

Session 9

Aesthetics and Political Philosophy

The Violent Particularity of Timbre as Sonic and Symbolist Ideal in New Music

Amy Bauer^{1†}

¹ Music Department, Claire Trevor School of the Arts, University of California, Irvine, Irvine, CA, USA

[†] Corresponding author: abauer@uci.edu

In his early writings, Gérard Grisey praises the “irreplaceable uniqueness” of each sound as a complex individual whose internal structure functions on multiple temporal levels. His description of way in which sound materializes time recalls Hegel's description of the way in which sound, in its “total form,” escapes the limits of spatiality. Grisey describes timbre's appeal for the composer as its “violently qualitative aspect,” those inherent qualities that distinguish each sound from its brethren, and leave a “violent impression in our memories.”[259] Similarly, for Hegel sound does not overcome the repose of the spatial object merely through negation, but—in the words of Peter Hanly—causes it to tremble. The inwardness of sound acquires a sensuous existence, and hence a kind of freedom, through its violent subjugation of matter. The individuation of vibration preserves in its material qualities an ideal; the vibrating sound becomes an independent subject, with soul-like manifestation. Similarly, it is sound as “an ensemble of dynamic characteristics in constant interaction” which Hughes Dufourt celebrates as the source and substance of spectral composition. Dufourt prioritizes in spectralism a music whose form arises directly from its material, opposing it to serial music, which—in an act of vicious subjugation—effaces the qualitative aspect of sound. This view is echoed by Tristan Murail, for whom the development of sounds as music-perceptual “atoms” follows in the tradition of Pierre Schaeffer's “sonic object”: sounds as a complex world into which we can enter and within which we can explore. Such “micro forms” transcend their function as timbre, and take on a lifelike appearance when combined and related in a manner that considers the aesthetic and psychological impact of the result. This paper will explore the philosophical ramifications of this “violent particularity” within two works that take complex, small-scale timbral objects from nature as models for sweeping, large-scale sonic designs.

Windsequenzen (1975/2002) by Peter Eötvös generates literal wind sequences evoked by the timbre of wind instruments played with closed keys and extreme positions, with percussion that evokes placid air currents, whirlwinds, and sea breezes. The greatest force of natural sound may paradoxically lie in the two movements that bookend the work titled *Windstille*. In *Windstille I* the flute is stopped with a cork while the keys are closed and the flute is overblown, to allow only whisper-soft, odd-numbered overtones played against overtones coaxed from a bass drum via fingernail (Figure 1). Murail's *Liber fulguralis* (2008) for electronics, piano, and chamber ensemble (with accompanying video by Hervé Bailly-Basin) by contrast employs recordings of actual thunder, electronic processing of that thunder, and instrumental forces to portray the dynamic profile of an elongated series of thunder peals lasting slightly over 28 minutes. The relative loudness of the first 15 seconds of thunder establishes a dynamic profile that will recur throughout the work, in a kind of fractal replication, while the opening chords reflect a spectral analysis of that same peal (Figure 2).

If Eötvös takes a playful approach to the timbral representation of nature, Murail's journey from natural sound to composition, through transformation, spectral interpolation, synthesis and manipulation, represents a kind of aesthetics of continuity in reverse. Murail calls this anamorphosis, which also characterizes our perception of monumental natural events; this approach has the ancillary effect of depersonalizing the work, of absenting the composer from the sounding music. We might characterize the sound images of *Windsequenzen* and *Liber fulguralis* as “violently particular” in their evocation of specific tones and intensities seldom heard in the concert hall. At the same time, they take the form of poetic metaphors which reveal symbolist roots: detached from an initial, natural model through translation, distortion or, as Murail avers, “treasons.” Such an approach invites historical associations to

earlier traditions of Japanese and French music that parsed the liminal thresholds between timbre and form, while encouraging philosophical reflections on the materiality of timbre as a sonic ideal.

Péter EÖTVÖS

Fl. in C
normal
gestimmt

sehr ruhig

46

[Grundton]

molto espr.

Fingernagel

höchster Oberton
1 cm

dunkel
10 cm

hell
1 cm

Gr. Tr.

großer
Filzschlegel

auflegen
zum
dämpfen
ohne
Schlag

ppp

(1 cm = vom Rand
entfernt)

sehr gleichmäßig, ohne Betonung,
das Obertonglissando ist wichtig

(10 cm = vom Rand
entfernt)

1 cm

20 cm

1 cm

ppp

ppp

klingen lassen

Figure 1: The first four mm from *Windstille I*, *Windsequenzen*, Peter Ętövs.

Diagram illustrating the G#0 Fundamental and C#1 Fundamental. The notation shows two staves (treble and bass clef) with various notes and accidentals. The G#0 Fundamental is shown on the left, and the C#1 Fundamental is shown on the right. The notes are labeled with numbers 30, 21, 13, 11, 9, and 8+.

Figure 2: The first four seven-note sustained chords in the instrumental ensemble as they reflect partials of a G#0 and C#1 fundamental, *Liber fulguralis*, Tristan Murail

References

- Bodhika, J.A.P. (2019). A brief study on thunder clap. *Applied Acoustics* 145. <https://doi.org/10.1016/j.apacoust.2018.09.018>
- Cagney, Liam. (2025). Tristan Murail's Musical Poetics. In Amy Bauer, Liam Cagney and Will Mason (Eds.), *The Oxford Handbook of Spectral Music* (pp. 269–309). London, UK: Oxford.
- Daydé, Emmanuel. Tristan Murail, Un Héros Trop Discret. *Artpress*. <https://www.artpress.com/2022/02/04/tristan-murail-heros-trop-discret/>
- Dufourt, Hughes. (2010). The Principles of Music and the Rationalization of Theory. In Irène Deliège and Max Paddison (Eds.), *Contemporary Music: Theoretical and Philosophical Perspectives* (19–50). London, UK: Routledge.
- Dufourt, Hughes. (2014). *La musique spectral. Une révolution épistémologique*. Paris: Éditions Delatour.
- Eötvös, Peter. *Windsequenzen* (1978/2002). Budapest, Hungary: UMP Editio Musica.

- Farkas, Zoltán. (2003). Program notes for Péter Eötvös: Intervalles-Interieurs,— Windsequenzen, UMZE Chamber Ensemble, Michael Svoboda, Klangforum Wien BMC records 092.
- Grisey, Gérard. (1978). Zur Entstehung des Klangs. Darmstädter Beiträge zur Neuen Musik, XVII, 73-79.
- Grisey, Gérard. (1987). Tempus ex machina. A composer's reflection on musical time. *Contemporary Music Review* 2:1, 239–275.
- Hanley, Peter. (2009). Hegel's Voice; Vibration and Violence. *Research in Phenomenology* 3, 359–73.
- Haselböck, Lukas. Zur »Klangfarbenlogik« bei Schönberg, Grisey und Murail. In Christian Utz (Ed.), *Klang und Wahrnehmung in der Musik des 20. und 21. Jahrhunderts, musiktheorien der gegenwart* 6 (pp. 137–162). Saarbrücken, Germany: Pfau.
- Hegel, G.F. (1971). *Philosophy of Nature: Encyclopaedia of the Philosophical Sciences II*. Trans. A.V. Miller (Oxford: Clarendon Press).
- Hirs, Rozalie und Bob Gilmore (Eds.), (2009). *Contemporary Compositional Techniques and OpenMusic*. Paris: Delatour.
- Hummel, Thomas (2005). ...um mir das Vergehen der Zeit zu versüßen. *MusikTexte* 107, 1–7.
- Lacroix, A., Farges, T., Marchiano, R., & Coulouvrat, F. (2018). Acoustical measurement of natural lightning flashes: Reconstructions and statistical analysis of energy spectra. *Journal of Geophysical Research: Atmospheres*, 123, 12,040–12,065. <https://doi.org/10.1029/2018JD028814>.
- Ledoux, Claude. (2000). From the philosophical to the Practical: An imaginary proposition concerning the music of Tristan Murail. *Contemporary Music Review*, 19:3, 41–65.
- Michel, Pierre. (2002). Entretien avec Tristan Murail. In Peter Szendy (Ed.), *Tristan Murail* (46–47). Paris: L'Harmattan.
- Murail, Tristan. (2005). The Revolution of Complex Sounds. *Contemporary Music Review* 24:2-3, 121–135.
- Murail, Tristan. (2005). Villeneuve-lès-Avignon Conferences, Centre Acanthes, 9–11 and 13 July 1992. *Contemporary Music Review* 24:2-3, 187–267.
- Murail, Tristan. (2008). Program notes for *Liber fulguralis*. Available at <https://www.henry-lemoine.com/fr/partitions-pour-ensemble/5455-liber-fulguralis.html>.
- Rigaudière, Pierre. (2022). Caught up in emotions. *Diapason*, 712, 12–13.
- Shambaugh, Christopher. (2023). Hegel's Philosophy of Sound." *Hegel Bulletin* 46/1: 1–24.

“Almost Unbelievably Rich”: Autistic Listening and Wandelweiser Music

Nathan Friedman^{1†}

¹ Department of Music, University of Chicago, Chicago, Illinois, USA

[†] Corresponding author: nfriedman@uchicago.edu

Since its origins in the early 1990s, the music of the international composers’ collective Wandelweiser has been known for its explorations of very quiet sounds over long durations, frequently subsumed by (usually lengthy) silences. Often extremely fragile and preoccupied with the borderlands of sound and silence (Epstein 2017; Margulis 2007; Melia 2011; Thurley 2015), Wandelweiser music produces contradictory effects. The seemingly tranquil sonic environment of this music is often undercut with enormous tension on the part of the performers and audience (Cooper 2021), and the apparent simplicity of its materials are revealed to be, in the words of Michael Pisaro-Liu, “almost unbelievably rich” in their timbral complexity (Barrett 2011; Pisaro 2009). In cultivating a listening experience marked by privation, with texture and volume reduced to a near-minimum, Wandelweiser music calls into question the specificity of the fundamental materials of Western music, effectively placing it under an aural microscope. This effect is particularly marked regarding timbre: the demands of playing so quietly for such extended durations strain the performers, rendering inevitable and then amplifying what would usually be considered sonic imperfections.

Drawing from the personal accounts of autistic writers such as Donna Williams (1998), philosopher Erin Manning (2013, 2016) has developed the concept of “autistic perception”, which she defines as “the opening, in perception, to the uncategorized, to the unclassified,” and that “rather than seeing the parts abstracted from the whole, autistic perception is alive with tendings that create ecologies before they coalesce into form” (2016, 14). There is substantial evidence that autistic people frequently do not hear in nearly so gestalt-oriented a way as neurotypicals, so common models of listening may apply much more weakly or not at all to their experience (Straus 2011). I argue that Wandelweiser music cultivates a comparable mode of perception that I call *autistic listening*, one that makes autistic perception palpable also for neurotypicals, albeit with more effort. With regards to timbre, the experience of listening to this music problematizes the individuation of instruments, allowing for an “ecological” mode of perception that floats between timbral wholes and parts, between attentive and preattentive states of listening (Blake 2019; Fales 2002). Using Robert Chapman’s (2023) Marxist model of neurodiversity, I connect autistic perception to the accelerating perceptual excesses that even neurotypicals experience in modernity. In calling attention to the “ecological” conditions of the concert experience and questioning of the fundamental “givens” of music in a relatively gentle way, Wandelweiser music provides a model for navigating the chaos of existing in the modern world, through which we may be able to learn from neurodiverse perspectives.

References

- Barrett, G. D. (2011). The Silent Network—The Music of Wandelweiser. *Contemporary Music Review*, 30(6), 449–470. <https://doi.org/10.1080/07494467.2011.676895>
- Blake, D. (2019). Timbre. In A. Rehding & S. Rings (Eds.), *The Oxford Handbook of Critical Concepts in Music Theory* (pp. 136–159). New York: Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780190454746.013.21>
- Chapman, R. (2023). *Empire of Normality: Neurodiversity and Capitalism*. London: Pluto Press.
- Cooper, E. (2021). The Strong Silent Type: Masculinity and Wandelweiser Music. *Tempo*, 75(296), 21–33. <https://doi.org/10.1017/S0040298220000923>

- Epstein, N. (2017). Musical Fragility: A Phenomenological Examination. *Tempo*, 71(281), 39–52. <https://doi.org/10.1017/S0040298217000432>
- Fales, C. (2002). The Paradox of Timbre. *Ethnomusicology*, 46(1), 56–95. <https://doi.org/10.2307/852808>
- Manning, E. (2013). *Always More Than One: Individuation's Dance*. Durham, NC: Duke University Press. <https://doi.org/10.1215/9780822395829>
- Manning, E. (2016). *The Minor Gesture*. Durham, NC: Duke University Press. <https://doi.org/10.1515/9780822374411>
- Margulis, E. H. (2007). Moved by Nothing: Listening to Musical Silence. *Journal of Music Theory*, 51(2), 245–276. <https://doi.org/10.1215/00222909-2009-003>
- Melia, N. (2011). Stille Musik—Wandelweiser and the Voices of Ontological Silence. *Contemporary Music Review*, 30(6), 471–495. <https://doi.org/10.1080/07494467.2011.676897>
- Pisaro, M. (2009, September 23). Wandelweiser. *Erstwords*. <http://erstwords.blogspot.com/2009/09/wandelweiser.html>
- Straus, J. N. (2011). *Extraordinary Measures: Disability in Music*. New York: Oxford University Press.
- Thurley, O. (2015). Disappearing Sounds: Fragility in the Music of Jakob Ullmann. *Tempo*, 69(274), 5–21. <https://doi.org/10.1017/S0040298215000339>
- Williams, D. (1998). *Autism and Sensing: The Unlost Instinct*. London: Jessica Kingsley Publishers.

Timbre's Beholders and the Racialization of Sensation in Concert Music Aesthetics

Colin Tucker^{1†}

¹ Colin Tucker Studio, Tsí Tkaròn:to/Toronto, Dish with One Spoon Treaty Territory

[†] Corresponding author: ColinLTucker@gmail.com

While much music scholarship has attended to the racialization of particular timbral characteristics (e.g. Chung 2024, Eidsheim 2019, Marshall 2015, Pasler 2006), scholars have yet to mark racializing references in canonical concert music texts on timbre, and to unpack the racialized positioning of the subject positions presupposed by these texts. As these lacunae are likely due in part to white epistemologies which situate the racializing discourse of canonical texts as an aberration, I proceed otherwise, by reading these texts through critical ethnic studies scholar Denise Ferreira da Silva's (2007) account of the racializing constitution of modern subjectivity.

This paper is from a book-in-progress of scholarly and creative interventions that mark the imperial whiteness of the concert hall's foundational Work-concept (Goehr 1992) through a focus on the Work's spectatorial beholder (Tucker 2026). This critical mapping proceeds through analysis of threshold (concert music) aesthetics and music theory texts' articulation of the spectator. I draw on da Silva's canonical analysis of how the modern white(ned) subject's interiority is, paradoxically, a relational proposition, and one predicated on a displacement of exteriority onto a racially-marked subject.

Decolonial scholar David Lloyd (2018) has read the *Critique of Judgment* by Immanuel Kant [1790] as a watershed instantiation of these relations in the aesthetic realm, noting how this text secures the racially-unmarked aesthetic subject's interiority by displacing the "pathological" realm of exteriority onto the racially-marked figure of the Savage, a move which I connect to later concert music aesthetics, particularly the threshold music philosophy of Eduard Hanslick (Rothfarb and Landerer [1854]). To paraphrase Silva, these texts' project of self-determined spectatorship might be understood as asserting that the aesthetic subject has sensory access to aesthetic practice, but is not determined by the latter's "pathological" exteriority, an assertion that the texts "prove" by displacing this exteriority onto racially-marked figures. I trace how the Kantian spectator achieves transparent sensory access to art through the "universally communicable" abstractions of Idea and Form, and how 19th century concert music aesthetics do so through a variety of paradigms predicated on the abstraction of pitch. Yet for music scholar Isabella van Elferen, timbre is "one of the great problems in musical epistemology" (2021, 72) due to its resistance to the taxonomizing typical of concert music discourse about pitch. In what follows, I trace two canonical "solutions" to this problem, and show how both are undergirded by a racializing displacement of exteriority similar to those characteristic of Kantian and Hanslickian aesthetics more broadly.

First, I trace across the long nineteenth century an Idealist "solution" to the timbre "problem," which displaces timbre's exteriority onto racially-marked figures, enabling the spectator to abstract vibration into pitch and thereby bracketing timbre. This move is codified in Kant's *Critique*, which equates "mere colour, e.g. the green of a grass plot" with "a mere tone... like that of a violin," and reduces both to the "barbaric" domain of "charm," which is later displaced onto the figure of the Savage, a figure made to embody the beginning of an imagined civilizing process ([1790], 43-44, 104). Thus the *Critique* secures the aesthetic subject's access to the interiority of Idea and Form through a racializing ontological cut that differentially aporions sensorial discipline (Schuller 2018) and temporality: the aesthetic subject is enabled to abstract vibration into pitch-Idea-Form *because* the Savage is reduced to the incapacity for the same, while the aesthetic subject is enabled to "refine" itself *because* the Savage is reduced to the distant past of civilization. Hanslick likewise defines timbre as pathological, which he also calls elemental,

contrasting “the enthusiasts [who] let themselves be carried away and rocked by the vibrations of tones” with the proper spectators who “contemplate them keenly” (Rothfarb and Landerer [1854], 82). As with many later writers on timbre, Hanslick foregrounds the racialization of temporality, displacing the exteriority of the elemental-pathological onto the racially-marked figure of the Primitive: “there is no doubt at all that music had a far more direct effect among ancient peoples than at present, because humanity in its primitive stages of development was more related to and subject to the *elemental* than in later times, when consciousness and self-determination come into their own” (86). This racializing ontology persists into the 20th century, with Arnold Schoenberg (whose personal library contained the *Critique*) articulating influential pitch-centered music theories and composition aesthetics over-against “a crude, naive pleasure in sheer colour... perhaps that of a child who enjoys striking matches, or the rather more primitive pleasure uncultivated peoples, or sections of the populace, derive from explosions and shooting” ([1931], 333). In summary, these writings do not per se dismiss matters of timbre; rather, they actively displace it onto racially-marked positions, and do so in a way that reiterates essentialized racializing apportionments of capacity. Therefore, the scholarly consensus that frames this move as a mere dismissal should be understood as whitewashing a pervasive racializing mechanism and thereby reproducing the active, interested regime of (un)knowing that philosopher and Black study scholar Charles Mills calls white ignorance (2007).

Second, I map how a similar racializing ontology also undergirds an influential Modernist “solution” to the timbre “problem.” This solution foregrounds timbre as such, reconciling this parameter’s exteriority with self-determined spectatorship by doubling down on the earlier Idealist “solution’s” racializing displacement of exteriority. Indeed, Van Elferen connects this Modernist “timbral imperative” to earlier Romantic aesthetics of timbre shaped by the Kantian sublime. The sublime, particularly in its Kantian articulation(s) has been widely analyzed as a racializing discourse (Armstrong 1996, Battersby 2013, Hall 1997), for instance in the *Critique*’s embodiment of the sublime’s ideal beholder in a (implicitly colonizing) warrior who is the “object of the greatest admiration,” “even to the savage” (76). And the Modernist timbral imperative’s underlying rhetoric reiterates racializing moves formally similar to those of Idealist paradigms. Van Elferen locates the Modernist timbral imperative in four key 20th century musicians: Schoenberg, Schaeffer, Stockhausen, and Boulez, all of whom foreground timbre’s exteriority by displacing it onto racially-marked positions. Pierre Boulez and Karlheinz Stockhausen explicitly exoticize their timbral experiments through instrumentation, gesture, texture and written/verbal discourse that both directly references Exoticized peoples and lands (Boulez [1967], Mendelssohn 2023, Tio 2000) and recalls earlier Western art music histories of Exoticism, of which “colorful” timbres and their heterogeneous combination have long been a hallmark (al-Tae 2010, Locke 2009). Somewhat differently, Schoenberg ends his famous *Theory of Harmony* with a Modernist “futuristic fantasy” of foregrounding timbre and organizing this parameter according to “a kind of logic entirely equivalent to that logic which satisfies us in the melody of pitches” ([1911], 421). Yet this passage traffics in civilizationist language oriented towards the “cultivation,” “development,” and “spirit” of a beholder who earlier in the text is articulated over-against “the primitive ear” which “hears the tone as irreducible, [when] physics recognizes it to be complex” (313), a formulation which repeats a well-worn racializing ontology of sensation that can be traced back at least to WH Wackenroder ([1799], 188). In parallel, in the *Treatise on Objects*, a watershed text for later contemporary art music, electronic music, sound art, and sound studies, Pierre Schaeffer grounds his elaboration of timbral nuance in a similar racializing ontology of sensation in which essentialized “civilizations are sealed off from each other,” on distinct temporal-evolutionary pathways ([1966], 27). The author in turn articulates his Modernist timbral project over-against “our primitive man” who is reduced to the “instinctive” (26), essentialized as “unable to develop harmonically,” and who thus “should have turned not only to rhythm but also to evolving and complex sounds” (482). Then, through a historical connection to the (Kantian) sublime, strategies of Exoticism, and/or written invocation of the Primitive, the Modernist timbral imperative updates rather than breaks with Idealist paradigms, thereby reconciling timbre’s perceptual complexity with self-determined spectatorship.

Altogether, I find that both Idealist and Modernist approaches to timbre posit specific and politically-situated beholders: a racially-unmarked, self-determined Spectator and a racially-marked, outer-determined Foil. Both approaches “solve” the former’s timbre “problem” by displacing it onto the latter. If racializing essentialisms underpin both paradigms’ reconciliation of timbre with interiority, this analysis points to the need to mark and disrupt similar racializing dynamics sedimented into both the predominant pitch-oriented paradigms of Western art music listening, performance, analysis, and composition, *and* timbre-oriented paradigms characteristic of contemporary art music, electronic music, sound art, and sound studies.

References

- Al-Tae, N. (2017). *Representations of the Orient in Western Music: Violence and Sensuality*. Routledge.
- Armstrong, M. (1996). "The effects of blackness": gender, race, and the sublime in aesthetic theories of Burke and Kant. *The Journal of Aesthetics and Art Criticism*, 54(3), 213-236.
- Battersby, C. (2013). *The sublime, terror and human difference*. Routledge.
- Boulez, P. (1986 [1967]). *Oriental Music: A Lost Paradise*. In *Orientations* (pp. 421-424). Faber and Faber.
- Chung, A. J. (2024). Early Modern Tuning, Temperament, and the Natural Philosophy of Empire. *Journal of Music Theory*, 68(2), 315-356.
- Da Silva, D. F. (2007). *Toward a global idea of race* (Vol. 27). U. of Minnesota Press.
- Dolan, E. I., & Rehding, A. (Eds.). (2021). *The Oxford handbook of timbre*. Oxford University Press.
- Eidsheim, N. S. (2019). *The race of sound: Listening, timbre, and vocality in African American music*. Duke University Press.
- Goehr, L. (1992). *The imaginary museum of musical works: An essay in the philosophy of music: An essay in the philosophy of music*. Clarendon Press.
- Hall, K. Q. (1997). Sensus communis and violence: A feminist reading of Kant's Critique of judgment. In *Feminist Interpretations of Immanuel Kant* (pp. 257-272). Penn State Univ. Press.
- Kant, I. (1914 [1790]). Translated by J.H. Bernard. *Critique of judgment*. MacMillan.
- Locke, R. (2009). *Musical Exoticism*. Cambridge University Press.
- Lloyd, D. (2018). *Under representation: The racial regime of aesthetics*. Fordham University Press.
- Marshall, M. L. (2015). Voce Bianca: Purity and whiteness in British early music vocality. *Women and Music: A Journal of Gender and Culture*, 19(1), 36-44.
- Mendelssohn, E. (2023). *White Musical Mythologies: Sonic Presence in Modernism*. Stanford University Press.
- Mills, C. (2007). White ignorance. *Race and epistemologies of ignorance*, 247, 26-31.
- Pasler, J. (2006). Theorizing Race in Nineteenth-Century France: Music as Emblem of Identity. *The Musical Quarterly*, 89(4), 459-504.
- Rothfarb, L, and C. Landerer (2018 [1854]). *Eduard Hanslick's On the Musically Beautiful: A New Translation*. Oxford University Press.
- Schaeffer, P. Translated by C. North and J. Dack (2017 [1966]). *Treatise on musical objects: An essay across disciplines*. Univ. of California Press.
- Schoenberg, A. (1978 [1911]). Translated by Roy Carter. *Theory of harmony*. Univ. of California Press.
- Schoenberg, A. Translated by L. Black. (1975 [1931]). *Instrumentation*. In *Style and Idea* (pp. 330-335). Univ. of California Press.
- Schuller, K. (2018). *The biopolitics of feeling: Race, sex, and science in the nineteenth century*. Duke University Press.

- Tio, Christina Ee Ming. "The avant garde and its' others': Orientalism in contemporary art music." PhD diss., University of Southampton, 2000.
- Tucker, C. (2026). The Concert Hall as Imperial Ecology: Protocols for Otherwise Musicking [manuscript in preparation].
- Van Elferen, I. (2021). Timbrality: The Vibrant Aesthetics of Tone Color. In *The Oxford handbook of timbre* (pp. 69-91). Oxford University Press.
- Wackenroder, WH. Translated by M.H. Schubert (1971 [1799]). "Fantasies on Art For Friends of Art." In *Confessions and Fantasies*. Penn State University Press.

Session 10

Timbre in Instrumental Performance

Physical Analysis of Sound Production in the Baroque Traverso: Geometry, Resonance, and Timbre Control

Patricio de la Cuadra^{1,2†}, Benoît Fabre³ and Augustin Ernoul^{3,4}

¹ Department of Electrical Engineering, ² Music Department, Pontificia Universidad Católica de Chile, Santiago, Chile

³ Sorbonne Université, Institut Jean le Rond d'Alembert (LAM), Paris, France; ⁴ Inria Paris Centre, Paris, France

[†] Corresponding author: pcuadra@uc.cl

Abstract

On a baroque transverse flute (*traverso*) and a modern Boehm flute the same written notes sound and feel strikingly different. This is, of course, partly a matter of musical style — but style and physics are not separable here: the geometry of the air column embodies an aesthetic and imposes a way of playing. Using a one-dimensional finite-element model of a corpus of historically-inspired traversos built by the author, we show that a flute's geometry lays down a *landscape* — where the octave stretches, and where the second harmonic aligns with a bore resonance so that a note “rings” — and that the player navigates this landscape with two continuous controls, jet speed and lip coverage. The traverso and the Boehm emerge as two design philosophies: the traverso sculpts a landscape with a strong “home-key” character (rich, flexible, deliberately uneven), the Boehm flattens it for chromatic uniformity. A digitised corpus occupies a structured *identity space*, indicating that timbral identity is designed and reproducible. All quantitative results here are simulated; impedance measurements and controlled recordings are the ongoing validation. We offer the framework as a computable bridge from the maker's bore to the player's experience and the listener's timbre.

1. The Puzzle, Reframed

The traverso held a central place in Western art music of the 17th and 18th centuries, yet compared with modern woodwinds it remains under-studied in quantitative timbre research. A practitioner knows that a traverso and a Boehm flute, built to play the same notes, feel and sound worlds apart — the traverso flexible and strongly coloured by key, the Boehm even and powerful. It is tempting to file this under “period style”, and style is genuinely part of it. But style and physics are not separate: in flute-type instruments the geometry of the air column does not merely fix pitch; it *embodies* an aesthetic and *imposes* a manner of playing (Coltman, 1976; Fletcher & Rossing, 1998; Fabre et al., 2012). The traverso's so-called imperfections — uneven octaves, notes that sing next to notes that are stuffy — are not flaws but a *designed landscape* that gives the instrument its character and its expressive flexibility.

