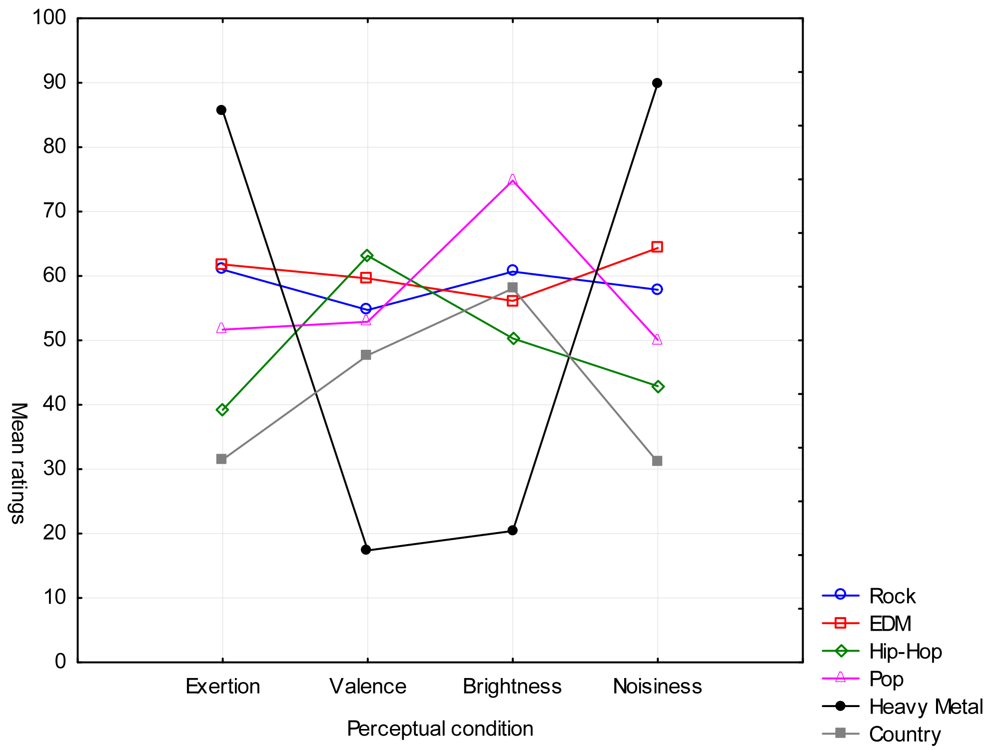


FIGURE A2.5: Summary graph of mean ratings for six genres in four perceptual conditions

Appendix 3

Experiment 3: Timbre and Motor Resonance: An fMRI Study

Introduction

As reviewed in Chapter 1, Overy and Molnar-Szakacs's "Shared Affective Motion Experience" (SAME) model suggests that "musical sound is perceived not only in terms of the auditory signal, but also in terms of the intentional, hierarchically organized sequences of expressive motor acts behind the signal. Thus, the expressive dynamics of heard sound gestures can be interpreted in terms of the expressive dynamics of personal vocal and physical gestures" (Overy & Molnar-Szakacs 2009, p. 492). In this experiment, the SAME model is put to the test in the context of timbre perception and cognition. Using fMRI, we examined the role of the human motor mirror neuron system in the processing of acoustically and perceptually "noisy" aspects of musical timbre (a- and p-noise, respectively) and its link to limbic activity indicative of affective response.

Stimuli

Twelve 2-second recordings of instrument and vocal timbres were created. Stimuli used in this experiment were the same as Experiment 1:

We recorded 3 electric guitar, 3 tenor saxophone, 3 *shakuhachi* flute, and 3 female vocal timbres. Within each category of sound producer, the three different signals were scaled according to increasing levels of "noisiness" defined with the musicians' assistance according to the specificities of each source: (1) the low-noise condition, (2) medium-noise condition, and (3) high-noise condition. For the electric guitar, condition 1 corresponded to a "clean" tone (as in jazz), condition 2 to moderate distortion (as in rock and Chicago blues), and condition 3 to heavy distortion (as in heavy metal and punk); for the saxophone, condition 1 corresponded to a regular tone (as in traditional jazz), 2 to slight growl (as in R&B and rock), and 3 to a harsh growl (as in free jazz); for the shakuhachi, condition 1 corresponded to a regular tone (*zennon*), 2 to slight air turbulence (*sasane*), and 3 to heavy air turbulence (*muraiki*); and for the voice, condition 1 corresponded to a sung vowel "O" (no vibrato), 2 to a sung vowel "A" with additional breathiness, and 3 to a sung vowel "E" with throat tension causing additional raspiness. The instruments and specific techniques selected reflect the primary timbral features discussed in the dissertation. They were also selected to represent a range of different material means of production, including electro-acoustic (electric guitar), reed (saxophone), air-stream (shakuhachi), and vocal production.

The pitch for all signals was B-flat3 (233 Hz), and loudness was equalized manually. Signals were recorded specifically for this experiment using Apple Logic software, a Neumann TLM-103 microphone, and a MOTU 896 interface at a sampling rate of 44.1 kHz. Signal transients were unaltered for maximum ecological validity. Additional details are as follows:

> Electric Guitar: Guitar signals were recorded with a 2000 Fender Deluxe Stratocaster (made in Mexico), performed by the author. A Dunlop medium thickness pick and D'Addario nickel-plated steel string (.10) were used. The amplifier was a Vox AC30C2 30-watt 2x12 combo, with all potentiometers set to "5" for balanced equalization. No additional modifications were made to the amplifier. This clean signal was edited in Apple GarageBand for the addition of digital distortion. GarageBand guitar presets were selected due to the program's ubiquity and easy reproducibility. For condition 1, the preset "Classic Blues" was selected; condition 2 used "Seattle Sound"; and condition 3 used "Heavy Riffing."
>
> Tenor Saxophone: Saxophone signals were recorded with a Selmer Super Balance Action tenor saxophone manufactured in 1951, performed by a professional jazz saxophonist. A Vandoren V16 ebonite mouthpiece and Vandoren ZZ #3 reed were used. For condition 2, a moderate growl was produced by performing a vocal growl in the back of the throat, similar to a hard "R" in French. For condition 3, an intense growl was produced by both vocally growling and humming an unspecified pitch while blowing.
>
> Shakuhachi: Shakuhachi signals were recorded with a 1.8 *ji-nashi* shakuhachi flute made by John Kaizan Neptune in Kyoto in the late 1990s or early 2000s, performed by the author. Condition 1 was a regular, unadorned *zennon* tone on the note *RO* (D4). In condition 2, the tongue partially perturbed the airstream to produce additional breathiness (*sasane* technique, or "bamboo grass blowing"). And in condition 3, the tongue fully perturbed the airstream to produce a pitch-occluding admixture of noise (*muraiki* technique, or "thrashing breath"). To conform the signals to the same frequency as the other instruments, the Audacity "Change Pitch" function was used to lower the signals from D4 to B-flat3.
>
> Vocals: Vocal signals were recorded with a female singer with formal musical training and professional performance background in opera and avant-garde concert music.

Subjects

Fifteen participants (8 female, 18–20 years old) were recruited from the UCLA community. Subjects had a range of musical backgrounds, from no formal training to over 10 years. Participants were ethnically diverse, right-handed, had normal or corrected-to-normal vision, no history of neuropsychiatric disorder, and were not taking psychoactive medication. Written informed consent was obtained from all participants and they received monetary compensation or course credit for their time. The study was approved by the UCLA Institutional Review Board.

Procedure

Subjects listened to the randomized stimuli while being scanned. A sound check prior to the functional scan (conducted with the scanner running) allowed subjects to adjust the headphone volume to a subjectively determined comfortable listening level averaging approximately 90 dB. Participants were then instructed to relax and keep their heads still while listening to the stimuli, and to keep their eyes open and their vision trained on a fixation cross presented through magnet-compatible LCD goggles. Each 2 s stimulus was repeated 6 times with 10 ms of silence between each onset (approximately 12 s total per timbre). After the presentation of all 12 stimuli, subjects were given a 16 s baseline period of silence. The full block was repeated 3 times.

MRI Data Acquisition, Preprocessing, and Statistics

Images were acquired on a Siemens 3T Trio MRI scanner. Functional runs comprised 231 T2*-weighted echoplanar images (EPIs) [repetition time (TR) 2000ms; echo time (TE) 28ms; flip angle=90°; 34 slices; slice thickness 4mm; matrix 64 x 64; FOV 192mm] sensitive to blood oxygenation-dependent (BOLD) contrast. To enable T1 equilibrium the first two volumes of each functional scan were automatically discarded before data collection commenced. Additionally, two sets of structural images were acquired for registration of functional data: a T2-weighted matched-bandwidth high-resolution scan with the same slice prescription as the EPI [repetition time (TR) 5000ms; echo time (TE) 34ms; flip angle=90°; 34 slices; slice thickness 4mm; matrix 128 x 128; FOV 192mm]; and a T1 weighted magnetization prepared rapid-acquisition gradient echo image (MPRAGE) [TR, 1900ms; TE 2.26ms; flip angle = 9°; 176 sagittal slices; slice thickness 1mm; matrix 256 x 256; FOV 250mm]. Audio stimuli were timed and presented with Presentation software through noise-cancelling, magnet-compatible SereneSound headphones. In addition, subjects wore disposable foam earplugs to help minimize scanner noise.

Image preprocessing and data analysis were performed with FSL version 3.1.8. Images were realigned to the middle volume to compensate for any head motion using MCFLIRT (Jenkinson et al., 2002). Volumes were then examined manually for gross motion artifacts that cannot be corrected with simple realignment. When motion artifacts were detected, a nuisance regressor for each affected volume was included in the general linear model. Data were temporally filtered with a high-pass filter cutoff of 100 seconds and spatially smoothed with a 8mm full width half maximum Gaussian kernel in three dimensions.

Statistical analyses were performed at the single subject level using a general linear model (GLM) with fMRI Expert Analysis Tool (FEAT, version 5.98). Contrasts specified included the following: (1) each of the 12 individual stimuli > baseline, (2) inter-instrument comparisons [Gtr 3 > Gtr1, Gtr 2>1, Gtr 3>1, etc.], (3) inter-timbre comparisons [all condition 3 > all condition 1, etc.], and (4) all timbres > baseline. First level contrast estimates were computed for each run and then registered to standard space (Montreal Neurological Institute, MNI) in three stages. The middle volume of each run of individual EPI data was registered first to the co-planar matched-bandwidth high-resolution T2-weighted image. Following this, the co-planar volume was registered to the T1-weighted MPRAGE. Both of these steps were carried out using FLIRT (affine transformations: EPI to co-planar, 6 degrees of freedom; co-planar to MPRAGE, 6 degrees of freedom) (Jenkinson & Smith, 2001; Jenkinson et al., 2002). Registration of the MPRAGE to

MNI space (FSL's MNI Avg152, T1 2x2x2mm) was carried out with FLIRT (affine transformation, 12 degrees of freedom) and refined using FNIRT (non-linear transformation). Contrast estimates for each subject were then computed treating each run as a fixed effect. Finally, a group-level analysis was performed to calculate a group mean for each contrast treating each subject as a random effect using FSL FLAME (FMRIB's local analysis of mixed effects) stage 1 and stage 2 (Beckmann, Jenkinson & Smith 2003; Woolrich et al., 2001). Except where noted, all images were thresholded at $Z > 2.3$ ($p<.01$), corrected for multiple comparisons using cluster-based Gaussian random field theory controlling family-wise error across the whole brain at $p<.05$.

Analysis and Results

Since each stimulus produced a unique signature of neural activity against the baseline silence, it would be impractical to display images for each condition; full results of activation can be found in the Addendum. However, certain images warrant review. Figure A3.1 compares the same coronal slice of Saxophone 2 and 1 (growled > not growled). As indicated here, Sax 2 shows activation in many sensorimotor and limbic areas not active in the low-noise condition. (For a complete report of regions of activity in each of the 12 timbre > baseline contrasts, see Addendum.)

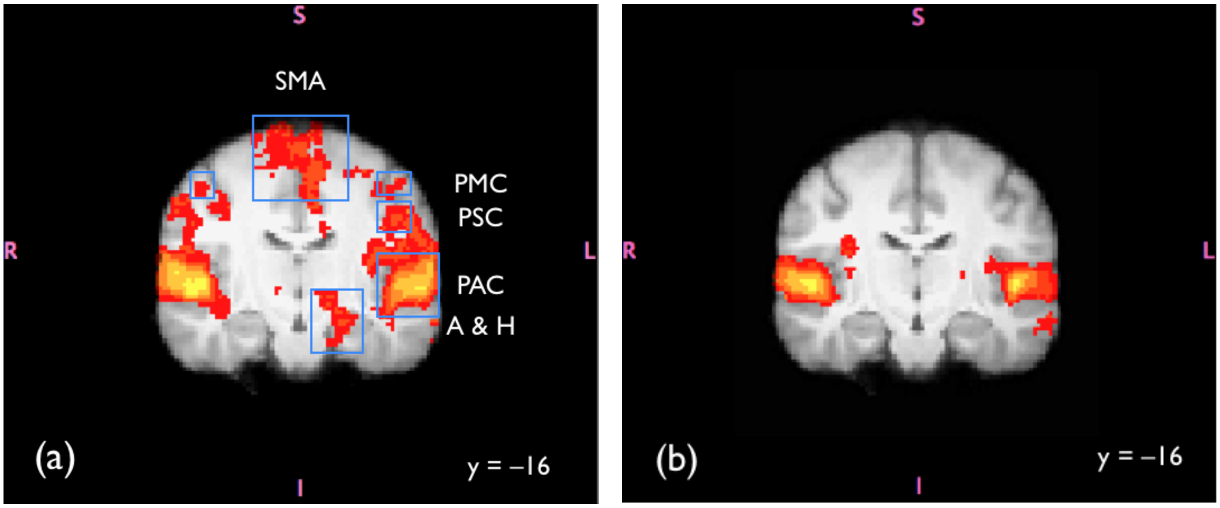


FIGURE A3.1: (a) Saxophone 2 > Baseline: activity in primary auditory cortices (PMC, Z = 5.87), left/right pre/primary motor cortex (PMC, 2.52), supplementary motor area (SMA, 4.89), primary somatosensory cortex (PSC, 3.40), and limbic areas (specifically left amygdala [A], 3.37, and hippocampus [H], 2.60). **(b) Saxophone 1 > Baseline**

Most contrasts between instruments and timbres yielded no clear results at the current threshold, but saxophone demonstrated significant differences between the three timbres. Figure A3.2 is a sagittal slice from the contrast between Sax 3 (the high-noise saxophone timbre) and Sax 1 (the low-noise condition) showing substantial cerebellum activity (Z = 3.66) in the noisy condition that is not present in the other timbre, as well as activity in the right cerebral cortex (3.33). When we contrast Sax 3 with Sax 2 (Fig. A3.3), moreover, we can see additional activity in cerebellum (3.27), primary motor cortex (3.71), and superior parietal lobule (4.12), as well as some notable activity in the visual cortex (3.73). Taken together, these contrasts suggest that noise in saxophone timbre induces greater activity in a range of motor, sensorimotor, and subcortical areas. Some of this activity is consistent with the "motor mirroring" hypothesis of timbre perception: different qualities of timbre—namely, higher degrees of noise—might cause the auditor to simulate the enhanced bodily exertion required to produce these qualities of timbre. Although it remains to be seen how somatotopically specific such a process of motor mirroring might be, it is suggestive nonetheless that regions associated with the classic mirror neuron system are active in the processing of isolated timbres.

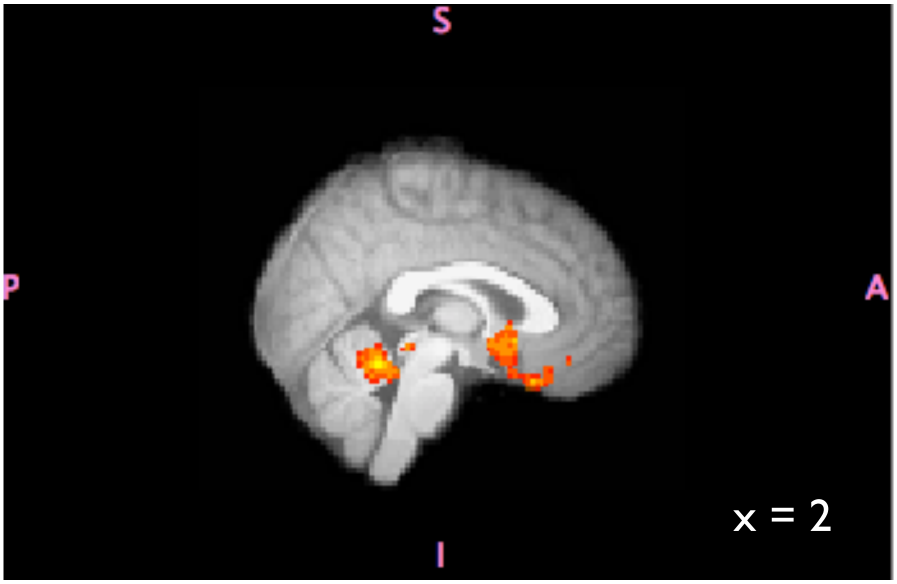


FIGURE A3.2: Sax 3 > Sax 1

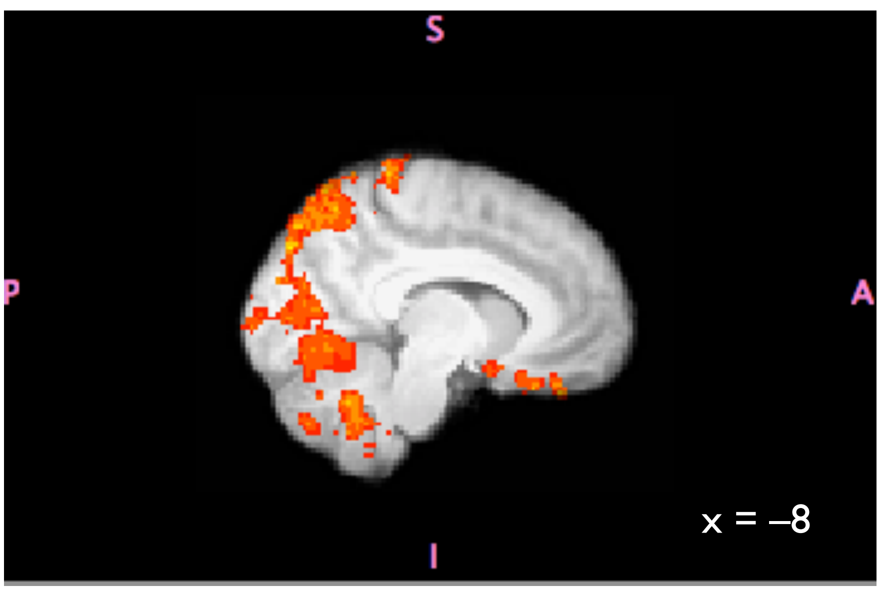


FIGURE A3.3: Sax 3 > Sax 2

Superior parietal lobule may also play a role in the perception of noisier timbre when collapsed across all instrument categories. As shown in Fig. A3.4, thresholded at Z = 1.7 (p<.05) there is significantly more activity in this region when subjects listened to Timbre 2 (the medium noisy condition) than Timbre 1 (low-noise) (Z max = 3.37).

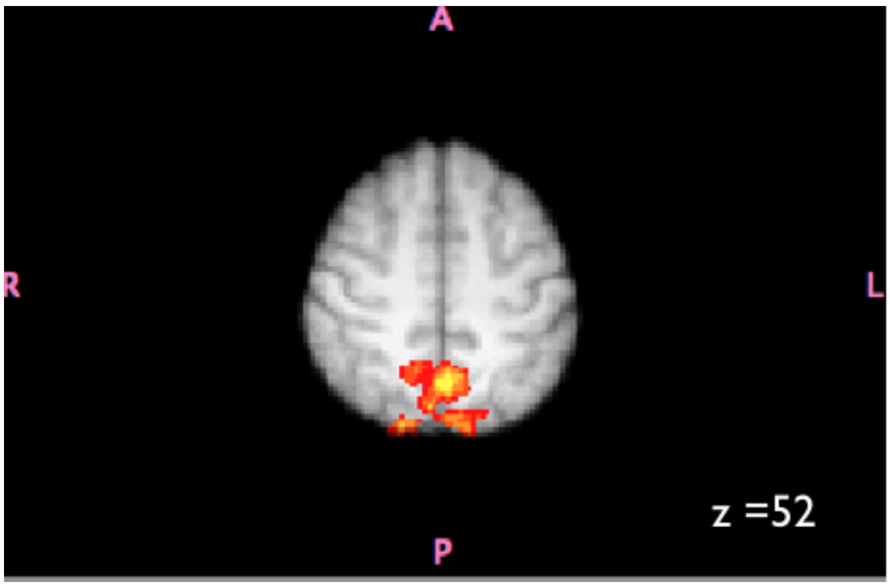


FIGURE A3.4: Timbre 2 > Timbre 1

Additionally, to evaluate the effect of different timbre conditions on vocal motor areas specifically, we performed a repeated measures ANOVA on beta estimates extracted from peak voxels in Rolandic operculum (left and right), Broca's area (left only), and pre/primary motor cortex (left and right) for all three timbres (Table A3.1). The Instrument factor had significant main effects in these regions, both alone and in interaction with the regions of interest [Instrument: F(3, 42) = 7.06, p=.0006; Instrument/vocal region activation: F(12, 168) = 3.52, p=.0001]. Timbre alone did not produce a significant main effect [F(2, 28) = 0.83, p=0.45], but in interaction with the regions of activation, the effect was "musicologically relevant" [F(8, 112) = 1.85, p=.07]. Activity in Rolandic operculum (left and right) decreased linearly between the "low-noise" and the "high-noise" conditions, and Broca's area exhibited a similar linear decrease. However, pre/primary motor cortex increased activation from the low to high noise timbres (Timbre 1 to Timbre 3). Though the right side of the cortex reached peak activation in Timbre 2, falling back to Timbre 1 levels in the third condition, the dominant, left pre/primary cortex showed a linear increase. A graph of this result can be found in Fig. A3.5; ramifications are discussed in greater detail in Chapter 2.