We make this landscape explicit and computable. Each instrument is represented as an axisymmetric duct assembled from digitised bore measurements (embouchure and side toneholes included) and solved with the open-source 1-D finite-element code OpenWind (Tournemenne & Chabassier, 2019). From the simulated input impedance we read two things a player cares about, developed in turn below. All numbers are *simulated* (S); measurement is the next step (§5).

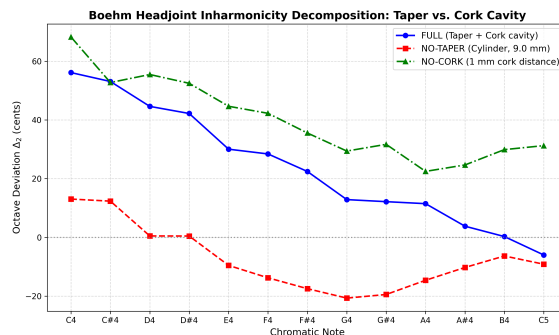


Figure 1: Simulated octave deviation Δ_2 of a tuned headjoint, decomposed. The conical taper sets the offset and the cork cavity sets the slope of the octave stretch across the chromatic register (cone $\approx +43$ c, cork ≈ -37 c, independent and additive). Simulated (OpenWind, 1-D).

2. From Geometry to the Landscape

Does the Octave Fall Where It Should?

A flute over-blows at the octave: the same fingering sounds in two registers. The central tuning compromise is how far the second register departs from a true 2: 1 octave,

$$\Delta_2 = 1200 \log_2 \frac{f_2}{2f_1} \quad [\text{cents}],$$

a temperament-independent number ($\Delta_2 = 0$ is a perfect octave). A bare cylinder bored with side holes is not harmonic: the inertia of the air in the open holes pushes the upper register flat by as much as -140 cents. Historically, makers restored the octaves by departing from a pure cylinder, adding a conical taper. Two opposite strategies appear in the record — a narrowing head-cone near the embouchure (the modern Boehm headjoint) and a contracting foot-cone toward the open end (the baroque traverso) — and our simulations show they are, to a useful approximation, *interchangeable*: either placement stretches the octave in the same direction, and one can be made to reproduce the other to within ~ 10 – 20 cents.

The Headjoint as Two Superposed Knobs.

In the modern headjoint the correction is carried by two independent, additive elements (Fig. 1): the conical taper sets an overall *offset* of the octave stretch ($+43$ cents at the bottom of the register), while the cork-stopper cavity behind the embouchure sets its *slope* across the register (about -37 cents at the top). They act on different aspects of the same curve, which is why small changes in the embouchure/headjoint region produce large, structured changes in response — the maker’s most sensitive controls live here.

3. From the Landscape to “Ring”, and the Player Who Navigates It

A well-placed octave does not guarantee that a note *speaks* well or sounds rich. In the self-sustained jet-resonator loop, what matters is whether the second harmonic injected by the jet falls on a bore resonance — Benade’s *cooperative regime*, the bridge from geometry to timbre. We summarise this with a measure of how well the second harmonic aligns, and it has a convenient property: it can be read from the resonator alone. The reason is a cancellation. The air jet acts with a phase lag proportional to frequency — physically a fixed convective travel time across the embouchure — so its lag at the octave is exactly twice its lag at the fundamental. Our alignment measure compares the octave with twice the fundamental,

so these two jet contributions are equal and subtract out (this follows from the self-sustained loop-phase condition of Auvray et al., 2012). What remains is set by the bore geometry, not by how hard or fast the player blows. A purely passive, geometric quantity therefore predicts an active, timbral property — where a note will be rich and easy versus stuffy and effortful.

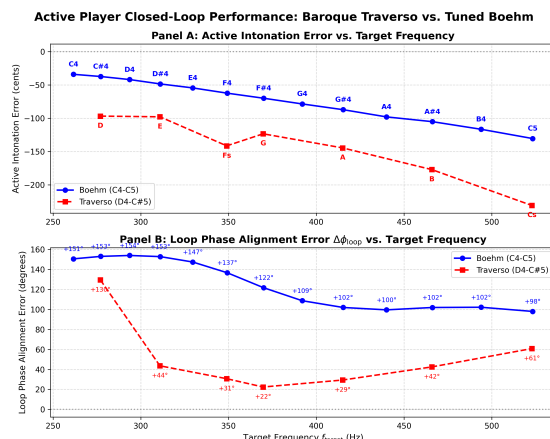


Figure 2: Simulated playability landscape across the register for a tuned modern flute and a baroque traverso. **Top:** active intonation error. **Bottom:** second-harmonic alignment; the traverso shows a central minimum (its “home key”), the modern flute is more uniform. Simulated.

Read across the register, this alignment traces qualitatively different profiles for the two design philosophies (Fig. 2). The baroque traverso shows a marked minimum in its central register (best support on its “home” notes, F#–G–A), rising toward the register extremes: a geometry-grounded handle on the long-standing distinction between notes that ring and notes that are stuffy, and on the instrument’s tonal character. The modern flute is comparatively uniform across the register — chromatic evenness in any key, at the cost of a pronounced home-key colour. The contrast is consistent with the two instruments’ musical purposes.

The Player Completes the Instrument.

The landscape is not destiny. The geometry leaves a residual, register-dependent slope in the octave stretch, and the player absorbs it with two continuous controls: the *speed of the air jet* and the *lip coverage* of the embouchure (rolling the flute in or out). In the model these co-vary smoothly across the scale; at the register break the required coverage rises noticeably, exactly the lip adjustment players describe when crossing into the second register. This two-control picture yields a concrete, testable prediction for recordings, and reframes a flute’s “timbre” not as a fixed spectral fingerprint but as an emergent property that the geometry sets up and the player realises.

4. A Corpus of Designed Identities

Does this character vary in structured ways across makers, or is it incidental? Projecting the author’s digitised corpus of historically-inspired traversos onto a small set of physically interpretable descriptors — the level and slope of the octave stretch, and the second-harmonic alignment — and reducing to two dimensions, the instruments do *not* scatter (Fig. 3). They cluster by maker lineage (Quantz-Baudin, Heinrich, Kirst, Wijne form coherent families), optimised or modified variants shift away from their originals, and a synthetic modern flute sits clearly apart. The first two dimensions capture 84% of the variation. This suggests that the passive octave structure and the playability landscape are *designed*

quantities — makers worked toward reproducible signatures — rather than incidental by-products, and that timbral identity is, in this sense, an emergent and reproducible property of the geometry.

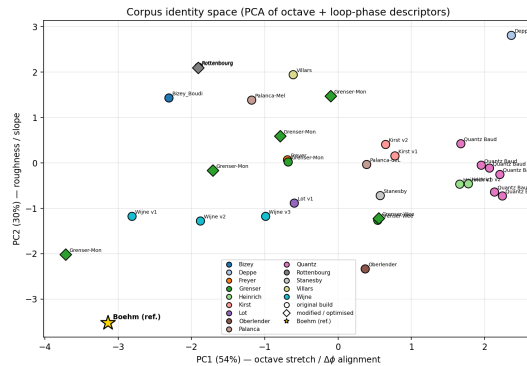


Figure 3: The author’s corpus in a two-dimensional “identity space” built from octave-stretch and second-harmonic-alignment descriptors. Colour marks maker lineage; circles are original builds, diamonds modified/optimised variants; the star is a synthetic Boehm reference. Families cluster. Simulated.

5. Discussion and Outlook

The three sections share one shape: a deliberate choice of internal geometry — a taper and a cork cavity, a maker’s bore lineage, a register-dependent target — propagates through the resonator into a quantity the player and listener ultimately experience. The framework makes that propagation explicit and computable, and in particular connects a passive, geometric descriptor to an active, perceptual one through a relation in which the player’s jet delay cancels.

The decisive next step is direct and already under way: whether geometries that align the second harmonic do in fact radiate richer upper partials and speak more readily is a question that only recorded sound can settle. Our ongoing work pairs the simulated descriptors reported here with input-impedance measurements and controlled recordings (fixed fingering, controlled blowing) on the same corpus, so that timbral effects driven by the instrument’s structure can be separated, as far as possible, from those arising from performance technique. We offer the framework not as a closed result but as a methodology and an invitation: the traverso is an instrument in which small geometric choices have large, structured acoustic consequences, making it a productive case study for treating timbre as an emergent property of geometry, resonance, and excitation.

References

- Auvray, R., Fabre, B., & Lagrée, P.-Y. (2012). Regime change and oscillation thresholds in recorder-like instruments. *Journal of the Acoustical Society of America*, 131(2), 1574–1585.
- Benade, A. H. (1976). *Fundamentals of musical acoustics*. Oxford University Press.
- Coltman, J. W. (1976). Sounding mechanism of the flute and organ pipe. *Journal of the Acoustical Society of America*, 60(3), 725–733.
- Fabre, B., Gilbert, J., Hirschberg, A., & Pelorson, X. (2012). Aeroacoustics of musical instruments. *Annual Review of Fluid Mechanics*, 44, 1–25.
- Fletcher, N. H., & Rossing, T. D. (1998). *The physics of musical instruments* (2nd ed.). Springer.
- Lefebvre, A., & Scavone, G. P. (2012). Characterization of woodwind instrument toneholes with the finite element method. *Journal of the Acoustical Society of America*, 131(4), 3153–3163.

- Nederveen, C. J. (1998). *Acoustical aspects of woodwind instruments* (rev. ed.). Northern Illinois University Press.
- Tournemenne, R., & Chabassier, J. (2019). A comparison of a one-dimensional finite element method and the transfer matrix method for the computation of wind music instrument impedance. *Acta Acustica united with Acustica*, 105(5), 838–849. <https://doi.org/10.3813/AAA.919364>

Individual Differences in Bow-Related Timbre Across 23 Recordings of Bach's Fourth Cello Suite Prelude

Clare Monfredo^{1†}

¹ Hunter College, City University of New York

[†] Corresponding Author: cmonfredo@gmail.com

ABSTRACT

In western classical music, performers are expected not only to realize composers' intentions faithfully from notated scores, but also to create an individual interpretation. To accomplish the latter, performers widely report that phrasing and timbre (particularly timbre as a result of articulation) are the main factors in individual interpretations. Despite the importance of these two factors, performance analyses that compare interpretations of a given piece have often overlooked timbre. Here I report findings, focusing on timbre and articulation, from quantitative and qualitative analyses of 23 commercial recordings of Bach's Fourth Cello Suite Prelude. I find that normalized RMS envelope and spectral centroid of performers' bow strokes (averaged by beat over measures) correspond to systematic differences in perceived stroke sustain and stroke quality – two critical elements of bow articulation known to indicate performance style – and thereby capture meaningful stylistic differences in interpretation. Notably, these two features do not correspond to standard categories of performance style (e.g., Baroque vs. modern). Together, these findings point to timbral analysis of player-specific articulation as an area in which to discover and understand differences in interpretation.

Introduction

Since the 19th century, classical music tradition has subscribed to the romantic ideology of *Werktreue*, or fidelity to the written work of composers. *Werktreue* asks performers to satisfy two competing demands: (1) to faithfully realize the composer's notated text, and (2) create an interpretation that is recognizably their own (Goehr, 1992). Performers of bowed string instruments cite both musical phrasing and timbre—particularly via sound quality and bowed articulation—as principal features of their personalized interpretations, especially in solo Bach (Leech-Wilkinson & Prior, 2017; Bylsma, 2001). In studies of interpretation, performance analysis has primarily focused on changes in tempo and dynamics as typical measurements of phrasing (e.g. Sung & Fabian, 2011). However, while timbre has been studied as a distinguishing musical feature of both different performers and specific musical techniques, it has rarely been studied as a primary feature of interpretation (Chudy, 2016; Traube, 2004; Llorens, 2023).

Here I present findings on interpretation from quantitative and qualitative analyses of articulation and timbre in 23 commercial recordings of Bach's Fourth Cello Suite Prelude made between 2000 and 2020 by 20 prominent cellists. The quantitative analyses of articulation assess two components of sound across different bow strokes: (i) RMS energy (envelope) and (ii) spectral centroid. Past work has identified correspondences between articulation and both RMS and spectral centroid of audio information. As a quantitative descriptor of sound, RMS has been used to classify articulation types (e.g., staccato versus legato) and measure decay in bowed instruments (Dalmazzo & Ramírez, 2019). Spectral centroid has been shown to generally correspond to tone color (e.g. "bright" versus "dark" tone) and, in bowed strings, specifically correlate with increased bow force (McAdams, 2019; Schoonderwaldt, 2009). Here I choose both measures: for RMS, I use time- and amplitude- normalization to enable comparing bow stroke shapes (for strokes both within and across recordings); for spectral centroid, I use raw and normalized – both time- and spectral- normalization – forms, where spectral-normalization is adopted to help control for differences in pitch and recording environment, and thus enable comparisons of sound quality.

Methods

Note segmentation. The 23 selected performers span ages, backgrounds (musical training and nationality), and period- and modern-instrument practices ([link to recordings and discography here](#)). Across all recordings, eighth note values (8 per measure across 89 measures) were manually notated as note onsets in Sonic Visualiser by repeated listening at 10, 20, and 50% speeds. Onset times were exported to CSV files. In this study, strokes were exclusively studied from the first 48 measures of the piece (creating a dataset of 23x48x8 total strokes).

Audio preprocessing and feature extraction. WAV sound files of the recordings were loaded as mono PCM at 44.1 kHz using librosa. Stroke onset times were imported from CSVs. For each recording, a single full-track spectral-centroid trace and a single RMS-energy trace were computed with the same short-time Fourier framing ($n_fft = 2048$, $hop = 256$, $center=True$, frequency resolution ≈ 21.5 Hz) via `librosa.feature.spectral_centroid` and `librosa.feature.rms`. Each trace was then segmented at the annotated onsets, yielding one centroid and one RMS time series per stroke (coinciding with beats).

Beat-position averaging and normalization. Within each performance, strokes were grouped by their metric position in measures, yielding 8 groups corresponding to eighth-note metric positions within bars. Each stroke's centroid (or RMS) trace was linearly resampled to 100 evenly spaced points spanning the stroke (normalized time spanning 0 to 1) and, for every beat group, averaged across strokes. For spectral centroid, two traces were then used for analysis: a raw form (in Hz) and a within-stroke z-scored form, computed by subtracting the per-stroke mean and dividing by the per-stroke standard deviation (done prior to resampling and averaging). For RMS, only a within-stroke z-scored form was used for analysis.

To compare the z-scored measures (spectral centroid or RMS) between players, the average value of the trace during the following time windows were calculated for each stroke: t_1 [0, 0.3), t_2 [0.3, 0.6), and t_3 [0.6, 0.9). Note that the last 0.1 of each stroke was disregarded given known noisy behavior at the end of segmented notes.

Qualitative assessments. Qualitative impressions were made from multiple listenings over extended time periods. Listenings were done at normal, 75%, and 50% speed. Articulations were evaluated based on amount of *stroke sustain*, or the relative length within the timeframe of the stroke in which the bow sounds to be activating the string, and type of *sound quality*, or the perceived amount of bow force exerted on the string. Sustain was judged on 4-part scale: short (significant breaks between bow changes), moderate released (breaks between changes), moderate long (slight breaks or clear demarcations between changes), and long (no breaks). Sound quality was judged on a 4-part scale: light, moderate, heavy, and very heavy. These two systems of evaluation were based on prior qualitative assessments of solo string recordings of Bach, and on my own experience as a professional cellist (Fabian, 2015).

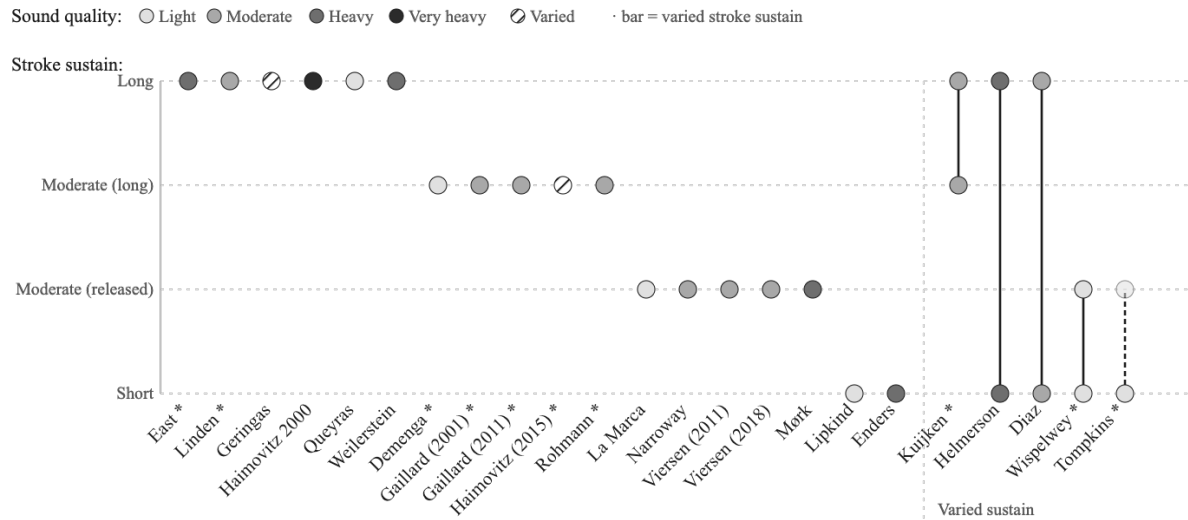


Table 1. Qualitative Impressions of Stroke Sustain and Sound Quality. Note that player's stroke sustain and sound quality are not correlated to performance practice. * Indicates Baroque performers.

Results

Two quantitative approaches reveal performer-specific bowing style. To quantify timbral differences across individual performers, two different sound measurements were used: RMS and spectral centroid. Further, to enable comparisons across individual strokes, stroke durations were time-normalized (0 to 1), and both spectral centroid and RMS were z-scored for each individual stroke.

Strikingly, in trace plots of both measurements, I found that performers could consistently show characteristic patterns, and also often showed characteristic patterns across beats (four example performers in **Figure 1**). In addition, these plots indicated that certain performers showed greater variability across beats than others (e.g. Haimovitz in **Figure 1**). These initial findings suggested that RMS, spectral centroid, or their combination, could capture qualitative timbral differences in bow use.

Given this possibility, I developed one such measure based on z-scored spectral centroid. From the trace of spectral centroid, a multi-peaked pattern could often be seen (e.g., in Queyras, Weilerstein, and Haimovitz in **Figure 1**). This was particularly notable since RMS generally lacked this multi-peaked structure (e.g. Enders in **Figure 1**; and most other players analyzed; not shown), instead showing approximately a single wave across a (time-normalized) stroke. To capture the multi-peaked structure, I quantified the average z-scored spectral centroid value for three different periods (early, middle, late; termed t1, t2, t3; see Methods for definitions) across strokes. Scatter plots of these values for performers showed clustering by performer (**Figure 2, left**; different points from a performer (in a given color) correspond to beat-specific strokes averaged across measures), highlighting the potential use of this measure to characterize timbre.

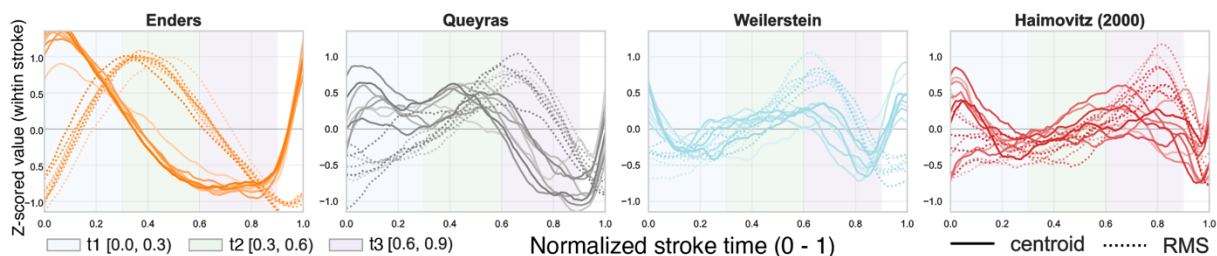


Figure 1. Timbral traces (RMS and spectral centroid) for four example performers. Lines: RMS envelope; dotted lines: spectral centroid. In qualitative evaluations, Enders (orange) uses short, heavy strokes; Queyras (grey), Weilerstein (blue), and Haimovitz (red) use long strokes of varying weights.

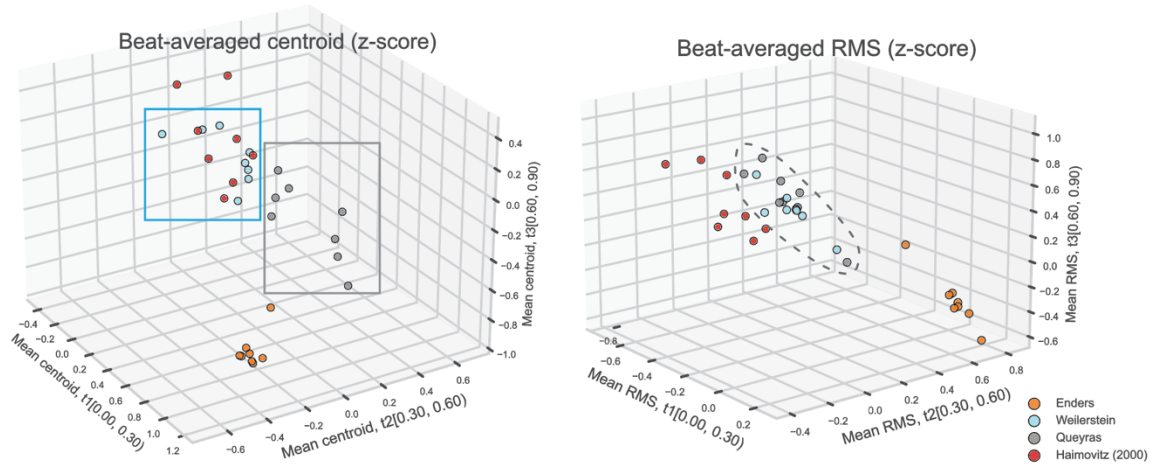


Figure 2. Summary features across four example performers. Left, RMS features; right, spectral centroid features. Features are the values of either the RMS or spectral centroid traces in specific time windows (beginning, middle, end) in the time-normalized stroke. Selected players are indicated by colors, where each point corresponds to a specific beat position in measure. Weilerstein and Queyras (blue and grey boxes, left) differ in spectral centroid but are very similar in RMS (dashed circle, right).

Though RMS showed less variable structure than spectral centroid, I found that RMS, as expected, did correspond to qualitatively judged stroke sustain. For example, Enders has relatively little sustain whereas Haimovitz has relatively high sustain; in terms of the RMS traces, this corresponds to an overtly earlier peak of RMS in Enders as compared to Haimovitz (**Figure 1**, dotted lines). Similar results were seen for other performers (not shown). To enable comparison to spectral centroid, I also quantified the (z-scored) RMS value for the same periods of time as stated above (early, middle, late; t1, t2, t3) (**Figure 2**, at right).

Quantitative signatures of stroke sustain and sound quality across performers. In performance practice categorizations (e.g. “modern” versus “historically-informed”) of bowed instruments, stroke sustain and sound quality are widely used to characterize performance style (Fabian, 2015).

The above analyses suggest two relationships between timbre and these performance practice characterizations. The first is between perceived stroke sustain and the time of the peak in RMS: in **Figure 2** at left, players with high t3 RMS values correspond to sustained players (Queyras, Weilerstein, Haimovitz), while low t3 RMS values correspond to un-sustained players (Enders). This was widely observed across players with different sustaining patterns. The second relationship is between perceived sound quality and the peak(s) in spectral centroid: in **Figure 2** at left (blue and grey boxes), Weilerstein (blue) and Queyras (grey) have diverging spectral centroid peaks, despite nearly identical RMS peaks (**Figure 2** at right, dashed circle). Queyras has consistently higher t2 and lower t3 centroid values compared to Weilerstein, corresponding to his lighter perceived tone. This difference in spectral centroid trace, along with others like it, suggests that placement of spectral centroid peaks within strokes could be an indicator of stroke sound quality. Together, spectral centroid and RMS provide highly specific, measurable accounts of stroke with which to study and characterize individual differences in interpretations.

Discussion

In this study, I found that RMS and spectral centroid allow for the quantification and comparison of audibly and stylistically significant features of bowed articulation and associated performance style. This approach differs from prior work where recordings are evaluated only qualitatively or through video recording and/or motion capture. Further, I find that by considering both RMS and spectral centroid for a given set of recordings, quantitative study of timbre can begin to provide more specific correspondences to qualitative assessments of stroke, and thus performance style. Broadly, these results indicate that timbral features can expand our understanding of individual musical interpretation.

References (selected)

- Bijlsma, A., Bartman, G., & Bach, J. S. (2001). *Bach, the Fencing Master: Reading Aloud from the First Three Cello Suites*. Bylsma.
- Chudy, M. (2016). *Discriminating music performers by timbre: On the relation between instrumental gesture, tone quality and perception in classical cello performance* [Doctoral dissertation, Queen Mary University of London].
- Dalmazzo, D., & Ramírez, R. (2019). Bowing gestures classification in violin performance: a machine learning approach. *Frontiers in Psychology*, 10, 344.
- Fabian, D. (2015). *A Musicology of Performance: Theory and Method based on Bach's Solos for Violin*. Open Book Publishers.
- Goehr, L. (1992). *The Imaginary Museum of Musical Works: An Essay in the Philosophy of Music*. Clarendon Press.
- Leech-Wilkinson, D., & Prior, H. M. (Eds.). (2017). *Music and Shape*. Oxford University Press.
- McAdams, S. (2019). The perceptual representation of timbre. In *Timbre: Acoustics, perception, and cognition* (pp. 23-57). Cham: Springer International Publishing.
- Schoonderwaldt, E. (2009). Mechanics and acoustics of violin bowing: Freedom, constraints and control in performance [Doctoral dissertation, KTH Royal Institute of Technology]. DiVA Portal.
- Traube, C. (2004). *An interdisciplinary study of the timbre of the classical guitar* [Doctoral dissertation, McGill University]. ProQuest Dissertations and Theses Global.

Timbre & Text: A Preliminary Study of Dietrich Fischer-Dieskau's Recordings of Schubert's "Gute Nacht"

Tracy Lipke-Perry^{1†}

¹ Crane School of Music, State University of New York at Potsdam, Potsdam, New York, United States

[†] Corresponding author: lipkeptd@potsdam.edu

Abstract

Dietrich Fischer-Dieskau's recordings of Schubert's *Winterreise* offer a unique window into the artistry and sensibilities of one of the world's elite baritones. Among the numerous available commercial recordings, seven studio recordings from 1955 to 1990 and three live performances from 1948, 1952, and 1953, are the focus of this study. Tracing the expressive variability of these performances provides insight into Fischer-Dieskau's world, the essence of a seemingly ageless voice, and the repeatability across performances and aesthetic discernment spanning a remarkable 42 years.

Of particular interest are parameters related to timbre, both because timbre may be altered for deliberate interpretive intent and because of the natural variations in voice quality across the lifespan. This study compares jitter, shimmer, and the ratio of harmonicity to noise in the first song of the cycle, "Gute Nacht," to discern expressivity trends directly related to timbre and text painting.

For each recording, LALALAI was used to isolate the vocal stem from the audio file. The voice signal was then imported into Praat for analysis with a pitch floor of 50 Hz and a pitch ceiling of 500 Hz. Exemplar data for three similar gestures across the 10 recordings of "Gute Nacht," were examined: A) the first phraselet of the song (i.e., first eight notes); B) the beginning of the second phrase (i.e., notes 17-24); and C) the analogous gesture beginning at the key change, 35 measures before the end. The onset and offset of each steady-state vowel sound was identified, facilitating identification of the average fundamental frequency and subsequent quantitative analysis, syllable by syllable. Coefficient of variation (CoV) and range were calculated for the 3 variables of interest across each of the 10 performances. Group average and standard deviation were then calculated across the three melodic gestures.

Key differences in shimmer and harmonicity to noise ratio were observed in the group data at the key change where the poem references a dream. Fischer-Dieskau systematically changed the timbre of his voice to heighten both the impact of the poem and Schubert's setting. This was one example of Fischer-Dieskau's interpretive compass on a formal scale, but it permeated the melodic line as well. A close examination of the performances syllable by syllable reveals correlates with metric structure, text painting, and interpretive phrasing. For example, greater jitter values were typically observed in the last half of each gesture, often coinciding with a downbeat highlighting a particularly salient word. The listener then perceives forward direction in the phrase and a longer line. The consistency of the CoV data overall point to general persistence of expressive interpretation correlating with the harmonic structure of the song, while subtle differences observed year-to-year within the jitter, shimmer and harmonicity to noise data suggest the prerogative of the artist to subtly vary interpretation over time. Among the variables sampled, it should be noted that no obvious age-related differences emerged.

This preliminary work lays the foundation for comprehensive analysis of the Fischer-Dieskau's *Winterreise* recordings and comparison with other artists', some of whom also recorded the work multiple times. The significance and implications are in embracing the intersections of science and art to better understand processes involved in creating art: conception, artist perception, performance, and listener perception as well as the potential for meaningful engagement with music and singing across the lifespan.

References

- Boersma, P. & Weenink, D. (2025). *Praat: doing phonetics by computer* (6.4.48) [Computer program]. <https://praat.org>
- Butte, C. J., Zhang, Y., Song, H., & Jiang, J. J. (2009). Perturbation and nonlinear dynamic analysis of different singing styles. *Journal of Voice*, 23(6), 647-652.
- Capobianco, S., Nacci, A., Calcinoni, O., Bruschini, L., Berrettini, S., & Bottalico, P. (2023). Assessing acoustic parameters in early music and romantic operatic singing. *Journal of Voice*, 39(4):1044-1052.
- Hakanpää T., Waaramaa T., Laukkanen, A. M. (2021). Comparing contemporary commercial and classical styles: Emotion expression in singing. *Journal of Voice*, 35(4):570-580.
- Lortie, C. L., Rivard, J., Thibeault, M., & Tremblay, P. (2017). The moderating effect of frequent singing on voice aging. *Journal of Voice*, 31(1), 112-e1.
- Sundberg, J., Iwarsson, J., & Hagegård, H. (1995). A singer's expression of emotions in sung performance. *Vocal fold physiology: Voice quality control*, 217-229.

Session 11

Sound Studies and Phenomenology

Listening to Demonstrations: Timbral Perspectives on Urban Crowd Sound

Kamyar Karimi^{1†}, Alice Jarry¹, Gabriel Vigliensoni¹

¹ Design and Computation Arts, Concordia University, Montréal, Québec, Canada

[†] Corresponding author: k_kamyar@live.concordia.ca

In this study, we present a literature review on how the sounds of public demonstrations are discussed in sound studies, soundscape research, and sound art. Following work in musical timbre perception (McAdams, 2013), we understand timbre as a multidimensional perceptual attribute shaped by spectral envelope, temporal attack and decay, spectral flux, inharmonicity, and modulation patterns, rather than as a residual “colour” of sound. We take this musical timbre framework as a starting point rather than a given model and ask what happens when it is applied to the sounds of urban crowds. Classic work on soundscapes and urban noise (Schafer, 1994; Truax, 1984) often situates demonstrations within broader acoustic environments but tends to treat them in terms of overall amplitude, contour, and semantic content. Media-historical accounts of listening (Sterne, 2003) further show how these environments are culturally framed, yet they rarely isolate timbre as an analytical category. In contrast to the controlled, often quasi-stationary instrumental sounds that underpin much timbre research, demonstration sound typically consists of mixtures of many unsynchronized and non-stationary sources, strong architectural reflections, and shifting configurations of listeners and sound-makers. In this context, the practical “object” of listening is not a single source but an emergent crowd texture, which complicates source-based models of timbre.

Drawing on LaBelle’s concept of acoustic justice (2020), we focus on how audibility and the right to make sound are unevenly distributed, and how certain sounds in public space are tolerated or suppressed. In demonstrations, this unevenness appears in the contrast between oppositional sounds such as chanting, drumming, whistling, and amplified speech, and sanctioned sounds such as sirens, police commands, public announcements, and other institutional signals. Within this framework, we review how listening practices and regulatory frameworks condition who is heard in demonstrations and how these sounds are framed as noise, signal, or expression. We then engage Goodman’s work on sonic warfare (2009) and Thompson’s analysis of noise and aesthetic judgment (2017) to show how specific timbral features contribute to different kinds of acoustic atmosphere, including tension, urgency, or cohesion. Together, these accounts suggest that timbre is a key parameter in how urban crowd sound is interpreted and managed.

We also examine evidentiary listening in forensic and activist sound practices. Works such as Abu Hamdan’s *Earshot* (2016), Ultra-red’s listening protocols and militant sound listening sessions (2015), and Yoganathan’s account of field recording as a practice of resistance and aural counterpublic formation (2021) show how recordings of firearms, voices, ambient city sound, and protest environments are used as materials for analysis, political articulation, and critical collective reflection. In these cases, timbral properties become analytically salient as a way of asking how a recording may be perceived. The sharpness and rise time of onsets in demonstration sound, the brightness of percussive instruments and police gunshots, the roughness of crowd noise, and the spectral stability or instability of voices become relevant to whether a sound is heard as clear or ambiguous, proximate or distant, and stable or fragile.

In the review we argue that evidentiary listening necessarily involves attention to such timbral dimensions, even when they are not named as timbre. A third theoretical perspective introduces phenomenological accounts of listening (Ihde, 2007) and research on soundwalks and urban listening (Jeon et al., 2013). This work highlights that listening is embodied, situated, and extended through environments and devices; viewed through a timbral lens, it opens a line of inquiry into how variations in

spectral density, envelope, and modulation patterns in crowd and signalling sounds might be felt in and around the body, in relation to space and the network of human and non-human actors present (Latour, 2005).

Across these strands, the review identifies a consistent gap: demonstration sound is often discussed in terms of volume, city noise regulations, and semantic content, while timbre tends to remain implicit or secondary. We argue that timbre is not incidental to the politics of urban crowd sound. It is one of the means through which collective sound becomes perceptible, interpretable, and governable. To make this claim analytically usable, we propose a timbre-sensitive taxonomy for studying demonstration sound. The six dimensions in Table 1 are synthesized from the literature reviewed above and are intended as a situated framework for listening organized around form, mediation, and political function. In doing so, we aim to provide a conceptual basis for future empirical and practice-based work on urban crowd sound in critical sound art and design.

Table 1: Analytic dimensions for a timbre-sensitive taxonomy of demonstration sound

<i>Dimension of analysis</i>	<i>What is being heard</i>	<i>Example Questions</i>	<i>Possible Examples</i>
Source	Who or what produces the sound	Is this voice, siren, loudspeaker, footstep, drone, traffic?	chant, megaphone, speech
Texture	How the sound behaves as a sonic mass or surface	Is it dense, rough, diffuse, layered, intermittent?	crowd wash, overlapping chants, street reverberation
Morphology	How the sound unfolds through time	Does it swell, interrupt, sustain, decay, repeat?	chant build-up, dispersing crowd
Timbral Qualities	Perceptual sound features	Is it bright, sharp, noisy, stable, unstable, harsh, muffled?	rough crowd noise, sharp onsets, compressed voice
Spatial Mediation	How space and devices shape the sound	Does architecture reflect it? Is it close, distant, filtered, amplified?	echoed chant, phone recording distortion, megaphone coloration
Social Function	What the sound does in the scene	Does it gather, warn, mask, intimidate, organize, saturate?	call-and-response chant, masking siren, directive speech

The six dimensions draw on different but related lineages. Source and timbral qualities derive most directly from timbre research and auditory perception, especially work concerned with perceptual features and source recognition (McAdams, 2013). Texture and morphology are informed by spectromorphological and soundscape approaches that attend to sonic mass, temporal unfolding, and emergent crowd form (Smalley, 1997; Schafer, 1994; Truax, 1984). Spatial mediation draws on soundscape, phenomenological, and media-historical accounts of how environments, architectures, and devices shape listening (Truax, 1984; Ihde, 2007; Sterne, 2003). Social function emerges from acoustic justice, sonic warfare, and activist sound practices, where the question revolves around what exactly

sound does functionally in a scene (LaBelle, 2020; Goodman, 2009; Abu Hamdan, 2016; Ultra-red, 2015; Yoganathan, 2021).

These dimensions overlap by design wherein texture refers to the organization of sonic mass, morphology to its temporal unfolding, and timbral qualities to the perceptual attributes through which those processes are heard (Smalley, 1997; McAdams, 2013). Rather than treating demonstration sound as a single-source event, this analytic instrument reframes urban crowd sound as a contested auditory formation whose timbral, spatial, and political dimensions must be analyzed together.

References

- Abu Hamdan, L. (2016). *Earshot* [Sound installation]. Portikus.
- Goodman, S. (2009). *Sonic warfare: Sound, affect, and the ecology of fear*. MIT Press.
- Ihde, D. (2007). *Listening and voice: Phenomenologies of sound* (2nd ed.). State University of New York Press.
- Jeon, J. Y., Hong, J. Y., & Lee, P. J. (2013). Soundwalk approach to identify urban soundscapes individually. *The Journal of the Acoustical Society of America*, 134(1), 803–812.
<https://doi.org/10.1121/1.4807801>
- LaBelle, B. (2020). *Acoustic justice: Listening, performativity, and the work of reorientation*. Bloomsbury Academic.
- Latour, B. (2005). *Reassembling the social: An introduction to actor-network-theory*. Oxford University Press.
- McAdams, S. (2013). Musical timbre perception. In D. Deutsch (Ed.), *The psychology of music* (3rd ed., pp. 35–67). Elsevier.
- Schafer, R. M. (1994). *The soundscape: Our sonic environment and the tuning of the world*. Destiny Books.
- Smalley, D. (1997). Spectromorphology: Explaining sound-shapes. *Organised Sound*, 2(2), 107–126.
<https://doi.org/10.1017/S1355771897009059>
- Sterne, J. (2003). *The audible past: Cultural origins of sound reproduction*. Duke University Press.
- Thompson, M. (2017). *Beyond unwanted sound: Noise, affect and aesthetic moralism*. Bloomsbury Academic.
- Truax, B. (1984). *Acoustic communication*. Ablex Publishing.
- Ultra-red. (2015). *Protocols for militant sound investigation* [Collective listening practice]. Ultra-red.
<https://www.ultrared.org>
- Yoganathan, N. (2021). Soundscapes of Resistance: Amplifying social justice activism and aural counterpublics through field recording-based sound practices. *Organised Sound*, 26(2), 201–210.
<https://doi.org/10.1017/S1355771821000248>

Timbre as Exposure: Fragility, Proximity, and the Audibility of the Body

Hee Seng Kye^{1,2†}

¹ Department of Composition, College of Music, Hanyang University, Seoul, Korea

² Music Research Center, Hanyang University, Seoul, Korea

† Corresponding author: hskye@hanyang.ac.kr

Introduction

This paper proposes that timbre can be understood as a form of *exposure*: a sonic condition in which the body becomes audibly present, vulnerable, and proximate to the listener. Timbre research has often emphasized identification, source recognition, and expressive nuance. The present study shifts attention to a different family of listening situations: those in which timbre renders the limits, effort, or fragility of the sounding body perceptible. Breath noise, instability, roughness, and reduced spectral richness are not only acoustic attributes; they are traces through which the body's vulnerability becomes legible to the ear.

The argument proceeds through three historically anchored case studies. The first examines late nineteenth-century vocal pedagogy and laryngology, in which timbre operated as a diagnostic medium for hidden bodily conditions. The second turns to early twentieth-century sound recording, which intensified the audibility of bodily artifacts and reconfigured what listeners expected timbre to reveal. The third addresses contemporary performance by disabled musicians, whose timbral practices foreground bodily presence as meaningful in itself, refusing the narrative of overcoming that has long shaped reception of disability in music (Quaglia, 2016; Straus, 2011). Through these cases, the paper proposes a concept of ethical listening that sustains proximity to the exposed body without seeking diagnostic closure or sentimental redemption (Kye, in press).

Case 1. Timbre as Diagnostic Medium: García, Mackenzie, and the Pathological Larynx

The category of timbre acquired specifically medical inflections in mid-nineteenth-century vocal pedagogy. Manuel García II's invention of the laryngoscope in 1854 made the singing larynx visible to its possessor and to others (Sterne, 2021). Less remarked is the parallel acoustic transformation: timbre, previously a matter of aesthetic taste, was becoming a perceptible index of bodily condition. García's *Treatise on the Art of Singing* opens its chapter on vocal quality with a frank diagnostic premise: "All uncultivated voices have, without exception, certain marked defects... Some, for instance, are tremulous; others nasal, guttural, veiled, harsh, &c." (García, 1840, p. 6). The taxonomy that follows is anatomically precise. Guttural timbre arises when "the tongue rises at its base," driving back the epiglottis; nasal timbre, from a relaxed soft palate; cavernous or hollow timbre, from the swelling of the tonsils; veiled sound, from a glottis "open, and yet has no brilliancy" (p. 6). Each defect is correlated with a specific anatomical malfunction, and each is, in principle, audible to the trained ear.

This convergence of aesthetic and diagnostic listening is extended in Morell Mackenzie's *Hygiene of the Vocal Organs* (1886). Physician to the Royal Society of Musicians and a practicing laryngologist whose patients included many of Victorian London's leading singers, Mackenzie correlates specific timbral irregularities with specific pathologies. Huskiness, on his account, results from a thickening of the vocal cords that indicates "chronic inflammation, consumption, &c." (Mackenzie, 1886, p. 134); a voice reduced almost to a whisper signals "the exhaustion of a wasting illness" (p. 126). Such practices belong to the longer history of medicalized listening that Jonathan Sterne (2003) traces to the early nineteenth-century stethoscope, through which the trained physician learned to apprehend internal bodily conditions as sonic configurations.

The conceptual point is sharper than the historical detail. The category of the pathological voice did not exist independently of these listening practices; it was constituted by them. To say that hoarseness, tremor, or breathiness “exposed” the diseased body misstates the relation: these qualities became audible *as* exposure only because the discourse had trained listeners to hear them that way. The discourse produced the audibility it claimed to discover.

A further dimension of Mackenzie’s text bears directly on this paper’s third case study. Beyond the diagnostic register, Mackenzie also celebrates the singer or actor whose skill renders “a naturally poor voice” so artfully managed “that its very defects are unperceived, or in some cases become beauties to the ear of the fervent admirer” (chap. VIII, p. 150). The “overcoming” frame, which contemporary disability studies has scrutinized as an ideological obstacle to ethical reception, is already operative here as a pedagogical ideal: the body’s audible irregularities are to be heard through the artist’s compensatory mastery, not on their own terms. The conceptual inheritance is significant. Judgments about timbral “purity,” “clarity,” or “freedom” carry this longer history of diagnostic and compensatory listening. Exposed timbres are heard against a normative anatomy of the unmarked voice and a normative narrative of its artful repair.

Case 2. The Microphone’s Body: Recording and Intensified Audibility

The advent of sound recording did not merely inherit this diagnostic-aesthetic listening; it transformed it. Early twentieth-century critics already understood that the phonograph altered what timbre revealed. In a 1929 review of Lotte Lehmann’s recording of Schumann’s *Frauenliebe und -leben*, the critic Herman Klein praised the performance while regretting “the constant hiss of the intake of breath” as “the only blemish” on an otherwise perfect achievement (cited in Beaudoin, 2024, pp. 13–14). The complaint registers something genuinely new. In the concert hall, Lehmann’s breath had been part of an integrated sonic presence; through the microphone, it had become a discrete, foregrounded object available to evaluative attention.

Recent work has reframed these “blemishes” as constitutive of recorded music itself. Richard Beaudoin (2024) argues that breath sounds, fingernail clicks, chair creaks, and effortful vocalizations belong to the musical work itself, constituting a “tracks-as-artworks” form of music that demands its own analytic vocabulary (p. 15). His categorization of audible exertion as “climactic” or “intimate,” and his close reading of pianist Hélène Grimaud’s “crossing of the threshold between breathing and vocalizing” (p. 156) in Beethoven, demonstrates that timbre in recorded contexts cannot be cleanly separated from the audible body that produces it. Nina Sun Eidsheim’s (2019) reframing of vocality around its “nonsonorous aspects” and listening as a culturally formed practice extends this argument. In Don Ihde’s (2007) phenomenology of mediated listening, acoustic technologies operate through “amplification/reduction” (p. 249): the microphone is no transparent conduit but reconfigures the spatial, dynamic, and proximal relations through which timbre is heard. The result is a sonic intimacy that nineteenth-century listeners would have found uncanny.

A century of engineering practice has worked to suppress what the microphone made possible. The technologies of “spectral cleaning” and “intelligent suppression” that Beaudoin (2024) traces are continuous with García’s pedagogical injunctions against guttural and nasal sound: both seek a body audibly subtracted from its own sounding. To attend to recordings as documents of bodily presence rather than transparent renderings of works is to recover, ethically as well as analytically, something the medium has been laboring to erase. The body has been audibly available, in a form intensified by the very technology that promised to deliver pure musical content.

Case 3. Disablist Timbre: Chesnutt, Mitchell, and the Refusal of Overcoming

A third configuration emerges in contemporary performance practice. Blake Howe (2016) describes “disablist music” as “a musical practice that rejects the normal performance body and instead molds its performance practices around the impairments of its performers” (p. 202). Rather than concealing or compensating for impairment, disablist music *audibilizes* it. Such audibilization participates in what Joseph Straus (2016) has theorized as an “aesthetics of disability,” a sensibility that finds expressive value in conditions previously coded as deviation. Two cases, one canonical to the disability-music literature and one contemporary, illustrate the timbral implications of this stance.

Vic Chesnutt, a wheelchair user from age eighteen following a car accident that left him with limited use of his hands, recorded across two decades with a voice and a guitar technique audibly shaped by reduced motor control and constrained breath support. His vocal timbre carries the texture of strained phonation: catches, drops, irregular vibrato, audible labor at the end of phrases. On *At the Cut* (2009), an album released months before his death, that roughness functions as expressive substance, the means through which the songs articulate vulnerability without requesting pity. His guitar registers an idiosyncratic touch that yields its own timbral signature. To hear these recordings is to encounter what Howe describes as the disabled performer’s “dual task”: performing music and performing disability, the two scripts becoming inseparable (p. 191, citing Straus, 2011).

Joni Mitchell’s post-aneurysm voice constitutes a second, more recent instance. Following a cerebral aneurysm in 2015 and a long recovery, Mitchell’s return to public performance at the 2022 Newport Folk Festival presented a transformed instrument: lower in tessitura, gravelly in texture, with audible effort in sustained phrases. A listener trained in conventional timbral norms may feel, alongside admiration for her resilience, a quieter discomfort at the distance between this voice and her earlier recordings. These bodily responses are what Thomas Csordas (1993) calls “somatic modes of attention,” culturally elaborated ways the body registers another’s audible condition. That discomfort is itself diagnostic, registering the difficulty of holding such a response open without translating it into either pity or “overcoming.”

The “overcoming” narrative, whose pedagogical genealogy reaches back at least to Mackenzie’s celebration of vocal art that “veils” defects, is the most persistent obstacle to ethical listening here. As Marianne Kielian-Gilbert (2016) argues, disability studies challenges the “neoliberal naturalization, the essentialism of progress, and narratives of triumph over, or of overcoming adversity” that structure conventional reception (p. 376). In musical terms, the overcoming script translates timbral irregularity into a prosthetic narrative: the disabled body is heard as a problem to be musically resolved (Quaglia, 2016). Chesnutt’s recordings and Mitchell’s performances refuse such resolution, sustaining audibility as a condition rather than a problem to be redeemed.

Conclusion: Ethical Listening as Sustained Proximity

In all three cases, timbre operates as a site of exposure where the body’s presence exceeds its communicative function. Drawing on phenomenology, sound studies, and theories of embodiment, alongside recent work that treats the voice as a historical and anthropological object rather than a natural given (Eidsheim, 2019; Weidman, 2006), the paper has argued that exposed timbres generate distinctive listening situations marked by heightened proximity, ethical ambiguity, and ethical responsibility. Listeners may experience empathy, discomfort, fascination, or voyeuristic attention: embodied responses elicited by timbral qualities themselves, prior to secondary interpretation or reflective judgment. Such listening is “drastic” rather than “gnostic” (Abbate, 2004), engaging the sonic event in its physicality without seeking decipherable meaning behind it.

The ethical ambiguity of such listening is a condition to be sustained, not a problem to be resolved. Ethical listening to exposed timbres involves neither diagnostic scrutiny nor sentimental identification, but a mode of listening attention that remains with discomfort without seeking to master or redeem it. Such listening refuses the interpretive closure that would stabilize the vulnerable body as an object of medical knowledge, aesthetic judgment, or inspirational narrative. It extends analysis beyond categorization and expressivity toward questions of proximity, vulnerability, and embodied responsibility in listening. From nineteenth-century pedagogy through twentieth-century recording to twenty-first-century disabled performance, timbre has remained a site at which the body becomes audibly present, ethically charged, and resistant to redemption.

References

- Abbate, C. (2004). Music—drastic or gnostic? *Critical Inquiry*, 30(3), 505–536.
- Beaudoin, R. (2024). *Sounds as they are: The unwritten music in classical recordings*. New York, USA: Oxford University Press.
- Csordas, T. J. (1993). Somatic modes of attention. *Cultural Anthropology*, 8(2), 135–156.
- Eidsheim, N. S. (2019). *The race of sound: Listening, timbre, and vocality in African American music*. Durham, USA: Duke University Press.
- García, M. (1924). *Garcia's treatise on the art of singing: A compendious method of instruction* (A. García, Ed.). London, UK: Leonard & Co. (Original work published 1840)
- Howe, B. (2016). Disabling music performance. In B. Howe, S. Jensen-Moulton, N. Lerner, & J. Straus (Eds.), *The Oxford handbook of music and disability studies* (pp. 191–209). New York, USA: Oxford University Press.
- Ihde, D. (2007). *Listening and voice: Phenomenologies of sound* (2nd ed.). Albany, USA: State University of New York Press.
- Kielian-Gilbert, M. (2016). Disabled moves: Multidimensional music listening, disturbing/activating differences of identity. In B. Howe, S. Jensen-Moulton, N. Lerner, & J. Straus (Eds.), *The Oxford handbook of music and disability studies* (pp. 371–399). New York, USA: Oxford University Press.
- Kye, H. S. (in press). Spirituality after damage: Auditory anamnesis and the posthuman musical body. In Z. Cochran & A. Couture (Eds.), *Routledge handbook of posthuman perspectives on music and spirituality*. London, UK: Routledge.
- Mackenzie, M. (1886). *The hygiene of the vocal organs: A practical handbook for singers and speakers*. London, UK: Macmillan.
- Quaglia, B. (2016). Musical prosthesis: Form, expression, and narrative structure in Beethoven's sonata movements. In B. Howe, S. Jensen-Moulton, N. Lerner, & J. Straus (Eds.), *The Oxford handbook of music and disability studies* (pp. 618–639). New York, USA: Oxford University Press.
- Sterne, J. (2003). *The audible past: Cultural origins of sound reproduction*. Durham, USA: Duke University Press.
- Sterne, J. (2021). *Diminished faculties: A political phenomenology of impairment*. Durham, USA: Duke University Press.
- Straus, J. N. (2011). *Extraordinary measures: Disability in music*. New York, USA: Oxford University Press.
- Straus, J. N. (2016). Representing the extraordinary body: Musical modernism's aesthetics of disability. In B. Howe, S. Jensen-Moulton, N. Lerner, & J. Straus (Eds.), *The Oxford handbook of music and disability studies* (pp. 729–755). New York, USA: Oxford University Press.
- Weidman, A. (2006). *Singing the classical, voicing the modern: The postcolonial politics of music in South India*. Durham, USA: Duke University Press.

Toward a Lachenmannian Phenomenology in the Music of Lachenmann, Saunders, and Iannotta

Zachary Bernstein^{1†}

¹Music Theory, Eastman School of Music, University of Rochester, Rochester, NY, USA

[†] Corresponding author: zbernstein@esm.rochester.edu

Abstract

Across many essays and lectures, Helmut Lachenmann (1996, 2021) has proposed a revisionist theory of sonic perception that is multi-dimensional, complexly hierarchical, fluid, subjective, temporal, and contextual. While Lachenmann describes this theory in reference to sound (*Klang*), it has much to offer the theorist of timbre (*Klangfarbe*). This paper will review aspects of Lachenmann's understanding of sound with the aim, initially, of providing an explication of his distinctive concepts, and to put him in dialogue with contemporary scholarship. I will then extract and demonstrate analytic principles from these concepts, with examples drawn from his own music and that of Rebecca Saunders and Clara Iannotta.

With a metaphor from Borges, Lachenmann draws attention to the ways subjective experience can group dissimilar elements into “families” of sounds. He extends this to note how those families have complex internal hierarchies, with different sounds manifesting extremes of different sonic aspects. In analysis, he points out how composers bring contrasting sounds into aural contact with one another via temporal progression and other aspects of musical context. He repeatedly finds that the result of sonic juxtaposition and temporal unfolding is that one instrument can be heard as a variant of another, as in his description of how a bell gradually transforms into a soprano's voice in Luigi Nono's *Il canto sospeso* (Lachenmann 2021, 243–44).

Lachenmann's approach lends a useful supplement to contemporary theories of timbral perception (e.g., Reymore 2021; see also Morrison 2024). His sonic families have only an oblique relation to the acoustic signal, and he demonstrates how subjective, in-time musical experience can complicate expectations about timbral categories. It is an irreducibly phenomenological approach as categories are redefined for each listener, in each piece, as the music unfolds (Yu 2022). It also presents a playful, experimental attitude toward source identification and auditory scene analysis. Finally, in its account of how sonic experience is shaped by the simultaneous apprehension of multiple events, it intersects with Hasegawa's (2018) exploration of timbre “as an emergent property of composite events.”