TABLE A3.1: Locations of peak voxels in five motor areas of interest

	Brodmann area	MNI coordinates *x*	*y*	*z*
Rolandic operculum (L)	43	−50	−6	6
Rolandic operculum (R)		48	−6	6
Pre/primary motor cortex (L)	4/6	−48	−6	44
Pre/primary motor cortex (R)		50	−8	40
Broca's area (L)	44/45	−50	14	10

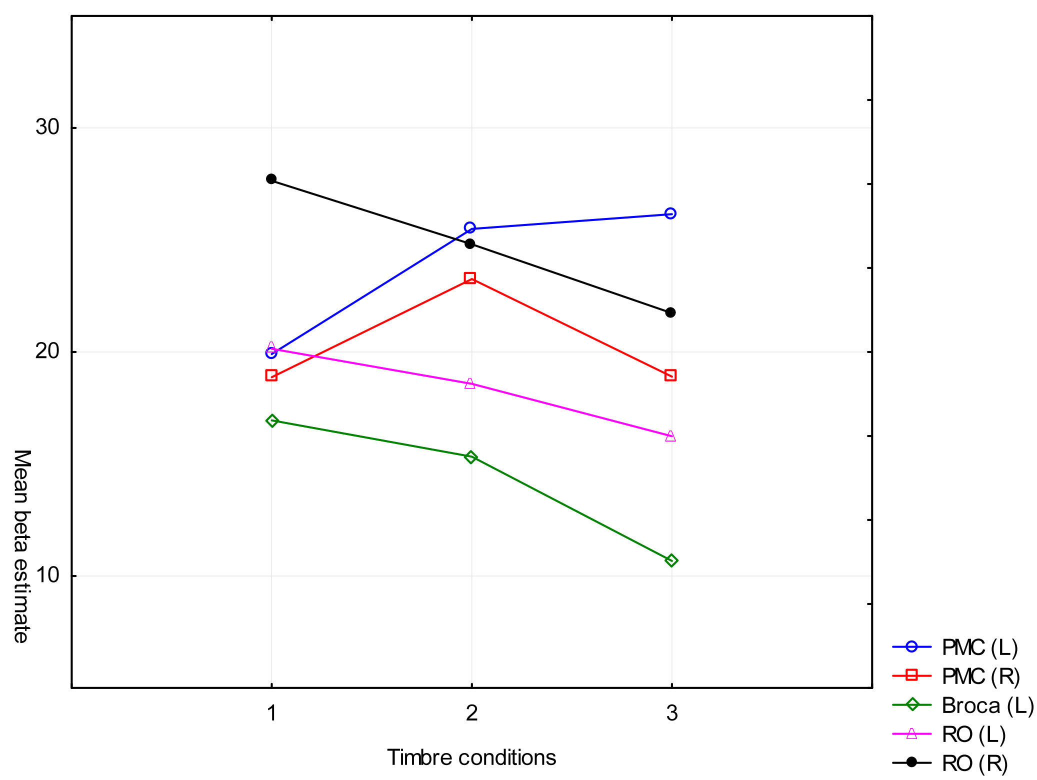


FIGURE A3.5: Main effects of timbre on activation of five motor areas of interest

Correlation Analysis with Behavioral Data

As enumerated above, functional imaging results indicated significant effects of instrument and timbre on activation in certain regions of interest. But this does not address the relationship between neurophysiology and conscious perception; brain activation alone is not sufficient evidence for experience, which can be readily quantified through behavioral measurements. To explore the relationship between neural activity and conscious evaluations of timbre, we carried out correlation analyses using two different techniques: (1) For the baseline contrasts, beta estimates for individual voxels with the highest local Z-score per cluster at the group level were correlated with behavioral data gathered from the same individuals (N = 10, see Appendix 1 for a complete summary). We opted for this method with baseline contrasts because performing a voxel-by-voxel correlation produced ambiguous results owing to the large, variance-dominating effects of the task. And (2), for the subtler contrasts that did not yield any significant results in the group analysis (e.g., Timbre 3 > Timbre 1), we performed a voxel-by-voxel correlation at the whole-brain level by adding behavioral data as explanatory variables to the FEAT models. The threshold for this analysis was set at Z = 1.7 (p<.05). Using two separate correlational approaches on baseline and more complex contrasts revealed more than would have been obtained from either method individually.

(1) Baseline contrasts (correlation of single voxel beta estimates):

First, comparing activation levels between the regions of interest on all task > baseline contrasts revealed entirely positive correlations between different structures of the brain, most of which at significant levels (*p*<.05). Thus, for example, left Broca's area (BA44/45) correlated positively (*r* = .19 to .42) with activity in all other regions except brainstem. This seems to indicate that the 12 timbre signals engage *all* these covarying areas of interest together, though the relative level of "perceptual load"—i.e., the total activation strength of these regions—differs considerably from signal to signal. The phenomenon of co-activation of multiple regions as a response to a stimuli is not unique to timbre processing, of course, but it is worthy of mention regardless.

Next, we correlated ratings for *bodily exertion*, *affective valence*, "brightness," and p-noise ("noisiness") with the beta values. Results with "musicologically relevant" correlations (*p*<.10) are shown in Table A3.2; statistically significant correlations (*p*<.05) are indicated with bold typeface.

TABLE A3.2: Correlations between behavioral data and activation in selected regions

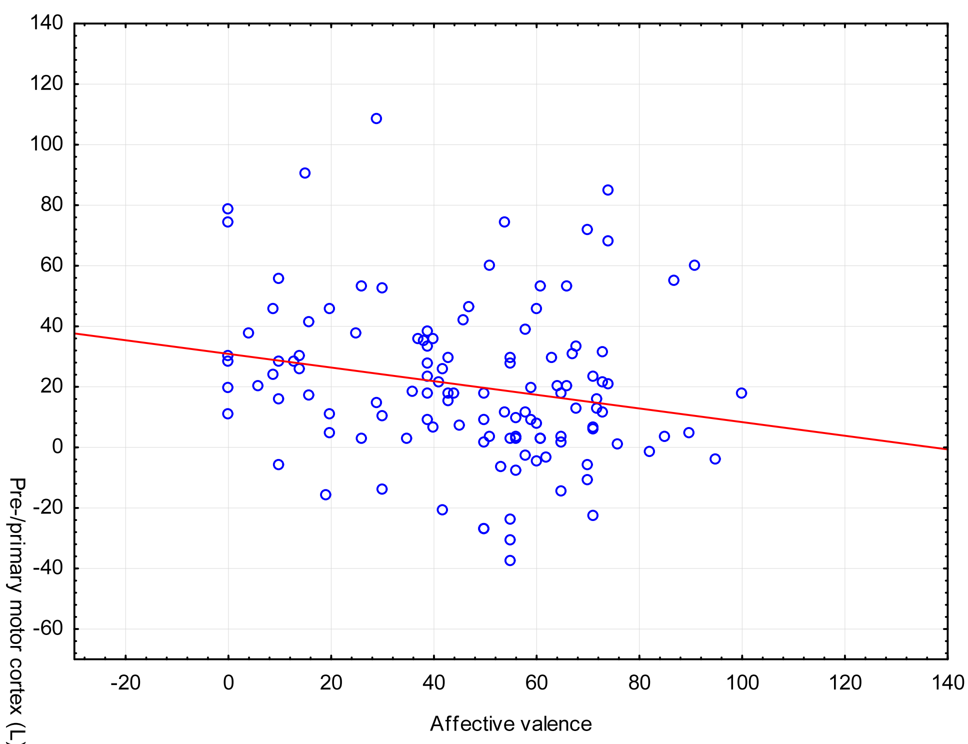
		MNI coordinates			
Behavioral variable	*Area*	*x*	*y*	*z*	*Correlation*
Bodily exertion	Auditory cortex (L)	−38	−38	12	.16
	Heschl's gyrus (R)	38	−26	12	.15
Affective valence	Auditory cortex (L)	−38	−38	12	**−.19**
	SMA	8	−2	66	.16
	Pre/primary motor cortex (L)	−48	−6	44	**−.21**
	Rolandic operculum (L)	−50	−6	6	**.19**
	Rolandic operculum (R)	48	−6	6	**.30**
Brightness	Auditory cortex (L)	−38	−38	12	**−.19**
	Inferior parietal lobule (L)	−48	−38	20	**−.22**
	Broca's area (L)	−50	14	10	−.18
	Brainstem	12	−28	−14	**−.23**
Noisiness	Rolandic operculum (R)	48	−6	6	**−.18**
	Heschl's gyrus (R)	38	−26	12	−.16

N = 10. Correlation analysis carried out with 120 individual behavioral ratings (10 subjects x 12 signals) and associated patterns of individual brain activation (using beta estimates for voxels with cluster Z-score maxima). Bold indicates significance at p<.05; regular typeface indicates "musicological relevance" at p<.10.

While none of the *r*-values were strong in absolute terms, especially when compared to correlations between behavioral variables (and between behavioral and acoustic data; see Appendix 1), certain notable trends emerged nonetheless. As indicated in the table, perception of *bodily exertion* implied by a timbre correlates with activity in auditory cortex (particularly Heschl's gyrus). This seems to suggest that more "embodied" timbres—i.e., sounds with a greater audible sense of the physical effort that goes into producing them—require more "work" to process in the STG. Since the perception of bodily exertion in timbre is related to greater high frequency energy (or spectral centroid), this activation is possibly attributable to recruitment of a greater area of the cortical surface as a result of tonotopic organization (Wessinger et al., 1997).

Affective valence is negatively correlated with auditory cortex, as well as left pre/primary motor cortex (BA4 & 6) activity. However, this same behavioral variable positively correlates with SMA and Rolandic operculum (in both hemispheres). The two scatterplots in Fig. A3.6 display this relationship. This result indicates that "disliking" a musical timbre is associated with greater activity in some motor areas (pre/primary motor cortex) and attenuated activity in others (SMA and Rolandic operculum). Since the ventral premotor cortex (BA6) has been shown to contain mirror neurons (Iacoboni et al., 1999), it may be argued from these data that motor mirroring of timbre in this region—at least with these 10 subjects and 12 signals—accompanies aversive affective response. The timbres with the lowest valence ratings, indeed, induced the greatest BOLD signal change in the pre- and primary motor cortices. Interestingly, *affective valence* positively correlates with SMA—which is involved in imagined movements—and Rolandic operculum, which researchers have found to be involved in listening to "pleasant" music, possibly because of its role in vocal production (Koelsch et al., 2006). Presumably, activation in passive music listening tasks is related to subvocalization, or "singing along" internally to the music (Cox, 2011), just as Rolandic operculum is involved in overt vocalization (Jeffries, Fritz & Braun 2003;

Brown, Ngan & Liotti 2008). It is suggestive that isolated instrument and vocal timbres that listeners *like* appear to activate the Rolandic operculum consistent with internalized vocal production, while suppressing activity in the pre/primary motor cortex. Further, p-noise ratings—how "noisy" or "not noisy" subjects perceive the timbres to be—are negatively correlated with Rolandic operculum (left hemisphere $r = -.12$, right hemisphere $r = -.18$), as graphed in Figs. A3.6 and A3.7. Taken together, this seems to indicate a role for Rolandic operculum, the primary larynx/phonation area of the premotor cortex, in the perception of timbre: timbres that have high affective valence scores and low noisiness scores activate this area to a greater degree than timbres with low valence and high p-noise scores. Put differently, listeners appear to exhibit more vocal motor resonance when hearing timbres they both *like* and consider "not noisy." This result and its ramifications for music perception and cognition are discussed in detail in Chapters 1 and 2.



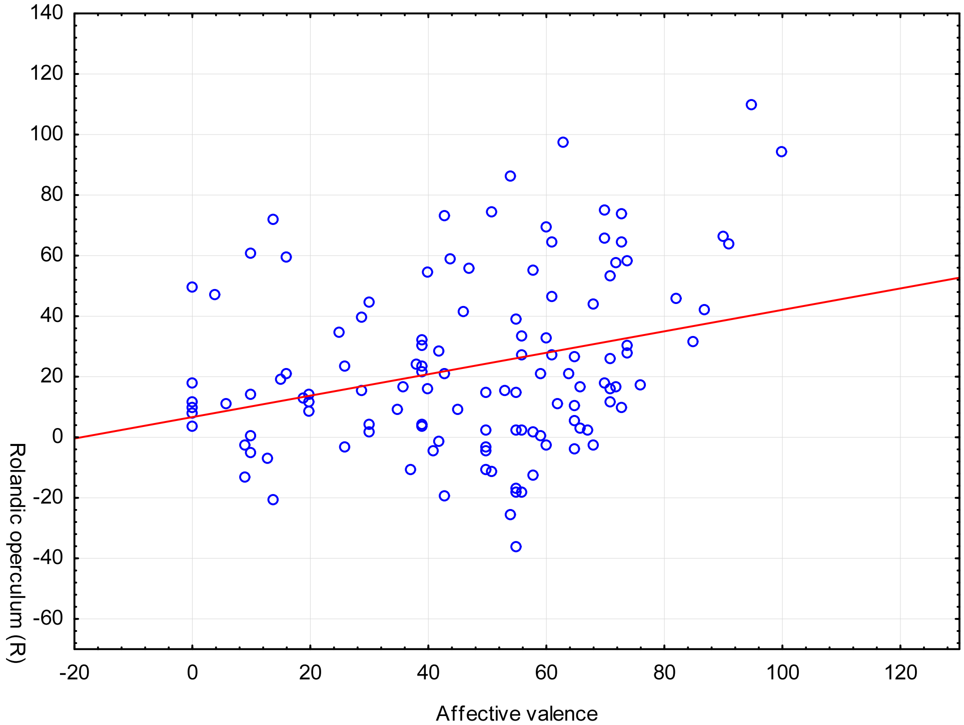


FIGURE A3.6: Correlations of pre-/primary motor cortex and Rolandic operculum with *affective valence* ratings

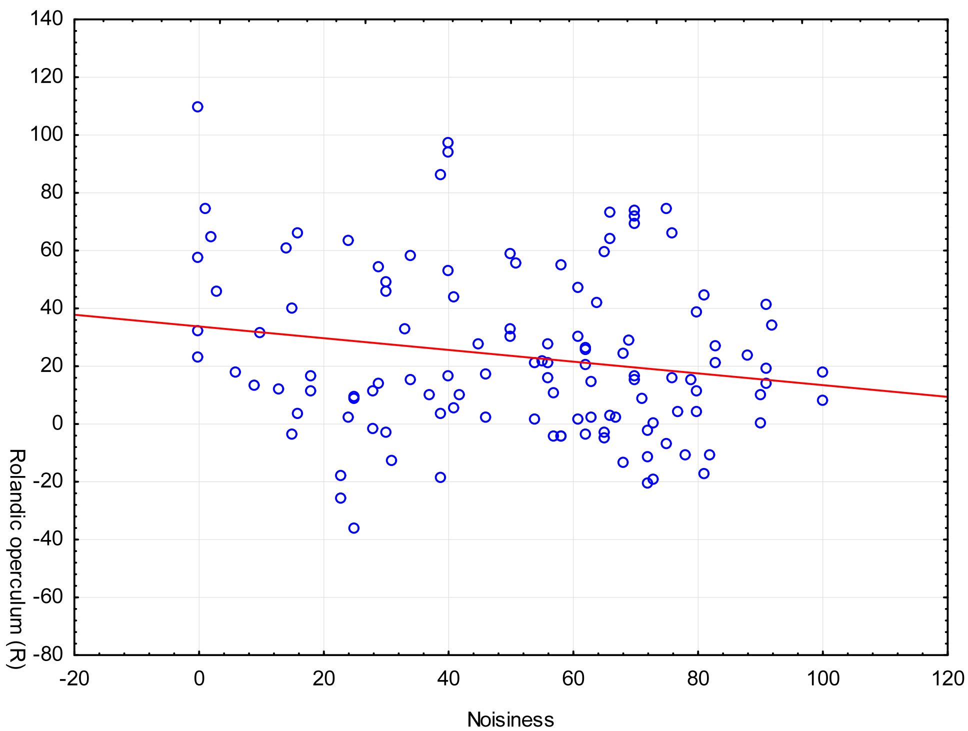
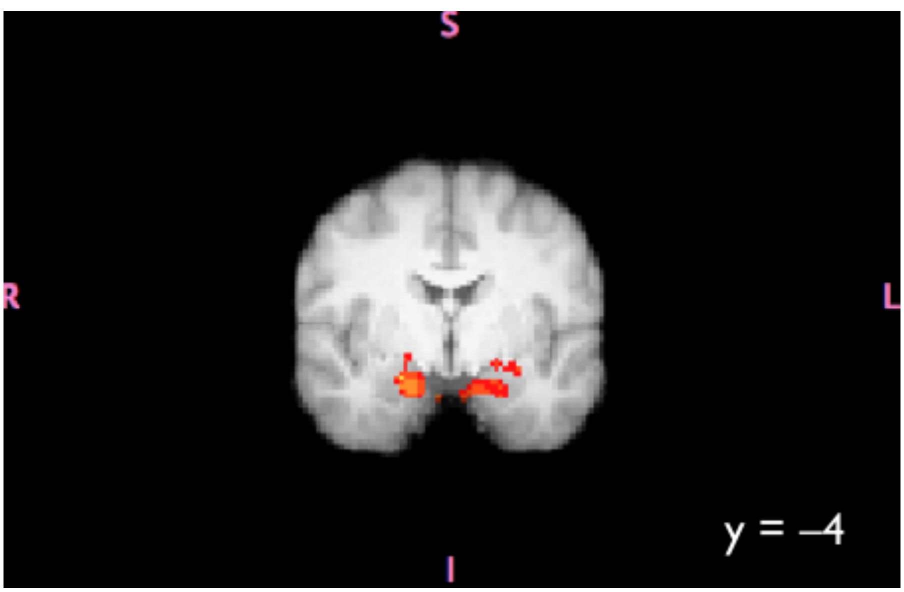


FIGURE A3.7: Correlation of Rolandic operculum (RO) with p-noise ratings

(2) Non-baseline contrasts (voxel-by-voxel correlation):

Next, we performed a voxel-by-voxel correlation on the more nuanced contrasts, including Timbre *x* > Timbre *y*, both within and between instruments.[1] For the perceived *bodily exertion* behavioral measure—i.e., ratings of the physical effort required to produce different qualities of timbre—significant correlations were found when contrasting all Timbre 3 (maximally noisy) > all Timbre 1 (not noisy). Activations were virtually entirely contained to subcortical structures, cerebellum, and limbic areas, with some activity in SPL. In particular, activity in amygdala (left and right) seemed to correlate with *bodily exertion* ratings: subjects who heard Timbre 3 as significantly more effortful than Timbre 1, in other words, also tended to exhibit greater activity in amygdala. Figure A3.8 shows this activation, which, at Z = 3.84, is the most robust correlation of the analysis.

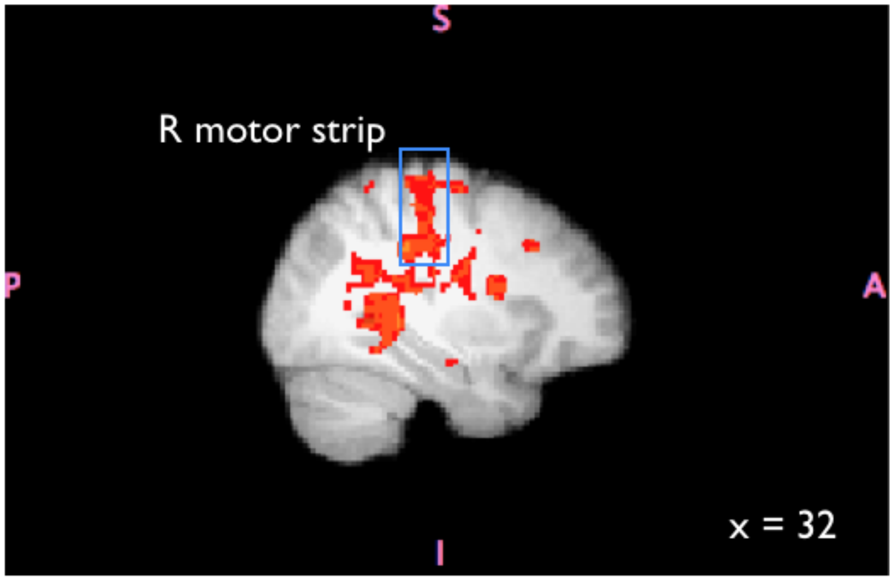


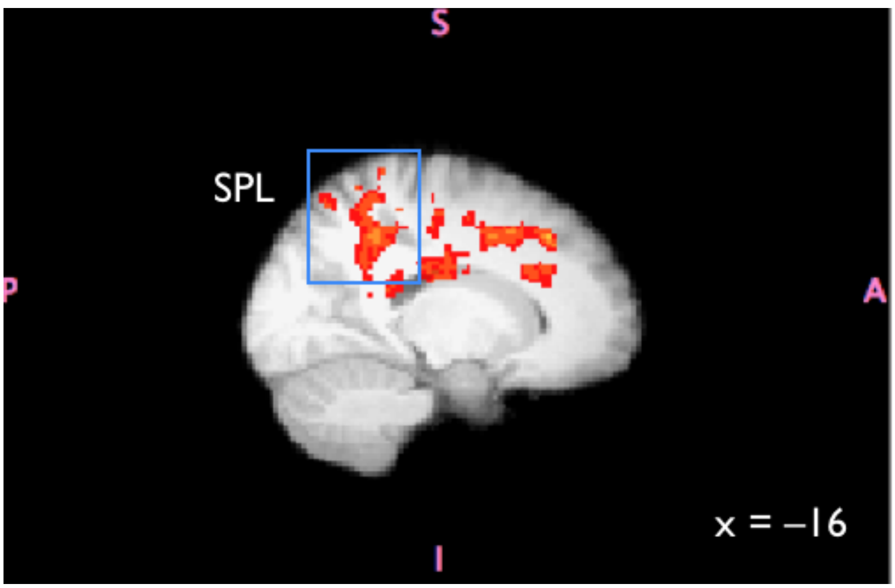
[1] Since raw behavioral data pertained only to each isolated signal, not to contrasts between signals, the explanatory variables for this correlation were calculated as the difference in ratings between one signal or set of signals and another.

FIGURE A3.8: Timbre 3 > Timbre 1, correlated with perceived *bodily exertion*

Correlating negative *affective valence* ratings (i.e., disliking) with the same contrast yielded no significant results.[2] However, adding the same *affective valence* regressor to the Timbre 2 > Timbre 1 contrast revealed increased activity in a variety of limbic (e.g., anterior insula, cingulum, parahippocampal gyrus, callosal body) and cortical regions, including right pre/primary motor cortex (Z = 2.50) and SMA (Z = 2.90). Unlike the *bodily exertion* correlation, subcortical structures were not involved. Interestingly, motor mirroring in these areas in relation to disliking is entirely right lateralized; Fig. A3.9 (top image) shows activations along the entire motor strip of the right hemisphere, while no activation difference was found in the left. It has long been suggested that emotion processing is right dominant (for a review, see Damasio, 1994 p. 140), and that processing of negative emotions in particular tends to be right-lateralized (Alfano & Cimino, 2008). The right dominance of these results is consistent with this hypothesis. We can also see significant action in SPL (middle image, Z max = 2.44), as well as parahippocampal gyrus (Blood et al., 1999). Taken together, this result converges with the single-voxel correlation to indicate that pre/primary motor cortex is involved in *negative* assessments of musical timbre.

[2] It is curious that none of these patterns appear in the 3 > 1 contrast, but since the signals were not systematically scaled for "noisiness" (a trade-off for greater ecological validity), we might simplify this along with the previous (and upcoming) correlations as a reflection of binary levels of noisy > not noisy, instead of the 3-level ordinal scale originally intended.