I will draw analytical examples from four works that demonstrate the value of attending to in-time redefinition of timbral categories. Lachenmann's *Schwankungen am Rand* works with the unlikely interplay of high strings, brass, pianos, electric guitars, and four thundersheets. The instruments are at times distinct; at other times, their sounds seem to merge. Strikingly, the thundersheets serve a mediating function, at times sounding, in the composer's words, as “radically deformed monster violins” (2021, 536). Lachenmann's *temA* and Saunders's *Skin* both demonstrate how vocal timbres can intersect with instrumental timbres. Both pieces further demonstrate the role of rhythm and gesture in shaping timbral categorization. Among many other sounds, *temA*, a thematization of breath (*Atem*), brings breathy vocal sounds into contact with extended techniques on the flute and cello. *Skin*, likewise, has the voice affiliate with a wide range of instrumental sounds emanating from its thirteen-piece ensemble: brass fluttertongue, wind glissandi, sustained string tones, and more (Hasegawa 2026, Saunders 2018). As the piece unfolds, the voice's changing timbral profile participates in new contexts, redefining its categorization as its

proceeds. Finally, Iannotta's *dead wasps in the jam jar (iii)* for string quartet, objects, and sine waves uses varied string techniques and modifications to create familial associations and sonic progressions in a context less governed by traditional instrumental categories (Iannotta 2026).

References

- Beavers, J. P. (2021). Ravel's Sound: Timbre and Orchestration in His Late Works. *Music Theory Online*, 27/1.
- Dubiel, J. (2017). Music Analysis and Kinds of Hearing-As. *Music Theory and Analysis*, 4/11, 233–42.
- Hasegawa, R. (2018). Timbre as Harmony—Harmony as Timbre. In E. I. Dolan & A. Rehding (Ed.), *The Oxford Handbook of Timbre* (pp. 525–551). New York, USA: Oxford University Press.
- Hasegawa, R. (2026). Timbre and Orchestration in Two Works by Rebecca Saunders for Soloist and Ensemble. In J. Delisle, R. Hasegawa, J. Noble, & M. Touizrar (Ed.), *The Oxford Handbook of Orchestration Studies* (pp. 574–598). New York, USA: Oxford University Press.
- Helgeson, A. (2013). What is Phenomenological Music, and What Does it Have to Do with Salvatore Sciarrino? *Perspectives of New Music*, 51/2, 4–36.
- Iannotta, C. (2026). Between Gesture and Sound. <https://claraianotta.com/between-gesture-and-sound/>.
- Lachenmann, H. (1996). *Musik als existentielle Erfahrung: Schriften 1996–1995*. Ed. Josef Häusler. Wiesbaden, Germany: Breitkopf & Härtel.
- Lachenmann, H. (2021). *Kunst als vom Geist beherrschte Magie: Texte zur Musik 1996 bis 2020*. Ed. Ulrich Mosch. Wiesbaden, Germany: Breitkopf & Härtel.
- Morrison, L. (2024). Timbre Space: On the Flat History of a Multidimensional Metaphor. *Music & Science*, 7, 1–16.
- Nonnenmann, R. (2000). Angebot durch Verweigerung: die Ästhetik instrumentalkonkreten Klangkomponierens in Helmut Lachenmanns frühen Orchesterwerken. Berlin, Germany: Schott.
- Reymore, L. (2021). Characterizing Prototypical Musical Instrument Timbres with Timbre Trait Profiles. *Musicae Scientiae* 26/3.
- Saunders, R. (2018.) Rebecca Saunders: analyse de Skin.” Video recording of lecture delivered at the IRCAM ManiFeste Academy, June 28, 2018. <https://med.ias.ircam.fr/x88c02c>.
- Yu, H. (2022). A Phenomenology of “Shimmer”: Klangtypen and Arpeggio in Liza Lim's *Invisibility* (2009). *Perspectives of New Music*, 60/2, 5–37.

Timbre as a Combination of Its Source and Contents: A Phenomenological Metaphor-Based Framework for Analysing Timbre in Musical Works

Matt Collins^{1†}

¹ Independent composer and researcher, London, United Kingdom

[†] Corresponding author: matt@mattcollins.uk

Within contemporary music analysis discourse, a variety of methodologies and technologies have been used to analyse timbre in musical works, with a keen interest now being taken in what was often an overlooked musical element in the past. The use of spectrograms is still one of the most common means by which to conduct such analyses (e.g., Latartara, 2008; Malloch, 2004; Siedenburger, 2009), although there is increasing recognition that the use of these alone does not fully capture how timbre functions and is perceived by listeners. While the approaches that attempt to respond to this issue are broad, what they share is the attempt to analyse timbre descriptively, and to situate these descriptions within a theoretical framework that draws from such disciplines as phenomenology, ethnography, sociology, or psychology (e.g., Beavers, 2021; Blake, 2012; Emmerson, 2016; Lavengood, 2020; Lochhead, 2023; Ramsay, 2016). One challenge that has emerged through these descriptive efforts is being able to ground an analysis explicitly within the chosen framework, instead of drifting into the use of idiosyncratic language.

In response to this challenge, this paper proposes a new theoretical framework for the analysis of timbre in musical works. This framework draws first on Walther-Hansen's theory of sound as a combination of the 'sound-in-itself' and the sound's apparent source (Walther-Hansen, 2020), from which this paper develops a concept of timbre that foregrounds this combination and the perceptual flux between these two models. Secondly, Wallmark and Kendall's application of Lakoff and Johnson's conceptual metaphor theory to timbre semantics (Lakoff & Johnson, 1980, 1999; Wallmark & Kendall, 2018) provides a basis for a kind of timbral analysis that, at least in part, relies on the use of descriptive metaphor. Underpinning the specific metaphors used is Walther-Hansen's encyclopaedia of common metaphorical descriptions of sound (Walther-Hansen, 2020). For the purposes of this framework, then, timbre is conceptualised from a phenomenological perspective as an ever-shifting combination of its source and contents, and culturally situated intersubjective metaphors are employed systematically as the means to describe said contents.

This paper demonstrates this framework by analysing the following contemporary mixed electroacoustic compositions: Priscilla McLean's 'Salt Canyons' from *The Inner Universe* (1983), Natasha Barrett's *RAS* (2000, rev. 2011), and Maggi Payne's *Scirocco* (1983) – building on analytical work first explicated in the author's Timbre 2020 proceedings paper (Collins, 2020). Through these timbral analyses, the framework is used to establish common compositional strategies for achieving timbral blend between the works' acoustic and electronic forces. Timbral blend is defined here as the capacity of sounds from two or more physical sources to coalesce so that a listener is not likely to determine that there is an aesthetic disconnect between them. The analyses proceed by describing and graphically representing the timbral threads that can be delineated within the works, whereby a timbral thread is defined as that which contains any sound(s) over time that could be reasonably perceptually categorised as being produced by one physical or imagined source. When a sound is explicitly identifiable as or suggestive of a kind of source, it is labelled as such, whereas when a sound's source is more ambiguous (i.e., tends towards the impression of a 'sound-in-itself'), metaphorical description is employed.

The first of four timbral threads that emerge and propagate during Priscilla McLean's 'Salt Canyons' (1983) for piano and tape is outlined here as a demonstration. The italicised metaphorical descriptors that follow are informed by Walther-Hansen's (2020) taxonomy of common sound metaphors, while the quoted source descriptions are from McLean's score for this work. The sounds heard in this thread appear

to originate either from the piano or have unidentifiable sources while nevertheless timbrally *jumping off* from the initial piano-like sounds. The thread begins with brief, fast repeated pitches on the piano which immediately *blur* into a *mass* of sound on the tape, described by McLean as a “choir” of overtones’ (1983, p. 51). This tape sound continues in the background until near the end of the work, altering subtly at three different points: becoming more ‘wind-like’ (1983, p. 52), then *thinner*, and finally, more *synthetic* – perhaps through electronic alteration. This thread typifies one of the ways in which McLean engenders timbral blend in this work: by combining live piano sounds, whose timbres bring the sounds’ sources into doubt due to the instrument’s preparation, with tape sounds that timbrally complement the live piano – in this thread’s case, the fundamental of the overtones heard on the tape corresponds with the repeated pitch on the piano, making it difficult to hear where one source ends and the other begins. Figure 1 provides an example of how the threads are graphically represented using a flowchart-type method.

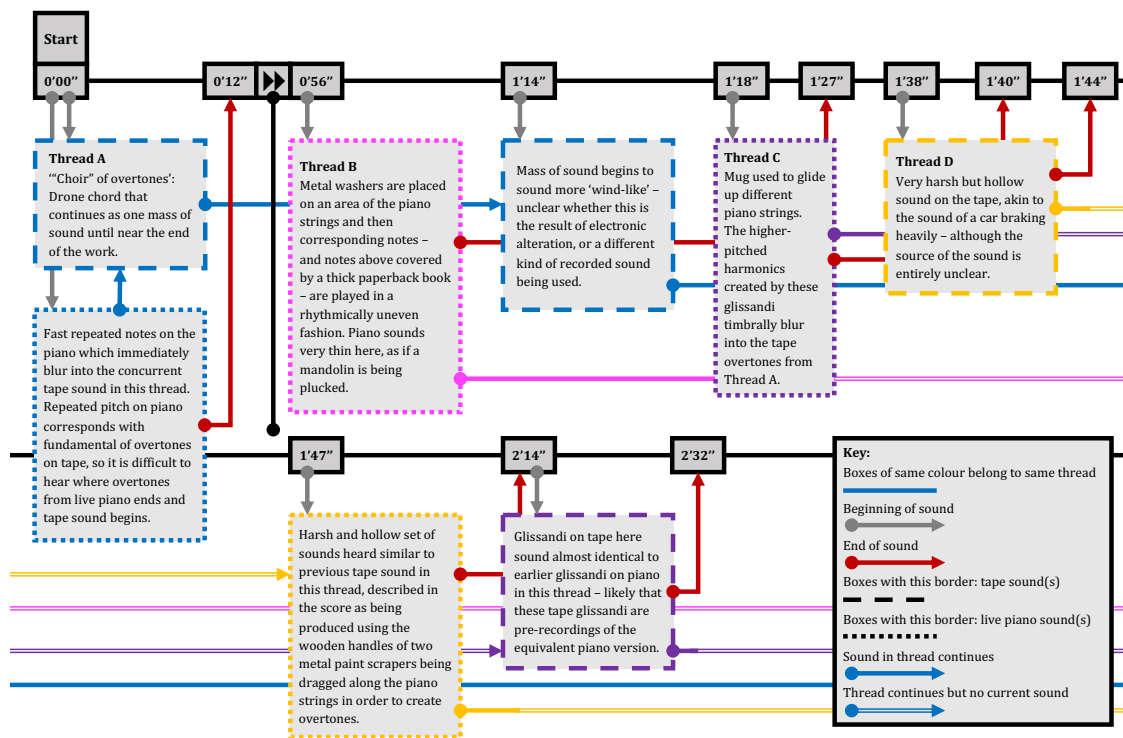


Figure 1: Simplified timbral analysis flowchart for part of Priscilla McLean's 'Salt Canyons'.

Across the analysed works, the most common compositional strategy to achieve timbral blend involves the use of timbres whose sources are challenging to identify. The acoustic sounds in these compositions primarily have sources that are either identifiable but not with complete certainty, unrecognisable even when heard in isolation, or would be recognisable were they not disguised in some way within the prevailing texture. The electronic sounds are principally one of two kinds: pre-recorded or electronically delayed live acoustic sounds that are otherwise subject to little or no apparent electronic alteration; or pre-recorded sounds of a seemingly real-world origin that are subject to different levels of perceptible electronic manipulation depending on the context. In all cases, it appears that the less recognisable a sound's source is, the more potential that sound has to blend with its acoustic or electronic counterpart.

This paper concludes by arguing that this framework enables a kind of analysis that is grounded in our everyday perceptual and descriptive experiences of timbre, and offers a practicable method through

which to communicate how timbre functions in a musical work and how it can be manipulated for specific aesthetic purposes in composition.

Acknowledgements

This paper builds on research undertaken during the author's DPhil at the University of Oxford, supported by the UK Arts and Humanities Research Council. The author would like to thank Gascia Ouzounian for her supervision of the research from which this work emerged, and for her feedback during the development of this project.

References

- Barrett, N. (2004). RAS [Audio recording by Puls & Ning Trio]. On *Works by Barrett, Hagen, Buene, Aagaard-Nilsen, Apollyon*. Euridice Turn Left AS.
<https://open.spotify.com/track/2crzeBbvoRnX0DA6fDysEu>
- Barrett, N. (2011). *RAS: Percussion quartet, computer and 8-channel sound* [Score]. NB noter.
- Beavers, J. P. (2021). Ravel's sound: Timbre and orchestration in his late works. *Music Theory Online*, 27(1). <https://doi.org/10.30535/mto.27.1.2>
- Blake, D. K. (2012). Timbre as differentiation in indie music. *Music Theory Online*, 18(2). <https://doi.org/10.30535/mto.18.2.1>
- Collins, M. (2020). Timbral threads: Compositional strategies for achieving timbral blend in mixed electroacoustic music. In A. Zacharakis, C. Saitis, & K. Siedenburger (Eds.), *Proceedings of the 2nd International Conference on Timbre* (pp. 34–37).
- Emmerson, S. (2016). The analysis of live and interactive electroacoustic music: Hans Tutschku – *Zellen-Linien* (2007). In S. Emmerson & L. Landy (Eds.), *Expanding the horizon of electroacoustic music analysis* (pp. 333–354). Cambridge University Press.
- Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. University of Chicago Press.
- Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh*. Basic Books.
- Latartara, J. (2008). Machaut's monophonic virelai *Tuit mi penser*: Intersections of language sound, pitch space, performance, and meaning. *Journal of Musicological Research*, 27(3), 226–253.
<https://doi.org/10.1080/01411890802212592>
- Lavengood, M. L. (2020). The cultural significance of timbre analysis: A case study in 1980s pop music, texture, and narrative. *Music Theory Online*, 26(3). <https://doi.org/10.30535/mto.26.3.3>
- Lochhead, J. (2023). Timbre realities: A phenomenological study of Liza Lim's *Extinction Events* and *Dawn Chorus* (2018). In J. De Souza, B. Steege, & J. Wiskus (Eds.), *The Oxford handbook of the phenomenology of music*. Oxford University Press.
<https://doi.org/10.1093/oxfordhb/9780197577844.013.16>
- Malloch, S. (2004). An exploration of timbre analysis: The game of sound in two performances of *Jeux Vénitiens*. *Musicae Scientiae*, 8(1), 53–81. <https://doi.org/10.1177/102986490400800103>
- McLean, B., & McLean, P. (2017). *The Inner Universe 5. Salt Canyons by Priscilla McLean* [Video]. YouTube. <https://www.youtube.com/watch?v=s2RBBP7VcFU>
- McLean, P. (1983). Salt Canyons. In *The inner universe* [Score]. MLC Publications.
- Payne, M. (1991). Scirocco [Audio recording]. On *Crystal*. Lovely Music, Ltd.
- Ramsay, B. (2016). Analysis of Foil by Autechre (from *Amber* (1994)). In S. Emmerson & L. Landy (Eds.), *Expanding the horizon of electroacoustic music analysis* (pp. 209–230). Cambridge University Press.

- Siedenburg, K. (2009). An exploration of real-time visualizations of musical timbre. In S. Baumann, J. J. Burred, A. Nürnberger, & S. Stober (Eds.), *Proceedings of the 3rd International Workshop on Learning Semantics of Audio Signals* (pp. 17–30).
https://bib.sebastianstober.de/lzas2009_proceedings.pdf#page=23
- Wallmark, Z., & Kendall, R. A. (2018). Describing sound: The cognitive linguistics of timbre. In E. I. Dolan & A. Rehding (Eds.), *The Oxford handbook of timbre* (pp. 578–608). Oxford University Press.
- Walther-Hansen, M. (2020). *Making sense of recordings: How cognitive processing of recorded sound works*. Oxford University Press.

Session 12

Hearing/Touching Synthetic Textures

Timbre Beyond Hearing: Auditory and Vibrotactile Interactions in the Perception of Timbre-Related Acoustic Features

Loonan Chauvette^{1,2†}, Anne Sophie Grenier^{1,3}, Philippe Albouy^{1,3}, Emily Coffey^{4,5,6,7},
Robert Zatorre^{5,6,7} and Andréanne Sharp^{1,2,5}

¹ CERVO Brain Research Centre, Québec, QC, Canada

² School of Rehabilitation, Université Laval, Québec, QC, Canada

³ School of Psychology, Université Laval, Québec, QC, Canada

⁴ Department of Psychology, Concordia University, Montreal, QC, Canada

⁵ Montreal Neurological Institute (MNI), McGill University, Montreal, QC, Canada

⁶ Centre for Research on Brain, Language and Music (CRBLM), McGill University, Montreal, QC, Canada

⁷ International Laboratory for Brain, Music and Sound Research (BRAMS), Montreal, QC, Canada

† Corresponding author: loonan.chauvette@cervo.ulaval.ca

Introduction

There is currently a growing interest in the development of auditory-to-vibrotactile sensory substitution and augmentation devices for in music composition (Armitage & Kia, 2013; Gunther & O’Modhrain, 2003), multimodal musical interfaces (Papetti & Saitis, 2018; Remache-Vinueza et al., 2021), assistive technologies (Flores Ramones & del-Rio-Guerra, 2023; Schulte et al., 2023), and virtual reality (Young et al., 2023). To date, most studies on explored auditory and vibrotactile perception of low-level acoustic features, such as frequency or intensity. However, effective development of auditory-tactile substitution and augmentation devices is now constrained by our limited understanding of vibrotactile perception and auditory-tactile integration of complex sounds containing multiple harmonics, spectral modulations, and amplitude variations. In hearing, these types of sound are apprehended via their unique timbre, but the question remains as to how the acoustic features that give rise to our auditory perception of timbre can be conveyed through the tactile modality. This study thus examines how spectral and temporal acoustic features associated with timbre are perceived across auditory, vibrotactile, and combined auditory–vibrotactile stimulation. Furthermore, there is also the questions of whether auditory and multimodal training can improve perceptual abilities for these timbre-related acoustic features and whether improved perceptual abilities lead to better outcomes in auditory scene analysis. As such, this research uses musicianship as a model for extensive auditory and multisensory training to explore how musical training influences unimodal and bimodal timbre perception abilities and its relationship to musical scene analysis.

Timbre is a multidimensional perceptual attribute shaped by the spectral and temporal acoustic properties of complex sounds. These same acoustic properties are also present in the vibrations encoded by our tactile system, leading current models of vibrotactile perception to explicitly mirror principles from timbre perception (Bensmaïa & Hollins, 2005; Russo et al., 2012). For instance, research the tactile perception often draw parallels between auditory timbre perception and tactile texture perception, which has given rise to the term “textural timbre” in the field of vibrotactile research (Yau et al., 2009). However, the analogies also go the other way around. Timbre semantics indeed often draws its vocabulary from crossmodal correspondences with other senses such as touch (e.g., texture, rough, smooth) (Saitis et al., 2020; Wallmark, 2019). These deep ties between auditory and vibrotactile perception are expected to some extent as acoustic waves are the same kind of signals as vibrations. Additionally, the way our auditory and tactile systems process these signals appears to be quite similar in many regards, as demonstrated by multiple findings that auditory and vibrotactile perception share several psychophysical characteristics (Merchel & Altinsoy, 2020; von Békésy, 1959). These parallels thus raise a fundamental question as to whether timbre is a modality-specific attribute of hearing, or if its representation can involve multimodal processing (Saitis et al., 2018). This question also holds practical

relevance, with a recent experiment revealing that the perception of some timbre dimensions can be modulated by vibrotactile stimulation (Verma et al., 2023), suggesting crossmodal contributions to the perceptual organization of timbre. However, key questions remain regarding 1) the relative effectiveness of auditory and tactile modalities in encoding acoustic correlates of timbre, 2) the multimodal integration of these acoustic features, 3) the influence of musicianship on bimodal timbre perception abilities, and 4) the relationship between multimodal timbre perception and musical scene analysis.

Method

Thirty non-musicians and twenty musicians with normal hearing and screened to ensure a clear perception of vibrotactile stimuli were included in the study. Non-musicians had ≤ 3 years of formal musical education outside of classes from the standard school curriculum, while musicians had ≥ 6 years of formal musical training. Discrimination thresholds were measured for 3 spectral acoustic features: number of harmonics, harmonic roll-off (related to spectral centroid and brightness), even-harmonic attenuation (associated with odd-to-even harmonic ratio and hollowness), and for 3 temporal features: attack time (related to attack quality), amplitude modulation depth, and amplitude modulation frequency (related to roughness or tremolo). Examples of these stimuli are presented in Fig. 1. Thresholds were estimated using an ABX paradigm with an adaptative staircase procedure. A subset of participants (22 non-musicians, 12 musicians) also completed the Music-in-Noise Test (MINT) (Coffey et al., 2019). For both tasks, stimuli were presented in auditory, tactile, and auditory-vibrotactile conditions via earphones and the Multichannel Vibrotactile Gloves (Chauvette et al., 2024) (Fig. 2).

Results

Participants could discriminate all timbre-related acoustic parameters using vibrotactile stimulations alone, demonstrating that these cues can be encoded by touch in addition to hearing. However, vibrotactile thresholds were significantly poorer for spectral features compared to auditory thresholds. In contrast, temporal features showed similar thresholds across both auditory and tactile modalities. Patterns of multimodal integration also appear to vary between groups, with non-musicians obtaining better auditory-tactile thresholds compared to auditory thresholds for temporal acoustic features, whereas musicians instead achieved better bimodal thresholds for spectral features. Overall, musicians generally outperformed non-musicians across all features and modality conditions, with the largest advantage in the auditory-vibrotactile condition, suggesting enhanced cross-modal integration. Finally, exploratory analyses showed that auditory and multimodal timbre perception abilities may be associated with music-in-noise performance, notably in musicians whose better MINT scores were partly mediated by their enhanced auditory timbre perception.

Conclusion

Our findings support the idea that touch, like hearing, can convey a range of timbre-related acoustic cues. While temporal acoustic features can easily be conveyed through touch, spectral features may require signal processing to be efficiently rendered as vibrotactile signals in assistive or musical devices. Musicianship was linked to enhanced timbre perception abilities, which seem to translate into improved musical scene analysis, suggesting that training may enhance perceptual benefits of such devices.

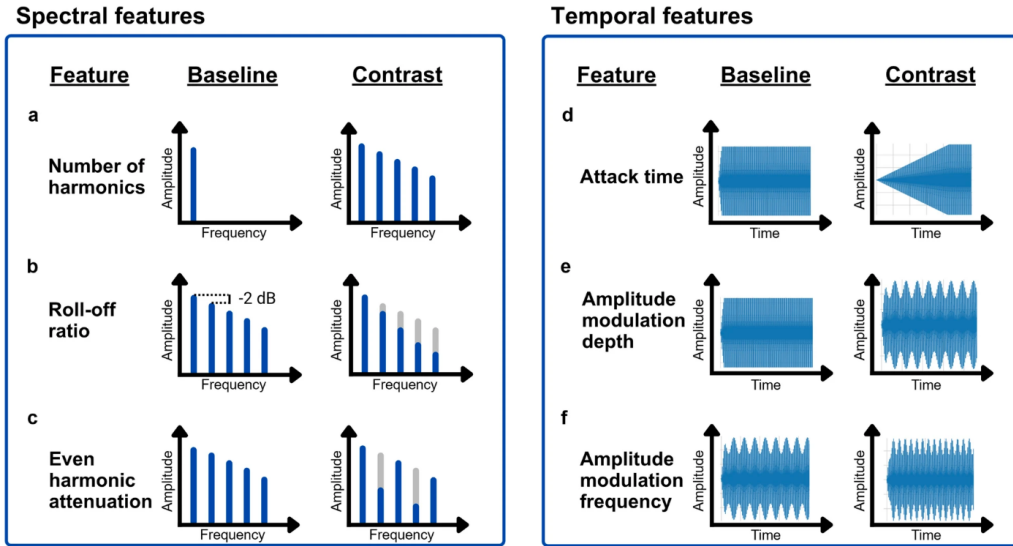


Figure 1: Example of stimuli used for the experiment.

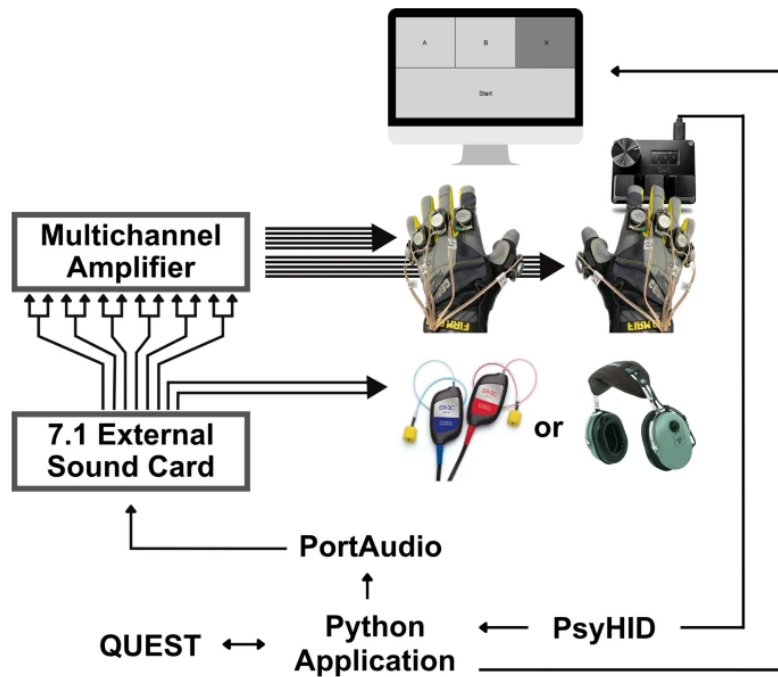


Figure 2: Experimental setup and apparatus used for the experiment.

References

- Armitage, J., & Kia, N. (2013, July 1). *Multimodal Music Composition*. Electronic Visualisation and the Arts (EVA 2013). <https://doi.org/10.14236/ewic/EVA2013.7>
- Bensmaïa, S., & Hollins, M. (2005). Pacinian representations of fine surface texture. *Perception & Psychophysics*, 67(5), 842–854. <https://doi.org/10.3758/bf03193537>

- Chauvette, L., Leprohon, É., Perron-Houle, L.-P., Pintat, V., Delnavaz, A., Voix, J., & Sharp, A. (2024). Multichannel Vibrotactile Glove: Validation of a new device designed to sense vibrations. *IEEE Transactions on Haptics*, 17(4), 913–923. <https://doi.org/10.1109/TOH.2024.3475740>
- Coffey, E. B. J., Arseneau-Bruneau, I., Zhang, X., & Zatorre, R. J. (2019). The Music-In-Noise Task (MINT): A Tool for Dissecting Complex Auditory Perception. *Frontiers in Neuroscience*, 13, 199. <https://doi.org/10.3389/fnins.2019.00199>
- Flores Ramones, A., & del-Rio-Guerra, M. S. (2023). Recent Developments in Haptic Devices Designed for Hearing-Impaired People: A Literature Review. *Sensors*, 23(6), 2968. <https://doi.org/10.3390/s23062968>
- Gunther, E., & O’Modhrain, S. (2003). Cutaneous Grooves: Composing for the Sense of Touch. *Journal of New Music Research*, 32(4), 369–381. <https://doi.org/10.1076/jnmr.32.4.369.18856>
- Merchel, S., & Altinsoy, M. E. (2020). Psychophysical comparison of the auditory and tactile perception: a survey. *Journal on Multimodal User Interfaces*, 14(3), 271–283. <https://doi.org/10.1007/s12193-020-00333-z>
- Papetti, S., & Saitis, C. (Eds.). (2018). *Musical Haptics*. Springer Nature. <https://doi.org/10.1007/978-3-319-58316-7>
- Remache-Vinueza, B., Trujillo-León, A., Zapata, M., Sarmiento-Ortiz, F., & Vidal-Verdú, F. (2021). Audio-Tactile Rendering: A Review on Technology and Methods to Convey Musical Information through the Sense of Touch. *Sensors*, 21(19), 6575. <https://doi.org/10.3390/s21196575>
- Russo, F. A., Ammirante, P., & Fels, D. I. (2012). Vibrotactile discrimination of musical timbre. *Journal of Experimental Psychology. Human Perception and Performance*, 38(4), 822–826. <https://doi.org/10.1037/a0029046>
- Saitis, C., Järveläinen, H., & Fritz, C. (2018). The Role of Haptic Cues in Musical Instrument Quality Perception. In S. Papetti & C. Saitis (Eds.), *Musical Haptics* (pp. 73–93). Springer International Publishing. https://doi.org/10.1007/978-3-319-58316-7_5
- Saitis, C., Weinzierl, S., von Kriegstein, K., Ystad, S., & Cuskey, C. (2020). Timbre semantics through the lens of crossmodal correspondences: A new way of asking old questions. *Acoustical Science and Technology*, 41(1), 365–368. <https://doi.org/10.1250/ast.41.365>
- Schulte, A., Marozeau, J., Ruhe, A., Büchner, A., Kral, A., & Innes-Brown, H. (2023). Improved speech intelligibility in the presence of congruent vibrotactile speech input. *Scientific Reports*, 13(1), 22657. <https://doi.org/10.1038/s41598-023-48893-w>
- Verma, T., Aker, S. C., & Marozeau, J. (2023). Effect of Vibrotactile Stimulation on Auditory Timbre Perception for Normal-Hearing Listeners and Cochlear-Implant Users. *Trends in Hearing*, 27, 23312165221138390. <https://doi.org/10.1177/23312165221138390>
- von Békésy, G. (1959). Similarities between hearing and skin sensations. *Psychological Review*, 66(1), 1–22. <https://doi.org/10.1037/h0046967>
- Wallmark, Z. (2019). Semantic Crosstalk in Timbre Perception. *Music & Science*, 2, 2059204319846617. <https://doi.org/10.1177/2059204319846617>
- Yau, J. M., Hollins, M., & Bensmaia, S. J. (2009). Textural timbre. *Communicative & Integrative Biology*, 2(4), 344–346. <https://doi.org/10.4161/cib.2.4.8551>
- Young, G. W., O’Dwyer, N., Vargas, M. F., Donnell, R. M., & Smolic, A. (2023). Feel the Music!—Audience Experiences of Audio–Tactile Feedback in a Novel Virtual Reality Volumetric Music Video. *Arts*, 12(4), Article 4. <https://doi.org/10.3390/arts12040156>

Empirical Study of the Semantics of Synthetic “Timbre-Textures”

Jason Noble^{1,6,†}, Francis Bourgoin², Takuto Fukuda^{3,6}, Caroline Traube^{4,6}, and Zachary Wallmark⁵

¹Département de musique, Université de Moncton, Moncton, NB, Canada

²Faculté des sciences de l'éducation, Université de Moncton, Moncton, NB, Canada

³Faculty of Music and Visual Arts, University of Pécs, Pécs, Hungary

⁴Faculté de Musique, Université de Montréal, Montréal, QC, Canada

⁵University of Oregon, Eugene, OR, USA

⁶Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT), Montréal, QC, Canada

[†] Corresponding author: jason.noble@umoncton.ca

Background

The recent surge of interest in timbre semantics has yielded a wellspring of new studies, contributing much to our understanding of a long-neglected subject. However, most studies have only examined isolated sounds, especially those of musical instruments. While the insights provided by such studies are highly valuable, the stimuli they use (usually single musical notes controlled for pitch, loudness, and duration) are far removed from ecologically valid musical experience. How can we assess experiences of semantic meaning in more musically stimulating contexts?

Our study begins to address this question in two ways: (1) inverting the usual task of timbre semantics studies so that rather than rating prepared stimuli on semantic scales our participants *produce* synthetic sounds based on semantic prompts; and (2) focusing on a musical aesthetic (sound mass) that emphasizes the total musical texture rather than its constituent parts (Noble & McAdams 2020). The advantage of (1) is that it frames the experiment as a musical exploration. The advantage of (2) is that it engages participants in a soundworld that more closely approximates real music. Our interactive approach draws on exciting precedents including Hayes et al (2022) and Wallmark & Nghiem (2019). It differs from them in adopting a granular synthesis-based approach in contrast to their FM synthesis-based approach. Granular synthesis is technically and conceptually important for sound mass, as demonstrated in the works of Iannis Xenakis, Barry Truax, and many later composers influenced by them.

Granular synthesis allows us to explore the grey area between timbre and texture (which, we believe, complements another grey area—much discussed by Spectralists—between timbre and harmony; e.g., Harvey 2000). It has been used to study the perception of naturalistic sound textures (e.g., wind, rain, crowd noises) by Schwarz and Caramiaux (2014); similar types of textures were also studied perceptually by McDermott and Simoncelli (2011). However, to the best of our knowledge, our study is the first to use granular synthesis to create explicitly *musical* textures in an experiment on semantics.

Methods

Stimuli and Interface

Participants manipulate four parameters of granular synthesis: (1) *spectral content* (ranging from “noisy” to “pure” and achieved by incrementally combining filtered bell sounds and pink noise in the production of grains), (2) *temporal density* (ranging from “rapid/continuous” to “slow/discontinuous” and achieved by covarying randomized grain length and grain duration within adjustable ranges), (3) *pitch ambitus* (the compass interval over which the grains are dispersed), and (4) *central pitch* (the lowest fundamental frequency of the grains). These parameters were chosen because they were hypothesized to be pertinent to semantic associations, and because they produce an overall range of soundworlds similar to classic examples of sound mass music by composers such as Xenakis, Ligeti, and Truax (Noble et al 2020).

The interface, developed in Max, consists of a control window with two 2-dimensional surfaces, each of which controls two of the aforementioned parameters. For example, in the first surface, the x axis maps onto spectral content and the y axis maps onto temporal density, so the position of a cursor on this surface determines values for both parameters simultaneously as the corresponding sound is produced in real time. There is also a button used by participants to advance to the next example, which is enabled only after they have interacted with both surfaces in each trial. A second window tracks the positions of the cursors along the four dimensions. Participants do not interact with this window; it is for display only.

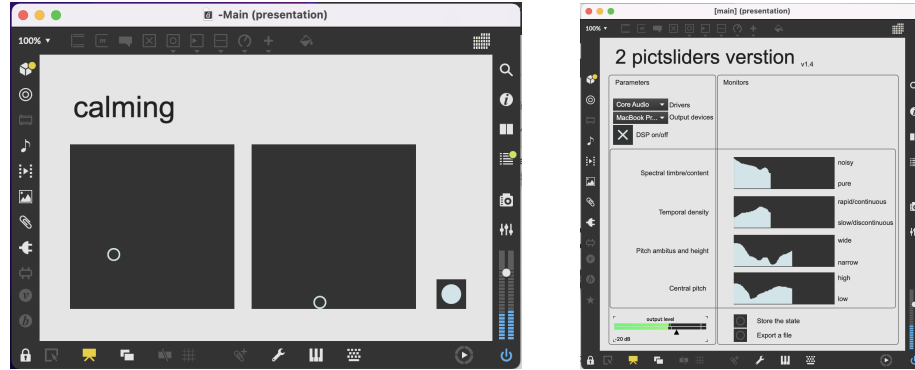


Figure 1: The control window and the display window of the interface used in the experiment.

Semantic Terms

Each trial presents a different semantic term, chosen from a list of 37 (Table 1). This list includes descriptive terms selected from the sound mass literature (e.g., ‘kaleidoscopic,’ ‘cloudy’) as well as classic terms of timbre semantics research (e.g., ‘bright,’ ‘rough’) and some common everyday affective terms (e.g., ‘overwhelming,’ ‘pleasant’). All terms were converted into adjectival form, even if they appeared in nominal form in the literature (e.g., ‘galaxy’ was converted to ‘galactic’). Three additional terms (‘liquid,’ ‘machine-like,’ ‘swarming’) were used in a practice phase.

Table 1: Semantic terms used as prompts in the experiment (alphabetical order).

Atmospheric	Complex	Fused	Kaleidoscopic	Rich	Thick
Bright	Dark	Galactic	Light	Rough	Thin
Brilliant	Dense	Heavy	Noisy	Round	Unpleasant
Busy	Dull	Hollow	Overwhelming	Sharp	Volatile
Calming	Exciting	Homogeneous	Pleasant	Smooth	Voluminous
Cloudy	Formless	Impenetrable	Pungent	Sweet	Warm
Cold					

Procedure

Following a practice phase of acclimatization to the sound synthesis task and experimental interface, participants receive the terms listed in Table 1 one at a time as prompts. For each trial, they move the cursor on each surface until they arrive at a timbre-texture that they feel effectively expresses the semantic prompt, at which point they press the button to advance to the next term. Captured data consist of the coordinates of the cursors on both surfaces (corresponding to values for each of the four parameters of granular synthesis) at the time that the advancement button is pressed.

Results

Participants

Data collection took place at Université de Montréal in 2023-2024. Twenty participants were recruited, consisting mostly of students and professors from the Musique numérique program at the university's Faculté de musique. Fluency in English was a requirement for participation. The participant pool on a whole had a high level of musical training and a high familiarity with electroacoustic sound production.

Global Data Analysis

Between-subject coherence was calculated using Interclass Correlation Coefficients (ICC) with the SPSS software. Globally, the four parameters or dimensions (spectral content, temporal density, pitch ambitus, central pitch) were all found to be significantly correlated between participants, indicating a generally coherent process of semantic association. For all dimensions, $p < .001$. The numbers in brackets represent 95% confidence intervals, using the standard alpha level of 0.05.

Dimension 1	Dimension 2	Dimension 3	Dimension 4
0.885 [0.823-0.932]	0.840 [0.754-0.906]	0.789 [0.677-0.876]	0.914 [0.869-0.950]

Figure 2: Global average values for each dimension. Shade indicates strength of intersubjective correlation, with darker green indicating stronger correlation.

Categorical Data Analysis

The semantic terms were grouped into categories by the researchers, following the precedent of Wallmark (2019). *Affect* terms emphasize emotional responses; *matter* terms emphasize comparison to physical properties; *cross-modal correspondences* terms emphasize comparison to experiences in non-auditory sense modalities; *metaphorical* terms emphasize more abstract relations, and *literal* terms emphasize concrete properties of sound. The terms were categorized as follows:

1. Affect (6): exciting, overwhelming, calming, pleasant, unpleasant, volatile
2. Matter (10): light, heavy, thick, thin, hollow, dense, voluminous, formless, round, impenetrable
3. Cross-modal correspondences (11): brilliant, dark, bright, smooth, sharp, warm, cold, pungent, sweet, rough, dull
4. Metaphorical (5): cloudy, atmospheric, kaleidoscopic, galactic, busy
5. Literal (5): rich, noisy, homogeneous, fused, complex

The data were analyzed according to these categories, providing a more nuanced picture of the process of association. Again, a highly coherent set of intersubjective correlations appears, with certain dimensions correlating more strongly with certain categories. Specifically, *affect* terms and *cross-modal correspondence* terms are most associated with dimension 4 (central pitch), *literal* terms and *matter* terms are most associated with dimension 1 (spectral content), and *metaphorical* terms are most associated with dimension 3 (pitch ambitus). All p values are $< .001$ unless otherwise indicated.

Categories	Dimension 1	Dimension 2	Dimension 3	Dimension 4
Affect	0.656 ^A [0.070-0.943]	0.840 [0.568-0.974]	0.809 [0.483-0.969]	0.947 [0.858-0.991]
Cross-modal correspondences	0.878 [0.743-0.961]	0.884 [0.755-0.963]	0.788 [0.551-0.932]	0.960 [0.912-0.988]
Literal	0.931 [0.795-0.992]	0.849 [0.554-0.982]	0.878 [0.638-0.985]	0.705 ^B [0.204-0.952]
Matter	0.913 [0.810-0.974]	0.671 ^C [0.279-0.902]	0.601 ^D [0.126-0.882]	0.893 [0.766-0.968]
Metaphorical	0.894 [0.686-0.987]	0.889 [0.671-0.987]	0.900 [0.704-0.988]	0.700 ^E [0.199-0.968]

^Ap=0.017 ^Bp=0.007 ^Cp=0.002 ^Dp=0.010 ^Ep=0.014

Figure 3: Average values for each dimension for each category. Shade indicates strength of intersubjective correlation, with darker green indicating stronger correlation.

Data Analysis for Individual Terms

Owing to the relatively small sample size, the robustness of the statistical accounts for individual terms is weaker, preventing us from drawing many strong conclusions about which specific configurations of parameters are associated with which semantic domains. We are in the process of collecting additional data to strengthen the statistical analysis for future publications. Nevertheless, a few correlations are strong enough to mention here. The participants' *hollow* and *smooth* sounds are characterized by low values on dimension 1 (pure spectral components with minimal noise content). Their *kaleidoscopic* sounds have high values on dimension 2 (rapid/continuous temporal density). Their *rough* and *busy* sounds have high values on dimension 3 (wide pitch ambitus). Their *bright* and *brilliant* sounds have high values on dimension 4 (high central pitch), and their *calming* sounds have low values on this dimension (low central pitch).

Conclusions and Future Research

To the best of our knowledge, this study represents the first use of interactive granular synthesis in the context of timbre semantics research. It supports the growing consensus that semantic associations with musical sound are not arbitrary and whimsical but exhibit a significant degree of intersubjective coherence.

The next step of this project is to collect data from another 20 participants, which we anticipate will enable us to produce robust statistical profiles for the individual semantic terms. This will also enable additional methods of statistical analysis such as Principal Component Analysis and Hierarchical Clustering Analysis, which may reveal underlying semantic dimensions not explicitly represented in the selected terms. Future experiments could expand the list of semantic terms studied and/or select different parameters of sound synthesis to explore different semantic associations and different sound worlds. Finally, this experiment could be conducted in different languages, to explore the potential effects of linguistic culture on semantic associations with musical sound.

This research has the potential to significantly expand the reach of the field by moving beyond isolated sounds to holistic musical textures, providing composers with an enriched understanding of semantic associations of timbre-textures that more closely approximate real musical experience.

References

- Harvey, J. (2000). Spectralism 1. *Contemporary Music Review* 19 (3): 11–14.
<https://doi.org/10.1080/07494460000640331>
- Hayes, B., Saitis, C., & Fazekas, G. (2022). Disembodied Timbres: a Study on Semantically Prompted FM Synthesis. PsyArXiv. <https://doi.org/10.17743/jaes.2022.0006>
- McDermott, J. H., & Simoncelli, E. P. (2011). Sound texture perception via statistics of the auditory periphery: evidence from sound synthesis. *Neuron*, 71(5), 926–940.
<https://doi.org/10.1016/j.neuron.2011.06.032>
- Noble, J., Thoret, E., Henry, M., & McAdams, S. (2020). Semantic dimensions of sound mass music. *Music Perception*, 38(2), 214–242. <https://doi.org/10.1525/mp.2020.38.2.214>
- Noble, J. & McAdams, S. (2020). Sound mass, auditory perception, and ‘post-tone’ music. *Journal of New Music Research* 49 (3): 231–251. DOI: [10.1080/09298215.2020.1749673](https://doi.org/10.1080/09298215.2020.1749673)
- Schwarz, Diemo & Caramiaux, Baptiste. (2014). Interactive Sound Texture Synthesis Through Semi-Automatic User Annotations. Proceedings of the *International Symposium on Computer Music Modeling and Retrieval*, pp. 372-392. 10.1007/978-3-319-12976-1_23.
- Wallmark, Z. (2019). A corpus analysis of timbre semantics in orchestration treatises. *Psychology of Music*, 47(4), 585–605. <https://doi.org/10.1177/0305735618768102>
- Wallmark, Z., Frank, R., & Nghiem, L. (2019). Creating novel tones from adjectives: An exploratory study using FM synthesis. *Psychomusicology: Music, Mind, and Brain* 29 (4): 188-199.
<https://doi.org/10.1037/pmu0000240>.

Gestural Control of Timbre With a Digital Musical Instrument

Gaël Moriceau^{1†}

¹ Laboratoire formes • ondes, Université de Montréal, Montréal, QC, Canada

[†] Corresponding author: gael.moriceau@umontreal.ca

Introduction

This presentation proposes an artistic investigation into how timbre can be shaped, transformed, and performed through embodied interaction with the T-Stick Digital Musical Instrument (DMI), combining a research-creation discussion with live demonstrations. Central aims include developing playing techniques, enabling real-time control of timbral and morphological attributes, and articulating a compositional approach rooted in the co-evolution of gesture, mapping, and sound structure. By situating timbre as primary compositional material, the project contributes to broader discussions on digital lutherie, embodied interaction, and the intertwined development of instruments, techniques, and musical works.

The T-Stick and Composition Process

The T-Stick is a cylindrical gestural controller developed at McGill University's Input Devices and Music Interaction Laboratory (IDMIL) since 2006 (Malloch & Wanderley, 2007). It integrates multisensor data streams, including accelerometers, gyroscopes, a pressure sensor, and a touch array, processed into high-level gestural descriptors reflecting whole-instrument manipulations such as shaking, brushing, and jabbing. The instrument supports an evolving repertoire and a growing performance practice (Meneses, 2022). In this project, the T-Stick is approached as a means for shaping timbre and morphology through embodied interaction, emphasizing how subtle variations in grip, orientation, contact, and energy produce continuously unfolding sonic behaviour.



Figure 1: Five T-Sticks Sopratinos (latest generation, built in 2025)

Composing for the T-Stick involves simultaneously assuming three interrelated roles: luthier, performer, and composer. As luthier, the instrument must be provided with a sound synthesis engine and mappings that connect sensor data to synthesis parameters. As performer, gestural explorations must be translated into reproducible playing techniques. As composer, the musical vocabulary formed by those techniques must be organized into a coherent structure. This process unfolds through iterative cycles of instrument building, exploratory improvisation, video documentation, mapping refinement, and musical structuring

(Figure 2). Improvisation sessions are regularly recorded in audio and video, creating documentation of performed gestures and produced sounds that allows for critical analysis and iterative refinement.

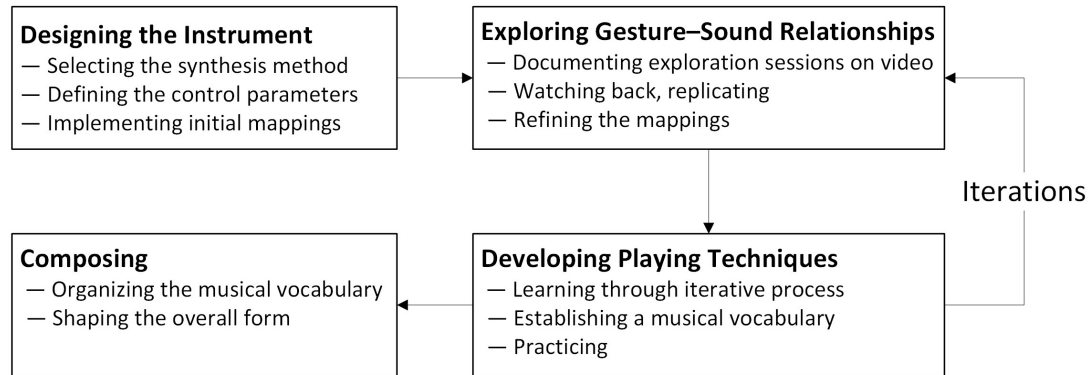


Figure 2: Composition Process.

Several examples of T-Stick instrumental gestures are available at: <https://vimeo.com/1014990304>.

Étude pour T-Stick et granulation

My first composition, *Étude pour T-Stick et granulation*, uses granulation to generate dense and evolving sound textures from a composite sound file. This file is built from three distinct timbral materials: a cymbal crash, varied percussions played backwards, and a low C played by a software synthesizer. Together they form a timbral space through which the performer can navigate in real time.

The inclination of the T-Stick determines the position within the sound file from which grains are extracted, offering an intuitive continuum from one timbral zone to another. The percentage of the instrument's surface covered by the hand controls grain density: gradually increasing contact area combined with pressure on the interface allows a transition from individual, very short noisy impulses to granular textures where spectral content emerges and sustained drones develop. This single gesture thus controls the morphological character of the sound, moving between what Schaeffer (2017) classifies as complex mass impulses and slowly evolving fixed-mass wefts (Figure 3).

Gestural control of reverberation is linked to the quantity of movement of the interface. Keeping the interface still after stopping a sound results in a short, dry decay. Conversely, moving the instrument just after releasing a sound extends its duration and adds reverberation colour. A large, quick rotational movement combined with high grain density creates a *Grosse note* in which all available timbres merge in a single dense sound object.

A video recording of the piece is available at: <https://youtu.be/l3HDQrApC50>.

Turbulences harmoniques

The second composition, *Turbulences harmoniques*, uses frequency modulation (FM) synthesis for direct control over spectral complexity. FM synthesis was chosen for the simplicity of its algorithm in generating complex spectra. Eleven touch sensors each trigger the same pitch with a distinct carrier-to-modulator (C/M) modulation ratio, and for any given pitch and ratio, the carrier frequency is calculated to

align the fundamental with the desired pitch (Truax, 1977). Four additional sensors control the octave, yielding a polyphonic instrument with a wide range of timbral possibilities (*Figure 4*).

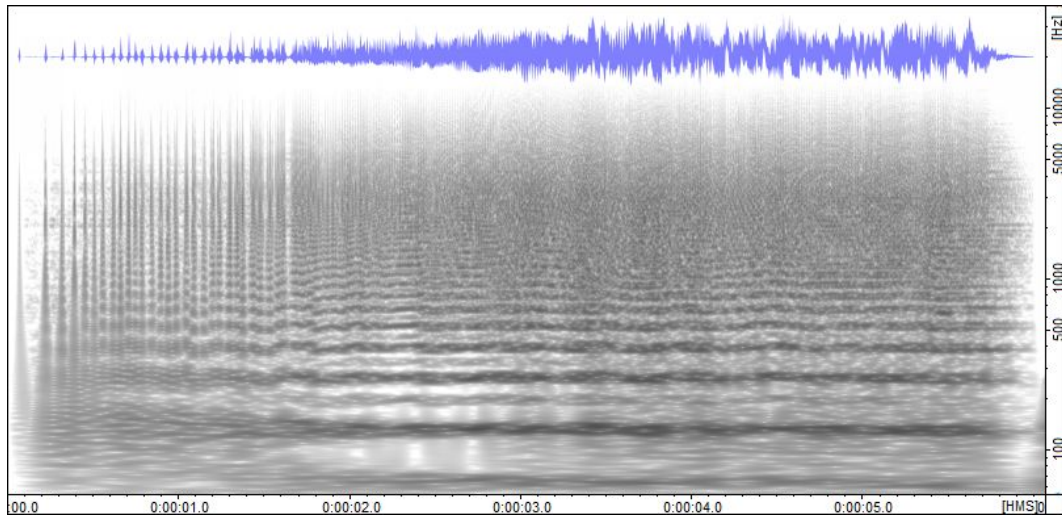


Figure 3: Transition from complex mass impulses to a slowly evolving fixed-mass weft.

The T-Stick’s inclination shapes both the dynamic envelope and timbral character: pointing upward produces a steep attack, short release, and high modulation index, resulting in a brighter, more complex spectrum; pointing downward yields a gentle attack, long release, and low modulation index, dulling the sound. This single gestural axis thus simultaneously controls articulation and spectral brightness, establishing a direct embodied link between the performer’s gesture and the timbral quality of the sound.

Several extended techniques emerged from improvisation sessions. A sustained rotation of the interface introduces irregular oscillations in the modulation index of each active voice, producing a slow timbral modulation analogous to a sweeping low-pass filter. The “sweep” technique consists of quickly triggering several notes at the same pitch in rapid succession, causing their partials to accumulate and merge into a unified, richer sound object (*Figure 5*). An unexpected discovery was the “stab”: a sudden strong acceleration along the instrument’s axis triggers a burst of partials, resulting in complex, noisy, and expressive sounds at the edge of control. These techniques illustrate how sustained practice with a single instrument can reveal unexpected timbral possibilities. A video recording of a live performance is available at: <https://vimeo.com/1083040417>.

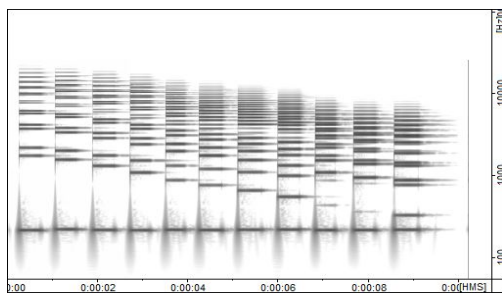


Figure 4: Short notes of brighter to darker timbre.

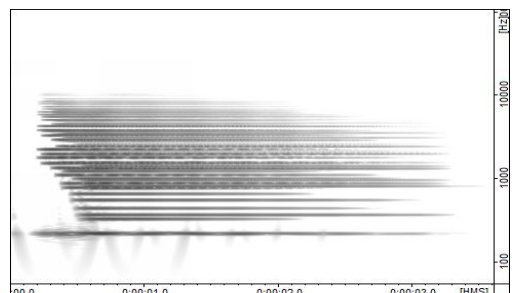


Figure 5: Fusion of timbres through a sweep gesture.

Configurable Filter Bank

The third sound engine, currently in development, uses a bank of narrow bandpass filters tuned to predetermined frequencies to define timbral presets. A flick of the wrist gesture, corresponding to a quick rotation around the interface's axis, triggers a switch between presets, dynamically reshaping the timbre in real time (*Figure 6*).

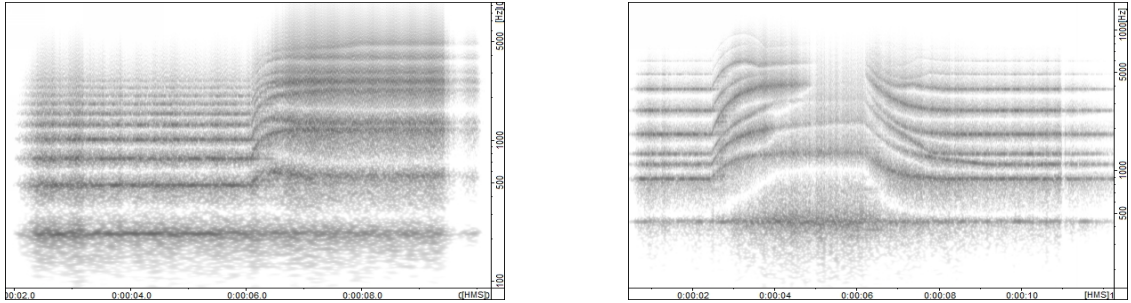


Figure 6: Two examples of timbre preset switch using a flick gesture.