S
R motor strip
P
A
x = 32
I


S
SPL
P
A
x = –16
I

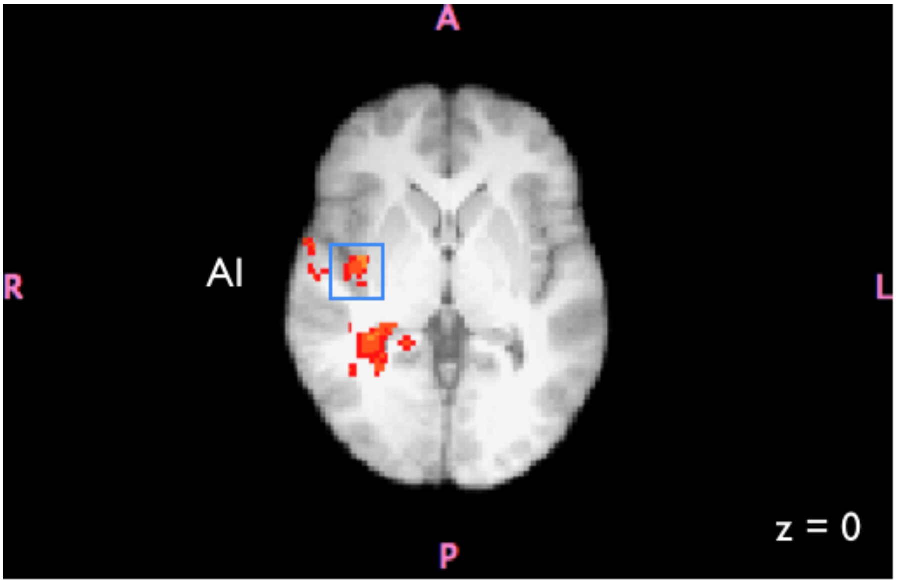


FIGURE A3.9: Timbre 2 > Timbre 1, correlated with negative *affective valence*

We posit that this activation pattern represents a motor component of the generation of negative affect, possibly reflected in the limbic activity cited above, specifically the right anterior insula (AI, bottom image, Z max = 3.01). The anterior insula is a well-documented emotion-processing area and a possible relay center between motor representations and limbic response (Carr et al., 2003; Wicker et al., 2003; for a review see Overy & Molnar-Szakacs, 2009 pp. 492-93). It has also been theorized to play an important role in musical affect, relaying mimetic motor activations to the limbic system for higher-order affective processing and appraisal (Molnar-Szakacs & Overy, 2006; Overy & Molnar-Szakacs, 2009). In addition to its involvement in *disliking* noisy timbre over not-noisy timbre when collapsed across sound generators (Timbre 2 > Timbre 1), moreover, we found that left anterior insula played a role specifically in negative affective response to vocal timbre (Vox 3 > Vox 1, Z = 2.33). Together, these results support Molnar-Szakacs and Overy's SAME hypothesis, suggesting that this region may function as a neural conduit between the MNS and limbic system. Further, this dataset provides added specificity to the SAME model: rather than simply being involved in the relay of motor representation and limbic reaction, these data imply a preferential role of anterior insula in the generation of *negatively valenced* affect. This area may play a role in disliking timbre, and possibly (by extension), in disliking *music* in general.

How do we make sense of this observed pattern? Music theorist Arnie Cox has suggested that liking or disliking music depends to a large degree on whether we like or dislike "what [the music] invites us to do" (Cox, 2011). In his "Mimetic Hypothesis," Cox contends that affect is governed by mimetic motor imagery, which is a fundamental aspect of all musical aesthesis. The experimental signals used here, however, are not *music per se*: they are simply isolated timbres with varying degrees of physical exertion required for production, and concomitantly, varying degrees of accompanying acoustical and perceptual noise. Without the usual temporal progression that marks musical expression (e.g., melodies, rhythms, etc.), motor imagery in this experimental paradigm can be seen as a product exclusively of the bodily acts implied in the production of each timbre: put differently, motor activity is embodied in the respective acoustic signatures, with acoustic noise strongly correlated with impressions of high arousal and bodily exertion (see Appendix 1). A correlation between disliking and activity in the right pre/primary motor cortex (and SMA) thus seems to point to a relationship between *intensity* of motor mirroring and intensity of negative affective response. In Cox's terms, it may indicate that these noisy timbres are "inviting" us to *do* something physically effortful—something associated with high arousal states, activity, and potency—that we do not necessarily want to do. To be clear, observing activity in pre/primary motor areas collapsed across four very different sound generators does not tell us about the specific locus of motor-mimetic engagement among listeners (for example, hands, mouth, legs, etc.): we cannot tell what specific muscular exertions are being mirrored. However, in the context of the Mimetic Hypothesis and the SAME model, we might conclude that noisy, disliked timbres are mirrored motorically in some unspecified capacity, and that this motor resonance is intimately linked to affective response. Locating the somatotopic specificities of said linkage is a task for further research.

Lastly, we added p-noise data at the whole-brain level to our model for the Timbre 2 > Timbre 1 contrast. The results are a hybrid of salient features from both the *bodily exertion* and negative *affective valence* images reviewed above. P-noise is significantly correlated with activity in limbic and subcortical areas—most notably the amygdala (A) and hippocampus (H)—*and* with right pre/primary motor cortex (Z = 2.43) and SMA (Z= 3.24). Figure A3.10 displays these activations: note also the activity in primary somatosensory cortex (Z = 2.39, to be discussed in next section). This indicates a possible relationship between certain motor areas and limbic response in the perception of "noisy" timbre, suggesting a functional link between motor mirroring of this specific perceptual feature of timbre and affective response. It is further significant that the right amygdala, as in Fig. A3.8, is the most active limbic structure (Z = 4.03, or p<.0001) of the contrast. The importance of the amygdala in affective response has long been known, particularly its role in *fear* responses and other negatively-valenced, high arousal emotions (LeDoux, 1998). This result indicates a possible association between amygdala and motor areas (pre/primary motor cortex and SMA) in the perception of timbral noisiness; that is, perceiving one isolated sound as "noisier" than another may involve the aforementioned regions.

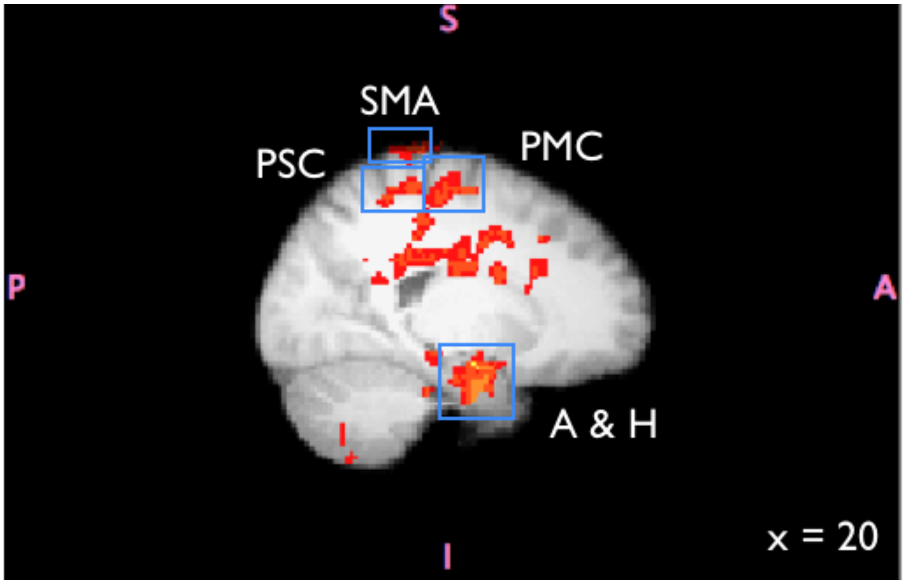


FIGURE A3.10: Timbre 2 > Timbre 1, correlated with p-noise

This result is consistent with the behavioral data collected and analyzed in Experiments 1 and 2. *Bodily exertion* and p-noise are high arousal, negatively-valenced percepts, so the greater limbic activity associated with them (particularly amygdala) should perhaps be expected. It is interesting to note, however, that *bodily exertion* correlates with amygdala activity, but not activity in actual motor areas. This might indicate that all participants, regardless the level of "bodily exertion" they reported hearing in these signals, exhibited activity in motor areas; reviewing activations in baseline contrasts (see Addendum), it is clear that most timbres elicited some statistically significant activity in motor regions. It is intriguing that behavioral measures of *bodily exertion* would relate most clearly to limbic areas associated with emotional processing (specifically negative, high arousal emotions) rather than to areas that code for motor imagery, planning, and execution. This is perhaps evidence for a link between the conscious perception of the bodily implications of a particular timbre (i.e., hearing a timbre as more or less imbued with physical effort, as reflected in behavioral responses) and the perhaps nonconscious affective reaction consistent with these said physical implications (i.e., activating regions responsible for processing high arousal emotions, as reflected in increased limbic—especially amygdala—activity). Such a hypothetical link bypasses motor systems to activate the more phylogenetically ancient regions of the brain.

Of course, with p-noise correlations, we see both motor and amygdala activity in one contrast. P-noise correlates reasonably strongly with both *bodily exertion* (r = .55) and *affective valence* (r = –.55) in Experiment 1, and here we see corroborating neural evidence: the perception that a given timbre is "noisy" is associated with increased motor activity (pre/primary motor cortex and SMA) as well as amygdala activity, possibly related to negative affective response. Although p-noise is related to these two qualities, moreover, it appears to rely on neural resources that encompass both motor mimesis *and* affective response; hearing "noise" in these natural signals enacts mechanisms of motor resonance—possibly because acoustical noise is an indicator of high arousal affective states, which in turn are more evolutionarily urgent—and the limbic responses consistent with the audition of high arousal states (and with experiencing such states oneself). This result might be tentatively interpreted to indicate that "noisy" musical timbre is both deeply embodied, via mechanisms of motor resonance, and also at the same time negatively affectively weighted, via right-dominance of limbic activity. The perceptual systems involved in evaluating "noise," therefore, may illustrate the interconnection between motor systems and emotional processing. Table A3.3 shows peak voxels and associated Z-scores for all the behavioral correlations discussed in this section.

TABLE A3.3: Peak activations in behavioral correlations

		MNI coordinates			
Regions	Brodmann area	*x*	*y*	*z*	Z-score
Fig. A3.8 (3>1 with bodily exertion)					

Amygdala (L)	25	−14	−4	−24	2.35
Amygdala (R)		22	−4	−18	3.84
Fig. A3.9 (2>1 with negative affective valence)					
Pre/primary motor cortex (R)	4/6	32	−24	48	2.50
SMA	6	4	−18	76	2.90
SPL (L)	7	−16	−40	44	2.44
Anterior insula (R)	33	38	−8	0	3.01
Fig. A3.10 (2>1 with noisiness)					
Pre/primary motor cortex (R)	4/6	28	−22	62	2.43
SMA	6	−10	−34	76	3.24
Primary somatosensory cortex (R)	3	24	−40	64	2.39
Amygdala (L)	25	−16	−10	−18	2.94
Amygdala (R)		20	−6	−18	4.03
Hippocampus (L)	28	−24	−10	−18	3.43
Hippocampus (R)		18	−6	−24	3.45

So far the interrelationship between motor and limbic systems in p-noise has been approached perceptually—"p-noise" denoting that which is perceived as somehow conforming to the semantic quality of "noisiness"—but how does BOLD signal change as measured by fMRI relate to acoustical noise (a-noise)? In the next and final analysis, we turn to this question.

Correlation Analysis with Acoustic Data

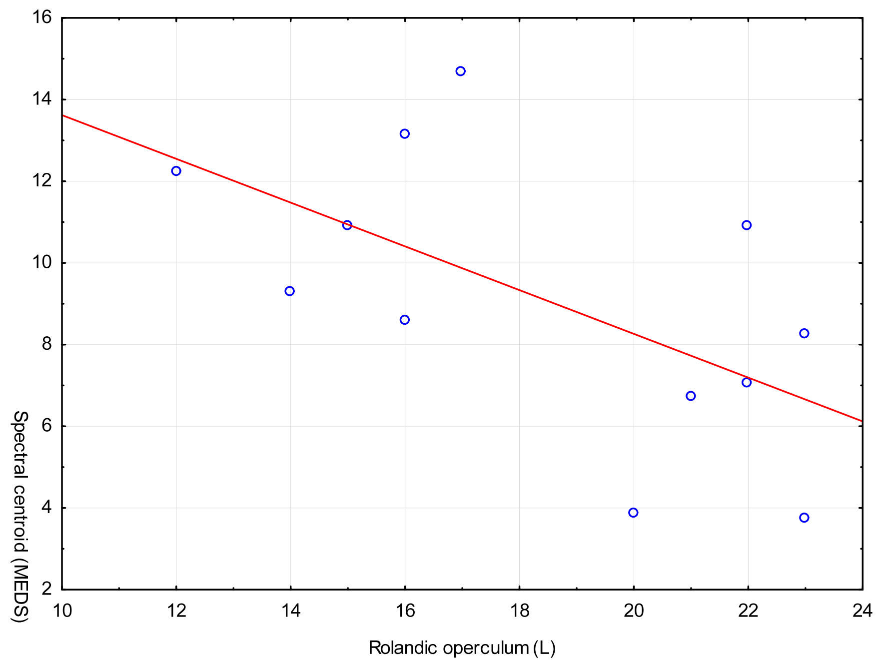
In order to understand what specific acoustic parameters correlate with neural activation, mean beta estimates for selected regions of interest were correlated with computationally extracted acoustic features of the stimuli (Table A3.4, see Appendix 1 for details). *Spectral centroid (SC)*, which is commonly experienced as "brightness," "nasality," or "sharpness," refers to the relative weight of high frequency energy in a signal (Hajda et al., 1997). As shown here, this parameter (as measured with two software programs, MIRtoolbox1.4 and MEDS) correlates positively with left auditory cortex (MEDS), and negatively with left Broca's area (MIR), left Rolandic operculum (MEDS), and cerebellum (both programs). P-noise can be linked to the qualities of *inharmonicity (I), spectral flatness (SF),* and *zero-cross rate (ZC)*; here too we can see significant positive correlations with STG (auditory cortex and Heschl's gyrus) and negative correlations with left Rolandic operculum. "Roughness," or sensory dissonance, exhibits no statistically significant correlations with beta estimates (though it does show a couple "musicologically relevant" ones).

TABLE A3.4: Correlations between activity in selected regions and acoustic parameters

	Brightness		*Noisiness*			*Sensory dissonance*
	SC (MIR)	*SC (MED)*	*I*	*SF*	*ZC*	*R*
Auditory (L)	.08	**.65**	.42	**.71**	.28	−.03
Auditory (R)	*−.53*	−.23	−.23	−.08	−.02	−.46
Heschl's gyrus (L)	−.19	.11	.36	.27	.05	.15
Heschl's gyrus (R)	−.07	.29	.43	**.62**	−.13	−.05
Rolandic operculum (L)	−.33	**−.60**	**−.76**	**−.69**	−.24	−.46
Rolandic operculum (R)	−.41	−.08	−.10	.23	−.23	*−.50*
Inferior parietal lobule (L)	−.02	.21	−.05	.12	.12	−.07
Inferior parietal lobule (R)	−.16	−.22	−.35	−.06	−.36	*−.52*
Primary somatosensory (L)	*−.53*	−.48	.02	−.26	−.24	.25
SMA	−.12	.17	.07	.24	.11	−.34
Pre-/primary motor (L)	−.09	0	−.15	−.33	.39	.22
Pre-/primary motor (R)	−.23	.05	.06	.07	.06	.07
Broca's area (L)	**−.64**	−.13	.18	.27	−.22	.01
Visual cortex (V1)	−.35	−.18	.36	.30	*−.51*	.16
Cerebellum	**−.74**	**−.60**	−.10	−.19	−.39	−.04
Amygdala (L)	−.26	−.21	−.41	−.41	.09	−.15
Brainstem	−.40	.03	.21	.29	−.08	−.20

N =15. Acoustic parameters extracted in MIRtoolbox1.4 and MEDS 2002-B-1 (see Appendix 1). Values for brain activity calculated from mean beta estimates. Bold entries p<.05, italicized entries p<.10. Acoustic abbreviations as follows: SC (MIR) = spectral centroid calculated with MIRtoolbox, SC (MED) = spectral centroid calculated with MEDS, I = inharmonicity, SF = spectral flatness, ZC = zero-cross rate, and R = roughness.

The region of greatest interest in these correlations, harkening back to the single-voxel correlations, is once again the Rolandic operculum. In the dominant hemisphere (left), this region is negatively correlated with all acoustic parameters of interest, half of which at significant levels. As discussed earlier, this vocal control region shows greater activity the more the listener likes the timbre under audition; conversely, it shows negative correlations with subjective judgments of "noisiness." These acoustic data corroborate the role of the Rolandic operculum in musical timbre processing. The acoustic correlates of p-noise—high spectral centroid, inharmonicity, a flat spectrum, and roughness—*negatively* associate with activity in this region, possibly indicating the suppression of vocal motor mimicry for these generally disliked timbral properties. This relationship is shown in the two plots of Fig. A3.11.



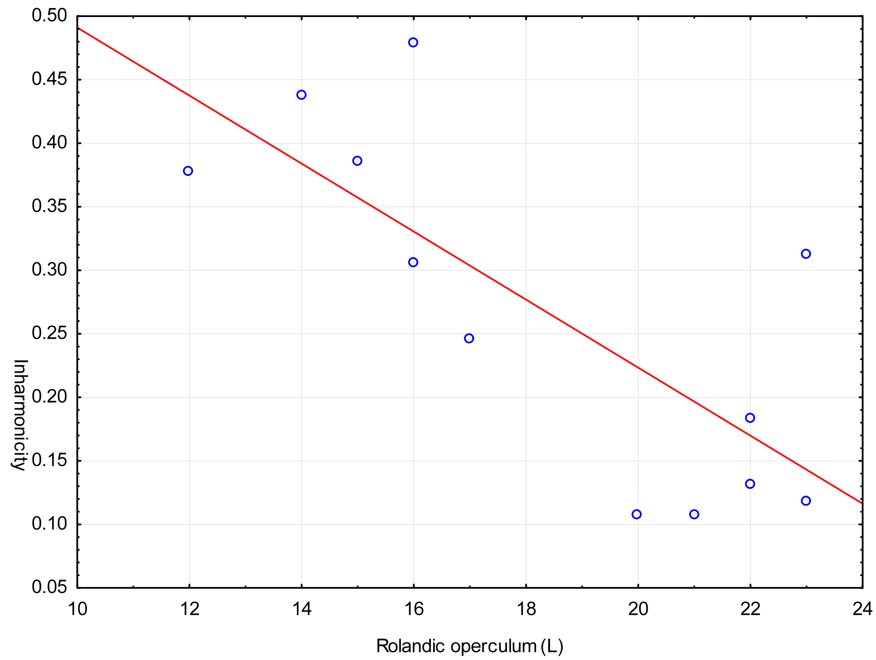


FIGURE A3.11: Rolandic operculum (L) correlated with spectral centroid and inharmonicity

Somatosensory Activation to Noisy Timbre

Finally, I want to return very briefly to the activation noted earlier in the somatosensory cortex. As discussed at length in Chapter 3, I hypothesize that not only does timbre involve motor mirroring; it also involves some level of *tactile* engagement. This engagement, moreover, is conceivably represented through both the audition of specific, noisy timbral qualities, and through the processing of metaphors commonly used to describe such qualities of sound. It is widely acknowledged that auditory and tactile processing share a special functional link (Schürmann et al., 2006). Further, tactile metaphors (e.g., "she had a *rough* day") have been shown to call upon the same somatosensory regions involved in tactile sensation (Lacey, Stilla & Sathian, 2012). My question in Chapter 3, then, is whether "rough" qualities of timbre—i.e., timbres commonly described using metaphors for tactile sensation, particularly negatively-valenced ones—actually draw upon somatosensory areas also implicated in the processing of both tactile sensation and metaphor.

To address this question, I used coordinates from Lacey, Stilla, & Sathian's paper (2012), which demonstrates activity in the parietal operculum (areas OP1 and OP3) when subjects both touch something and process tactile metaphor. Thresholded at Z = 1.7, we can see that these precise regions are indeed involved in a number of contrasts, all of which are left-lateralized: Sax 3 > Sax 1 implicates OP1 (Fig. A3.12); Sax 2 > Sax 1 correlated with perceived "noisiness" involves OP1 and OP3; and Timbre 2 > Timbre 1 correlated with negative *affective valence* involves OP3 (Fig. A3.13). Table A3.5 shows coordinates for the center of gravity of each activation, with associated Z-score. Correlation analysis with acoustic data yielded inconclusive results. A full discussion of the implications of this finding can be found in Chapter 3.

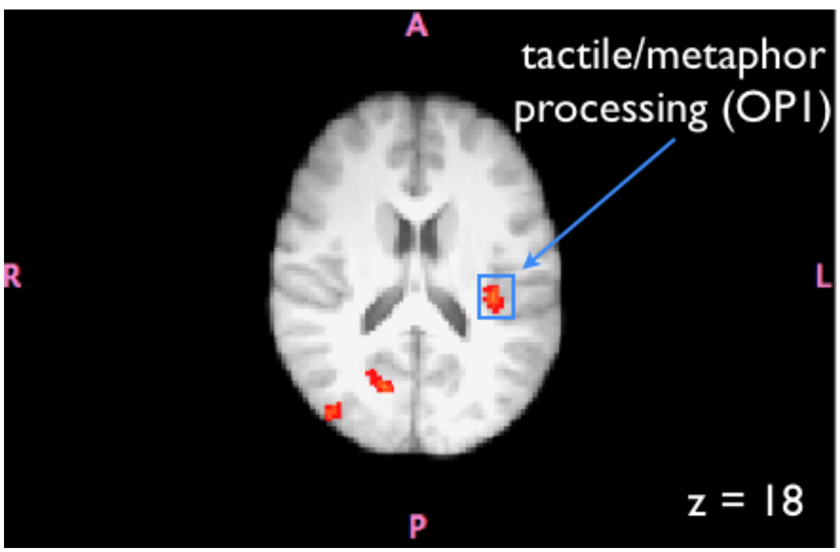


FIGURE A3.12: Growled Saxophone (Sax 3) > Regular Saxophone (Sax 1)

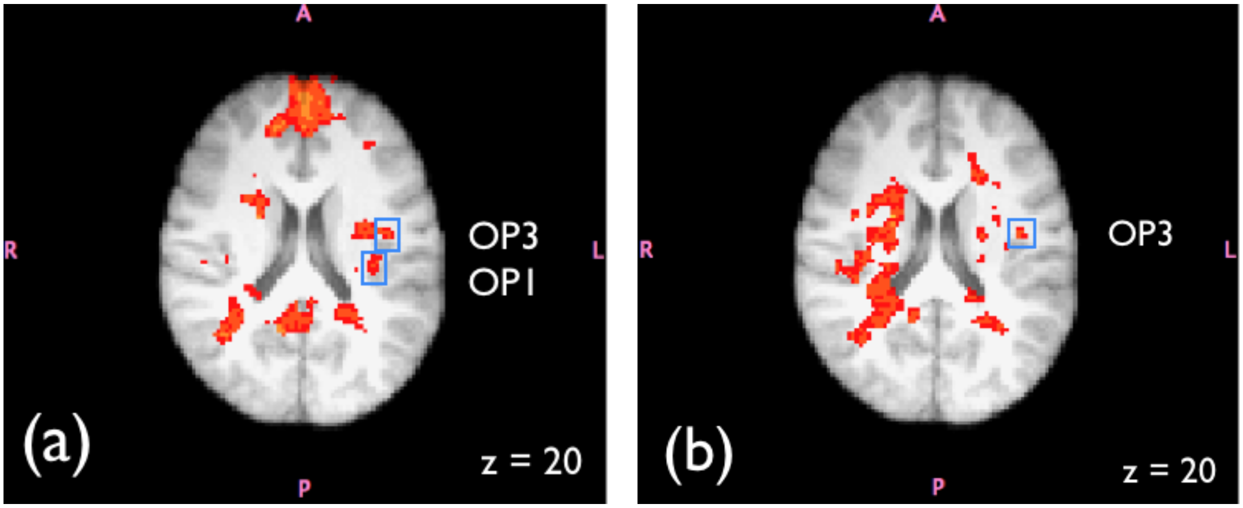


FIGURE A3.13: (a) Saxophone 2 > Saxophone 1 correlated with p-noise; (b) Timbre 2 > Timbre 1 correlated with *negative affect*

TABLE A3.5: Locations of peak voxels in areas of tactile sensation/metaphor overlap

	MNI coordinates			
	x	*y*	*z*	Z-score
Sax 3 > Sax 1				
OP1 (L)	−36	−26	18	2.71
Sax 2 > Sax 1 (with noisiness)				
OP1 (L)	−36	−26	20	1.89
OP3 (L)	−42	−10	20	2.42
Timbre 2 > Timbre 1 (with negative affective valence)				
OP3 (L)	−40	−8	20	2.50

Threshold for correlations set at $Z>1.7$ ($p<.05$). Coordinates for tactile sensation and tactile metaphor processing overlap derived from Lacey, Stilla, & Sathian 2012.