Perspectives

These three studies share a common thread: timbre is not a fixed property of a DMI but an emergent result of the ongoing negotiation between gesture, mapping, and sound structure. Keeping mappings open-ended, preserving some indeterminacy between gesture and sound, has proven essential to sustaining timbral discovery over time. Future directions include composing for an ensemble of T-Sticks, where questions of timbral orchestration and blend come to the fore, and developing a notation system capable of representing the gestural nuances that shape timbre in performance.

References

- Malloch, J., & Wanderley, M. M. (2007). The T-Stick: From musical interface to musical instrument. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, 66–69.
- Meneses, E. A. L. (2022). *Iterative design in DMIs and AMIs: Expanding and embedding a high-level gesture vocabulary for T-Stick and GuitarAMI* (Doctoral dissertation). McGill University, Montréal.
- Schaeffer, P., North, C., & Dack, J. (2017). *Treatise on musical objects: An essay across disciplines*. University of California Press.
- Truax, B. (1977). Organizational techniques for C:M ratios in frequency modulation. *Computer Music Journal*, 1(4), 39–45.

Poster Session: Day 2

Towards Embodied and Inclusive Crossmodal Experiences of Timbre and Music: User Perspectives of a New Vibrotactile Assistive Device

Alex Bégin^{1,2}, Loonan Chauvette^{1,2†}, Jérémie Voix^{3,4}, Matthieu J. Guitton^{1,5}, and Andréanne Sharp^{1,2}

¹ CERVO Brain Research Centre, Quebec, QC, Canada

² School of Rehabilitation, Université Laval, Québec, QC, Canada

³ Montreal Neurological Institute (MNI), Université McGill, Montreal, Qc, Canada, H3A 2B4

⁴ École de technologie supérieure (ÉTS), Université du Québec, Montréal, QC, Canada, H3C 1K3

⁵ Département d'ophtalmologie et d'ORL-chirurgie cervico-faciale

† Corresponding author: loonan.chauvette@cervo.ulaval.ca

Introduction

Due to similarities between auditory and vibrotactile perception (Merchel & Altinsoy, 2020; von Békésy, 1959), vibrotactile devices have been proposed as tools for rendering music through touch to improve musical engagement and accessibility (Païsa et al., 2023; Papetti & Saitis, 2018; Siedenburg et al., 2024). Nonetheless, research on vibrotactile input to timbre perception is still emerging. Vibrotactile stimulation can allow musical instrument discrimination (Russo et al., 2012) and can contribute to representation of timbre in normal-hearing and cochlear implants user (Verma et al., 2023). Recently, the Multichannel Vibrotactile Gloves have been proposed as a validated research platform to study auditory and vibrotactile interactions (Chauvette et al., 2024). This device has been used to investigate multisensory perception of timbre-related acoustic parameters (Chauvette et al., 2025), but the qualitative experience of users and its social acceptability remain unexplored. Understanding how different user groups engage with vibrotactile music, that is how they describe the tactile feel of familiar songs and how they evaluate the usability of vibrotactile technologies, is essential to develop new devices that support inclusive and meaningful musical engagement. This study investigates the user experience of familiar songs presented via the Multichannel Vibrotactile Gloves and the technology's social acceptability in musicians and non-musicians with and without hearing loss.

Method

A total of 129 participants listened to familiar songs of their choice with the Multichannel Vibrotactile Gloves, with participants selecting songs across different genres (e.g., classical, jazz, hip-hop, metal). Songs were played through the gloves only, which contain a total of 10 voice-coil actuators located on the superior side of the proximal phalanges. The audio signal was unprocessed other than up mixing the stereo signal for multichannel presentation using VoiceMeeter Potato v.3.1.2 (VB-Audio, voicemeeter.com). With this setup, participants could both feel vibrations that fell in the skin's frequency range (around 40 Hz to 1000 Hz) and hear the sounds emitted from the vibrotactile actuators through air conduction. The sample included non-musicians (n = 86) and musicians (n = 43). Among these participants, 11 had a hearing loss (five hearing aid users, one cochlear implant user). Participant completed an adapted version of the *Quebec User Evaluation of Satisfaction with Assistive Technology* (Demers et al., 2000; Mao et al., 2010) composed of eight items ranking the technology's size, weight, safety, comfort, efficiency, sensation, relevance, and novelty on a scale from 1 to 5, and indicating which three of these eight items they judge to be the most important for them. Participants further answered four open-ended written questions regarding their experience:

- What are your general impressions of the technology?
- How would you judge the general experience of feeling sounds as vibrations on your hand?
- If you had a similar technology, would you use it? Why or why not?

- What could make the technology more attractive for regular usage?

A mixed method was used for analysis, using quantitative analysis for questionnaire scores, and a lexical and thematic analysis from the written responses.

Results

User ratings were generally positive, with average ratings above 80% for all items. Lexical and thematic analysis revealed three main themes:

1) Embodiment and immersion: Participants often described the experience as an embodied form of musical experience that felt richer or more detailed than purely auditory listening. Many participants further stated that the experience evoked similar feelings as listening to music in live settings or live performance for musicians.

« J'ai adoré ! C'était comme si j'étais à un concert de musique "live", ou même comme si j'étais moi-même la musicienne (dans le cas spécifique de la guitare). »

I loved it! It was as if I was at a "live" music concert, or as if I was myself the musician (specifically for the guitar).

- 30 y/o female bilateral hearing aid user and amateur dancer, listening to pop and folk songs

2) The complementary nature of sound and vibration: Users reported that the gloves provided additional information that helped to perceive nuances in musical structure and stylistic differences between genres.

« C'est très variable dépendant du style de musique écouté. Pour le pop et le prog », ça marchait vraiment bien. Pour les balades, c'était plus distrayant (car diffus, juste une vibration continue), pas de plus-value. Pour le jazz, c'était vraiment bien (même que je devinais la ligne de basse facilement alors qu'elle était peu audible), par contre au solo c'était diffus. »

It's highly variable depending on the style of music. For pop and prog, it works very well. For ballads, it's more distracting (more vague, like a continuous vibration), no real added value. For jazz, it's really good (I could even easily approximate the bassline even if it was barely audible), however solo felt diffuse.

- 50 y/o female amateur musician with normal hearing, listening to pop, prog rock, and jazz

3) Interest and curiosity: participants often described the multimodal music listening experience as intriguing, surprising or unlike typical listening experiences.

« Intéressant. Ça demande de la concentration et elle [la musique] varie beaucoup entre les oreilles et les vibrations, mais c'est impressionnant comme ça peut être complémentaire. »

Interesting. It requires concentration and it [the music] varies a lot between the ears and the vibrations, but it's impressive how it can be complementary.

- 45 y/o male musician with normal hearing, listening to pop music

Overall, the device was seen as useful and complementary to auditory listening, with explicit mentions of its relevance for people with hearing impairments. However, potential improvements were noted regarding connectivity, style, and portability.

Conclusion

Findings suggest that vibrotactile rendering can provide a rich, engaging and socially acceptable means of experiencing music. The main perceived benefits mostly related to affective dimensions like embodiment, immersion, and feeling closer to the music, supporting recent findings that vibrotactile input can boost musical engagement, notably through increasing feelings of “being part” of the music (Siedenburg et al., 2024). Participants also reported distinct tactile sensations across genres, drawing parallels to polyphonic models of timbre (Alluri & Toiviainen, 2010). This result is consistent with previously reported vibrotactile recognition of emotions in music (Sharp et al., 2019, 2020) and uses of vibrotactile stimulation to convey mood in film music (Branje et al., 2014; Mazzoni & Bryan-Kinns, 2016). Results also demonstrate the interest of users for novel musical experiences that still pertain to familiar songs. Design priorities for vibrotactile devices were also highlighted, such as ergonomics, portability, and connectivity.

References

- Alluri, V., & Toiviainen, P. (2010). Exploring perceptual and acoustical correlates of polyphonic timbre. *Music Perception*, 27(3), 223–242. <https://doi.org/10.1525/mp.2010.27.3.223>
- Branje, C., Nespoil, G., Russo, F., & Fels, D. I. (2014). The Effect of Vibrotactile Stimulation on the Emotional Response to Horror Films. *Computers in Entertainment*, 11(1), Article 4. <https://doi.org/10.1145/2543698.2543703>
- Chauvette, L., Leprohon, É., Perron-Houle, L.-P., Pintat, V., Delnavaz, A., Voix, J., & Sharp, A. (2024). Multichannel Vibrotactile Glove: Validation of a new device designed to sense vibrations. *IEEE Transactions on Haptics*, 17(4), 913–923. <https://doi.org/10.1109/TOH.2024.3475740>
- Chauvette, L., Sophie Grenier, A., Albouy, P., Coffey, E., Zatorre, R., & Sharp, A. (2025). Auditory and vibrotactile interactions in perception of timbre acoustic features. *Scientific Reports*, 15(1), 38055. <https://doi.org/10.1038/s41598-025-21908-4>
- Demers, L., Weiss-Lambrou, R., & Ska, B. (2000). Item Analysis of the Quebec User Evaluation of Satisfaction with Assistive Technology (QUEST). *Assistive Technology*, 12(2), 96–105. <https://doi.org/10.1080/10400435.2000.10132015>
- Mao, H.-F., Chen, W.-Y., Yao, G., Huang, S.-L., Lin, C.-C., & Huang, W.-N. W. (2010). Cross-cultural adaptation and validation of the Quebec User Evaluation of Satisfaction with Assistive Technology (QUEST 2.0): the development of the Taiwanese version. *Clinical Rehabilitation*, 24(5), 412–421. <https://doi.org/10.1177/0269215509347438>
- Mazzoni, A., & Bryan-Kinns, N. (2016). Mood Glove: A haptic wearable prototype system to enhance mood music in film. *Entertainment Computing*, 17, 9–17. <https://doi.org/10.1016/j.entcom.2016.06.002>
- Merchel, S., & Altinsoy, M. E. (2020). Psychophysical comparison of the auditory and tactile perception: a survey. *Journal on Multimodal User Interfaces*, 14(3), 271–283. <https://doi.org/10.1007/s12193-020-00333-z>
- Paisa, R., Nilsson, N. C., & Serafin, S. (2023). Tactile displays for auditory augmentation—A scoping review and reflections on music applications for hearing impaired users. *Frontiers in Computer Science*, 5, 1085539. <https://doi.org/10.3389/fcomp.2023.1085539>

- Papetti, S., & Saitis, C. (Eds.). (2018). *Musical Haptics*. Springer Nature. <https://doi.org/10.1007/978-3-319-58316-7>
- Russo, F. A., Ammirante, P., & Fels, D. I. (2012). Vibrotactile discrimination of musical timbre. *Journal of Experimental Psychology. Human Perception and Performance*, 38(4), 822–826. <https://doi.org/10.1037/a0029046>
- Sharp, A., Bacon, B. A., & Champoux, F. (2020). Enhanced tactile identification of musical emotion in the deaf. *Experimental Brain Research*, 238(5), 1229–1236. <https://doi.org/10.1007/s00221-020-05789-9>
- Sharp, A., Houde, M.-S., Bacon, B.-A., & Champoux, F. (2019). Musicians Show Better Auditory and Tactile Identification of Emotions in Music. *Frontiers in Psychology*, 10, 1976.
- Siedenburg, K., Bürgel, M., Özgür, E., Scheicht, C., & Töpken, S. (2024). Vibrotactile enhancement of musical engagement. *Scientific Reports*, 14(1), 7764. <https://doi.org/10.1038/s41598-024-57961-8>
- Verma, T., Aker, S. C., & Marozeau, J. (2023). Effect of Vibrotactile Stimulation on Auditory Timbre Perception for Normal-Hearing Listeners and Cochlear-Implant Users. *Trends in Hearing*, 27, 23312165221138390. <https://doi.org/10.1177/23312165221138390>
- von Békésy, G. (1959). Similarities between hearing and skin sensations. *Psychological Review*, 66(1), 1–22. <https://doi.org/10.1037/h0046967>

Play and Recognition in Mixed Instrumental Works

Nicholas Ma^{1†}

¹ Schulich School of Music, McGill University, Montréal, Québec, Canada

[†] Corresponding author: long.t.ma@mail.mcgill.ca

Introduction

Play has long occupied a significant position in the history of Western art music. From the lively bird-call imitations of Janequin's *Le chant des oiseaux* to the humor in Mozart's *A Musical Joke*, the mischievous narrative gestures of Strauss' *Till Eulenspiegel's Merry Pranks*, and the grotesque theatrical absurdity of Ligeti's *Le Grand Macabre*, musical play has consistently emerged through gesture, timbre, theatricality, and semiotic association. More recent works, including Alexander Schubert's *HELLO* (2014) and Ben Nobuto's *SERENITY 2.0* (2021), extend this lineage into digitally mediated compositional territory. Yet despite this historical continuity, the relationship between play, timbre, and gesture in contemporary mixed instrumental practice remains comparatively undertheorized, particularly in works shaped by digital media and internet culture.

This paper proposes an emerging taxonomy of mechanisms through which playful musical gesture operates in mixed instrumental works, drawing upon the semiological framework of Jean-Jacques Nattiez, the perceptual theories of Stephen McAdams, and the spectromorphological writings of Denis Smalley. Through a survey of recent works including Schubert's *HELLO*, Nobuto's *SERENITY 2.0*, and my own mixed-media composition *For You (but not really for You)*, I argue that play and humor in such works emerge through four mechanisms: morphological exaggeration, contextual incongruity, expectation subversion, and absurdity through technical excess.

Gesture, Timbre, and Cultural Recognition

A central premise of this paper is that playful musical meaning emerges through recognition operating across multiple perceptual levels. McAdams' model of auditory perception provides a useful starting point. As McAdams observes, "auditory grouping creates the basic building blocks of musical structure in the mind of the listener, which then give rise to emergent perceptual qualities that can acquire musical functions and engender meaning within the musical context" (McAdams, Goodchild, and Soden, 2022, Section 1.3). At the level of auditory grouping, listeners organize continuous acoustic input into coherent sonic events, such as fused electroacoustic-instrumental gestures or temporally unified sound attacks (e.g. glissandi, impacts, or sampled audio fragments).

These grouped events give rise to emergent perceptual properties including timbral quality, pitch contour, density, and temporal envelope. At this stage, a squeak or abrupt pitch bend is not yet perceived as comedic. Instead, it is perceived as possessing particular sonic behaviors such as instability, elasticity, or rupture, which in turn may suggest exaggerated motion or deviation from normative acoustic expectations.

Smalley's framework of gestural surrogacy (Smalley, 1997, p. 112) is relevant here, particularly his observation that listeners decode spectromorphological motion by inferring a gestural cause from a sound in reverse back to the proprioceptive and psychological experience of physical action. Spectromorphologies which have implied gestural causes that are exaggerated, impossible, or physically incongruous may predispose listeners toward comic or playful interpretations. However, this predisposition is not identical to comedic meaning. A rapidly descending glissando carries the embodied

trace of something falling or deflating but whether that quality registers as humorous depends on the context. Smalley identifies sounds requiring such references as transcontextual (Smalley, 1997, p. 110), where meaning is extrinsically linked to knowing the source. Humor in contemporary mixed music frequently operates through such intrinsic associations. The comedic function of a distorted laughing sound effect is not inherent to the acoustic signal; it depends on collective cultural conditioning within digitally mediated environments. The same acoustic event may be perceived as playful or neutral depending on what cultural webs of association the listener brings to it.

This is reinforced by Nattiez's semiological tripartition, where meaning is actively constructed in the esthetic process instead of being directly transmitted from composer to listener (Nattiez, 1990, pp. 16-18). Nattiez defines meaning as arising when an object is placed in relation to "a collection of other objects that belong to his or her experience of the world" (Nattiez, 1990, p. 17). This dynamic also aligns with contemporary meme theory. As Serpa and Santos (2022) argue, memes derive communicative power through their perceived triviality, functioning as mechanisms of cultural transmission through repetition, humor and shared recognition. In this sense, internet-informed timbral gestures operate semiotically, where their humor emerges through communal familiarity and the activation of shared interpretive frameworks rather than through acoustic properties alone. Hence for listeners raised within internet culture, certain timbres now function almost as topic signifiers. Distorted clipping, hyper-compressed bass hits, or popular audio samples such as the "bruh" sound effect possess socially learned meanings. Just as earlier musical traditions developed recognizable gestures associated with hunting or pastoralism, digitally mediated listening cultures have produced new timbral conventions associated with irony, absurdity, awkwardness, and play.

An Emerging Taxonomy of Playful Gesture

Drawing from these frameworks, I propose several overlapping mechanisms through which play may emerge in mixed instrumental works. These categories constitute an emerging taxonomy of playful gesture but are not exhaustive. Each category is analytically distinct and may multiple operate simultaneously in a single gesture.

1. Morphological Exaggeration

Playful gestures can exaggerate physical motion through abrupt contour, unstable pitch behavior, rapid bends, cartoon-like elasticity, or excessive dynamic articulation. Through Smalley's gestural surrogacy logic, such sonic events exceed what the corresponding physical gesture would normally generate. This tendency can already be observed historically in works such as *Till Eulenspiegel's Merry Pranks*, where the registrally wide horn motifs and theatrical orchestral gestures construct mischief through recognizable gestural caricature.

Many instances of morphological exaggeration appear in Alexander Schubert's audio-visual mixed work *HELLO*. In Schubert's recorded performance (Schubert, 2014), the passage between 1:08-2:00 features the ensemble playing along with the electronics to perform exaggerated rising gestures and frequent sudden interruptions. Between 1:49-1:52 these interruptions suddenly cease and turn into a stretched out sustained texture, temporarily suspending the previously fragmented gestural momentum before being interrupted by a repeatedly looped clapping gesture at 1:54-1:56. The exaggerated stretched out motion appears again from 1:56-2:01, before it is followed by the loud claps reappearing faster than before, cumulating in a winding-down motion. The cartoon-like elasticity and range of gestures all invoke embodied schemas of physical action pushed beyond their realistic limits. The resulting humor emerges through this excessive deformation of expected gestural trajectories.

2. Contextual Displacement

Humor often emerges when culturally familiar materials are displaced into unexpected aesthetic contexts (Carroll, 2013, pp. 19-25), for example Ligeti's car horn prelude in *Le Grand Macabre*, which functions as a semiotic collision between a mundane urban sound object being transplanted into operatic space.

Internet-informed mixed works frequently intensify this process. Viral meme audio fragments or compressed media artifacts become compositional material whose humor depends on their recognizability within shared digital culture. For example, Ben Nobuto's *SERENITY 2.0* cycles through fragments drawn from different cultural domains including 80s Japanese pop, YouTube vlogs, and Baroque music (Nobuto 2021). From a Nattiezian perspective, play here emerges esthetically through the listener's recognition of incompatible interpretive networks colliding simultaneously. Spoken audio samples also provide contextual incongruity. For example, at 4:26 in Nobuto's score-video recording (Score Follower, 2022), an audio sample of television personality James Corden saying "Reggie, do you have a question for our guest this evening?" starts the following section of music. While the sentence on its own may be neutral, a playful effect arises from semiotically displacing a late-night television aesthetic into the context of a contemporary concert work.

3. Expectation Subversion

A fundamental mechanism of humor involves the interruption or destabilization of expectation (Bevis, 2013, pp. 49-50). In musical contexts, this may occur formally, gesturally, or semiotically. In mixed-media works, abrupt cuts, awkward silences, or perceived technical failures often create tension through uncertainty (Kemper, 2023).

My own mixed-media composition *For You (but not really for You)* for saxophone quartet, electronics, and video deploys this mechanism as a structural device. The work opens with a dense montage of unsettling voices, fragmented media clips, and an accumulating spectromorphological texture. Approximately 20 seconds into the work, this texture cuts entirely to silence. The interruption falls at a point where the opening material has established sufficient momentum to make this sudden stop disorienting. This silence also creates uncertainty as to whether a technical malfunction has occurred in the electronics, thus heightening audience attention through expectation destabilization. After several seconds, this silence is broken by the opening hook of Rick Astley's 1987 single *Never Gonna Give You Up*. The comedic effect arises primarily through the radically disproportionate resolution of the preceding tension. Rather than receiving the anticipated continuation or cumulation of the opening material, the listener instead encounters an internet meme abruptly intruding into the formal discourse of the concert work. However, this subversion depends upon the extrinsic cultural recognition of the listener as the quotation itself possesses no intrinsic comedic acoustic property. Instead, its humorous function derives from the internet history of bait-and-switch deception through which the song became culturally associated with playful misdirection. In Smalley's terms, the sound functions through extrinsic rather than intrinsic meaning.

4. Absurdity Through Technical Excess

Using complex technology to produce trivial outcomes results in the listener experiencing the perceptual sense of a mismatch between the sophistication of means and the blandness of the source or result. Humor may emerge from the disproportion between these two fields.

The second movement of *For You (but not really for You)* engages this mechanism directly. Rapid saxophone bends are fused with cat meows through granular and spectral techniques to create hybrid

timbral objects sharing similar perceptual contours and motion trajectories. Drawing upon McAdams' framework, these fused auditory groupings generate emergent perceptual properties that listeners may interpret as playful, unstable, or absurd. Importantly, this results in a tension between compositional rigor and aesthetic triviality, as the advanced processing techniques are not deployed toward transcendence or abstraction, but just toward banal internet artifacts. This aligns with Serpa and Santos' (2022) observation on internet culture where distinctions between seriousness and triviality are collapsed.

Conclusion

Play in contemporary mixed instrumental composition increasingly emerges through interactions between gesture, timbre, perception, and culturally conditioned recognition systems shaped by digital media environments. By synthesizing the perceptual frameworks of McAdams, the semiology of Nattiez, and the spectromorphological theories of Smalley, this paper proposes an emerging taxonomy for understanding how playful gesture functions within internet-informed compositional practice. Such gestures often rely upon exaggerated morphology, contextual displacement, expectation subversion, and absurdity through technical excess.

As digitally mediated listening environments continue to reshape collective auditory experience, composers may increasingly engage with internet-derived gestural languages as legitimate compositional material. Future research may further examine how audiovisual presentation and video-based media reshape contemporary musical semiotics and modes of reception. This taxonomy points toward an expanded field of musical analysis where sound, image, and technological mediation are treated as co-constitutive elements of musical meaning rather than separate domains.

References

- Bevis, M. (2013). *Comedy: A very short introduction*. Oxford University Press.
<https://doi.org/10.1093/actrade/9780199601714.001.0001>
- Carroll, N. (2014). *Humour: A very short introduction* (First edition). Oxford University Press.
- Kemper, J. (2023). Glitch, the Post-digital Aesthetic of Failure and Twenty-First-Century Media. *European Journal of Cultural Studies*, 26(1), 47–63. <https://doi.org/10.1177/13675494211060537>
- Ma, N. (2026). *For You (but not really for You)* [Unpublished composition for saxophone quartet, electronics, and video].
- McAdams, S., Goodchild, M., & Soden, K. (2022). A Taxonomy of Orchestral Grouping Effects Derived from Principles of Auditory Perception. *Music Theory Online*, 28(3).
<https://mtosmt.org/issues/mto.22.28.3/mto.22.28.3.mcadams.html>
- Nattiez, J.-J. (1990). *Music and discourse: Toward a semiology of music*. Princeton University Press.
- Nobuto, B. (2021). *SERENITY 2.0*. Ben Nobuto. <https://bennobuto.com/serenity-20>
- Score Follower. (2022, July 11). *Ben Nobuto—SERENITY 2.0 [m/ score]* [Video recording].
<https://www.youtube.com/watch?v=ufGRThgUtiQ>
- Schubert, A. (2014, July 30). *Alexander Schubert—HELLO [Decoder Ensemble]* [Video recording].
<https://www.youtube.com/watch?v=fDD4S1QxbmU>
- Serpa, S., & Santos, A. (2026, March 31). *THE SOCIAL LIFE OF MEMES: CULTURE, COMMUNICATION, AND POWER IN THE GENERATIVE DIGITAL AGE*. ResearchGate.
<https://doi.org/10.5281/zenodo.122.12265>
- Smalley, D. (1997). Spectromorphology: Explaining sound-shapes. *Organised Sound*, 2(2), 107–126.
<https://doi.org/10.1017/S1355771897009059>

Measuring Consistency in Explainable Synthetic Timbre Annotators

Peter McCabe^{1†}, Alec Krsek¹, and Patrick J. Donnelly¹

¹College of Engineering, Oregon State University, Corvallis, Oregon, United States of America

[†] Corresponding author: mccabepe@oregonstate.edu

Understanding the perception of timbre is a fundamental problem in numerous disciplines including psychoacoustics (Carron et al., 2017), sound design (Seetharaman & Pardo, 2016), and machine listening (Kim et al., 2020). Unfortunately, due to the prohibitive cost of undertaking human listening studies, research in measuring timbral perception is severely limited. Using Large Language Models (LLMs) as an approximation for human data is a burgeoning practice across the scientific community as it allows practitioners to avoid expensive and time-consuming human trials, and an increasing interest has developed in how well LLMs “understand” concepts like timbre (Marjeh et al., 2024). However, LLM data are inherently biased and noisy: LLMs are black-box models designed for language modeling, and any desired data must also be describable by language commonly found in the LLM’s training data. Prior work has been restricted to comparing textual descriptions and their associations with sounds (Saitis & Siedenburg, 2023) or using audio-language models trained on significantly smaller datasets (Tian et al., 2025). However, even in the largest and most used audio-language datasets, such as WavCaps (Mei et al., 2024), timbral descriptors make up only a tiny fraction of the total vocabulary. Additionally, practitioners cannot explain the “reasoning” behind the data generated by LLMs. Motivated by these challenges, our study seeks to find a scalable method for utilizing the capabilities of LLMs for timbral analysis on real audio that generates easily explainable timbre ratings that simulate the behavior of human annotators.

We are inspired by the work of Saitis and Siedenburg (2023) in which they compare ChatGPT’s associations with instruments and timbral characteristics to those of humans (Reymore et al., 2023). In our work, rather than directly prompting LLMs to rate descriptions of sounds, we explore sourcing timbre ratings indirectly by prompting LLMs for Python code that automatically assigns timbre ratings to audio files. Each generated program uses combinations of familiar audio features to calculate timbre ratings in a given range or scale. These programs then act as synthetic annotators. Under this scheme, rather than relying only on linguistic associations between timbre and subjects, our method benefits from both the language and code data used to train LLMs. Additionally, because the LLM-generated programs rely on known audio features, the rating process is entirely explainable, controllable, and deterministic. We investigate if abstracting the rating mechanism away from direct textual associations and towards code based on explainable audio features affects its self-consistency and similarity to human judgments.

Data Generation Process

To answer these research questions, we employ the AudioCommons (AC) Timbral Characterisation Dataset (Pearce et al., 2019), as it contains human-annotated audio timbre ratings both on a bounded scale from 0-100 and binary assignment. As shown in Table 1, this dataset features eight words with human ratings: *Booming*, *Bright*, *Deep*, *Hard*, *Reverb*, *Rough*, *Sharp*, and *Warm*. We prompt both Claude and two versions of ChatGPT—GPT5 and the GPT4 family—to generate 10 functions each for both binary and continuous scales, leading to 60 distinct Python functions. We base our prompt on prior work (Saitis & Siedenburg, 2023) and adjust the scale and wording for binary or continuous as needed. The continuous prompt is: “Please write a Python function using *librosa* to rate how well the descriptions provided in the following (separated by semi-colons) apply to the sound of the audio clip on a scale from 0 to 100 where 0 is “does not describe at all”, 50 is “describes moderately well” and 100 is “describes extremely well”: 1) booming; 2) bright; 3) deep; 4) hard; 5) reverb; 6) rough; 7) sharp; 8) warm. The function should return a list or vector where each item corresponds to attributes 1-8 (in matched order).”

We standardize the functions’ output and correct minor mistakes that prevent code execution, but leave them otherwise unchanged. We process all audio samples, for a total of 1297 unique ratings.

Table 1: AudioCommons timbre ratings dataset summary covering eight different labels.

	Booming	Bright	Deep	Hard	Reverb	Rough	Sharp	Warm
Files	102	210	218	206	420	212	165	184
Raters	6	13	15	17	1	15	6	13
Scale	0-100	0-100	0-100	0-100	0/1	0-100	0-100	0-100

Statistical Analysis

First, we examine the LLM raters from each group. In Figure 1, we see that the LLM ratings have a wide range of correlations. Some axes are highly correlated across LLMs and measurement scales (*Booming*, *Bright*, *Deep*, *Rough*), some are highly correlated in the continuous scale and less correlated in the binary scale (*Sharp*, *Warm*), and others are not correlated or negatively correlated in either scale (*Hard*, *Reverb*). This indicates that some LLM raters are more similar than others, and the timbre dimension and measurement scale can affect consistency. To test the distributional similarity between ratings, we perform a Kolmogorov–Smirnov test, and find that despite some strong associations, even the highly correlated ratings are unlikely to have been drawn from the same distribution ($p < 0.001$). However, by comparing the ratios between the average inter-rater variance and the average inter-group variance across all audio samples, we find the variance between raters within the same group strictly dominates the cross-group variance. In other words, the discrepancies within LLM raters were on average greater than those between LLMs. Therefore, it is unlikely that the ratings produced by the different LLMs are significantly different enough from one another to inform future engineering or experimental design decisions.

Next, we compare synthetic raters with real humans. In Table 2, we show the correlation between the average LLM and human raters for both continuous and binary scales. To compare binary LLM and continuous human ratings, we discretize ratings with values greater than 50 to 1 and those below 50 to 0. Note that we do not alter the scales for the correlation calculation between continuous LLM ratings and binary human ratings present in the *Reverb* subset. We see that the same axes—apart from *Rough*—that had moderate-to-high correlations between continuous LLM ratings also correlate with human ratings to a similar extent. The axes with poor LLM-LLM correlation (*Hard*, *Reverb*) are also poorly correlated with human ratings. Additionally, we see that binary ratings do not strongly correlate with discretized human ratings. Even the well-correlated axes like *Bright* and *Sharp* drop significantly when comparing binary values. However, the only axis with naturally binary human ratings, *Reverb*, saw an increase in correlation when measuring on the same scale. This could imply that the discretization method we selected is biased. Supporting this observation, we measured the agreement between the binary LLM *Reverb* ratings and binned continuous LLM *Reverb* ratings and found only 20% agreement.

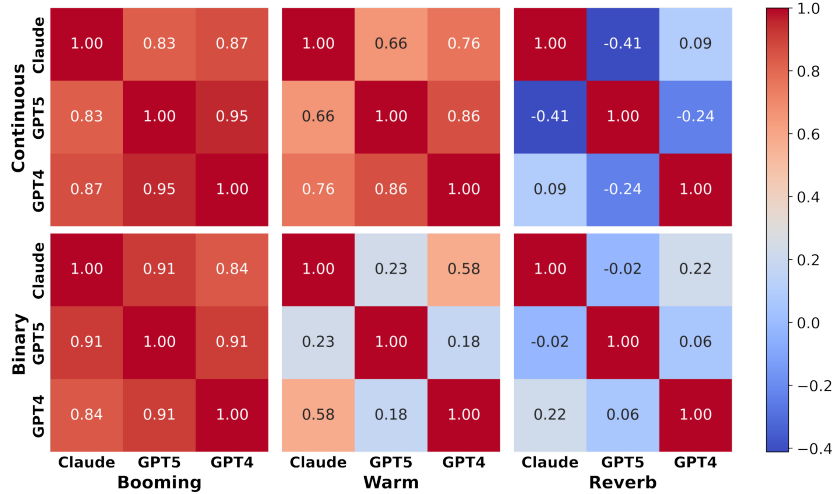


Figure 1: Selection of correlation matrices between continuous and binary LLM ratings.

Additionally, we found that the agreement between the human and binary LLM Reverb ratings decreased from 58% to 55% when using discretized continuous ratings. This supports a conclusion that the binary LLM ratings have a different underlying generating function than simply discretizing the continuous values at the natural midpoint. Therefore, if binary data is required, directly generating binary data may be preferred. Finally, we find that the LLM raters have strictly greater variances than human raters on all measured axes. We perform a Wilcoxon signed-rank test with respect to the continuous rater variances in both groups and find a statistically significant ($p < 0.001$) difference in the variance distributions, suggesting that the two groups do not come from the same underlying population, and are therefore not directly interchangeable. Notably, the Hard and Rough axes had both the lowest correlation between LLM and human ratings and the greatest difference in variances between rater groups. Based on this observation we calculated the correlation between cross-group correlations and variances ratios, but could not conclude a correlation between these two values was statistically significant ($r = 0.43$, $p = 0.167$).

Table 2: Correlations between LLM and human raters across the eight timbre labels. *The average Hard value was a constant 0, leading to an undefined correlation.

Scale	Booming	Bright	Deep	Hard	Reverb	Rough	Sharp	Warm
0-100	0.62	0.71	0.79	0.49	0.19	0.27	0.74	0.70
0/1	-0.15	0.26	0.50	N/A*	0.28	0.17	0.24	0.38

Conclusion

In this work, we introduce a method for efficiently utilizing LLMs for timbral analysis on real audio data. Our method produces synthetic raters that are entirely explainable and controllable. We measure rater consistency and find that raters vary more within LLMs than between LLMs, and that continuous ratings are highly correlated for most timbre dimensions. We measure similarity to human judgment across multiple rating scales, and find that continuous LLM timbre ratings exhibit at least weak correlation with human raters across all measured timbre dimensions, but have markedly higher variances than humans. We find that prompting annotators for binary outputs produces more human-like ratings than post-hoc

binning of continuous ratings, but we also see that binary values do not align with continuous ratings well. Overall, this method shows promise for the automatic rating of timbre qualia as a procedural, explainable process. We find LLM raters to be high-variance approximations of human raters capable of aligning with human perception to a moderate extent, even with no expert intervention. While the LLM raters were aligned unevenly across labels, our approach allows harmful or degenerate raters responsible for this high variance to be edited or removed to increase alignment with human ratings. As future work, we will explore expert and automatic methods to filter out these harmful raters while maintaining alignment with human data.

References

- Carron, M., Rotureau, T., Dubois, F., Misdariis, N., & Susini, P. (2017). Speaking About Sounds: A Tool For Communication On Sound Features. *J. of Design Research*, 15(2), 85.
<https://doi.org/10.1504/JDR.2017.086749>
- Kim, K. L., Lee, J., Kum, S., Park, C. L., & Nam, J. (2020). Semantic Tagging of Singing Voices in Popular Music Recordings. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 28, 1656–1668. <https://doi.org/10.1109/TASLP.2020.2993893>
- Marjeh, R., Sucholutsky, I., Van Rijn, P., Jacoby, N., & Griffiths, T. L. (2024). Large Language Models Predict Human Sensory Judgments Across Six Modalities. *Scientific Reports*, 14(1), 21445.
<https://doi.org/10.1038/s41598-024-72071-1>
- Mei, X., Meng, C., Liu, H., Kong, Q., Ko, T., Zhao, C., Plumbley, M. D., Zou, Y., & Wang, W. (2024). WavCaps: A ChatGPT-Assisted Weakly-Labelled Audio Captioning Dataset for Audio-Multimodal Research. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 32, 3339–3354.
<https://doi.org/10.1109/taslp.2024.3419446>
- Pearce, A., Brookes, T., & Mason, R. (2019). *Timbral Characterisation Tool v0.2 Development Dataset* [Dataset]. Zenodo. <https://doi.org/10.5281/zenodo.2545496>
- Tian, H., Lattner, S., & Saitis, C. (2025). Assessing the Alignment of Audio Representations With Timbre Similarity Ratings. In *Proceedings of the First MiniCon Conference*. ISMIR 2025 Hybrid Conference. https://ismir2025program.ismir.net/poster_283.html
- Reymore, L., Noble, J., Saitis, C., Traube, C., & Wallmark, Z. (2023). Timbre Semantic Associations Vary Both Between and Within Instruments: An Empirical Study Incorporating Register and Pitch Height. *Music Perception*, 40(3), 253–274. <https://doi.org/10.1525/mp.2023.40.3.253>
- Saitis, C., & Siedenburg, K. (2023). When ChatGPT Talks Timbre. *Proceedings of the Third International Conference on Timbre*, 73–77.
- Seetharaman, P., & Pardo, B. (2016). Audealize: Crowdsourced Audio Production Tools. *Journal of the Audio Engineering Society*, 64(9), 683–695. <https://doi.org/10.17743/jaes.2016.0037>

Learning Vocal Timbre Similarity for Personalized Singer Matching and Song Recommendation

Jiaxing Yu¹, Yingping Cao¹, Ziyi Huang¹, Yuhang Jin¹, Shuyu Li¹, and Kejun Zhang^{1,2†}

¹ College of Computer Science and Technology, Zhejiang University, Hangzhou, Zhejiang, China

² Innovation Center of Yangtze River Delta, Zhejiang University, Jiaxing, Zhejiang, China

[†] Corresponding author: zhangkejun@zju.edu.cn

Abstract

Vocal timbre is an important yet underexplored factor in personalized music recommendation, particularly for singing practice and vocal performance. Most existing systems focus on listening behavior, metadata, or audio-content descriptors aligned with genre and style (Tzanetakis & Cook, 2002; Schedl, 2019). For listeners these signals are well matched to the task, but for users who sing along, rehearse, or perform, the relevant unit of similarity is the human voice rather than the production style of the track. Two singers may share a genre yet have voices that are incompatible in tessitura or spectral identity, and a learner whose own timbre resembles a particular professional singer is more likely to find that singer's repertoire ergonomically singable. Vocal timbre therefore offers a natural axis for personalization that current systems overlook.

This paper proposes a timbre-aware recommendation framework that identifies singers with similar vocal timbre to a target user and suggests suitable songs to improve singing comfort and performance quality. The framework is grounded in the Popular Hooks dataset (Wu et al., 2024), a multimodal corpus of 38,694 popular musical hooks, totaling about 249 hours of music. Each hook is paired with synchronized MIDI, audio extracted from a YouTube music video, lyrics, and the video itself. Hooks are restricted to a 4/4 time signature and span the major popular-music genres (pop, rock, electronic, R&B, dance, country, hip-hop/rap, jazz, classical, others), several languages, and diverse artist nationalities. Annotations cover artist identity, tonality, structure, genre, region, and emotion (Russell's 4Q). The system extracts timbre-related acoustic features from user recordings and the dataset, including spectral and formant-based descriptors that capture resonance structure such as the singer's formant (Sundberg, 1987), and deep speaker embeddings transferred from speaker recognition (Snyder et al., 2018; Desplanques et al., 2020), and computes normalized similarity scores to enable interpretable singer matching. Based on the similarity ranking, top-matched singers are selected, and their song catalogs are analyzed using vocal range, pitch distribution, and melodic difficulty (Eerola et al., 2006) to recommend vocally compatible tracks. By integrating vocal timbre modeling with singer similarity analysis and song suitability estimation, this work bridges music information retrieval and vocal pedagogy, offering practical support for karaoke, singing practice, and personalized music discovery.

The contributions of this paper are: (a) a timbre-aware singer matching and song recommendation framework formulated on the Popular Hooks dataset (Wu et al., 2024); (b) a hybrid vocal-timbre representation that fuses interpretable spectral and formant descriptors with deep speaker embeddings; and (c) a song-suitability ranking layer that uses vocal range, pitch distribution, and melodic difficulty to translate singer similarity into actionable, vocally comfortable recommendations.

References

Desplanques, B., Thienpondt, J., & Demuyne, K. (2020). ECAPA-TDNN: Emphasized channel attention, propagation and aggregation in TDNN based speaker verification. arXiv preprint arXiv:2005.07143.

- Eerola, T., Himberg, T., Toivainen, P., & Louhivuori, J. (2006). Perceived complexity of western and African folk melodies by western and African listeners. *Psychology of Music*, 34(3), 337–371.
- Schedl, M. (2019). Deep learning in music recommendation systems. *Frontiers in Applied Mathematics and Statistics*, 5, 44.
- Snyder, D., Garcia-Romero, D., Sell, G., Povey, D., & Khudanpur, S. (2018, April). X-vectors: Robust DNN embeddings for speaker recognition. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 5329–5333). IEEE.
- Sundberg, J. (1987). *The Science of the Singing Voice*. DeKalb, IL: Northern Illinois University Press.
- Tzanetakis, G., & Cook, P. (2002). Musical genre classification of audio signals. *IEEE Transactions on speech and audio processing*, 10(5), 293-302.
- Wu, X., Wang, J., Yu, J., Zhang, T., & Zhang, K. (2024, July). Popular hooks: A multimodal dataset of musical hooks for music understanding and generation. In *2024 IEEE International Conference on Multimedia and Expo Workshops (ICMEW)* (pp. 1–6). IEEE.

Investigating "Timbre Tagging" of Pitch Structure in a Statistical Learning Paradigm

Adam Andrew Garrow^{1†} and Marcus Thomas Pearce¹

¹ School of Electronic Engineering and Computer Science, Queen Mary University of London, London, United Kingdom

[†] Corresponding author: a.a.garrow@qmul.ac.uk

Introduction

Statistical learning (SL) is a crucial mechanism in auditory and music perception (Pearce, 2018) and is linked to style-dependent means by which listeners internalise and deploy structural musical knowledge (Huron, 2006). While SL research has focused primarily on pitch structure, there is also evidence for the role of acoustic cues such as timbre. Listeners can use timbral cues to accurately recognise music genres in sub-second timespans before other structural cues are available (Gjerdingen & Perrott, 2008). Instrumental timbre cues can also direct genre-specific retrieval of music expectations, e.g., cadences (Vuvan & Hughes, 2019), tonal hierarchies (Vuvan & Hughes, 2021), and goodness-of-fit ratings for chordal sequences (Jimenez et al., 2025).

Van Hedger et al. (2022) found instrument-like sounds attenuated SL of pitch regularities from an unfamiliar musical scale versus artificial timbres, as the former may cue existing melodic expectations that interfere with internalisation of target pitch structures. However, it is unclear how extensively such pattern learning is “tagged” to the timbre for learning and if new timbre cues assist or interfere with knowledge retrieval about learned pitch structures.

Aims and Hypotheses

We aimed to determine if timbres from an auditory learning context aid retrieval of statistical pitch structures and if new timbre cues interfere with this process; we used listeners’ accuracy in identifying novel and familiar pitch sequences as a measure of SL.

We expected (1) a two-way interaction between sequence-wise (mis)match to the learned timbre and within-trial timbre (mis)match between target and foil sequences such that pitch-based SL would be most enhanced for target sequences in a learned timbre and foil sequences in a new timbre (the “supportive” condition), and most attenuated in the opposite configuration (the “disruptive” condition); and (2) listeners would be underconfident relative to their true pitch sequence recognition accuracy in supportive timbral contexts, and overconfident relative to their accuracy in disruptive timbral contexts. We also recorded listeners’ musical perceptual abilities to see if this further impacted timbral effects on SL.

Method

Eighty online participants heard an extended and statistically structured monophonic pitch sequence in one of two artificial “reference” timbres, differing in brightness. Participants then completed a 96-trial recognition task – each trial played a short target sequence sharing the long exposure sequence’s statistical structure and a short foil sequence without this structure. Participants chose the sequence they thought had a pitch pattern closest to the exposure sequence. All stimuli used 5 scale degrees of the Bohlen-Pierce scale. Sequences paired in a trial could each either (1) match or contrast the reference timbre of the exposure sequence, and (2) play in the same timbre or contrasting timbres, forming four possible within-trial conditions (counterbalanced across trials).

Participants also rated their confidence in the correctness of their answer in each trial as “guess”, “somewhat confident”, or “extremely confident”. Participants’ musical perceptual abilities were scored via the perceptual abilities (PA) subscale of the Goldsmiths Musical Sophistication Index (Müllensiefen et al., 2014).

To analyse our data, we fitted a series of binomial generalised linear mixed models (GLMMs) and, where relevant, performed planned contrasts of select estimated marginal means (EMMs) sourced from these models’ fits (Table 1).

Table 1: The series of GLMMs and corresponding planned contrasts of select EMMs.

Full Model	Planned Contrast
$accuracy \sim target_timbre * match_in_trial + (1 participant) + (1 sequence_1) + (1 sequence_2)$	Supportive vs. pooled matching-timbre trials
	Disruptive vs. pooled matching-timbre trials
$accuracy \sim target_timbre * match_in_trial * perceptual_abilities + (1 participant) + (1 sequence_1) + (1 sequence_2)$	N/A
$accuracy \sim target_timbre * match_in_trial * confidence + (1 participant)$	Supportive vs. pooled matching-timbre trials
	Disruptive vs. pooled matching-timbre trials

Results

Per-participant task accuracies by trial type are shown in Figure 1, with a further breakdown of these performances by response confidence in Figure 2.

The first GLMM showed a significant main effect of target and foils’ within-trial match/contrast ($\beta = -0.059$, $SE = 0.073$, $z = -8.061$, $p < .001$) and a two-way interaction between this factor and whether the target’s timbre matched the reference timbre ($\beta = 0.885$, $SE = 0.107$, $z = 8.293$, $p < .001$) on recognition accuracy. As expected, planned contrasts revealed that compared to trials whose sequences had matching timbres, accuracy was greater in trials with the supportive timbral configuration ($OR = 1.34$, $SE = 0.091$, $z = 4.347$, $p < .001$) and worse in trials with the disruptive timbral configuration ($OR = 0.555$, $SE = 0.035$, $z = -9.398$, $p < .001$).

The second GLMM showed no significant main effect of participants’ self-reported perceptual abilities ($\beta = 0.019$, $SE = 0.071$, $z = 1.126$, $p = .260$) nor any interactive effects with timbral factors of trials ($\beta = -4.70 \times 10^3$ to 5.94×10^5 , $SE = 1.15 \times 10^2$ to 1.68×10^2 , $z = -0.395$ to 0.004 , $p \geq .693$) on overall task accuracy, indicating listeners’ performances were influenced by associative timbral cues via a generalised mechanism.

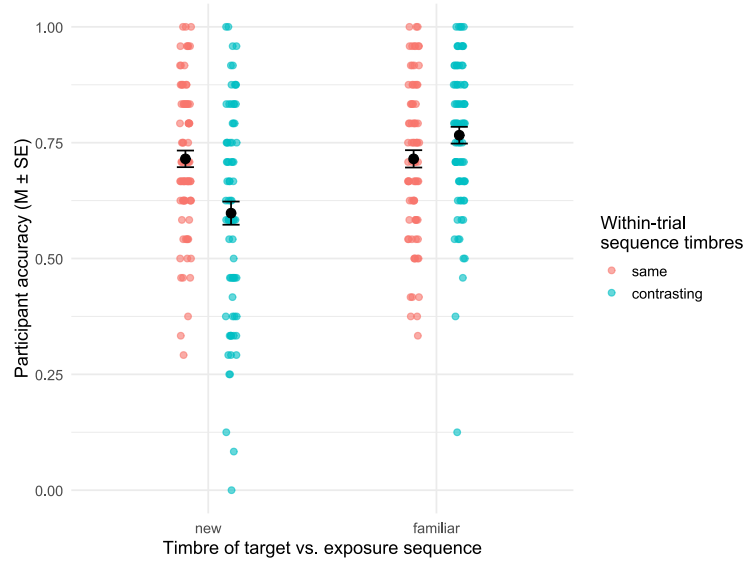


Figure 1: Accuracy of participant responses by trial type factors.

In the final GLMM, all significant timbral effects on accuracy from the first model were maintained and there was a three-way interaction effect between the target's timbre, within-trial timbral (mis)match, and linear effect of self-reported confidence ($\beta = 0.842$, $SE = 0.276$, $z = 3.045$, $p = .002$) on trial accuracy. Planned contrasts further revealed that compared to the matching-timbres trials baseline, accuracy was mildly improved in supportive trials at the moderate confidence level ($OR = 1.23$, $SE = 0.107$, $z = 2.408$, Holm-adjusted $p = .048$) and strongly attenuated in disruptive trials at moderate ($OR = 0.523$, $SE = 0.042$, $z = -8.116$, Holm-adjusted $p < .001$) and high ($OR = 0.321$, $SE = 0.057$, $z = -6.353$, Holm-adjusted $p < .001$) confidence levels.

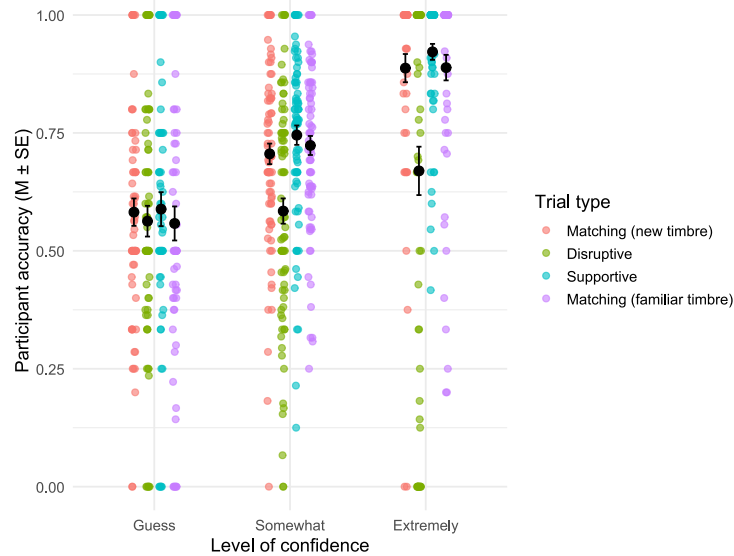


Figure 2: Accuracy of participant responses by trial type factors and self-reported confidence.

Discussion

This study investigated whether learned associations between timbre and statistically structured pitch sequences affected retrieval of learned pitch structure. We found mildly supportive and strongly disruptive effects of timbral cues on recognition memory performance when retrieving a target pitch structure in timbrally contrasting listening contexts, with these effects occurring irrespective of listeners' existing perceived musical perceptual abilities. Finally, timbral cues partially altered listeners' confidence versus their true accuracy in the task.

These findings reveal a previously unseen role of timbre in distinguishing “in-style” and “out-of-style” pitch sequences via top-down association, acting as a fallback mechanism listeners may choose to utilise if regularities explicitly identified in the pitch structure alone are deemed insufficient for making a correct decision. This study provides an important foundation upon which to probe how listeners cognitively compartmentalise a wide variety of musical styles and genres experienced in everyday listening.

Future work should expand upon the current study by testing the strength of timbral tags when listeners must learn and identify multiple pitch structures simultaneously. Adapting our recognition task paradigm to incorporate real musical instruments would also clarify the ecological validity of our findings.

References

- Gjerdingen, R. O., & Perrott, D. (2008). Scanning the dial: The rapid recognition of music genres. *Journal of New Music Research*, 37(2), 93–100. <https://doi.org/10.1080/09298210802479268>
- Huron, D. B. (2006). Genres, schemas, and firewalls. In *Sweet anticipation: Music and the psychology of expectation* (pp. 203–218). MIT Press. <https://doi.org/10.7551/mitpress/6575.003.0013>
- Jimenez, I., Kuusi, T., Ojala, J., Czedik-Eysenberg, I., Harrison, P., & Chander, A. (2025). Effect of timbre on goodness-of-fit ratings of short chord sequences. *Music Perception*, 42(5), 481–504. <https://doi.org/10.1525/mp.2025.2397481>
- Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). The musicality of non-musicians: An index for assessing musical sophistication in the general population. *PLoS ONE*, 9(2), Article e89642. <https://dx.plos.org/10.1371/journal.pone.0089642>
- Pearce, M. T. (2018). Statistical learning and probabilistic prediction in music cognition: Mechanisms of stylistic enculturation. *Annals of the New York Academy of Sciences*, 1423(1), 378–395. <https://doi.org/10.1111/nyas.13654>
- Van Hedger, S. C., Johnsrude, I. S., & Batterink, L. J. (2022). Musical instrument familiarity affects statistical learning of tone sequences. *Cognition*, 218, Article e104949. <https://doi.org/10.1016/j.cognition.2021.104949>
- Vuvan, D. T., & Hughes, B. (2019). Musical style affects the strength of harmonic expectancy. *Music & Science*, 2, 1–9. <https://doi.org/10.1177/2059204318816066>
- Vuvan, D. T., & Hughes, B. (2021). Probe tone paradigm reveals less differentiated tonal hierarchy in rock music. *Music Perception*, 38(5), 425–434. <https://doi.org/10.1525/mp.2021.38.5.425>

Everyday Sounds as Emotional Catalysts: A Research-Creation Study in Contemporary Music

Jonas Regnier^{1†}, Yifan Huang¹, and Stephen McAdams¹

¹ CIRMMT, Schulich School of Music, McGill University, Montréal, Québec, Canada

[†] Corresponding author: jonas.regnier@mail.mcgill.ca

Everyday sounds such as cooking noises and baby cries are deeply embedded in lived experience. These sounds are not only acoustically rich but also highly identifiable as sources, often carrying strong emotional associations. Early pioneers such as Pierre Schaeffer (1948) explored their potential in *musique concrète*. Later soundscape composers, including Barry Truax, emphasized the emotional power of environmental sounds to evoke memory (Truax, 1995). Despite their artistic use, relatively little empirical work has examined whether, and how, everyday sounds enhance emotional experience in contemporary mixed music. This study addresses this gap through a three-phase research-creation approach integrating qualitative inquiry, compositional practice, and perceptual testing.

Specifically, we asked: (1) which everyday sounds listeners associate with emotionally meaningful experiences; (2) how such sounds can be transformed into compositional materials in a mixed work for trumpet and live electronics; and (3) whether listeners perceive and feel the intended emotional effects when these sounds are embedded in a contemporary musical context.

In the first phase, semi-structured interviews were conducted with 12 participants with and without training in contemporary music. Participants were asked to recall everyday sounds associated with emotional meaning. Interviews were initially guided by five emotion categories: sadness, love, anger, fear, and joy. These five categories were used as an elicitation scaffold rather than as an exhaustive model of musical emotion, helping participants recall emotionally meaningful situations while still allowing them to describe mixed or complex affective experiences. The interview protocol included ten questions concerning sounds that were part of participants' everyday lives, sounds carrying strong emotional meaning, and sounds they associated with love, joy, anger, sadness, and fear. The interviews were analyzed qualitatively by emotion category, and the resulting sound list informed the subsequent compositional phase.

Emotion	Sounds
Love	Raking sounds/leaves, cooking sounds (boiling water), squeaky floorboard (footsteps on wood), rubbing sounds (velvet), sea wave sounds, purring cat
Joy	Bird sounds, animal sounds (purring cat), applause, river, splash in pool, instrumental sounds (piano)
Anger	Industrial noises (percussion), shattered glass, slammed door, human screams, whistling kettle
Sadness	Rain, crunching snow, rubbing sounds (fabric), wind chimes
Fear	Tire noises (car braking), nail scraping, thunder, alarm, car engine, bike sounds, silence

Figure 1: List of everyday sounds derived from participant interviews and organized by emotion that were used in the fixed media electronics of “Wistful Fragments”.