Brief Conclusion

These findings suggest three broad conclusions: first, different instruments and timbral qualities (from not noisy to noisy) produce different neural fingerprints, and components of the motor system—likely including the mirror neuron system—are involved in this differentiation. Significant limbic, paralimbic, and subcortical activation also suggest that perception of different timbral qualities is marked with a range of affective qualities. Timbre, in other words, is not an abstract musical building block that only acquires emotional meaning when structured into larger musical units; different timbral qualities, owing in part to motor mimicry, are inherently loaded with different valences. Second, the voice appears to be far and away the most neurally intensive instrument of these stimuli to process. In all covarying regions observed in baseline contrasts, vocal signals produced the most significant activation (particularly Voice 1 and Voice 2), followed by Saxophone 2 and 3 (the growled, noisier saxophone conditions).

Finally, and most importantly for the thesis of vocal-based motor mimicry of timbre, it appears that motor areas associated with vocal production—Rolandic operculum, Broca's area (specifically in the dominant, left lobe), and pre/primary motor cortex—are sensitive to changes in timbre, even when the instrument category is collapsed. Evidence for this comes in four forms: (1) measurement of signal change, (2) single-voxel baseline correlation with behavioral ratings, (3) voxel-by-voxel correlation of more subtle contrasts with behavioral ratings, and (4) correlation with acoustic parameters. First, ANOVA revealed "musicologically relevant" ($p = .07$) differences in beta estimate means for all five vocal regions between the three timbre conditions: Rolandic operculum and Broca's area, regions specifically connected to vocal production, decreased activity while pre/primary motor cortex, a more general motor area active in speech and many other functions, increased activity.

Second, Rolandic operculum was positively correlated with affective valence (left $r = .19$, right $r = .30$), while pre/primary motor cortex was negatively correlated ($r = -.21$) in single-voxel correlations of beta estimates in the baseline contrasts. Since the noisier conditions received lower mean *affective valence* ratings (i.e., liking or disliking) than less noisy timbres, this would seem to indicate that noisier, less-liked timbres drive down activation in the Rolandic operculum, a finding that is corroborated by a negative correlation with p-noise ratings (right $r = -.18$) for this region, while driving up activity in pre/primary motor cortex. These correlations are certainly not strong, but they are significant nonetheless ($p<.05$), especially in concert with these other pieces of evidence.

Third, correlation at the whole-brain level with behavioral data generated from the non-baseline contrasts revealed a connection between p-noise/*negative affect* and activity in pre/primary motor cortex (specifically right hemisphere). This activity was accompanied by activity in amygdala and anterior insula, indicating a possible relay between these areas of the "emotional" brain and the perception of negative, noisy timbre. Rolandic operculum activity did not reach significance in this analysis.

And fourth, a-noise parameters that correlate strongly with p-noise—high spectral centroid, inharmonicity, spectral flatness, zero-cross rate, and roughness—are negatively correlated with Rolandic operculum activation (particularly the left hemisphere), including significant correlations with spectral centroid (MEDS) ($r = -.60$), inharmonicity ($r = -.76$), and flatness ($r = -.69$).[3]

Taken together, we can interpret these data to indicate that listening to acoustically and perceptually "noisy" timbres—in electric guitar, saxophone, shakuhachi, and particularly the singing voice—seems to have a suppressive effect on motor mimetic properties of the vocal production system, specifically the Rolandic operculum (and to a lesser extent Broca's area). Not coincidentally, these same "noisy" timbres elicit lower *affective valence* ratings from listeners. In short, these results may indicate that motor resonance in Rolandic operculum and Broca's area *increases* with liking; and, because liking is negatively correlated with perceived noisiness ($r = -.39$,

[3] Broca's area exhibited a strong negative correlation with spectral centroid as calculated in the MIRtoolbox ($r = -.64$), but all other acoustic parameters showed inconclusive correlations. Similarly, correlations between pre/primary motor cortex and acoustic parameters were ambiguous.

p<.0001), vocal mirroring (i.e., subvocalization) thus appears to *decrease* with p-noise. On the other hand, pre/primary motor cortex exhibits the opposite effect, as demonstrated in both single-voxel correlations of beta estimates and voxel-by-voxel correlations. (Activity in the these motor regions is associated with negative *affective valence* and with p-noise.) Although the role of Rolandic operculum has been previously demonstrated on the level of hearing full pieces of "pleasant" music (Koelsch et al., 2006), this is the first study to report its significance at the level of isolated timbres.

Lastly, voxel-by-voxel correlations with behavioral data suggest an important role for the amygdala (particularly right) in perceptions of *bodily exertion* and p-noise, and the anterior insula in negative *affective valence.* These limbic structures, preferentially associated with the processing of fear, disgust, surprise, and other negative, high arousal emotions, appear to modulate significantly with these three perceptual variables. In connecting motor activation (in pre/primary motor cortex and SMA) with the limbic system, these correlations may suggest a functional link between these two systems in the perception of noisy timbre.

Certain aspects of these results go against our *a priori* hypothesis: motor resonance does not appear to uniformly increase with increased a- and p-noise. In it contingent upon the stimuli: for example, saxophone timbres *do* induce a response consistent with the hypothesized model, while vocal timbres do not. It is also contingent upon the specific motor region under observation: these results tentatively suggest a functional difference between specifically-*vocal* motor areas (Rolandic operculum and Broca's area), and more general-purpose motor areas (pre/primary motor cortex and SMA). Averaged across all the stimuli employed in this study, however, it seems that "noisy timbre"—a perceptual and acoustic quality with negative affective valence—might have the effect of attenuating certain mechanisms of embodied music cognition, at least on the neural level. This raises many more questions than it answers: for instance, what role (if any) do timbral reactions and appraisals play in the patterns of activation observed here? Does the positive correlation between these vocal motor regions and "liking" indicate that the listener might somehow *choose* to resonate with liked, less noisy timbres while rejecting the invitation for motor mimicry with disliked, noisy timbres? What comes first, the neural pattern of motor mirroring consistent with a certain reaction, or the conscious appraisal of said reaction? And do these results hold in the cognition of timbre in real, ecologically valid situations of listening to music ("polyphonic timbre," see Alluri et al., 2012) as opposed to only isolated timbres? These questions remain topics for further musicological deliberation and empirical investigation.

Acknowledgments

I would like to thank Marco Iacoboni and Katy Cross for the invaluable guidance they have provided over the course of this project. It would have been inconceivable without their support. Thanks are also due to my RA, Marita Meyer, for her invaluable help; Leo Moore, for all manner of trouble-shooting and general advice; and the rest of the Iacoboni Laboratory, in addition to the staff at the UCLA Ahmanson Lovelace Brain Mapping Center. The project was funded through a UCLA Council for Research and Office of the Vice-Chancellor for Research Transdisciplinary Seed Grant.

ADDENDUM: Areas activated by each baseline contrast (timbre > baseline)

	Brodmann area	MNI coordinates x	y	z	Z-score
Electric Guitar 1					
Primary auditory cortex (L)	22	−38	−38	12	5.84
Primary auditory cortex (R)		62	−16	8	5.92
Heschl's gyrus (L)		−38	−26	12	3.44
Heschl's gyrus (R)		38	−26	12	3.51
Rolandic operculum (L)	43	−50	0	4	2.58
Rolandic operculum (R)		50	0	4	3.37
Primary somatosensory cortex (L)	3	−36	−18	36	3.46
Inferior parietal lobule (L)	39/40	−44	−38	24	3.39
Inferior parietal lobule (R)		44	−34	22	2.63
Supplementary motor area	6	2	6	62	4.94
Pre/primary motor cortex (L)	4/6	−52	−6	48	3.35
Primary visual cortex V1	17	−2	−84	6	3.71
Cerebellum		26	−68	−50	3.36
Thalamus (L)		−14	−28	0	2.67
Hippocampus (L)		−22	−18	−16	2.60
Brainstem		−12	−28	−12	3.56
Electric Guitar 2					
Primary auditory cortex (L)	22	−40	−38	12	5.32
Primary auditory cortex (R)		56	−18	4	5.73
Heschl's gyrus (L)		−38	−28	12	4.17
Heschl's gyrus (R)		36	−28	12	3.44
Rolandic operculum (L)	43	−50	−2	4	2.56
Rolandic operculum (R)		48	0	4	2.55
Primary somatosensory cortex (L)	3	−40	−24	42	3.41
Inferior parietal lobule (L)	39/40	−44	−40	20	3.35
Inferior parietal lobule (R)		44	−36	20	3.35
Supplementary motor area	6	4	−2	66	4.24
Pre/primary motor cortex (L)	4/6	−50	−6	46	3.32
Cerebellum		−22	−60	−20	4.11
Thalamus (L)		−10	−22	4	3.43
Hippocampus (L)		−18	−40	4	2.63
Amygdala (L)		−16	−12	−16	3.38
Brainstem		−12	−28	−12	2.62
Electric Guitar 3					
Primary auditory cortex (L)	22	−46	−18	4	5.74
Primary auditory cortex (R)		54	−18	0	5.82
Heschl's gyrus (L)		−38	−28	10	5.02
Heschl's gyrus (R)		40	−24	10	4.25
Rolandic operculum (L)	43	−50	−2	4	2.57
Rolandic operculum (R)		50	−2	4	3.35
Inferior parietal lobule (L)	39/40	−52	−34	18	3.34
Inferior parietal lobule (R)		52	−34	18	3.47
Supplementary motor area	6	−4	4	64	3.51
Broca's area (L)	44/45	−48	14	6	2.65
Saxophone 1					
Primary auditory cortex (L)	22	−48	−18	2	5.55
Primary auditory cortex (R)		54	−10	0	5.52

Heschl's gyrus (L)		−36	−26	12	3.31
Heschl's gyrus (R)		40	−26	12	4.81
Rolandic operculum (L)	43	−50	−6	4	3.96
Rolandic operculum (R)		48	−6	4	3.22
Superior frontal gyrus (L)	10	−8	52	36	3.98
Inferior parietal lobule (L)	39/40	−56	−36	14	4.74
Inferior parietal lobule (R)		58	−26	14	4.02
Supplementary motor area	6	−4	12	62	3.26
Brainstem		−8	−28	−12	1.85
Saxophone 2					
Primary auditory cortex (L)	22	−52	−24	2	5.85
Primary auditory cortex (R)		46	−16	2	5.87
Heschl's gyrus (L)		−40	−26	12	4.23
Heschl's gyrus (R)		40	−26	12	4.27
Rolandic operculum (L)	43	−50	−6	6	3.36
Rolandic operculum (R)		48	−6	6	3.37
Superior frontal gyrus (L)	10	−10	40	48	3.36
Primary somatosensory cortex (L)	3	−48	−26	48	3.36
Primary somatosensory cortex (R)		38	−12	34	2.68
Inferior parietal lobule (L)	39/40	−48	−40	20	3.40
Inferior parietal lobule (R)		−52	−34	18	3.36
Supplementary motor area	6	8	−2	66	4.89
Pre/primary motor cortex (L)	4/6	−48	−6	44	4.23
Pre/primary motor cortex (R)		50	−8	40	3.45
Primary visual cortex V1	17	−2	−76	6	3.45
Broca's area (L)	44/45	−50	14	10	3.42
Cerebellum		24	−80	−46	3.34
Thalamus (L)		−10	26	6	2.66
Thalamus (R)		14	−26	4	2.67
Hippocampus (L)		−24	−26	−12	3.38
Amygdala (L)		−20	−12	−14	3.37
Brainstem		−6	−28	−32	3.28
Saxophone 3					
Primary auditory cortex (L)	22	−40	−38	12	5.54
Primary auditory cortex (R)		54	−18	0	5.63
Heschl's gyrus (L)		−40	−26	12	4.09
Heschl's gyrus (R)		40	−26	12	3.39
Rolandic operculum (L)	43	−50	−6	8	3.30
Rolandic operculum (R)		50	−6	8	3.31
Superior frontal gyrus (L)	10	−14	64	16	3.95
Primary somatosensory cortex (L)	3	−40	−26	42	4.13
Primary somatosensory cortex (R)		40	−18	46	3.31
Inferior parietal lobule (L)	39/40	−48	−38	16	4.05
Inferior parietal lobule (R)		44	−38	22	1.93
Supplementary motor area	6	0	−6	44	4.15
Pre/primary motor cortex (L)	4/6	−52	−6	44	4.05
Pre/primary motor cortex (R)		48	−6	50	3.33
Broca's area (L)	44/45	−44	14	18	3.39
Cerebellum		10	−66	−30	4.09
Thalamus (L)		−10	−30	4	3.36
Hippocampus (L)		−28	−26	−18	2.62
Amygdala (L)		−24	−16	−12	2.52
Brainstem		−6	−36	−22	2.63

Shakuhachi 1					
Primary auditory cortex (L)	22	−40	−36	12	5.93
Primary auditory cortex (R)		58	−30	10	5.40
Heschl's gyrus (L)		−40	−26	12	4.26
Heschl's gyrus (R)		40	−26	12	4.30
Rolandic operculum (L)	43	−50	−6	8	2.59
Rolandic operculum (R)		50	−6	8	2.59
Superior frontal gyrus (L)	10	−4	52	34	4.16
Inferior parietal lobule (L)	39/40	−48	−38	16	5.82
Inferior parietal lobule (R)		52	−34	22	4.29
Supplementary motor area	6	0	−6	70	3.78
Pre/primary motor area (L)	4/6	−56	−6	48	3.28
Thalamus (L)		−8	−26	0	3.48
Thalamus (R)		14	−26	−2	3.48
Hippocampus (L)		−18	−26	−12	3.30
Amygdala (L)		−20	−12	−14	3.39
Brainstem		−6	−36	−10	3.51
Shakuhachi 2					
Primary auditory cortex (L)	22	−50	−22	2	5.64
Primary auditory cortex (R)		50	−14	2	5.39
Heschl's gyrus (L)		−40	−26	12	3.36
Heschl's gyrus (R)		40	−26	12	4.88
Rolandic operculum (L)	43	−50	−6	8	2.56
Rolandic operculum (R)		50	−6	8	3.30
Primary somatosensory cortex (R)	3	62	−2	24	3.26
Inferior parietal lobule (L)	39/40	−48	−36	22	3.27
Inferior parietal lobule (R)		48	−36	22	3.41
Supplementary motor area	6	0	−2	60	4.20
Shakuhachi 3					
Primary auditory cortex (L)	22	−40	−36	12	6.33
Primary auditory cortex (R)		42	−24	12	4.54
Heschl's gyrus (L)		−40	−26	12	4.51
Heschl's gyrus (R)		40	−26	12	4.56
Rolandic operculum (L)	43	−50	−6	8	2.62
Rolandic operculum (R)		50	−6	6	2.59
Superior frontal gyrus (L)	10	−14	50	36	3.50
Inferior parietal lobule (L)	39/40	−42	−34	16	5.39
Inferior parietal lobule (R)		44	−36	24	3.67
Supplementary motor area	6	2	−20	60	3.91
Primary visual cortex V1	17	4	−98	2	2.58
Broca's area (L)	44/45	−48	34	8	3.52
Cerebellum		16	−92	−32	3.41
Thalamus (L)		−12	−30	2	2.71
Thalamus (R)		16	−22	2	3.67
Hippocampus (L)		−16	−38	2	3.53
Parahippocampal gyrus (L)		−24	−30	−26	3.47
Brainstem		12	−28	−8	2.74
Voice 1					
Primary auditory cortex (L)	22	−46	−20	2	5.79
Primary auditory cortex (R)		50	−16	2	5.03
Heschl's gyrus (L)		−40	−26	12	4.20
Heschl's gyrus (R)		40	−26	12	3.46
Rolandic operculum (L)	43	−50	−6	4	4.89

Rolandic operculum (R)		50	−6	8	3.37
Superior frontal gyrus (L)	10	−10	54	36	4.09
Primary somatosensory cortex (L)	3	−40	−26	48	3.39
Primary somatosensory cortex (R)		40	−26	48	1.88
Inferior parietal lobule (L)	39/40	−48	−38	20	4.14
Inferior parietal lobule (R)		48	−34	22	3.42
Supplementary motor area	6	−6	12	62	4.16
Pre/primary motor area (L)	4/6	−48	−6	46	3.42
Pre/primary motor area (R)		52	−6	48	4.16
Primary visual cortex V1	17	−12	−96	−6	3.40
Broca's area (L)	44/45	−50	22	24	3.32
Cerebellum (R)		28	−74	−34	4.92
Thalamus (L)		−10	−32	2	4.21
Thalamus (R)		20	−20	8	3.47
Hippocampus (L)		−16	−8	−20	3.38
Hippocampus (R)		28	−16	−18	4.16
Amygdala (L)		−24	−12	−12	4.96
Amygdala (R)		20	−12	−14	3.38
Ventral striatum (L)		−6	10	−4	2.60
Brainstem		12	−28	−14	3.46
Voice 2					
Primary auditory cortex (L)	22	−50	−24	4	5.80
Primary auditory cortex (R)		50	−18	4	5.85
Heschl's gyrus (L)		−40	−26	12	3.44
Heschl's gyrus (R)		40	−26	12	4.27
Rolandic operculum (L)	43	−50	−6	8	3.38
Rolandic operculum (R)		50	−6	8	2.59
Superior frontal gyrus (L)	10	−2	32	48	2.51
Primary somatosensory cortex (L)	3	−42	−32	44	2.58
Primary somatosensory cortex (R)		42	−14	30	2.69
Inferior parietal lobule (L)	39/40	−48	−38	18	4.18
Inferior parietal lobule (R)		48	−32	24	2.58
Supplementary motor area	6	0	6	64	4.04
Pre/primary motor area (L)	4/6	−52	−2	46	3.36
Primary visual cortex V1	17	−2	−76	12	2.66
Broca's area (L)	44/45	−48	28	8	5.06
Cerebellum (R)		16	−84	−32	4.16
Thalamus (L)		−10	−32	2	3.45
Thalamus (R)		14	−20	10	3.48
Hippocampus (L)		−20	−12	−20	2.60
Hippocampus (R)		20	−12	−20	2.59
Amygdala (L)		−24	−12	−12	3.41
Brainstem		−4	−38	−12	4.28
Voice 3					
Primary auditory cortex (L)	22	−46	−24	6	5.46
Primary auditory cortex (R)		54	−16	8	5.41
Heschl's gyrus (L)		−40	−26	12	4.03
Heschl's gyrus (R)		40	−26	12	4.07
Rolandic operculum (L)	43	−50	−6	8	3.99
Rolandic operculum (R)		50	−6	8	2.56
Inferior parietal lobule (L)	39/40	−48	−40	20	4.02
Inferior parietal lobule (R)		48	−34	20	4.04
Supplementary motor area	6	10	−12	76	4.12
Pre/primary motor area (L)	4/6	−48	−2	44	4.01

Broca's area (L)	44/45	−48	26	12	3.32
Insula (L)	52	−40	−22	− 2	3.99
Insula (R)		42	−20	−6	2.66
Thalamus (L)		−10	−32	2	3.32
Amygdala (L)		−16	−10	−14	3.97
Brainstem		−14	−28	−12	3.16

N = 15. Z-scores given for local cluster maxima at a threshold of Z = 1.7 (p<.05).

Appendix 4

Auxiliary Experiments

Summary and Purpose

This study consisted of four distinct perceptual experiments meant to supplement the data gathered from Experiment 2, including an exploratory experiment on the relationship between timbre perception and social identity. Because these research questions were confirmatory and exploratory, no specific *a priori* hypotheses were advanced. The individual tasks were as follows:

a) *Genre identification*: This experiment sought to determine what percentage of listeners could identify the "correct" genre when listening to brief excerpts of popular music recordings. It was designed as a post-hoc verification of the Experiment 2 stimuli set, as well as a replication of findings from the literature (Gjerdingen & Perrott, 2008).

b) *"Intensity" ratings*: This experiment aimed to evaluate how subjects rate the 18 stimuli on a bipolar scale of *not intense–very intense*, denoting degree of emotional intensity (positive or negative) implied by the brief excerpts. The *intensity* measure was meant to supplement the other perceptual variables queried in Experiment 2.

c) *Comparison of affective valence ratings of short and long excerpts*: Experiment 2 revealed relationships between *affective valence*—the quality of "liking" or "disliking" music—and the other perceptual variables, in addition to the acoustic parameters of the stimuli. In this task, ratings of these short excerpts (400 ms) were compared with longer versions of the same music (4 s). The principle aim was to determine if affect ratings varied significantly depending upon length of the excerpt. Previous studies found that length of duration has a negligible effect on a variety of identification tasks (Schellenberg, 1999; Gjerdingen & Perrott, 2008; Krumhansl, 2010), but this finding has not been extended to the question of perceived preference. Results are presented in the Introduction to the dissertation.

d) *Timbre and identity:* This exploratory experiment sought to relate the experience of "identifying with" a particular music—i.e., the feeling that a particular genre of music somehow represents one's own emotions, attitudes, preferences, and life experience—to the experience of social identity, i.e., identifying with the types of people who are closely associated with particular genres (Rentfrow & Gosling, 2003, 2007; Rentfrow, McDonald & Oldmeadow, 2009). Results play into the broader discussion of timbre and empathy in the brief Epilogue to the dissertation.

Stimuli

Stimuli used for Experiments 4a–d were the same as Experiment 2:

> Eighteen brief excerpts of popular music recordings were created using Audacity software. Excerpts were selected to represent six genres of contemporary popular music: (1) Rock, (2) Electronic Dance Music, or EDM, (3) Hip-Hop/Rap, (4) Pop, (5) Heavy Metal, and (6) Country. Each genre was thus represented by three excerpts. Excerpts were 400 ms in duration—the "medium-length" condition in the previously cited studies—with 10 ms amplitude ramps on both ends of the signal. Loudness was equalized manually.
>
> Excerpts were chosen based on the criterion of "genre representativeness": they were selected to include basic acoustic elements of the "timbral environments" (Ferrer, 2011) with which their representative genres are commonly associated. For example, rock and heavy metal excerpts included prominent distorted electric guitar and vocals; EDM included heavy bass and harsh synthesizer timbres associated with the sub-genre of dubstep; rap included sample-based beat production and rhythmic, "rapped" vocals; pop included a highly polished production style, vocals, and synthesizers; and country included common signifiers of "twang," including vocals, fiddle, and slide guitar (Neal, forthcoming). In order to isolate the timbral components of different singing styles without semantic confounds, excerpts with vocals were selected to exclude any identifiable words. Sources for the stimuli were recorded from the mid-1950s to the present: again, the criterion of "genre representativeness" was applied in determining year of re-

cording. Rock excerpts, for example, were all from iconic songs from the "classic" era (late 1960s to early 1970s) due to high genre identification levels achieved in pilot testing (as opposed to contemporary rock, which exhibited ambiguity of identification in pilot tests); country excerpts originate in source recordings from the 1950s and 60s for the same reason (contemporary country, which is more stylistically mongrel than its mid-century predecessors, yielded highly ambiguous identification results in pilot tests). The process of selecting excerpts was admittedly highly subjective and non-scientific, drawing upon the author's listening history, knowledge of popular genres as a performer, and musicological training. Excerpt sources are listed in Appendix 2, Table 2.1.

Excerpts were brief enough that no melodic and harmonic elements, structural implications, or rhythmic patterns were discernable. A vague sense of tempo (e.g., fast or slow) could be observed in certain excerpts, but no other details were perceptible besides timbre, texture, and artifacts of the recording that might provide stylistic cues and time-period of release. Such "thin slices" (Krumhansl, 2010) were desired in order to isolate the elements of polyphonic timbre under experimental observation.

Subjects

Twenty-eight participants were recruited from the UCLA community. The study group consisted of 17 females and 11 males between the ages of 18 and 22 (age M = 19.6, SD = 1.31). The subjects self-reported their number of years of formal musical training (M = 4.20 years, SD = 4.99); all subjects were non-music majors with musical backgrounds ranging from no formal experience to 15 years. The order of the four experiments was randomized. Each subject was paid $10 in cash after completion of both experiments.

Procedure

Experimental control was achieved through Music Experiment Development System (MEDS 2002-B-1) software (Kendall, 1990-2002). Subjects listened to the mono signals through high-quality headphones at a comfortable listening level, approximately 65 dB SPL, in a quiet room with minimal visual distractions. The experiment was preceded by a "practice" run in which participants listened to all 18 signals (randomized).

The experiment consisted of four separate ratings tasks that were presented in a random order. Presentation of all signals was similarly randomized:

a) *Genre identification*: Subjects were asked to select one genre that best represented each signal from the following list: (1) Rock, (2) Electronic Dance Music [EDM], (3) Hip-Hop, (4) Pop, (5) Metal, and (6) Country. They were given the following prompt:

> In this experiment, please choose a musical genre from the following list that best represents each excerpt: Rock, Electronic Dance Music (EDM), Rap, Pop, Metal, and Country. You must scroll down to see the complete list of genres. Click OK to begin.

b) *Intensity ratings*: Subjects were asked to rate the perceived degree of "emotional intensity" in each signal with the use of a numbered horizontal rating scale (0-100) with bipolar labels (*not intense–very intense*) consistent with the semantic differential paradigm (Osgood, Suci & Tannenbaum, 1957). The word "intensity" was selected due to its ubiquity in common usage; it was also selected because of its relative paucity in affect studies, most of which opt for "tension," "arousal," "energy," and similar terms (Eerola, Ferrer & Alluri, 2012). Subjects were asked to rate perceived intensity irrespective of positive or negative valence. They were given the following prompt:

> This experiment explores the perceived emotional intensity of very short music excerpts. On a scale from "not intense" to "very intense," please rate the intensity of feeling conveyed in each clip (whether positive or negative). Please use the full scale in your responses. Click OK to continue.

c) *Comparison of affective valence ratings between long and short excerpts:* Participants were asked to evaluate the *affective valence* of each 400 ms signal on a numbered horizontal rating scale (0-100) with bipolar labels (*strongly dislike-strongly like*). *Affective valence* has been shown to equate to musical *preference* (Eerola, Ferrer & Alluri, 2012). Following this rating task, subjects were presented with longer versions (4 s) of the same musical excerpts and asked to rate them on the same *valence* scale. They were given the following two prompts:

> Next, please tell us how much you like or dislike each of these excerpts. Rate the same clips of music on a scale of "strongly dislike" to "strongly like." Please use the full scale in your judgments. Click OK to begin.
>
> Finally, you will listen to longer versions of the excerpts. Once again, please tell us how much you like or dislike each clip, from "strongly dislike" to "strongly like." Use the full scale in your judgments. Click OK to begin.

d) *Timbre and Identity:* Subjects were asked to rate the degree to which they "identify" with the each musical excerpt (400 ms) on a 5-point, Likert-modeled ordinal scale with the following options: (1) strongly do not identify, (2) do not identify, (3) no opinion, (4) identify, (5) strongly identify. Following this task, they were asked to rate the degree to which they identify with the "social group associated with the music" on the same 5-point scale. They were given the following instructions:

> This next experiment evaluates your subjective response to the excerpts by assessing your level of identification both with the music itself and the community of people with whom you associate the music. You will be asked to rate how much each clip reflects your personality and preferences, from "strongly do not identify" to "strongly identify." Click OK to begin.
>
> Next, you will be asked to rate your level of identification with the social group associated with the music. Do you feel you belong to the group who likes this music? The options are the same as the last experiment ("strongly do not identify" to "strongly identify"). Click OK to continue.

The duration of the full experiment was approximately 5-10 minutes.

Analysis and Results

(a) Genre Identification

Nominal responses were organized into frequency tables and tabulated based on the percentage of "correct" responses per signal.[1] A "correct" response was defined as a correspondence between the genre selected and the genre the signal was intended to represent.

The results are shown in Table A4.1. Hip-Hop, Pop, Metal, and Country all exhibited high degrees of consistency (greater than 87%), with Metal and Country showing the most "correct" responses (with 95% and 92%, respectively). EDM excerpts 1 and 3 show high "correct" responses, but it is clear that excerpt 2 (21%) was perceived mainly as Pop, not EDM. This is possibly the result of the presence of a female voice in this excerpt, although the track was produced by a major EDM DJ (Kaskade); the other two excerpts did not feature vocals. This low rating skews EDM representativeness results downward. As a genre, Rock exhibited by far the most perceptual ambiguity at 59% correct. While still well above chance—which would yield approximately 17% correct—it is significantly lower than other genres. Pilot testing ($N = 5$) showed these excerpts of classic rock to be identifiable, and EDM 2 to be heard *as* EDM, but this effect did not translate into the experiment. (Excerpts of contemporary rock, including Green Day and Coldplay, were used in a pilot but failed to yield even 30% accuracy.) The most common "incorrect" genre selected for the Rock excerpts was Country, indicating the possibility that the timbral markers of contemporary country music are perceptually similar to classic rock.

TABLE A4.1: Genre representativeness of stimuli

Excerpt	% Correct	Mean % correct for genre
Rock 1	.64	
Rock 2	.50	*.59*

[1] I place scare quotes around the word "correct" to draw attention to the complexity surrounding the issue of genre representativeness. For more, see Gjerdingen (2008, p. 97)

Rock 3	.64	
EDM 1	.96	
EDM 2	.21	*.69*
EDM 3	.89	
Hip-Hop 1	.96	
Hip-Hop 2	.89	*.91*
Hip-Hop 3	.89	
Pop 1	.79	
Pop 2	.93	*.87*
Pop 3	.89	
Metal 1	.89	
Metal 2	.96	*.95*
Metal 3	1.0	
Country 1	.93	
Country 2	.96	*.92*
Country 3	.86	

(N = 28)

Taken together, these results both confirm the genre representativeness of the stimuli (with the exceptions outlined above) and replicate earlier findings regarding the ability of listeners to categorize music by genre even after very short exposure (Gjerdingen & Perrott, 2008; McLucas, 2010). For the purposes of this dissertation, these results affirm the perceptual significance of timbre in musical appraisal. Since the excerpts are too brief to convey melodic, harmonic, or rhythmic information, one might conclude that listeners were responding to a "timbral fingerprint" for each genre. Although it is not the goal of this study to quantify the acoustic correlates of genre (Tzanetakis & Cook, 2002), it is suggestive nonetheless that timbre alone is capable of triggering such accurate evaluations in so short a time interval.

(b) Intensity Ratings

Acceptable inter-rater consistency for the *intensity* factor was found (Cronbach α = .74), and no subjects reported confusion performing the rating task. In a repeated measures ANOVA with the other two interval scales of the experiment (the affective valence comparison experiment), significant differences were discovered: for the Perceptual Quality factor (3 levels) main effect, $F(2,54) = 17.62$, $p<.00001$; for Genre (6 levels), $F(5, 135) = 11.85$, $p<.00001$; and for Examples (3 levels), $F(2, 54) = 2.44$, $p=.10$, which predictably did not reach significance, replicating Experiment 2. Interactions between Perceptual Quality and Genre were also significant. These two factors were spherical after correction (with Greenhouse-Geisser, Perceptual Quality $\varepsilon = .90$ and Genre $\varepsilon = .74$, both of which $p<.0001$). Results were confirmed with post hoc testing (Tukey HSD).

Next, *intensity* means for all signals were correlated with the means for the perceptual and acoustic variables gathered in Experiment 2 (see Appendix 2). These results are summarized in Table A4.2.

TABLE A4.2: *Intensity* means for all signals correlated with Experiment 2 perceptual and acoustic means

Perceptual	Correlation (*r*)	*Acoustic*	Correlation (*r*)
Exertion	**.95**	SC (MIR)	.31
Valence	**−.61**	SC (MED)	**.52**
Brightness	**−.73**	I	.02
Noisiness	**.98**	SF	.25
Happiness	−.31	ZC	**.57**
Sadness	**−.74**	R	**.55**
Anger	**.88**		
Fear	.43		
Tenderness	**−.88**		

Note: *Intensity ratings are independent from others. Intensity $N = 28$, others $N = 35$.*

As seen here, correlations with the perceptual variables were robust. For instance, *intensity* correlated with *exertion* nearly perfectly (*r* = .95), and p-noise ("noisiness") even closer still (*r* = .98). Moreover, the same binary valence schema discussed in Appendix 2, where positive and negative percepts tend to be grouped together, is active here as well. Positive qualities (*valence,* "brightness,*"* *happiness, sadness,* and *tenderness*) are negatively correlated with *intensity*, while negative qualities (*exertion,* p-noise*, anger,* and *fear*) are positively correlated, as shown in Fig. A4.1. This indicates that *intensity* is a negatively valenced percept in relation to these signals, and at very high strengths of correlation. In fact, with near perfect correlations with *exertion* and p-noise, it is tempting to conclude that the word "intensity" in this context is functioning as a cognate for some of these other terms. There is likely redundancy in these variables. Nonetheless, it is striking the extent to which the semantic quality of *intensity*—defined as "emotional intensity" and "intensity of feeling," and specified as consisting of either positive or negative emotions in the prompt—tracks with the perceptual variables under consideration. It is less clear how *intensity* maps onto acoustics, but significant correlations with each of the three perceptual domains of *brightness* (SC (MED)), *noise* (ZC), and *roughness* (R) were found.

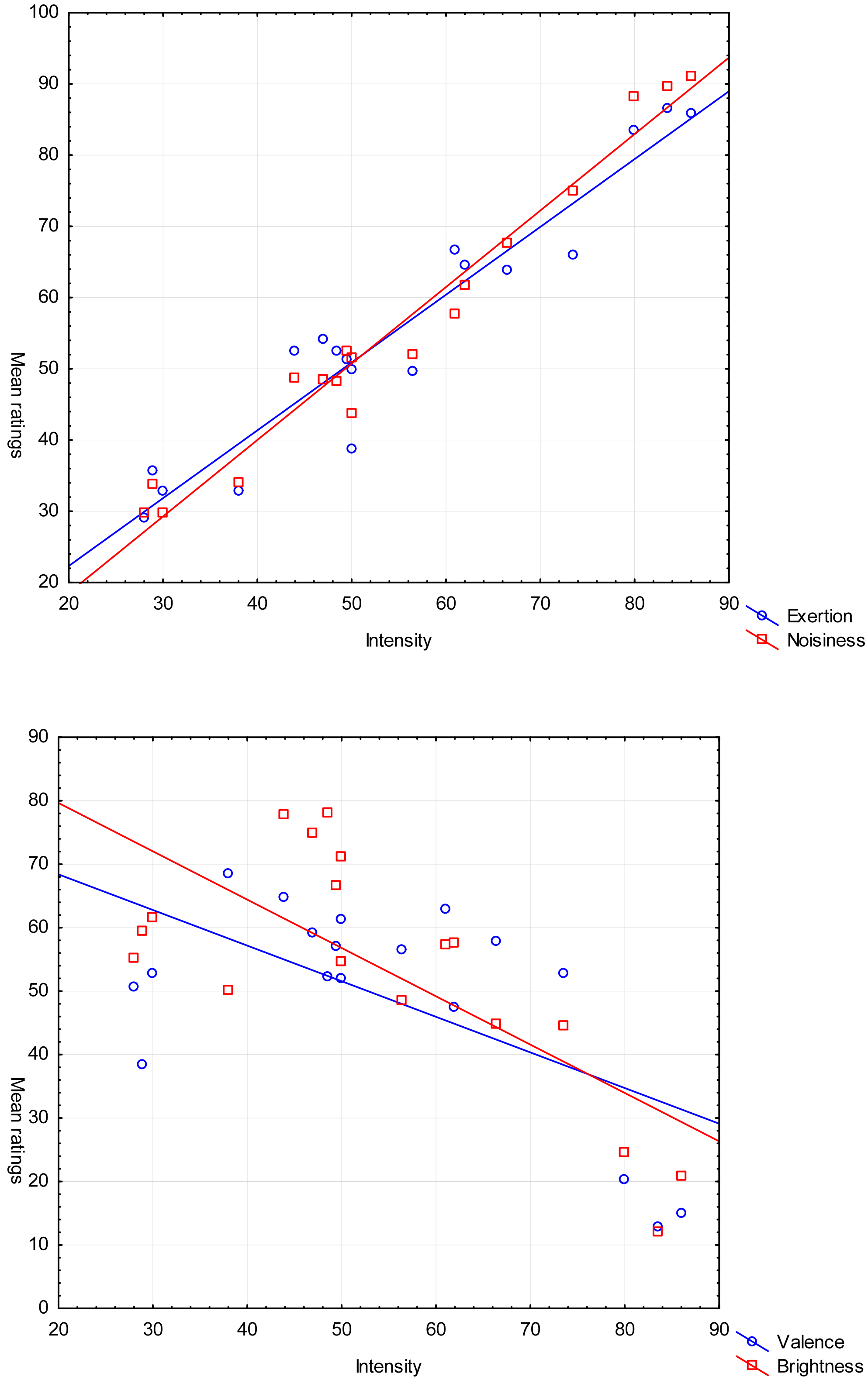


FIGURE A4.1: Scatterplots of *intensity* ratings correlated with negative and positive qualities (above and below, respectively)

(c) Comparison of Affective Valence Ratings between Short and Long Conditions

Inter-rater consistency of *affective valence* ratings for both conditions (short [400 ms] and long [4 s]) was acceptable (Cronbach α = .73). When broken down separately, the short condition produced much lower agreement (α = .43) than the long (α = .71), which might be expected due to presumably higher clarity of recognition in the longer task.

A repeated measure ANOVA was carried out to evaluate the difference between the two conditions. Sphericity was confirmed for all main effects (Mauchly's test) without need for correction. As determined in Experiment 2, the Genre factor (6 levels) reached significance at $F(5, 135) = 25.48$, $p<.00001$. However, no other factors in the ANOVA produced a main effect with statistical significance. As observed in the *intensity* ratings task above (as well as Experiment 2) the Examples factor failed to reach significance [$F(2, 54) = 0.87$, $p = .42$]. This was expected. Interestingly, though, differences in *valence* mean ratings between the two length conditions similarly failed to attain significance [$F(1, 27) = 1.19$, $p = .28$]. Moreover, interactions involving the Valence factor were not significant either [Valence and Genre, $F(5, 135) = 1.48$, $p = .20$; Valence and Examples, $F(2, 54) = 1.10$, $p = .34$]. The three-way interaction between Valence, Genre, and Examples proved significant, but the Genre factor contributed the most to the interaction [$F(10, 270) = 3.55$, $p = .0002$]. Results were confirmed in post hoc testing (Tukey HSD).

Figure A4.2 shows the negligible differences in mean *valence* scores between the two conditions. There is slight drift in mean ratings among all genres, but none is significant. This was a surprising result for two main reasons. First, it might be expected that familiarity would play a role in changing mean ratings in the longer, 4 s condition. Many of the songs excerpted in the experiment were very well know (see Table A2.1 in Appendix 2 for a complete list), including "Purple Haze" (Jimi Hendrix), "Whole Lotta Love" (Led Zeppelin), "Gin and Juice" (Snoop

Dogg), and "Poker Face" (Lady Gaga), among others. Since musical liking tends to increase with familiarity (McDermott, 2012), it could reasonably be assumed that the longer condition would yield higher ratings than the 400 ms condition. The experiment regrettably did not consider recognizability as a variable for either conditions, however, so it is only conjecture that longer excerpts are more recognizable than short counterparts (or even that the long versions were recognized in the first place).

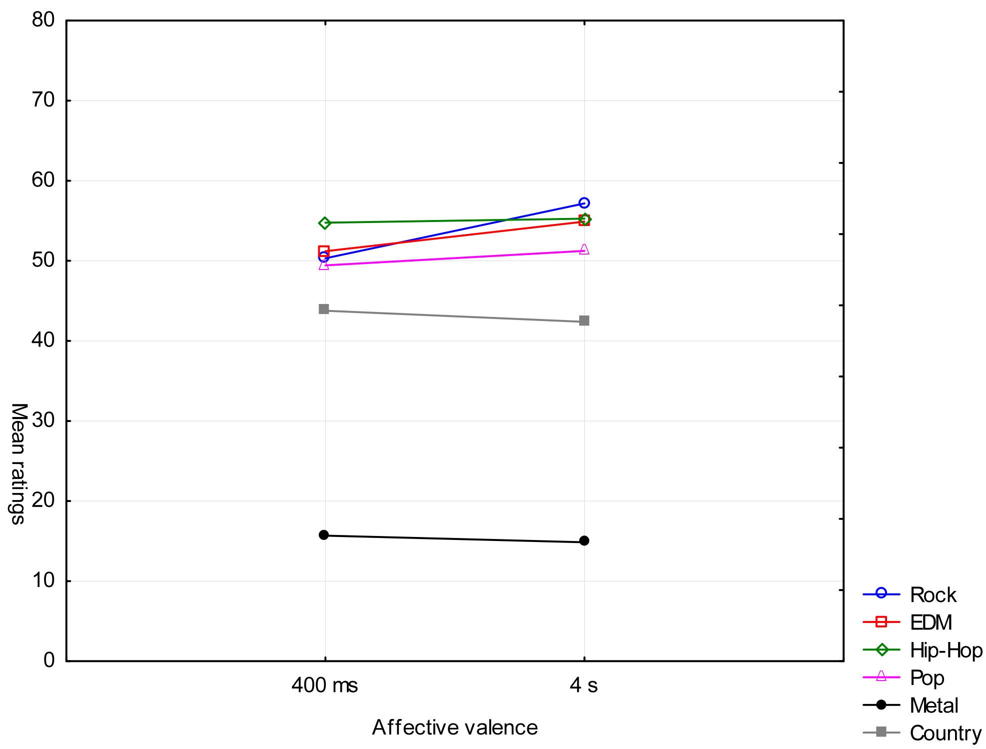
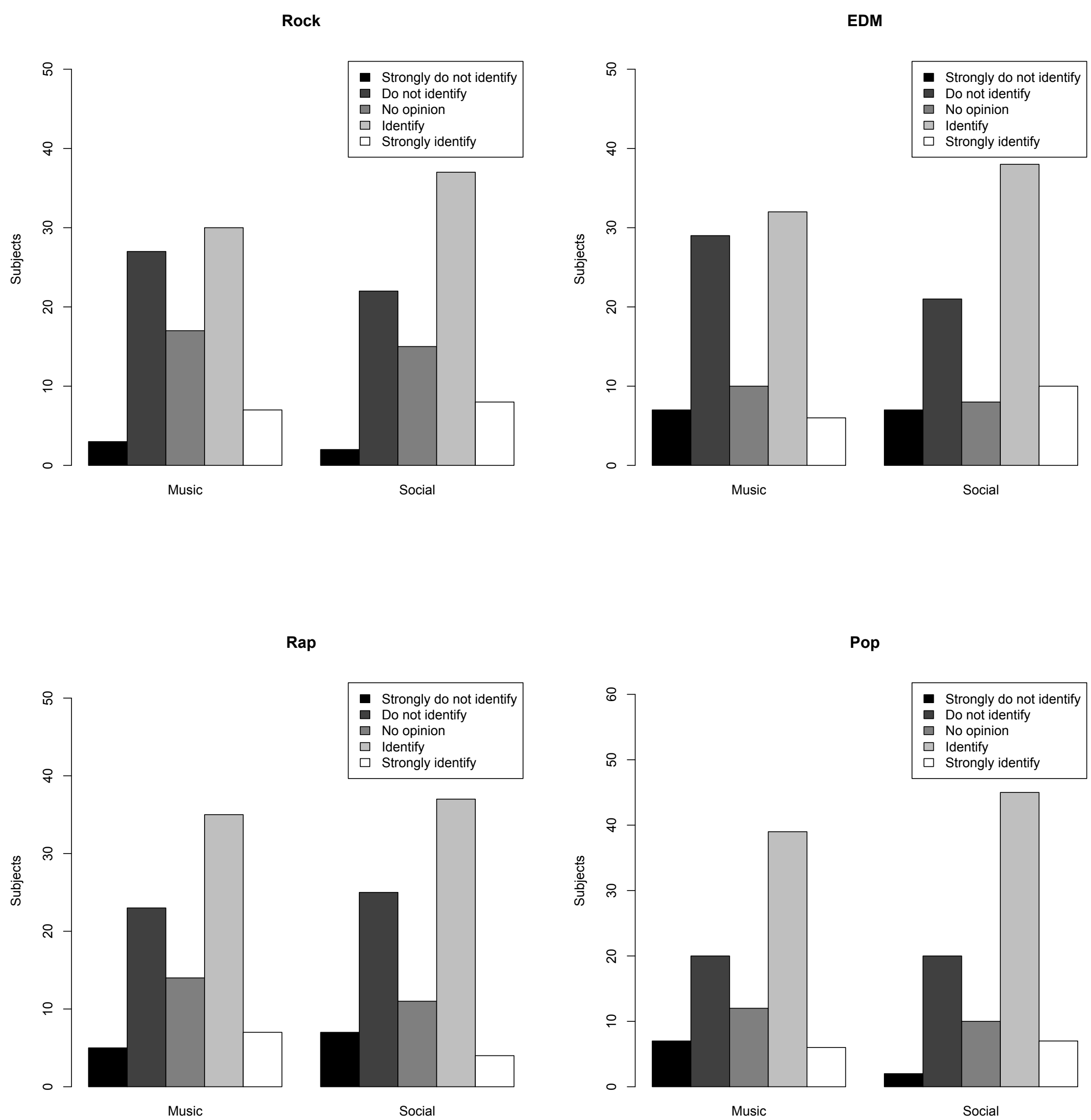


FIGURE A4.2: Insignificant differences between two time conditions (p = .28)

The second novel conclusion suggested by these results is that *timbre*, prior to pitch relationships, rhythm, and other musical elements, is the preeminent driver of affective appraisal. Admittedly this claim is something of an abstraction, since music is rarely listened to in 400 ms intervals, nor is timbre separable from other aspects of the music. Nonetheless, the short excerpts are too brief to convey any melodic, harmonic, rhythmic, or formal information; timbre is therefore the primary attribute influencing behavioral responses. If the additional context of the long version—all the melodic, harmonic, and rhythmic information lacking in the short excerpts—*does not* affect evaluations of liking and disliking, then it possibly stands to reason that affective judgments of music are propelled by perceptions of timbre. 400 ms is enough time to condition a particular appraisal, which is not modified significantly by having access to more of the musical picture. Put differently, snap judgments of polyphonic timbre tend to be "right" insofar as that they accurately mirror affective valence responses to the music. Bigand et al. (2005) reached a similar conclusion with the use of 1 s stimuli, but this timeframe is long enough to sustain much more melodic, harmonic, and rhythmic information; in short, though timbre doubtless plays a major role in their results, it cannot be said to account for most of the effect. This result is in line with neurophysiological research on music and emotion induction (Peretz et al., 1998; Goydke et al., 2004), as well as the literature on the neural processing of timbre (Tervaniemi et al., 1997; Platel et al., 1997). A full discussion of its ramifications can be found in Part I of the dissertation.

(d) Timbre and Identity

To determine significance of ordinal data, a dependent-sample T-test and a Wilcoxon sign-rank test were performed. Both procedures produced similar results: differences in mean ratings between musical identification and social identification were not significant among Rock, EDM, Hip-Hop, and Pop. This suggests that "musical identification" and "social identification" are roughly synonymous to most participants. Metal and Country social identification ratings, however, were significantly different from the musical identification means of the genres above. Further, musical identification means in the Metal genre differed substantially from all social identification categories ($p<.0001$), indicating polarization of these ratings (76% of responses to metal indicated "strongly do not identify"). Differences in Country musical identification ratings reached significance in comparison to Rock, Pop, and Metal social identification ratings.



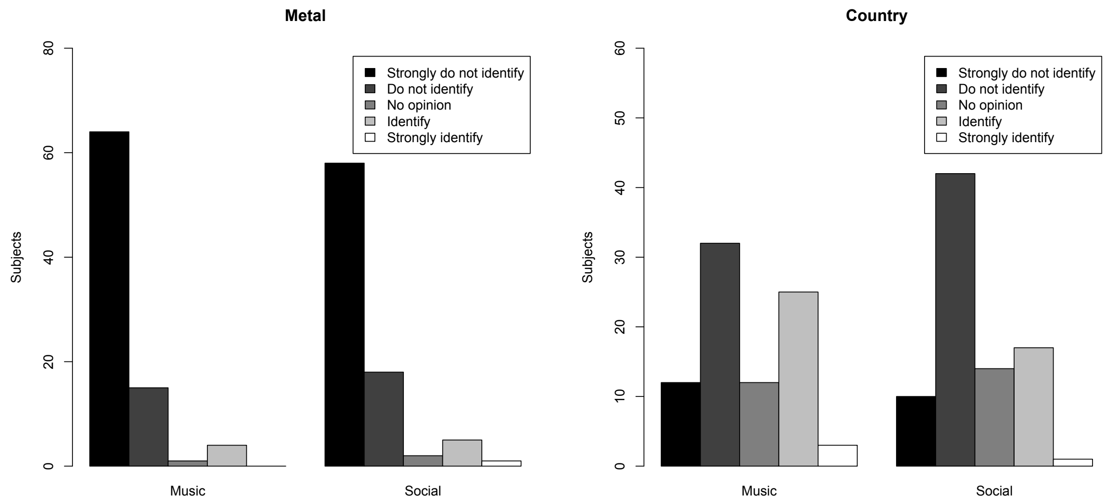


FIGURE A4.3: Musical and social identification ratings for six popular music genres

These data—which are summarized by genre in Figure A4.3, then correlated in Table A4.3—can be interpreted in different ways. First, the correspondence between musical and social identification can be assessed at the level of individual excerpts. Phrased as a question, does identifying with a 400 ms clip of music correlate with identifying with the social group associated with that excerpt? The results indicate high levels of correlation in one-to-one correspondence between individual excerpts: r = .32 at the low end (the only insignificant correlation of the group) to r = .86 on the high (M = .70). These are, predictably, the highest correlations in the data set.

TABLE A4.3: Musical identification correlated with social identification (by genre)

Musical identification (MI) *Social Identification (SI)*

SI	Rock	EDM	Hip-Hop	Pop	Metal	Country
MI						
Rock	**.63**	−.20	−.20	−.19	.20	.21
EDM	−.01	**.39**	−.14	.18	.03	.15
Hip-Hop	−.12	−.15	**.49**	−.11	−.08	−.16
Pop	−.07	.14	−.12	**.64**	−.19	−.08
Metal	.25	−.17	−.04	−.15	**.56**	.14
Country	.25	.08	−.30	.15	.12	**.61**

(N = 28)

Second, we can average ratings across the three excerpts within each genre to assess the correspondence between these two types of identification at the genre level. As shown in Table A4.3, identifying with a musical genre, as represented in the form of 400 ms excerpts, correlates with identifying with the genre's associated social group (all bold entries indicate significance at $p<.05$).

Third, differences between musical and social identification ratings across genres can be assessed: phrased as a question, do people who identify strongly with one type of music identify with social groups associated with *other* genres of music? Reviewing Table A4.3, it is apparent that MI (musical identification) scores do not strongly correlate with SI scores (social identification) from other genres. This is true across the data set: no significant correlations across genre are present. For example, participants who identify with Rock (MI and SI) do not strongly identify with social groups associated with EDM (r = −.20), Hip-Hop (−.20), Pop (−.19), Metal (.20), or Country (.21). However, people who identify with Rock (MI) associate more strongly with metal and country fans (which are positively correlated) than with EDM, Hip-Hop, and Pop aficionados (negative correlations). No negative correlations reach significance; the highest, Country music people (MI) identifying with Hip-Hop fans (SI), reaches only r = −.30. In other words, there are no cases in this data set where strong musical identification is strongly correlated with *negative* social association.

Issues of genre and social identity might seem far removed from questions of timbre and affect. But if specific polyphonic timbres can communicate systematically observable patterns of embodiment, affective valence, semantic associations ("brightness" and "noisiness"), emotional "intensity," primary emotions conveyed, and genre, and if this evaluative process can occur in the proverbial blink of an eye, with very little room for the contribution of other musical elements, then timbre perception is a socially- (and thus ethically- and politically-) valenced act. Certain timbral qualities—for example, the "twang" of country music or the harsh, guttural vocals of metal—are instantly recognizable markers for the social groups they represent. I will take up the interaction of timbre and social perception, specifically in regard to the concept of empathy, in further research.

Bibliography

Aglioti, Salvatore, and Mariella Pazzaglia. "Representing Actions Through Their Sound." *Experimental Brain Research* 206 (2010): 141-51.

Ake, David. *Jazz Cultures*. Berkeley and Los Angeles: University of California Press, 2002.

———. *Jazz Matters: Sound, Place, and Time Since Bebop*. Berkeley and Los Angeles: University of California Press, 2010.

Alfano, K.M., and C.R Cimino. "Alteration of Expected Hemispheric Asymmetries: Valence and Arousal Effects in Neuropsychological Models of Emotion." *Brain and Cognition* 66 (2008): 213-20.

Alluri, Vinoo, P. Toiviainen, I.P. Jääskeläinen, E. Glerean, M. Sams, and E. Brattico. "Large-scale Brain Networks Emerge from Dynamic Processing of Musical Timbre, Key, and Rhythm." *Neuroimage* 59 (2012): 3677-89.

Alluri, Vinoo, and Petri Toiviainen. "Effect of Enculturation on the Semantic and Acoustic Correlates of Polyphonic Timbre." *Music Perception* 29, no. 3 (2012): 297-310.

———. "Exploring Perceptual and Acoustical Correlates of Polyphonic Timbre." *Music Perception* 27, no. 3 (2010): 223-41.

American Standard Acoustical Terminology. New York: American Standards Association, 1960.

Anderson, Iain. *This is Our Music: Free Jazz, The Sixties, and American Culture*. Philadelphia: University of Pennsylvania Press, 2007.

Attali, Jacques. *Noise: The Political Economy of Music*. Translated by Brian Massumi. Minneapolis: University of Minnesota Press, 1985.

Aucouturier, J.J. "Dix expériences sur la modélisation du timbre polyphonique." Ph.D. Diss., Universite Paris 6, 2006.

Aziz-Zadeh, Lisa, M. Iacoboni, E. Zaidel, S. Wilson, and J.C. Mazziotta. "Left Hemisphere Motor Facilitation in Response to Manual Action Sounds." *European Journal of Neuroscience* 19 (2004): 2609-12.

Baddeley, Alan, and Robert Logie. "Auditory Imagery and Working Memory." In *Auditory Imagery*, edited by D. Reisberg. Hillsdale, NJ: Lawrence Erlbaum, 1992.

Bailes, Freya. "Timbre as an Elusive Component of Imagery for Music." *Empirical Musicology Review* 2, no. 1 (2007): 21-34.

Bailey, Peter. "Breaking the Sound Barrier: A Historian Listens to Noise." *Body and Society* 2, no. 2 (1996): 49-66.

Balkwill, Laura-Lee, and William Forde Thompson. "A Cross-Cultural Investigation of the Perception of Emotion in Music: Psychophysical and Cultural Cues." *Music Perception* 17, no. 1 (1999): 43-64.

Balzano, Gerard J. "What are Musical Pitch and Timbre?". *Music Perception* 3, no. 3 (1986): 297-314.

Bangert, Marc, and E. Altenmüller. "Mapping Perception to Action in Piano Practice: A Longitudinal DC-EEG Study." *BMC Neuroscience* 4 (2001): 26-36.

Bangert, Marc, T. Peschel, G. Schlaug, M. Rotte, D. Drescher, H. Hinrichs, H.-J. Heinze, and E. Altenmüller. "Shared Networks for Auditory and Motor Processing in Professional Pianists: Evidence from fMRI Conjunction." *Neuroimage* 30 (2006): 917-26.

Baraka, Imamu Amiri. *The LeRoi Jones/Amiri Baraka Reader*. New York: Basic, 2000.

Barsalou, Lawrence W. "Grounded Cognition." *Annual Review of Psychology* 59 (2008): 617-45.

Barthes, Roland. *Image, Music, Text.* Translated by Stephen Heath. New York: Hill and Wang, 1977.

Bataille, Georges. *Erotism: Death and Sensuality*. Translated by Mary Dalwood. San Francisco: City Lights, 1986.

Becker, Judith. *Deep Listeners: Music, Emotion, and Trancing*. Bloomington: University of Indiana Press, 2004.

Beckmann, C., M. Jenkinson, and S.M. Smith. "General multi-level linear modelling for group analysis in FMRI." *Neuroimage* 20 (2003): 1052-63.

Beckwith, Karl. "'A Human Being This Once Resembled': Bodily Transgression in Extreme Metal Music." Unpublished paper (2006).

Belin, Pascal. "Voice Processing in Human and Non-Human Primates." *Philosophical Transactions of the Royal Society B* 361 (2006): 2091-107.

Belin, Pascal, R.J. Zatorre, P. Lafaille, P. Ahad, and B. Pike. "Voice-Selective Areas in the Human Auditory Cortex." *Nature* 403 (2000): 309-12.

Berger, Harris. "Death Metal Tonality and the Act of Listening." *Popular Music* 18, no. 2 (1999): 161-78.

Berger, Harris, and Cornelia Fales. "'Heaviness' in the Perception of Heavy Metal Guitar Timbres: The Match of Perceptual and Acoustic Features over Time." In *Wired for Sound: Engineering and Technologies in Sonic Culture*, edited by Paul D. Greene and Thomas Porcello. Middleton, CN: Wesleyan University Press, 2005.

Berger, Kenneth W. "Some Factors in the Recognition of Timbre." *Journal of the Acoustical Society of America* 36, no. 10 (1964): 1888-91.

Berkman, Franya. "Appropriating Universality: The Coltranes and 1960s Spirituality." *American Studies* 48, no. 1 (2007): 41-62.

Bey, Hakim. "The Utopian Blues." In *Sounding Off!: Music as Subversion/Resistance/Revolution*, edited by Ron Sakolsky and Fred Wei-han Ho. Brooklyn: Autonomedia, 1995.

Bharucha, Jamshed J., Meagan Curtis, and Kaivon Paroo. "Varieties of Musical Experience." *Cognition* 100, no. 1 (2006): 131-72.

Bigand, Emmanuel, S. Vieillard, F. Madurell, J. Marozeau, and A. Dacquet. "Multidimensional Scaling of Emotional Responses to Music: The Effect of Musical Expertise and of the Duration of the Excerpts." *Cognition & Emotion* 19, no. 8 (2005): 1113-39.

Björk, E.A. "The Perceived Quality of Natural Sounds." *Acustica* 57 (1985): 185-88.

Blake, David K. "Timbre as Differentiation in Indie Music." *Music Theory Online* 18, no. 2 (2012).

Bloechl, Olivia. *Native American Song at the Frontiers of Early Modern Music*. Cambridge: Cambridge University Press, 2008.

Blood, Anne J., R.J. Zatorre, P. Bermudez, and A.C. Evans. "Emotional Responses to Pleasant and Unpleasant Music Correlate with Activity in Paralimbic Brain Regions." *Nature: Neuroscience* 2, no. 4 (1999): 382-87.

Blumstein, Daniel T., Gregory A. Bryant, and Peter Kaye. "The Sound of Arousal in Music is Context-Dependent." *Biology Letters* 8 (2012): 744-47.

Blumstein, Daniel T., and C. Récapet. "The Sound of Arousal: The Addition of Novel Nonlinearities Increases Responsiveness in Marmot Alarm Calls." *Ethology* 115 (2009): 1074-81.

Bogue, Ronald. "Violence in Three Shades of Metal: Death, Doom and Black." In *Deleuze and Music*, edited by Ian Buchanan and Marcel Swiboda. Edinburgh: Edinburgh University Press, 2004.

Borch, D.Z., J. Sundberg, P.A. Lindestad, and M. Thalen. "Vocal Fold Vibration and Voice Course Aperiodicity in 'Dist' Tones: A Study of a Timbral Ornament in Rock Singing." *Logopedics Phoniatrics Vocology* 29 (2004): 147-53.

Borgo, David. "Between Worlds: The Embodied and Ecstatic Sounds of Jazz." *Open Space* 5 (2003).

———. "Reverence for Uncertainty: Chaos, Order, and the Dynamics of Musical Free Improvisation." Ph.D. Diss., UCLA, 1999.

———. *Sync or Swarm: Improvising Music in a Complex Age*. New York and London: Continuum, 2005.

Bregman, Alfred S. *Auditory Scene Analysis: The Perceptual Organization of Sound*. Cambridge, MA: MIT Press, 1990.

Brown, Steven, E. Ngan, and M. Liotti. "A Larynx Area in the Human Motor Cortex." *Cerebral Cortex* 18 (2008): 837-45.

Bryson, Bethany. "'Anything But Heavy Metal': Symbolic Exclusion and Musical Dislikes." *American Sociological Review* 61 (1996): 884-99.

Buccino, Giovanni, S. Vogt, A. Ritzl, G.R. Fink, K. Zilles, H.-J. Freund, and G. Rizzolatti. "Neural Circuits Underlying imitation Learning of Hand Actions: An Event-Related fMRI Study." *Neuron* 42 (2004): 323-34.

Burnett, Charles. "Sound in the Middle Ages." In *The Second Sense: Studies in Hearing and Musical Judgement from Antiquity to the Seventeenth Century*, edited by Charles Burnett, Michael Fend and Penelope Gouk. London: University of London Press, 1991.

Butler, Robert A. "The Relative Influence of Pitch and Timbre on the Apparent Location of Sound in the Median Sagittal Plane." *Perception and Psychophysics* 14, no. 2 (1973): 255-58.

Cabrera, D. "'Psysound': A computer program for psychoacoustical analysis." *Proceedings of the Australian Acoustical Society Conference, Melbourne* (1999): 47-54.

Caclin, Anne, E. Brattico, M. Tervaniemi, R. Näätänen, D. Morlet, M.-H. Giard, and S. McAdams. "Separate Neural Processing of Timbre Dimensions in Auditory Sensory Memory." *Journal of Cognitive Neuroscience* 18, no. 12 (2006): 1959-72.

Caclin, Anne, S. McAdams, B.K. Smith, and M.-H. Giard. "Interactive Processing of Timbre Dimensions: An Exploration with Event-related Potentials." *Journal of Cognitive Neuroscience* 20, no. 1 (2008): 49-64.

Callan, Daniel E., V. Tsytsarev, T. Hanakawa, A. Callan, M. Katsuhara, H. Fukuyama, and R. Turner. "Song and Speech: Brain Regions Involved with Perception and Covert Production." *Neuroimage* 31 (2006): 1327-42.

Calvo-Merino, B., D.E. Glaser, J. Grèzes, R.E. Passingham, and P. Haggard. "Action Observation and Acquired Motor Skills: An fMRI Study with Expert Dancers." *Cerebral Cortex* 15 (2005): 1243-49.

Carr, Laurie, M. Iacoboni, M.-C. Dubeau, J.C. Mazziotta, and G.L. Lenzi. "Neural Mechanisms of Empathy in Humans: A Relay from Neural Systems for Imitation to Limbic Systems." *Proceedings of the National Academy of Sciences of the USA* 100, no. 9 (2003): 5497-502.

Carson, Anne. *Glass, Irony and God*. New York: New Directions, 1995.

Cavarero, Adriana. *For More Than One Voice: Toward a Philosophy of Vocal Expression*. Translated by Paul A. Kottman. Palo Alto, CA: Stanford University Press, 2005.

Chartrand, Jean-Pierre, and Pascal Belin. "Superior Voice Timbre Processing in Musicians." *Neuroscience Letters* 405 (2006): 154-67.

Chartrand, Tanya, and John A. Bargh. "The Chameleon Effect: The Perception-Behavior Link and Social Interaction." *Journal of Personality and Social Psychology* 76, no. 6 (1999): 893-910.

Chen, Jer-Ming, John Smith, and Joe Wolfe. "Experienced Saxophonists Learn to Tune Their Vocal Tracts." *Science* 319 (2008): 776.

Chidester, Phillip. "Through the Blood: Death Metal and the Rhetoric of Song as a Transcendent Discursive/Presentational Form." Ph.D Diss., University of Kansas, 2004.

Clark, Andy. *Being There: Putting Brain, Body, and World Together Again*. Cambridge, MA: MIT Press, 1998.

———. *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*. New York: Oxford University Press, 2008.

Clarke, Eric F. *Ways of Listening: An Ecological Approach to the Perception of Musical Meaning*. Oxford: Oxford University Press, 2005.

Clifton, Thomas. *Music as Heard: A Study in Applied Phenomenology*. New Haven: Yale University Press, 1983.

Clynes, Manfred. *Sentics: The Touch of Emotions*. New York: Doubleday, 1977.

Cobussen, Marcel. *Thresholds: Rethinking Spirituality Through Music*. Aldershot, UK: Ashgate, 2008.

Cogan, Robert. *New Images of Musical Sound*. Cambridge, MA: Harvard University Press, 1984.

Cox, Arnie. "Embodying Music: Principles of the Mimetic Hypothesis." *Music Theory Online* 17, no. 2 (2011).

———. "The Metaphoric Logic of Musical Motion and Space." Ph.D. Diss., University of Oregon, 1999.

———. "The Mimetic Hypothesis and Embodied Musical Meaning." *Musicae Scientiae* 5, no. 2 (2001): 195-212.

Crowder, Robert C. "Auditory Memory." In *Thinking in Sound: The Cognitive Psychology of Human Audition*, edited by Stephen McAdams and Emmanuel Bigand. Oxford: Clarendon Press, 1993.

———. "Imagery for Musical Timbre." *Journal of Experimental Psychology: Human Perception and Performance* 15, no. 3 (1989): 472-78.

Cruz, Jon. *Culture on the Margins: The Black Spiritual and the Rise of American Cultural Interpretation*. Princeton, NJ: Princeton University Press, 1999.

Damasio, Antonio. *Descartes' Error: Emotion, Reason, and the Human Brain*. New York: Penguin, 1994.

———. *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*. New York: Mariner, 2000.

Darke, Graham. "Assessment of Timbre Using Verbal Attributes." Paper presented at the Conference on Interdisciplinary Musicology, Montreal, Quebec, Canada, 2005.

Darwin, Charles. *The Descent of Man, and Selection in Relation to Sex*. London: John Murray, 1871.

De Souza, Jonathan. "Musical Instruments, Bodies, and Cognition." Ph.D. Diss., University of Chicago, 2013.

"Death Metal: A Documentary." Grimoire, 2004.

DeMichael, Don. "John Coltrane and Eric Dolphy Answer Their Critics." *Down Beat*, April 12, 1962, 20-23.

di Pellegrino, G., L. Fadiga, L. Fogassi, V. Gallese, and G. Rizzolatti. "Understanding Motor Events: A Neurophysiological Study." *Experimental Brain Research* 91 (1992): 176-80.

Dibben, Nicola. "What Do We Hear, When We Hear Music? Music Perception and Musical Material." *Musicae Scientiae* 5, no. 2 (2001): 161-94.

Dolan, Emily I. *The Orchestral Revolution: Haydn and the Technologies of Timbre*. Cambridge: Cambridge University Press, 2012.

Douglas, Mary. *Purity and Danger: An Analysis of the Concepts of Pollution and Taboo*. London and New York: Routledge Classics, 2002.

Drost, Ulrich C., Martina Rieger, and Wolfgang Prinz. "Instrument Specificity in Experienced Musicians." *The Quarterly Journal of Experimental Psychology* 60, no. 4 (2007): 527-33.

Drott, Eric. "Free Jazz and the French Critic." *Journal of the American Musicological Society* 61, no. 3 (2008): 541-81.

Dumas, Henry. *Echo Tree: The Collected Short Fiction of Henry Dumas*. New York: Coffee House Press, 2003.

Dunn, Sam. "Metal: A Headbanger's Journey." Warner Home Videos, 2005.

Echard, Richard. *Neil Young and the Poetics of Energy*. Bloomington, IN: University of Indiana Press, 2005.

Eerola, Tuomas, Rafael Ferrer, and Vinoo Alluri. "Timbre and Affect Dimensions: Evidence from Affect and Similarity Ratings and Acoustic Correlates of Isolated Instrument Sounds." *Music Perception* 30, no. 1 (2012): 49-70.

Eerola, Tuomas, and Jonna K. Vuoskoski. “A Review of Music and Emotion Studies: Approaches, Emotion Models, and Stimuli.” *Music Perception* 30, no. 3 (2013): 307-40.

Eidsheim, Nina Sun. “Sensing Voice: Materiality and the Lived Body in Singing and Listening.” *Senses & Society* 6, no. 2 (2011): 133-55.

———. “Synthesizing Race: Towards an Analysis of the Performativity of Vocal Timbre.” *TRANS-Transcultural Music Review* 13 (2009).

———. “Voice as a Technology of Selfhood: Towards an Analysis of Racialized Timbre and Vocal Performance.” Ph.D. Diss., University of California–San Diego, 2008.

Eitan, Zohar, and Inbar Rothschild. “How Music Touches: Musical Parameters and Listeners’ Audio-Tactile Metaphorical Mappings.” *Psychology of Music* 39, no. 4 (2010): 449-67.

Erickson, Robert. *Sound Structure in Music*. Berkeley and Los Angeles: University of California Press, 1975.

Fadiga, Luciano, L. Craighero, G. Buccino, and G. Rizzolatti. “Speech Listening Specifically Modulates the Excitability of Tongue Muscles: A TMS Study.” *European Journal of Neuroscience* 15 (2002): 399-402.

Fales, Cornelia. “Listening to Timbre During the French Enlightenment.” Paper presented at the Conference on Interdisciplinary Musicology, Montreal, Quebec, Canada, 2005.

———. “The Paradox of Timbre.” *Ethnomusicology* 46, no. 1 (2002): 56-95.

———. “Short-Circuiting Perceptual Systems: Timbre in Ambient and Techno Music.” In *Wired for Sound: Engineering and Technologies in Sonic Culture*, edited by Paul D. Greene and Thomas Porcello. Middleton, CN: Wesleyan University Press, 2005.

Fales, Cornelia, and Stephen McAdams. “The Fusion and Layering of Noise and Tone: Implications for Timbre in African Instruments.” *Leonardo Music Journal* 4 (1994): 69-77.

Feld, Steven, Aaron A. Fox, Thomas Porcello, and David Samuels. “Vocal Anthropology.” In *A Companion to Linguistic Anthropology*, edited by Alessandro Duranti. Maldon, UK: Blackwell, 2004.

Ferrer, Rafael. “Embodied Cognition Applied to Timbre and Music Appreciation: Theoretical Foundation.” *British Postgraduate Musicology* 10 (2009).

———. “Timbral Environments: An Ecological Approach to the Cognition of Timbre.” *Empirical Musicology Review* 6, no. 2 (2011): 64-74.

Ferrer, Rafael, and Tuomas Eerola. “Semantic Structures of Timbre Emerging from Social and Acoustic Descriptions of Music.” *Journal of Audio, Speech, and Music Processing* 11 (2011): 1-16.

Fink, Robert. "'Arrows of Desire': Long-Range Linear Structure and the Transformation of Musical Energy." Ph.D. Diss., University of California–Berkeley, 1994.

Fink, Robert, Mindy LaTour O'Brien, and Zachary Wallmark, eds. *The Relentless Pursuit of Tone: Timbre in Popular Music*. New York: Oxford University Press, forthcoming.

Fitch, W.T., J. Neubauer, and H. Herzel. "Calls Out of Chaos: The Adaptive Significance of Nonlinear Phenomena in Mammalian Vocal Production." *Animal Behavior* 63 (2002): 407-18.

Fodor, Jerry A. *The Modularity of the Mind*. Cambridge, MA: MIT Press, 1983.

Forster, Chris. *Musical Mathematics: On the Art and Science of Acoustic Instruments*. San Francisco Chronicle, 2010.

Fox, Ted. *In the Groove: The People Behind the Music*. New York: St. Martin's Press, 1986.

Foxe, J.J., G.R. Wylie, A. Martinez, C.E. Schroeder, D.C. Javitt, D. Guifoyle, W. Ritter, and M.M. Murray. "Auditory-Somatosensory Multisensory Processing in Auditory Association Cortex: An fMRI Study." *Journal of Neurophysiology* 88, no. 1 (2002): 540-43.

Freedberg, David, and Vittorio Gallese. "Motion, Emotion and Empathy in Aesthetic Experience." *Trends in Cognitive Sciences* 11, no. 5 (2007): 197-203.

Friedheim, Philip. "Wagner and the Aesthetics of the Scream." *19th-Century Music* 7, no. 1 (1983): 63-70.

Frith, Simon. *Performing Rites: On the Value of Popular Music*. Cambridge, MA: Harvard University Press, 1998.

Gabrielsson, Alf, and Patrik N. Juslin. "Emotional Expression in Music Performance: Between the Performer's Intention and the Listener's Experience." *Psychology of Music* 24, no. 68 (1996): 68-91.

Gallese, Vittorio, L. Fadiga, L. Fogassi, and G. Rizzolatti. "Action Recognition in the Premotor Cortex." *Brain* 119, no. 2 (1996): 593-609.

Gallese, Vittorio, and George Lakoff. "The Brain's Concepts: The Role of the Sensory-Motor System in Conceptual Knowledge." *Cognitive Neuropsychology* 22, no. 3-4 (2005): 455-79.

Gaver, William W. "How Do We Hear in the World?: Explorations in Ecological Acoustics." *Ecological Psychology* 5, no. 4 (1993): 285-313.

Gazzola, Valeria, Lisa Aziz-Zadeh, and Christian Keysers. "Empathy and the Somatotopic Auditory Mirror System in Humans." *Current Biology* 16 (2006): 1824-29.

Gibiat, Vincent, and Michèle Castellengo. "Period Doubling Occurences in Wind Instruments Musical Performance." *Acustica* 86 (2000): 746-54.

Gibson, James J. *The Ecological Approach to Visual Perception*. Boston: Houghton-Mifflin, 1979.

———. *The Senses Considered as Perceptual Systems*. Boston: Houghton Mifflin, 1966.

Giordano, Bruno L., and Stephen McAdams. "Sound Source Mechanics and Musical Timbre Perception: Evidence from Previous Studies." *Music Perception* 28, no. 2 (2010): 155-68.

Girard, René. *Violence and the Sacred*. Translated by Patrick Gregory. Baltimore: Johns Hopkins University Press, 1977.

Gitler, Ira. "Double View of Coltrane 'Live'." *Down Beat*, April 26, 1962, 29.

Gjerdingen, Robert O., and David Perrott. "Scanning the Dial: The Rapid Recognition of Music Genres." *Journal of New Music Research* 37, no. 2 (2008): 93-100.

Gladwell, Malcolm. *Blink: The Power of Thinking Without Thinking*. New York: Little, Brown and Co., 2005.

Godøy, Rolf Inge. "Motor-Mimetic Music Cognition." *Leonardo Music Journal* 36, no. 4 (2003): 317-19.

Godøy, Rolf Inge, and Marc Leman, eds. *Musical Gestures: Sound, Movement, and Meaning*. New York and London: Routledge, 2010.

Goodchild, Megan. "Towards a Perceptually-Based Theory of Orchestral Gestures." Paper presented at the AMS/SEM/SMT Joint Conference, New Orleans, LA, 2012.

Gowensmith, William Neil, and Larry J. Bloom. "The Effects of Heavy Metal Music on Arousal and Anger." *Journal of Music Therapy* 34 (1997): 33-45.

Goydke, Katja N., E. Altenmüller, J. Möller, and T. Münte. "Changes in Emotional Tone and Instrumental Timbre are Reflected by the Mismatch Negativity." *Cognitive Brain Research* 21 (2004): 351-59.

Gracyk, Theodore. *Rhythm and Noise: An Aesthetics of Rock*. Raleigh-Durham: Duke University Press, 1996.

Grahn, Jessica A. "Neural Mechanisms of Rhythm Perception: Current Findings and Future Perspectives." *Topics in Cognitive Science* 4 (2012): 585-606.

Green, Joshua, and J. Haidt. "How (and Where) Does Moral Judgment Work?". *Trends in Cognitive Sciences* 6, no. 12 (2002): 517-23.

Gregory, Andrew H. "Timbre and Auditory Streaming." *Music Perception* 12, no. 2 (1994): 161-74.

Grey, De Sayles. "John Coltrane and the 'Avant-Garde' Movement in Jazz History." Ph.D. Diss., University of Pittsburgh, 1986.

Grey, John M. "An Exploration of Musical Timbre." Ph.D. Dissertation, Stanford University, 1975.

———. "Multidimensional Perceptual Scaling of Musical Timbres." *Journal of the Acoustical Society of America* 61 (1977): 1270-77.

Gridley, Mark C. "Misconceptions in Linking Free Jazz with the Civil Rights Movement." *College Music Symposium* 47 (2007): 139-55.

Gridley, Mark C., and Robert Hoff. "Are Music Perceptions Biased by Priming Effects of Journalism?". *Psychology Journal* 7, no. 2 (2010): 55-60.

———. "Do Mirror Neurons Explain Misattribution of Emotions in Music?". *Perceptual & Motor Skills* 102, no. 2 (2006): 600-02.

———. "Who's Actually Angry, John Coltrane or His Critics?". *Psychology Journal* 4, no. 4 (2007): 153-60.

Guest, Steve, C. Catmur, D. Lloyd, and C. Spence. "Audiotactile Interactions in Roughness Perception." *Experimental Brain Research* 146 (2002): 161-71.

Hahn, Tomie. *Sensational Knowledge: Transmitting Japanese Dance and Music*. Middleton, CN: Wesleyan University Press, 2007.

Hailstone, Julia C., R. Omar, S.M.D. Henley, C. Frost, M.G. Kenward, and J.D. Warren. "It's Not What You Play, It's How You Play It: Timbre Affects Perception of Emotion in Music." *The Quarterly Journal of Experimental Psychology* 62, no. 11 (2009): 2141-55.

Hajda, John. "The Effect of Dynamic Acoustical Features on Musical Timbre." In *Analysis, Synthesis, and Perception of Musical Sounds*, edited by James Beauchamp. New York: Springer, 2007.

Hajda, John, R.A. Kendall, E.C. Carterette, and M.L. Harshberger. "Methodological Issues in Timbre Research." In *Perception and Cognition of Music*, edited by Iréne Deliége and John Sloboda. East Sussex: Psychology Press, 1997.

Halmrast, Tor, K. Guettler, R. Bader, and R.I. Godøy. "Gesture and Timbre." In *Musical Gestures: Sound, Movement, and Meaning*, edited by Rolf Inge Godøy and Marc Leman. New York: Routledge, 2010.

Halpern, Andrea R., R.J. Zatorre, M. Bouffard, and J.A. Johnson. "Behavioral and Neural Correlates of Perceived and Imagined Musical Timbre." *Neuropsychologia* 42 (2004): 1281-92.

Handel, Stephen. *Listening: An Introduction to the Perception of Auditory Events*. Cambridge, MA: MIT Press, 1989.

———. "Timbre Perception and Auditory Object Identification." In *Hearing*, edited by Brian C. J. Moore. San Diego: Academic Press, 1995.

Harker, Brian. *Jazz: An American Journey*. Upper Saddle River, NJ: Pearson, 2005.

Harrell, Jack. "The Poetics of Destruction: Death Metal Rock." *Popular Music and Society* 18, no. 1 (1994): 91-103.

Haslinger, B., P. Erhard, E. Altenmüller, U. Schroeder, H. Boecker, and A.O. Ceballos-Baumann. "Transmodal Sensorimotor Networks during Action Observation in Professional Pianists." *Journal of Cognitive Neuroscience* 17, no. 2 (2005): 282-93.

Hauser, Marc D. "The Sound and the Fury: Primate Vocalizations as Reflections of Emotion and Thought." In *The Origins of Music*, edited by Nils J. Wallin, Björn Merker, and Steven Brown. Cambridge, MA: MIT Press, 2000.

Hegarty, Paul. *Noise/Music: A History*. New York: Continuum, 2007.

Helmholtz, Hermann von. *On the Sensations of Tone as a Physiological Basis for the Theory of Music*. Translated by Alexander Ellis. New York: Dover, 1867/1954.

Hentoff, Nat. Liner notes to John Coltrane's *Meditations*. Impulse Records A-9110, 1996, CD. Originally released in 1965.

Hester, Karlton. "The Melodic and Polyrhythmic Development of John Coltrane's Spontaneous Composition in a Racist Society." Ph.D. Diss., CUNY, 1990.

Hickok, Gregory, B. Buchsbaum, C. Humphries, and T. Muftuler. "Auditory-Motor Interaction Revealed by fMRI: Speech, Music, and Working Memory in Area Spt." *Journal of Cognitive Neuroscience* 15, no. 5 (2003): 673-82.

Howard, David M., and James Angus. *Acoustics and Psychoacoustics*. 2nd ed. Amsterdam: Focal Press, 2001.

Howes, David, ed. *Empire of the Senses: The Sensual Culture Reader*. Oxford and New York: Berg, 2005.

Huettel, Scott A., Allen W. Song, and Gregory McCarthy. *Functional Magnetic Resonance Imaging*. 2 ed. Sunderland, MA: Sinauer, 2008.

Hunter, David. *Guitar Effects Pedals: The Practice Handbook*. New York: Backbeat, 2004.

Huron, David. "The New Empiricism: Systematic Musicology in a Postmodern Age." Paper presented at the Ernest Bloch Lectures, University of California, Berkeley, 1999.

———. *Sweet Anticipation: Music and the Psychology of Expectation*. Cambridge, MA: MIT Press, 2006.

———. "Toward a Theory of Timbre." Paper presented at the 12th Annual Conference of Music Theory Midwest, Cincinnati, OH, 2001.

Iacoboni, Marco. *Mirroring People: The Science of Empathy and How We Connect with Others*. New York: Picador, 2009.

Iacoboni, Marco, R.P. Woods, M. Brass, H. Bekkering, J.C. Mazziotta, and G. Rizzolatti. "Cortical Mechanisms of Human Imitation." *Science* 286 (1999): 2526-28.

Iverson, Paul, and Carol L. Krumhansl. "Isolating the Dynamic Attributes of Musical Timbre." *Journal of the Acoustical Society of America* 94, no. 5 (1993): 2595-603.

Iyer, Vijay. "Microstructures of Feel, Macrostructures of Sound: Embodied Cognition in West African and African-American Musics." Ph.D. Diss., University of California–Berkeley, 1998.

James, Clive. *Cultural Amnesia: Necessary Memories from History and the Arts*. New York: Norton, 2007.

James, Kieran. "From 'The Undead Will Feast' to 'The Time to Kill is Now': Frankfurt School and Freudian Perspectives on Death-Metal." *Musicology Australia* 31 (2009): 17-39.

James, William. *The Principles of Psychology*. Vol. 2, New York: Dover, 1890/1950.

Jenkinson, M., P. Bannister, J.M. Brady, and S.M. Smith. "Improved optimisation for the robust and accurate linear registration and motion correction of brain images." *Neuroimage* 17, no. 2 (2002): 825-41.

Jenkinson, M., and S.M. Smith. "A global optimisation method for robust affine registration of brain images." *Medical Image Analysis* 5 (2001): 143-56.

Johnson, Mark. *The Body in the Mind: The Bodily Basis of Meaning, Imagination, and Reason*. Chicago: University of Chicago Press, 1987.

———. *The Meaning of the Body: Aesthetics of Human Understanding*. Chicago: University of Chicago Press, 2007.

Johnson, Mark, and Steve Larson. "'Something in the Way She Moves'—Metaphors of Musical Motion." *Metaphor and Symbol* 18, no. 2 (2003): 63-84.

Johnstone, Tom, and Klaus R. Scherer. "The Effects of Emotions on Voice Quality." *Proceedings of the 14th International Congress of Phonetic Sciences, San Francisco* (1999): 2029-32.

Jost, Ekkehard. *Free Jazz*. New York: Da Capo, 1981.

Juslin, Patrik N. "Communicating Emotion in Music Performance: A Review and Theoretical Framework." In *Music and Emotion: Theory and Research*, edited by Patrik N. Juslin and John A. Sloboda. Oxford and New York: Oxford University Press, 2001.

Juslin, Patrik N., and Petri Laukka. "Communication of Emotions in Vocal Expression and Music Performance: Different Channels, Same Code?". *Psychological Bulletin* 129, no. 5 (2003): 770-814.

Juslin, Patrik N., and Daniel Västfjäll. "Emotional Responses to Music: The Need to Consider Underlying Mechanisms." *Behavioral and Brain Sciences* 31 (2008): 559-621.

Kahn, Ashley. *The House that Trane Built: The Story of Impulse Records*. New York: Norton, 2006.

Kahn, Jennifer. "Can You Call a 9-Year-Old a Psychopath?" *New York Times Magazine*, 11 May, 2012.

Kahn-Harris, Keith. *Extreme Metal: Music and Culture on the Edge*. Oxford and New York: Berg, 2007.

Kalakoski, Virpi. "Musical Imagery and Working Memory." In *Musical Imagery*, edited by R.I. Godøy and H. Jørgensen. Lisse: Swets & Zeitlinger, 2001.

Kaplan, Jonas T., and Marco Iacoboni. "Getting a Grip on Other Minds: Mirror Neurons, Intention Understanding, and Cognitive Empathy." *Social Neuroscience* 1, no. 3-4 (2006): 175-83.

Kendall, Roger A. Music Experiment Development System (MEDS) Version 2002-B-1.

———. "The Role of Acoustic Signal Partitions in Listener Categorization of Musical Phrases." *Music Perception* 4, no. 2 (1986): 185-213.

Kendall, Roger A., and Edward C. Carterette. "Difference Thresholds for Timbre Related to Spectral Centroid." Paper presented at the 4th ICMPC, Montreal, Quebec, Canada, 1996.

———. "Perceptual Scaling of Simultaneous Wind Instrument Timbres." *Music Perception* 8, no. 4 (1991): 369-404.

———. "Verbal Attributes of Simultaneous Wind Instrument Timbres: I. von Bismarck's Adjectives." *Music Perception* 10, no. 4 (1993): 445-67.

———. "Verbal Attributes of Simultaneous Wind Instrument Timbres: II. Adjectives Induced from Piston's 'Orchestration." *Music Perception* 10, no. 4 (1993): 469-501.

Kendall, Roger A., 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.

Kendall, Roger A., and Pantelis N. Vassilakis. "Perception and Acoustical Analyses of Traditionally Orchestrated Triads vs. Non-Traditional Counterparts." Paper presented at the 2nd Pan-American/Iberian Meeting on Acoustics, Cancun, Mexico, 2010.

Keysers, Christian, E. Kohler, M.A. Umiltà, L. Nanetti, L. Fogassi, and V. Gallese. "Audiovisual Mirror Neurons and Action Recognition." *Experimental Brain Research* 153 (2003): 628-36.

Kivy, Peter. *Introduction to a Philosophy of Music*. New York: Oxford University Press, 2002.

Koelsch, Stefan, T. Fritz, D.Y. von Cramon, K. Müller, and A.D. Friederici. "Investigating Emotion with Music: An fMRI Study." *Human Brain Mapping* 27 (2006): 239-50.

Koelsch, Stefan, and Walter A. Siebel. "Towards a Neural Basis of Music Perception." *Trends in Cognitive Sciences* 9, no. 12 (2005): 578-84.

Kofsky, Frank. *Black Nationalism and the Revolution in Music*. New York: Pathfinder, 1970.

———. *John Coltrane and the Jazz Revolution of the 1960s*. New York: Pathfinder, 1998.

Kohler, Evelyne, C. Keysers, M.A. Umiltà, L. Fogassi, V. Gallese, and G. Rizzolatti. "Hearing Sounds, Understanding Actions: Action Representation in Mirror Neurons." *Science* 297, no. 846 (2002): 846-48.

Kosko, Bart. *Noise*. New York: Viking, 2006.

Kreiman, Jody, and Diana Sidtis. *Foundations of Voice Studies: An Interdisciplinary Approach to Voice Production and Perception*. Malden, MA: Wiley-Blackwell, 2011.

Krumhansl, Carol L. "Plink: 'Thin Slices' of Music." *Music Perception* 27, no. 5 (2010): 337-54.

Kumar, Sukhbinder, K. von Kriegstein, K. Friston, and T.D. Griffiths. "Features versus Feelings: Dissociable Representations of the Acoustic Features and Valence of Aversive Sounds." *The Journal of Neuroscience* 32, no. 41 (2012): 14184-92.

Lacey, Simon, Randall Stilla, and Krish Sathian. "Metaphorically Feeling: Comprehending Textural Metaphors Activates Somatosensory Cortex." *Brain and Language* 120 (2012): 416-21.

Lakoff, George, and Mark Johnson. *Metaphors We Live By*. Chicago: University of Chicago Press, 1980.

———. *Philosophy in the Flesh*. New York: Basic Books, 1999.

Larkin, Philip. *All What Jazz*. New York: Farrar, Straus & Giroux, 1970.

Larson, Steve. *Musical Forces: Motion, Metaphor, and Meaning in Music*. Bloomington, IN: University of Indiana Press, 2012.

Lartillot, O., and P. Toiviainen. "A Matlab toolbox for musical feature extraction from audio." *Proceedings of the 10th International Conference on Digital Audio Effects* (2007): 237-44.

Laukka, Petri, Patrik N. Juslin, and Roberto Bresin. "A Dimensional Approach to Vocal Expression of Emotion." *Cognition & Emotion* 19, no. 5 (2005): 633-53.

Le Guin, Elisabeth. *Boccherini's Body: An Essay in Carnal Musicology*. Berkeley and Los Angeles: University of California Press, 2005.

LeDoux, Joseph. *The Emotional Brain: The Mysterious Underpinnings of Emotional Life*. New York: Simon & Schuster, 1998.

Lem, Nolan. Email communication with the author, May 7, 2010.

Leman, Marc. *Embodied Music Cognition and Mediation Technology*. Cambridge, MA: MIT Press, 2007.

Levinas, Emmanuel. *Otherwise Than Being: or, Beyond Essence*. Translated by Alphonso Lingis. The Hague and Boston: M. Nijhoff, 1981.

———. *Totality and Infinity: An Essay on Exteriority*. Translated by Alphonso Lingis. Pittsburgh: Duquesne University Press, 1969.

Lewis, George E. "Improvised Music Since 1950: Afrological and Eurological Forms." *Black Music Research Journal* 16, no. 1 (1996): 91-119.

Leydon, Rebecca. "Clean as a Whistle: Timbral Trajectories and the Modern Musical Sublime." *Music Theory Online* 18, no. 2 (2012).

Liberman, A.M., F.S. Cooper, D.P. Shankweiler, and M. Studdert-Kennedy. "Perception of the speech code." *Psychological Review* 74, no. 6 (1967): 431-61.

Lichte, W. F. "Attributes of Complex Tones." *Journal of Experimental Psychology* 28 (1941): 455-80.

Lidov, David. *Is Language a Music? Writings on Musical Form and Signification*. Bloomington, IN: Indiana University Press, 2005.

———. "Mind and Body in Music." *Semiotica* 66, no. 1 (1987): 66-97.

Liebman, David. *Developing a Personal Saxophone Sound*. Medfield, MA: Dorn, 1989.

———. "John Coltrane's Meditations Suite: A Study in Symmetry." *Annual Review of Jazz Studies* 8 (1996): 167-80.

Lipps, Theodor. *Ästhetik*. Berlin: B. G. Teubner, 1907.

Logie, Robert, and J. Edworthy. "Shared Mechanisms in the Processing of Verbal and Musical Materials." In *Imagery*, edited by D.G. Russell, D. Marks and J. Richardson. Dunedin, NZ: Human Performance Associates, 1986.

Lotze, M., G. Scheler, H.R.M. Tan, C. Braun, and N. Birnbaumer. "The Musician's Brain: Functional Imaging of Amateurs and Professionals During Performance and Imagery." *Neuroimage* 20 (2003): 1817-29.

Lucretius. *On the Nature of the Universe*. Translated by R.E. Latham. London and New York: Penguin, 1951.

Margulis, Elizabeth Hellmuth, L.M. Mlsna, A.K. Uppunda, T.B. Parrish, and P.C.M. Wong. "Selective Neurophysiologic Responses to Music in Instrumentalists with Different Listening Biographies." *Human Brain Mapping* 30 (2009): 267-75.

Mathieu, Bill. "Review of *Ascension*." *Down Beat*, June 23, 1965.

Maurer, D., T. Pathman, and C.J. Mondloch. "The Shape of Boubas: Sound-Shape Correspondences in Toddlers and Adults." *Developmental Science* 9, no. 3 (2006): 316-22.

McAdams, Stephen. "Recognition of Sound Sources and Events." In *Thinking in Sound: The Cognitive Psychology of Human Audition*, edited by Stephen McAdams and Emmanuel Bigand. Oxford: Clarendon Press, 1993.

McAdams, Stephen, S. Winsberg, S. Donnadieu, G. De Soete, and J. Krimphoff. "Perceptual Scaling of Synthesized Musical Timbres: Common Dimensions, Specificities, and Latent Subject Classes." *Psychological Research* 58 (1995): 177-92.

McAleer, Phil, Alexander Todorov, and Pascal Belin. "How Do You Say 'Hello'? Personality Impressions from Brief Novel Voices." *PLOS One* (12 March 2014).

McClary, Susan. *Feminine Endings: Music, Gender, and Sexuality*. Minneapolis: University of Minnesota Press, 1991.

McClary, Susan, and Robert Walser. "Theorizing the Body in African-American Music." *Black Music Research Journal* 14, no. 1 (1994): 75-84.

McDermott, Josh H. "Auditory Preferences and Aesthetics: Music, Voices, and Everyday Sounds." In *Neuroscience of Preference and Choice: Cognitive and Neural Mechanisms*, edited by Raymond Dolan and Tali Sharot. London: Academic, 2012.

McLucas, Anne Dhu. *The Musical Ear: Oral Tradition in the USA*. Aldershot, UK: Ashgate, 2010.

Mead, Andrew. "Bodily Hearing: Physiological Metaphors and Musical Understanding." *Journal of Music Theory* 43, no. 1 (1999): 1-19.

Mende, W., H. Herzel, and K. Wermke. "Bifurcations and Chaos in Newborn Infant Cries." *Physics Letters A* 145 (1990): 418-24.

Menon, Vinod, D.J. Levitin, B.K. Smith, A. Lembke, B.D. Krasnow, D. Glazer, G.H. Glover, and S. McAdams. "Neural Correlates of Timbre Change in Harmonic Sounds." *Neuroimage* 17 (2002): 1742-54.

Merleau-Ponty, Maurice. *Phenomenology of Perception*. Translated by Colin Smith. London and New York: Routledge, 1962.

Middlebrooks, J.C., and D.M. Greenhaw. "Sound Localization by Human Listeners." *Annual Review of Psychology* 42, no. 1 (1991): 135.

Middleton, Richard. *Studying Popular Music*. Berkshire, UK: Open University Press, 1990.

Miller, Steve. *Experimental Design and Statistics*. London: Routledge, 1974.

Molnar-Szakacs, Istvan, and Katie Overy. "Music and Mirror Neurons: From Motion to 'E'motion." *Social Cognitive and Affective Neuroscience* 1, no. 3 (2006): 235-41.

Moore, Allan. *Rock: The Primary Text*. Surrey: Ashgate, 2001.

Moore, B.C.J., R. Peters, and R.W. Peters. "Thresholds for the Detection of Inharmonicity in Complex Tones." *Journal of the Acoustical Society of America* 77 (1985): 1861-67.

Mudrian, Albert. *Choosing Death: The Improbable History of Death Metal and Grindcore*. Los Angeles: Feral House, 2004.

Mukamel, Roy, A.D. Ekstrom, J. Kaplan, M. Iacoboni, and I. Fried. "Single-Neuron Responses in Humans during Execution and Observation of Actions." *Current Biology* 20, no. 8 (2010): 750-56.

Senate Committee on Governmental Affairs. *Music Violence: How Does it Affect Our Chidren?*, 105th Congress-395, 1997.

Nattiez, Jean-Jacques. *Music and Discourse: Towards a Semiology of Music*. Translated by Carolyn Abbate. Princeton: Princeton University Press, 1991.

Neal, Jocelyn. "The Twang Factor in Country Music." In *The Relentless Pursuit of Tone: Timbre in Popular Music*, edited by Robert Fink, Mindy LaTour O'Brien and Zachary Wallmark. New York: Oxford University Press, forthcoming.

Nisenson, Eric. *Ascension: John Coltrane and His Quest*. New York: St. Martin's Press, 1993.

Noë, Alva. *Action in Perception*. Cambridge, MA: MIT Press, 2004.

Obadike, Mendi. "Low Fidelity: Stereotyped Blackness in the Field of Sound." Ph.D. Diss., Duke University, 2005.

Osgood, C.E., G.J. Suci, and P.H. Tannenbaum. *The Measurement of Meaning*. Urbana: University of Illinois Press, 1957.

Overy, Katie, and Istvan Molnar-Szakacs. "Being Together in Time: Musical Experience and the Mirror Neuron System." *Music Perception* 26, no. 5 (2009): 489-504.

Panksepp, J., and J.B. Panksepp. "The Seven Sins of Evolutionary Psychology." *Evolution and Cognition* 6 (2000): 108-31.

Pantev, Christo, L.E. Roberts, M. Schulz, A. Engelien, and B. Ross. "Timbre-specific Enhancement of Auditory Cortical Representations in Musicians." *Neuroreport* 12, no. 1 (2001): 169-74.

Patel, Aniruddh D. *Music, Language, and the Brain*. New York: Oxford University Press, 1999.

Peeters, Geoffroy, B.L. Giordano, P. Susini, N. Misdariis, and S. McAdams. "The Timbre Toolbox: Extracting Audio Descriptors from Signals." *Journal of the Acoustical Society of America* 130, no. 5 (2011): 2902-16.

Peeva, D., B. Baird, O. Izmirli, and D. Blevins. "Haptic and Sound Correlations: Pitch, Loudness and Texture." Paper presented at the Eighth International Conference on Information Visualization, London, 2004.

Peretz, Isabelle, L. Gagnon, and B. Bouchard. "Music and Emotion: Perceptual Determinants, Immediacy, and Isolation after Brain Damage." *Cognition* 68 (1998): 111-41.

Peretz, Isabelle, and Robert J. Zatorre, eds. *The Cognitive Neuroscience of Music*. Oxford and New York: Oxford University Press, 2003.

Phillipov, Michelle. *Death Metal and Music Criticism: Analysis at the Limits*. Plymouth, UK: Lexington, 2012.

Phillips, John L. *How to Think About Statistics*. 6 ed. New York: Henry Holt, 1999.

Pitt, Mark A., and Robert G. Crowder. "The Role of Spectral and Dynamic Cues in Imagery for Musical Timbre." *Journal of Experimental Psychology: Human Perception and Performance* 18, no. 3 (1992): 728-38.

Platel, Hervé, C. Price, J.-C. Baron, R. Wise, J. Lambert, R.S.J. Frackowiak, B. Lechevalier, and F. Eustache. "The Structural Components of Music Perception: A Functional Anatomical Study." *Brain* 120 (1997): 229-43.

Plomp, Reinier. *Aspects of Tone Sensation: A Psychophysical Study*. London: Academic Press, 1976.

Plomp, Reinier, and W.J.M Levelt. "Tonal Consonance and Critical Bandwidth." *Journal of the Acoustical Society of America* 38 (1965): 548-60.

Popescu, M., A. Otsuka, and A.A. Ioannides. "Dynamics of brain activity in motor and frontal cortical areas during music listening: a magnetoencephalographic study." *Neuroimage* 21, no. 4 (2004): 1622-38.

Porter, Lewis. *John Coltrane: His Life and Music*. Ann Arbor: University of Michigan Press, 2000.

Pressnitzer, D., S. McAdams, S. Winsberg, and J. Fineberg. "Perception of Musical Tension for Non-Tonal Orchestral Timbres and its Relation to Psychoacoustic Roughness." *Perception and Psychophysics* 62 (2000): 66-80.

Pulvermüller, Friedemann. “Brain Mechanisms Linking Language and Action.” *Nature Neuroscience* 6 (2005): 576-82.

Purcell, Natalie. *Death Metal Music: The Passion and Politics of a Subculture*. Jefferson, NC: McFarland, 2003.

Radano, Ronald. *Lying up a Nation: Race and Black Music*. Chicago: University of Chicago Press, 2003.

Ramachandran, V.S., and Edward Hubbard. “Synaesthesia: A Window into Perception, Thought and Language.” *Journal of Consciousness Studies* 8 (2001): 3-34.

Ratliff, Ben. *Coltrane: The Story of a Sound*. New York: Farrar, Straus and Giroux, 2007.

Rentfrow, Peter J., and Samuel D. Gosling. “The Content and Validity of Music-Genre Stereotypes among College Students.” *Psychology of Music* 35, no. 2 (2007): 306-26.

———. “The Do Re Mi’s of Everyday Life: The Structure and Personality Correlates of Music Preferences.” *Journal of Personality and Social Psychology* 84, no. 6 (2003): 1236-56.

Rentfrow, Peter J., Jennifer A. McDonald, and Julian A. Oldmeadow. “You Are What You Listen To: Young People’s Stereotypes about Music Fans.” *Group Processes & Intergroup Relations* 12, no. 3 (2009): 329-44.

Repp, Bruno H. “Music as Motion: A Synopsis of Alexander Truslit's (1938) *Gestaltung und Bewegung in der Musik*.” *Haskins Laboratories Status Report on Speech Research* 111-112 (1992): 262-78.

Repp, Bruno H., and Günther Knoblich. “Action Can Affect Auditory Perception.” *Psychological Science* 18, no. 6 (2007): 6-7.

Ribot, Marc. “Earplugs.” In *Arcana: Musicians on Music*, edited by John Zorn. New York: Granary Books, 2000.

Ricoeur, Paul. *Hermeneutics and the Human Sciences: Essays on Language, Action and Interpretation*. Translated by John B. Thompson. Cambridge: Cambridge University Press, 1981.

Rilke, Rainer Maria. *The Duino Elegies*. Translated by Leslie Norris and Alan Keele. Rochester, NY: Camden House, 1923/1993.

Rioux, Vincent, and Daniel Västfjäll. “Analyses of Verbal Descriptions of the Sound Quality of a Flue Organ Pipe.” *Musicae Scientiae* 6, no. 1 (2002): 55-79.

Rizzolatti, Giacomo, and Michael A. Arbib. “Language within Our Grasp.” *Trends in Neuroscience* 21 (1998): 188-94.

Rizzolatti, Giacomo, and Laila Craighero. “The Mirror-Neuron System.” *Annual Review of Neuroscience* 27 (2004): 169-92.

Rizzolatti, Giacomo, L. Fadiga, V. Gallese, and L. Fogassi. "Premotor Cortex and the Recognition of Motor Actions." *Cognitive Brain Research* 3 (1996): 131-41.

Rizzolatti, Giacomo, and Corrado Sinigaglia. *Mirrors in the Brain: How Our Minds Share Actions and Emotions*. Translated by Frances Anderson. Oxford: Oxford University Press, 2008.

Robson, Colin. *Experiment, Design and Statistics in Psychology*. Harmondsworth, U.K.: Penguin Education, 1973.

Rosenbaum, Ron. "Bob Dylan Interview." *Playboy*, March, 1978.

———. "'That Wild Mercury Sound': Tangled Up in Bob Again." *New York Observer*, 28 May, 2001.

Rousseau, Jean-Jacques. *Essay on the Origin of Languages and Writings Related to Music*. Translated by John T. Scott. Hanover, NH: University Press of New England, 1781/1998.

Samson, Severine, and Robert J. Zatorre. "Contribution of the Right Temporal Lobe to Musical Timbre Discrimination." *Neuropsychologia* 32, no. 2 (1994): 231-40.

Samson, Severine, Robert J. Zatorre, and James O. Ramsay. "Multidimensional Scaling of Synthetic Musical Timbre: Perception of Spectral and Temporal Characteristics." *Canadian Journal of Experimental Psychology* 51, no. 4 (1997): 307-15.

Sandell, Gregory J. "Macrotimbre: Contribution of Attack and Steady State." *Journal of the Acoustical Society of America* 103, no. 5 (1998).

Saslaw, Janna. "Forces, Containers, and Paths: The Role of Body-Derived Image Schemas in the Conceptualization of Music." *Journal of Music Theory* 40, no. 2 (1996): 217-43.

Schellenberg, E. Glenn, Paul Iverson, and Margaret C. McKinnon. "Name that Tune: Identifying Popular Recordings from Brief Excerpts." *Psychonomic Bulletin & Review* 6, no. 4 (1999): 641-46.

Scherer, Klaus R. "Appraisal Considered as a Process of Multilevel Sequential Checking." In *Appraisal Processes in Emotion: Theory, Methods, Research*, edited by Klaus R. Scherer, Angela Schorr and Tom Johnstone. New York: Oxford University Press, 2001.

———. "Expression of Emotion in Voice and Music." *Journal of Voice* 9, no. 3 (1995): 235-48.

———. "Vocal Affect Expression: A Review and a Model for Future Research." *Psychological Bulletin* 99, no. 2 (1986): 143-65.

Scherer, Klaus R., and James S. Oshinsky. "Cue Utilization in Emotion Attribution from Auditory Stimuli." *Motivation and Emotion* 1, no. 4 (1977): 331-46.

Scherer, Klaus R., Angela Schorr, and Tom Johnstone, eds. *Appraisal Process in Emotion: Theory, Methods, Research*. New York: Oxford University Press, 2001.

Scherer, Klaus R., and Marcel R. Zentner. "Emotional Effects of Music: Production Rules." In *Music and Emotion: Theory and Research*, edited by Patrik N. Juslin and John A. Sloboda. Oxford and New York: Oxford University Press, 2001.

Schoenberg, Arnold. *Theory of Harmony*. Translated by R.E. Carter. London: Faber, 1911/1978.

Schubert, Emory, and Joe Wolfe. "Does Timbral Brightness Scale with Frequency and Spectral Centroid?". *Acta Acustica* 92 (2006): 820-25.

Schürmann, M., G. Caetano, Y. Hlushchuk, V. Jousmäki, and R. Hari. "Touch Activates Human Auditory Cortex." *Neuroimage* 30, no. 4 (2006): 1325-31.

Scott, S.K., A.W. Young, A.J. Calder, D.J. Hellawell, J.P. Aggleton, and M. Johnson. "Impaired Auditory Recognition of Fear and Anger Following Bilateral Amygdala Lesions." *Nature* 385 (1997): 254-57.

Scruton, Roger. *The Aesthetics of Music*. New York: Oxford University Press, 1997.

Seashore, Carl E. *Psychology of Music*. New York: Dover, 1938/1967.

Segnini Sequera, Rodrigo. "On the Vocalization of Non-Speech Sounds: Implicit Mechanisms and Musical Applications." Ph.D. Diss., Stanford University, 2006.

Sethares, William. *Tuning, Timbre, Spectrum, Scale*. London: Springer, 1998.

Shapiro, Lawrence. *Embodied Cognition*. London and New York: Routledge, 2010.

Shepherd, John, and Peter Wicke. *Music and Cultural Theory*. Cambridge: Polity, 1997.

Shove, Patrick, and Bruno H. Repp. "Musical Motion and Performance: Theoretical and Empirical Perspectives." In *The Practice of Performance: Studies in Musical Interpretation*, edited by John Rink. Cambridge: Cambridge University Press, 1995.

Simpkins, Cuthbert. *Coltrane: A Biography*. New York: Herndon House, 1975.

Slawson, Wayne. *Sound Color*. Berkeley and Los Angeles: University of California Press, 1985.

Slingerland, Edward. *What Science Offers the Humanities: Integrating Body and Culture*. Cambridge: Cambridge University Press, 2008.

Small, Christopher. *Musicking: The Meanings of Performing and Listening*. Middletown, CN: Wesleyan University Press, 1998.

Smith, J.D., D. Reisberg, and M. Wilson. "Subvocalization and Auditory Imagery: Interactions between Inner Ear and Inner Voice." In *Auditory Imagery*, edited by D. Reisberg. Hillsdale, NJ: Lawrence Erlbaum, 1992.

Smith, J.D., M. Wilson, and D. Reisberg. "The Role of Subvocalization in Auditory Imagery." *Neuropsychologia* 33, no. 11 (1995): 1433-54.

Smith, Julie Dawn. "Perverse Hysterics: The Noisy Cri of Les Diaboliques." In *Big Ears: Listening for Gender in Jazz Studies*, edited by Nichole T. Rustin and Sherrie Tucker. Durham, NC: Duke University Press, 2008.

Spellman, A. B. Liner notes to John Coltrane's *Ascension*. Impulse Records AS-95, 2000, CD. Originally released in 1965.

Spencer, Herbert. "The Origin and Function of Music." In *Literary Style and Music*. New York: Philosophical Library, 1857/1951.

Stelzner, Chuck. "Death Metal/Throat Vocal Analysis." Unpublished paper.

Such, David. *Avant-garde Jazz Musicians: Performing "Out There"* Iowa City: University of Iowa Press, 1993.

Sundberg, Johan. *The Science of the Singing Voice*. DeKalb, IL: Northern Illinois University Press, 1987.

Sylvan, Robin. *Tracing the Spirit: The Religious Dimensions of Popular Music*. New York: NYU Press, 2002.

Tagg, Philip. "Towards a Definition of 'Music'." Institute of Popular Music: University of Liverpool, 2002.

Tervaniemi, Mari, I. Winkler, and R. Näätänen. "Pre-Attentive Categorization of Sounds by Timbre as Revealed by Event-Related Potentials." *Neuroreport* 8 (1997): 2571-74.

Théberge, Paul. *Any Sound You Can Imagine: Making Music/Consuming Technology*. Middletown, CN: Wesleyan University Press, 1997.

Thelen, Esther, G. Schöner, C. Scheier, and L. Smith. "The Dynamics of Embodiment: A Field Theory of Infant Perseverative Reaching." *Behavioral and Brain Sciences* 24 (2001): 1-86.

Thomas, J. C. *Chasin' the Trane*. New York: Da Capo, 1975.

Todd, Neil P. McAngus, and Frederick W. Cody. "Vestibular Responses to Loud Dance Music: A Physiological Basis of the 'Rock and Roll Threshold'?". *Journal of the Acoustical Society of America* 107, no. 1 (2000): 496-500.

Toiviainen, Petri, M. Tervaniemi, J. Louhivuori, M. Saher, M. Huotilainen, and R. Näätänen. "Timbre Similarity: Convergence of Neural, Behavioral, and Computational Approaches." *Music Perception* 16, no. 2 (1998): 223-41.

Tomlinson, Gary. "Evolutionary Studies in the Humanities: The Case of Music." *Critical Inquiry* 39, no. 4 (2013): 647-75.

———. *The Singing of the New World: Indigenous Voice in the Era of European Contact*. Cambridge: Cambridge University Press, 2009.

Tsai, Chen-Gia, L-C Wang, S-F Wang, Y-W Shau, and T-Y Hsiao. "Aggressiveness of the Growl-Like Timbre: Acoustic Characteristics, Musical Implications, and Biomechanical Mechanisms." *Music Perception* 27, no. 3 (2010): 209-21.

Turino, Thomas. "Signs of Imagination, Identity, and Experience: A Peircian Semiotic Theory for Music." *Ethnomusicology* 42, no. 2 (1999): 221-55.

Turner, Victor. *The Ritual Process: Structure and Anti-Structure*. London: Routledge, 1969.

Tynan, John. "Take Five." *Down Beat,* November 23, 1961, 40.

Tzanetakis, George, and Perry R. Cook. "Musical Genre Classification of Audio Signals." *IEEE Transactions on Speech and Audio Processing* 10, no. 5 (2002): 293-302.

Umiltà, M.A., E. Kohler, V. Gallese, L. Fogassi, L. Fadiga, C. Keysers, and G. Rizzolatti. "I Know What You are Doing: A Neurophysiological Study." *Neuron* 31 (2001): 155-65.

Vaneechoutte, M., and J.R. Skoyles. "The Memetic Origin of Language: Modern Humans as Musical Primates." *Journal of Memetics—Evolutionary Models of Information Transmission* 2 (1998).

Varela, Francisco J., Evan Thompson, and Eleanor Rosch. *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge, MA: MIT Press, 1991.

Vassilakis, Pantelis N. "Auditory Roughness as a Means of Musical Expression." In *Selected Reports in Ethnomusicology: Perspectives in Systematic Musicology* edited by Roger A. Kendall and Roger W.H. Savage. Los Angeles: UCLA Ethnomusicology, 2005.

———. "Perceptual and Physical Properties of Amplitude Fluctuation and their Musical Significance." Ph.D. Diss., UCLA, 2001.

Vassilakis, Pantelis N., and Roger A. Kendall. "Psychoacoustics and Cognitive Aspects of Auditory Roughness: Definitions, Models, and Applications." *Proceedings of Human Vision and Electronic Imaging* 15, no. 7527 (2010).

von Bismarck, G. "Timbre of Steady Sounds: A Factorial Investigation of Its Verbal Attributes." *Acustica* 30 (1974): 147-72.

Waksman, Steve. *Instruments of Desire: The Electric Guitar and the Shaping of Musical Experience*. Cambridge, MA: Harvard University Press, 1999.

Wallmark, Zachary. "'I Would Like to Die': Constructions of the Eroticized Body in the Monteverdi Madrigal." Paper presented at the Material and Imagined Bodies Conference, Brown University, Providence, RI, 2010.

Wallmark, Zachary, and Marco Iacoboni. "Embodied Listening and Musical Empathy: Perspectives from Mirror Neuron Research." Paper presented at the AMS/SEM/SMT Joint Conference, New Orleans, LA, 2012.

Walser, Robert. "The Body in the Music." *College Music Symposium* 31 (1991): 117-25.

———. *Running with the Devil: Power, Gender, and Madness in Heavy Metal Music*. Hanover, NH: University Press of New England, 1993.

Ward, Jamie, Brett Huckstep, and Elias Tsakanikos. "Sound-Color Synaesthesia: To What Extent Does it Use Cross-Model Mechanisms Common to Us All?". *Cortex* 42, no. 2 (2006): 264-80.

Warren, Jane E., D.A. Sauter, F. Eisner, J. Wiland, M.A. Dresner, R.J.S. Wise, S. Rosen, and S.K. Scott. "Positive Emotions Preferentially Engage an Auditory-Motor 'Mirror' System." *The Journal of Neuroscience* 26, no. 50 (2006): 13067-75.

Washington, Michael Spence. "Beautiful Nightmare: Coltrane, Jazz, and American Culture." Ph.D. Diss., Harvard University, 2001.

Weinstein, Deena. *Heavy Metal: A Cultural Sociology*. San Francisco: Jossey-Bass, 1991.

Welding, Pete. "Double View of Coltrane 'Live'." *Down Beat*, April 26, 1962, 29.

Wessel, David L. "Timbre Space as a Musical Control Structure." *Computer Music Journal* 3, no. 2 (1979): 45-52.

Wessinger, C.M., M.H. Buonocore, C.L. Kussmaul, and G.R. Mangun. "Tonotopy in Human Auditory Cortex Examined with Functional Magnetic Resonance Imaging." *Human Brain Mapping* 5 (1997): 18-25.

Wicker, Bruno, C. Keysers, J. Plailly, J.-P. Royet, V. Gallese, and G. Rizzolatti. "Both of Us Disgusted in My Insula: The Common Neural Basis of Seeing and Feeling Disgust." *Neuron* 40 (2003): 655-64.

Wilson, Olly. "The Heterogeneous Sound Ideal in African-American Music." In *Signifyin(g), Sanctifyin', and Slam Dunking: A Reader in African-American Expressive Culture*, edited by Gena Dagel Caponi. Amherst: University of Massachusetts Press, 1999.

Wilson, Stephen M., and Marco Iacoboni. "Neural Responses to Non-Native Phonemes Varying in Producibility: Evidence for the Sensorimotor Nature of Speech Perception." *Neuroimage* 33 (2006): 316-25.

Wilson, Stephen M., A.P. Saygin, M.I. Serreno, and M. Iacoboni. "Listening to Speech Activates Motor Areas Involved in Speech Production." *Nature Neuroscience* 7, no. 7 (2004): 701-02.

Windsor, W. Luke. "Data Collection, Experimental Design, and Statistics in Music Research." In *Empirical Musicology: Aims, Methods, Prospects*, edited by Eric F. Clarke and Nicholas Cook. Oxford: Oxford University Press, 2004.

Woideck, Carl, ed. *The John Coltrane Companion*. New York: Schirmer, 1998.

Zacharakis, Asterios, Konstantinos Pastiadis, and Joshua D. Reiss. "An Interlanguage Study of Musical Timbre Semantic Dimensions and Their Acoustic Correlates." *Music Perception* 31, no. 4 (2014): 339-58.

Zagorski-Thomas, Simon. "The Spectromorphology of Recorded Popular Music: The Shaping of Sonic Cartoons through Record Production." In *The Relentless Pursuit of Tone: Timbre in Popular Music*, edited by Robert Fink, Mindy LaTour O'Brien and Zachary Wallmark. New York: Oxford University Press, forthcoming.

Zatorre, Robert J., Joyce L. Chen, and Virginia B. Penhune. "When the Brain Plays Music: Auditory-Motor Interactions in Music Perception and Production." *Nature Reviews: Neuroscience* 8 (2007): 547-58.

Zbikowski, Lawrence M. *Conceptualizing Music: Cognitive Structure, Theory, and Analysis*. New York: Oxford University Press, 2002.

———. "Music, Language, and Multimodal Metaphor." In *Multimodal Metaphor*, edited by Charles Forceville and Eduardo Urios-Aparisi. Berlin: Mouton de Gruyter, 2009.

———. "Musicology, Cognitive Science, and Metaphor: Reflections on Michael Spitzer's *Metaphor and Musical Thought*." *Musica Humana* 1, no. 1 (2009): 81-104.