The sounds listed in Figure 1 were either sounds mentioned by several listeners or sounds with individual associations. They were used in the subsequent composition stage. Composer Jonas Regnier created an 11-minute work for trumpet and live electronics (surround 7.1), *Wistful Fragments* (2024) in which everyday sounds were recorded and integrated into both the electronic tape and were mimicked in some trumpet parts. The binaural reduction audio recording of this composition can be heard following this link: <https://soundcloud.com/jonas-regnier-113013741/wistful-fragments-2024>.

The electronics of the composition contained a fixed media part (composed on ProTools and Logic) using all everyday sounds mentioned above, as well as live effects (triggered through a MaxMSP patch) to help the trumpet blend better spatially and timbrally with the fixed media part.

Wistful Fragments is formally divided into five domains, each associated with an intended emotion (Figure 2). Each domain is further sub-divided into a minimum of two sections, which are not successive in the final form. Across sections, the relationship between the trumpet and everyday sounds was systematically varied. In some sections, the trumpet was written in an emotionally neutral manner; in others, it aligned with the emotional intention of the everyday sounds using established musical cues for emotional expression (Juslin & Laukka, 2003). This design enabled comparisons between emotional effects of everyday sounds when not reinforced by emotionally congruent trumpet writing and those arising from their interaction with acoustic instrumental cues.

Time interval	Emotion	Everyday life Sounds	Trumpet emotional cues	Transition
0'00 - 1'27"	Sadness (neutral)	Rain, metallic percussion	/	/
1'28" - 2'41"	Love (neutral)	Ocean waves, raking leaves, boiling water, whispers	/	No
2'42" - 3'37"	Anger	Slamming door, shattering glass, screaming voice	Short loud sounds, screaming, fast gliss.	No
3'38" - 4'42"	Love	Purring cat, boiling water	Air noise, legato medium-pitched notes	Yes
4'43" - 5'44"	Sadness	Crunching snow, rubbed fabric sounds	Multiphonics, singing with gliss.	Yes
5'45" - 6'45"	Fear (neutral)	Thunder, silences	/	No
6'46" - 7'13"	Joy (neutral)	Bird sounds, distant applause	/	No
7'14" - 7'59"	Anger (neutral)	Industrial noises, whistling kettle	/	Yes
8'00" - 9'00"	Fear	Alarm, traffic sounds, tire noises, sirens	Fast vibrato/trills, fast repeated notes	Yes
9'01" - 10'12"	Joy	Bird sounds, schoolyard, kids playing, piano, river sounds	Staccato high notes	Yes
10'13" - end	Sadness (neutral)	Rain, wind chimes	Down gliss, air noise	Yes

Figure 2: General form of *Wistful Fragments*, with corresponding sections, emotions, everyday life sounds, trumpet emotional cues, and potential presence of a transition with the previous section.

The final phase examined two aspects of listeners' responses: whether the intended emotions were perceived at the section level, and how strongly the piece induced felt emotional intensity over time. Forty-one general-audience participants in Montreal, with no exposure to the piece, took part in a perceptual experiment. After passing a hearing test, participants heard a binaural reduction of the piece on headphones in a sound-proof controlled environment. They were first asked to continuously rate their felt emotional intensity on a slider from no emotional intensity to very high emotional intensity throughout the entire piece. They then evaluated their perceived emotions for each section using the five target categories, with multiple ratings making it possible to capture mixed and shifted emotions.

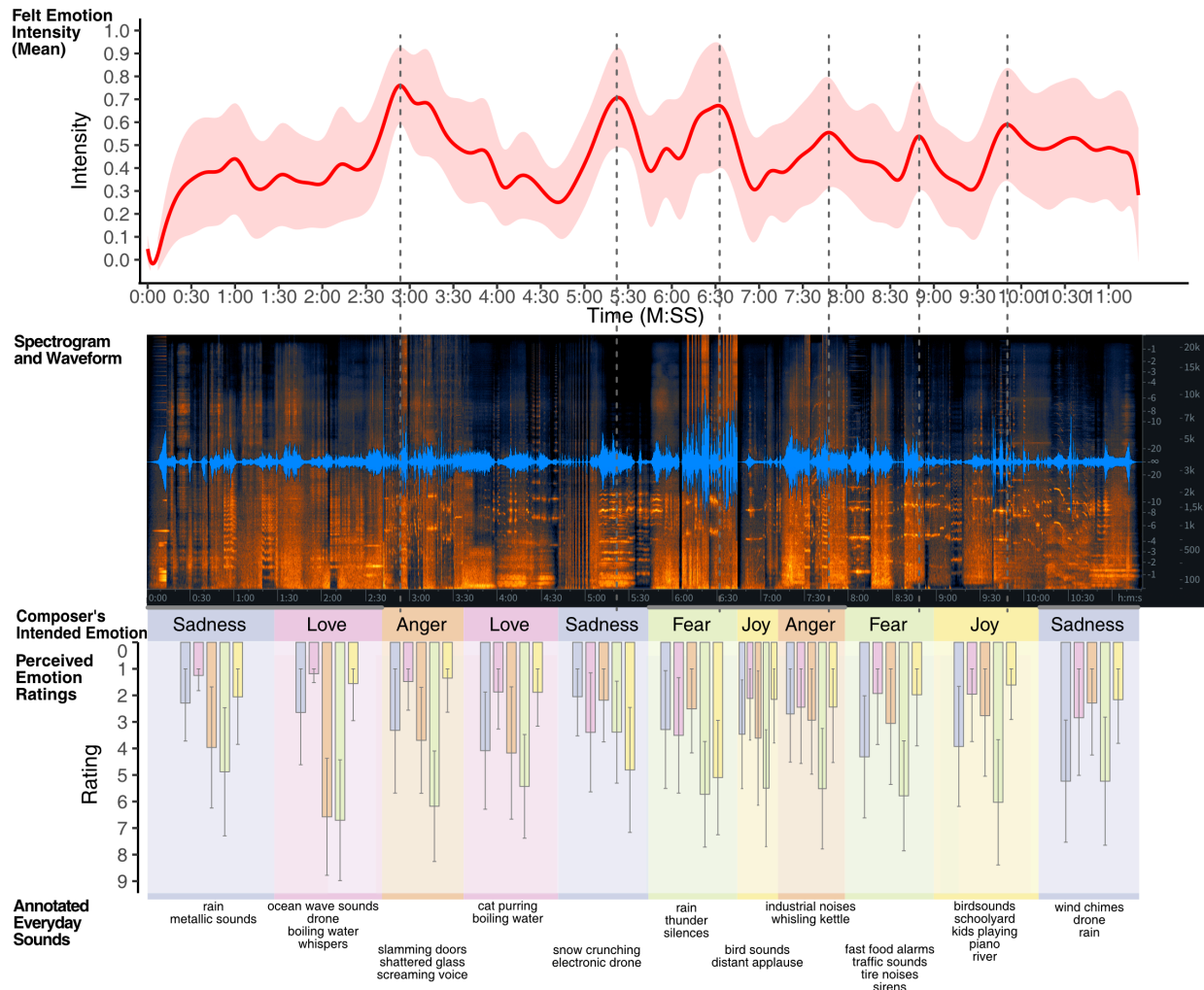


Figure 3: Listener responses to *Wistful Fragments*. Top: time-varying felt emotional intensity averaged across participants. Shaded areas indicate the standard deviation. Middle: waveform and spectrogram of the binaural reduction used in the experiment. Bottom: mean perceived-emotion ratings for each emotionally intended excerpt, with corresponding everyday sound materials.

Continuous felt-intensity ratings were smoothed and treated as time-varying functions for functional data analysis (Figure 3, top panel). The highest intensity peak occurred during a human shout, the only explicitly vocal sound in the piece. This suggests that vocal source identity, abrupt onset, and high-arousal associations may have jointly contributed to listeners' heightened emotional engagement. Strong

responses were also observed in sections where everyday sounds were not reinforced by emotionally congruent trumpet writing, indicating that everyday sound materials can contribute substantially to emotional engagement even without congruent instrumental reinforcement. Notably, a quieter segment in which the trumpet writing was emotionally aligned with the everyday sound material elicited intensity levels comparable to those of a louder segment with relatively neutral trumpet writing. This comparison suggests that emotional congruence between instrumental and everyday sound materials may enhance felt intensity beyond what would be predicted by sound level alone.

Section-level perceived-emotion ratings showed clearer correspondence with the intended categories for negative emotions than for positive ones. Fear was prominent across multiple sections, possibly reflecting listeners' unfamiliarity with contemporary music or its association with fearful situations in film and television. Sadness sometimes overlapped with joy or love, suggesting that listeners may have aesthetically reinterpreted certain everyday sounds within the musical context rather than mapping them onto a single basic-emotion category, perhaps evoking nostalgia.

In post-experiment questionnaires, approximately 70% of participants reported that the everyday sounds enhanced their emotional experience "very much" or "extremely," and no participant reported that they diminished the experience. These responses suggest that the everyday sounds were experienced as active contributors to the emotional character of the piece.

These findings suggest that everyday sounds in contemporary mixed music function not merely as referential sound objects, but as emotionally active timbral materials whose effects may depend on their timbre qualities, source identity, and interaction with instrumental writing. By linking interviews, composition, and perceptual testing, this study offers a research-creation model for examining how lived sonic experience can be transformed into musical emotion.

Acknowledgements

This project was supported by a CIRMMT Student Award and a grant from the Canadian Social Sciences and Humanities Research Council. We are grateful to Bennett Smith for programming the experimental interface used in the perceptual experiment.

References

- Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*, 129(5), 770–814.
<https://doi.org/10.1037/0033-2909.129.5.770>
- Schaeffer, P. (1948). *Cinq études de bruits*. Paris, Studio d'Essai, RTF.
- Truax, B. (1995). Sound in Context: Soundscape Research and Composition at Simon Fraser University. *Proceedings of the International Computer Music Conference (ICMC)*, Banff, AB.
<http://hdl.handle.net/2027/spo.bbp2372.1995.001>

Grounding Musical Intimacy in Timbre: Exploring a Conceptual Model of Intimate Listening

Yifan Huang^{1†}, Martha de Francisco¹, and Stephen McAdams¹
¹ Schulich School of Music, McGill University, Canada, Québec, Canada

[†]Corresponding author: yifan.huang3@mail.mcgill.ca

Musical intimacy has long been a recurring concern in discussions of listening experience. It is widely discussed across musical practices and listening contexts, yet it remains an underdefined concept in music research. Existing scholarship has approached musical intimacy primarily through qualitative description and critical interpretation of practitioners' intuitive knowledge (e.g., Accornero, 2021; Kraugerud, 2020), with timbre often treated as an implicit dimension rather than a systematically examined perceptual factor. This study addresses this gap by developing a practitioner-informed conceptual framework of musical intimacy, with a particular focus on the role of timbre in intimacy perception.

Method. The study comprises individual 1.5–2-hour semi-structured interviews, each incorporating open-ended questions and a brief listening-rating task, with 14 experienced musicians and audio professionals spanning composition, conducting, performance, and recording/production. Participants reflected on their listening experiences and on the strategies they use to create or shape musical intimacy in their own practice. Within each session, participants evaluated a curated set of seven musical excerpts drawn from classical chamber music, contemporary electroacoustic music, electronic music, traditional Chinese guqin solo music, jazz vocal performance, black metal, and pop music. Excerpts were rated on a 5-point Likert scale of perceived intimacy, after which participants verbalized the sonic and musical features informing their judgments. The present preliminary analysis draws on the first seven completed interviews: one contemporary composer, one contemporary ensemble conductor, one pianist, one classical guitarist, and three audio engineers/producers. The remaining seven interviews are scheduled for completion before the conference and will be incorporated into the final presentation. Transcripts were analyzed using reflexive thematic analysis, with iterative coding between the interview corpus and the listening-task verbalizations.

Preliminary findings. Across the seven excerpts, participants showed the strongest convergence at the two perceptual extremes: the excerpt perceived as most intimate, and the excerpt perceived as least intimate. The excerpt most consistently rated as intimate was a jazz female vocal performance. Participants attributed its intimacy to the convergence of several features: an extremely close vocal perspective produced by close miking, a warm and soft vocal timbre, slow and flexible timing, emotionally moving lyrics, and the perceived sincerity of the singer's delivery. By contrast, the excerpt most consistently rated as not intimate was a black metal excerpt. It was described as fast, dense, intense, and outwardly projected, with harder instrumental attacks and lyrical content marked by strongly negative imagery. Several interviewees contrasted intimacy with music that felt aggressive and extroverted. This contrast was not reducible to aesthetic dislike: some participants appreciated aggressive or metal-related styles but still described this particular excerpt as largely unrelated to intimacy. These responses suggest that the opposite of musical intimacy may be music perceived as forceful, aggressive, and externally projected.

A working conceptual model. Thematic analysis on open-ended questions points toward a core-to-context model that distinguishes what musical intimacy is from how it is perceptually realized and contextually interpreted. The model is organized into four nested layers, moving outward from a conceptual core to broader musical and listener-based conditions, as shown in Figure 1.

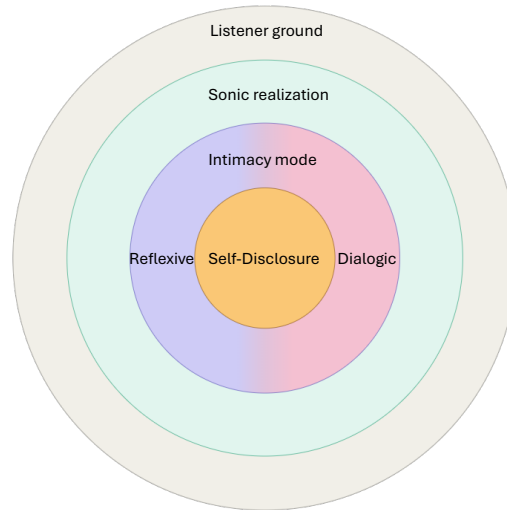


Figure 1: Preliminary conceptual model of Musical Intimacy

The innermost layer is a perceived disposition of self-disclosure on the part of the music or musician. Across roles, participants repeatedly associated intimacy with authenticity, genuineness, vulnerability, openness, and proximity or closeness. These qualities appeared to function as a recurring conceptual core for musical intimacy: for sound to be heard as intimate, it often had to be perceived as giving access to something personal, sincere, and relatively unguarded. One interviewee articulated this as “musicians first discovering oneself, then being willing to express it”.

A second layer concerns the listener’s mode of engagement with this perceived self-disclosure. Two modes recurred across interviews. The first may be described as **reflexive intimacy**, in which the music seems to enter the listener’s inner world and occasions an introspective dialogue. In this mode, participants described the feeling that the music was narrating the listener’s own story, surfacing latent feelings, or creating space for self-reflection. The second may be described as **dialogic intimacy**, in which the music or performer appears to address the listener directly, as if telling a story in a sincere one-on-one exchange. This dialogic mode resembles what some participants associated with confession, direct address, or the sense of a performer communicating personally rather than publicly. Notably, participants suggested that such one-on-one intimacy can occur even within concerts attended by many people. This indicates that musical intimacy may not be reducible to physically private settings; rather, it may depend on the perceived quality of address and relation enacted through sound.

A third layer concerns the musical and sonic realization of intimacy. Participants articulated this layer primarily through the interrelated handling of **time** and **space**, where timbre enhances the perception of intimacy in both dimensions. Time was repeatedly described as a consequential dimension in the creation and perception of intimacy. At the macro-temporal level, participants emphasized rhythmic freedom, restraint, natural pacing, and the deliberate “leaving of space to think,” a temporal approach that allows music to tell its story at an appropriate pace. Time, however, was not discussed only at the level of rhythm, tempo, or phrasing. Participants also referred to micro-temporal qualities at the level of timbre itself: softer attacks and more extended or lingering releases were often described as strengthening intimacy. Practitioners described concrete strategies operating across these temporal scales, such as a

guitarist altering plucking angle to change attack without changing dynamic level, or a conductor softening rehearsal language to allow performers greater expressive latitude in phrasing.

Space was articulated as the question of where the story is being told and how the listener is positioned in relation to it. Participants, particularly engineers and producers, emphasized perceived acoustic proximity. Cozy, smaller perceived rooms with close and warm sound were frequently described as supporting intimacy. However, participants did not reduce intimacy to small rooms or dry sound alone. Several noted that intimacy can also arise in larger perceived spaces, provided that sufficient proximal sonic detail is preserved. The presence of breath, friction, embodied noise also helps with intimate feelings. Engineers described strategies of microphone placement and selection oriented toward capturing precisely these proximal timbral details.

The outermost layer is the listener ground: cultural background, musical training, genre familiarity, personal associations, and linguistic access modulate how the inner layers are perceived. Several participants noted that music with deep biographical significance can acquire an intimacy that no other performance can substitute for. Thus, our model suggests that musical intimacy arises from shareable sonic and musical affordances at its core, modulated by the personal, cultural, and autobiographical meanings of the listener's ground.

Contribution. This study models musical intimacy as a perceived quality of self-disclosure, realized through practitioners' shaping of time, space, and timbre and interpreted through different modes of listener engagement. By organizing practitioner-derived descriptions into a conceptual framework, it offers a conceptual foundation for subsequent psychoacoustic and perceptual experiments. Future studies may examine how targeted manipulations of timbre give rise to different modes of perceived musical intimacy.

Acknowledgements

This research was supported by a grant from the Canadian Social Sciences and Humanities Research Council.

References

- Accornero, G. (2021). Shocking intimacy: Techniques, technologies, and aesthetics of amplification in Clara Iannotta's "Intent on Resurrection." *Sound Stage Screen*, 1(2), Article 2. <https://doi.org/10.54103/sss15487>
- Kraugerud, E. (2020). *Come closer: Acousmatic intimacy in popular music sound* [Doctoral thesis, University of Oslo]. <https://www.duo.uio.no/handle/10852/83699>

Creating a Timbral Monster: The Superclarinet of Martino's Triple Concerto

Zac Brennan^{1†}

¹ Music and Dance, University of Massachusetts, Amherst, Massachusetts, United States of America

[†] Corresponding author: zbrennan@umass.edu

Donald Martino devised a “Superclarinet” for his Triple Concerto for B-flat Clarinet, Bass Clarinet, and Contrabass Clarinet and Orchestra (1978), defining this instrument as “a six-octave gargantuan who would use the concerto as a world in which to romp and play with the ‘Superfriends’” (Martino 1978). The Superclarinet is a “transformation” of unaltered, “innate personalities” and given its voice through extremely detailed, precise notation and performance directions. It is each soloist’s articulation that brings forth their timbral characterization, with pitch and rhythm serving to further give shape to Martino’s gargantuan. The concerto’s three Superfriends are also divided on timbral lines, scored together by instrument family: strings, winds, and percussion and keyboard groups. Martino carefully controls what scholars have described as timbral “excess,” as he notates and describes the attack and release of each passage, while highlighting tonal qualities resulting from the distinct material construction and dynamic capacity of the soprano, bass, and contrabass clarinet (Van Elferen, 2021; Vélez, 2021). Ultimately, the listener perceives the Superclarinet as a multifaceted, but still unified phenomenon through timbre, as “aesthetic attention itself” (Dolan 2013).

At the heart of the Triple Concerto are the interactions among the individual soloists and Superfriends. Each of the three clarinets, respectively described by Martino as “elegant,” “poetic,” and “obstreperous,” is instructed to maintain its assigned character, even as a larger composite instrument is constructed in their aggregate through seamless hocketing and careful phrase construction. Reveling in the timbral variety of this monstrous instrument, Martino observed that “the naturalness of their interaction would melt their differences.” The timbral “gargantuan” at once distinguishes, rather than shies away from, its inherent contrasts, all the while presenting a nuanced whole. Martino’s understanding of the varied colors of each clarinet across wide pitch and dynamic ranges empowers him to highlight their differences, while always reinforcing a broader sense of coalition at the precise moment where the central illusion of monstrous unity might break.

This paper explores the role of timbre in constructing the Superclarinet and the Superfriends in two specific passages of the Triple Concerto: the first movement’s opening and its cadenza. I analyze how Martino fabricates the composite instrument of three clarinets and negotiates the most virtuosic and importantly unaccompanied passages, providing a framework for soloistic feats. In the Triple Concerto, Martino deploys timbre in its most archetypal fashion, as “musical difference,” yet crucially here, also as a gargantuan agent of musical coalescence. (Van Elferen, 2021).

Martino introduces his “six-octave gargantuan” in a manner befitting a cinematic monster. A triple forte percussion roll accompanied by a low register arm cluster in the piano decrescendos and evolves into recognizable pitch with the introduction of the bass clarinet playing a low D-natural near silently before rising up a twelfth following a hairpin, the first appendage of the Superclarinet to make an appearance. Steve Mackey described this as reminiscent of “some primordial creature emerging from the Big Bang,” adding, “Never before or since have I heard such a liquid connection between two notes—more of a cross-fade than an ordinary legato. As a clarinet player himself, Don knew how the 2nd note was contained in the first note and coaxed it out under cover of the decaying thunderclap” (Mackey, 2005). The third note of the bass clarinet’s opening line is a B-natural, a ninth higher than the A preceding it. The three notes, in three registers, are played over a hairpin, the chalumeau D-natural intensifying to the throat A-natural and backing off to near silence on the final clarino B. The three registers lay out the

zones occupied by the soloists, played on a single clarinet (Fig. 1). The rest of the work can then be seen as an elaboration of this opening: a single line encompassing three distinct ranges.

Figure 1: Martino, *Triple Concerto*, Opening mm. 1-2

Martino takes his time unveiling his creation. The second clarinet to enter is the soprano clarinet in the ninth bar, and soprano and bass play alternately separate and homophonic passagework for most of the opening. The soprano's opening material is lyrical, sounding an inversion of the bass clarinet's initial notes, first on E-flat and jumping down an eleventh to A-flat. Countering the extremely soft dynamics and incremental appearances, the contrabass clarinet enters with an extended, virtuosic passage accompanied by piano and strings supplying interjecting stabs. A rhythmically condensed variation on an inverted retrograde of the bass clarinet's opening phrase is transposed down a minor second, as the contrabass opens on a low B-flat, moving up a second and a diminished fifth, then descends to a C-flat in the span of five thirty-second notes, in contrast to the bass clarinet's leaps over the space of three whole notes.

The three introductions to the Superclarinet's tendrils illustrate Martino's intentions for his monster. The "poetic" bass enters first and establishes the ranges, narrating the arc of the work about to unfold and presenting a simple line that is elaborated on in differing ways by its partners. The "elegant" soprano follows, lyrically decorating the bass clarinet's opening before joining its poetic counterpart in a unified hocket against the upper winds. The "obstreperous" contrabass then embellishes the opening in almost disobedient fashion, spitting it out at speed before going on a romp, scattering the ensemble in its path. By bar 52, the first significant combined line between the soloists occurs and all three personalities are condensed under a single complex whole.

The *Triple Concerto*'s cadenza emphasizes the Superclarinet in its most clear iteration when the soloists perform in near monophony between bars 225–278. Martino sets up the cadenza with a pointillistic, orchestral compound line reminiscent of Webern (Zeller, 2022). Each clarinet plays a separate line of their own over this, as the ensemble gradually gives way to the soloists.

The cadenza (Fig. 2) draws from core ideas introduced at the movement's opening. The contrabass clarinet begins by jumping a perfect fifth and then a ninth beginning on the same pitch class D-natural

that opens the Concerto, then plays a leap-filled gymnastic line reminiscent of its own opening solo before giving way to the soprano. The soprano's figure features similar rhythms and articulations to the contrabass but the player is instructed to perform them in a different manner. Martino wrote of his approach to notated expression that, "In my own music I attempt as best I can to notate nuances-but a mechanistic reproduction is furthest from my mind!" (Martino, 1966). The contra is given *pieno, con larghezza, ansioso, and calmandosi* to assist the player in delivering the arc Martino is looking for from the 2-bar passage. On the contrary, the soprano is only instructed to perform its line *elegante*. The soprano gives way to a lyrical bass clarinet passage played extremely softly, which is rudely interrupted by the contra taking off on a *sdegnoso* sojourn, pinballing between fortississimo and pianississimo.

What preserves the integrity of the Superclarinet is the efficient transitions between instruments. The rhythmic unit of each hocket—that is, how long each clarinet plays before yielding to the next—diminutes as the cadenza ramps up in intensity and the care with which the transitions are handled becomes more crucial. The sub-personalities of the Superclarinet begin to break down somewhat as the cadenza proceeds; the three clarinets are fully clashing by m. 259 and it takes a fermata-rest in m. 261 to restore monophony. The remainder of the cadenza features a back and forth between intense contrapuntal passagework before a *legato* bass clarinet passage at bar 274 yields to a *staccatissimo* answer in the contra and the cadenza, and first movement, conclude in a resulting silence.

Martino creates a convincing multi-soloist concerto by playing with the timbral possibilities permitted by the abundance of soloists. *Klangfarben*-esque lines with evolving articulations allow the three clarinets to be timbrally idiomatic while still unified.

Figure 2: Martino, *Triple Concerto*, Cadenza, mm. 225-236

References

- Boros, J., and Martino, D. (1991). A Conversation with Donald Martino. *Perspectives of New Music*, 29(3), 212–78.
- Boros, J. (1991). Donald Martino's Fantasy Variations: The First Three Measures. *Perspectives of New Music* 29(2), 280–93.

- Dolan, E. I. (2013). *The Orchestral Revolution: Haydn and the Technologies of Timbre*. Cambridge University Press.
- Kyr, R. (1991). Point/Counter-Point: Donald Martino's Radical Statement of Mind and Soul. *Perspectives of New Music* 29(2), 382–92.
- Leong, D., et. al. (2019). *Performing Knowledge: Twentieth Century Music in Analysis and Performance*. Oxford University Press.
- Mackey, S. (2005, December 15). Obituary: Donald Martino (1931-2005). NewMusicUSA. <https://newmusicusa.org/nmbx/obituary-donald-martino-19312005/>.
- Martino, D. (1972). Exotic Winds. *The Musical Times*, 113(1555), 865.
- Martino, D. (1966). Notation in General-Articulation in Particular. *Perspectives of New Music* 4(2), 47–58.
- Martino, D. (1961). The Source Set and Its Aggregate Formations. *Journal of Music Theory*, 5(2), 224–73.
- Martino, D. (1978.) [Liner Notes]. In *Triple Concerto/Notturmo*. Albany Records TR168, <https://www.dramonline.org/albums/donald-martino-triple-concerto-notturmo>.
- Nicholls, D. (1992). Donald Martino: A Survey of His Recent Music. *Music & Letters* 73(1), 75–79.
- Rothstein, W. (1980). Linear Structure in the Twelve-Tone System: An Analysis of Donald Martino's 'Pianississimo.' *Journal of Music Theory* 24(2), 129–65.
- Van Elferen, I. (2021). Timbrality: The Vibrant Aesthetics of Tone Color. In E. I. Dolan and A. Rehding (Eds.), *The Oxford Handbook of Timbre* (pp. 68-91), Oxford University Press.
- Vélez, D. V. (2021). The Matter of Timbre: Listening, Genealogy, Sound. In E. I. Dolan and A. Rehding (Eds.), *The Oxford Handbook of Timbre* (pp. 22-51), Oxford University Press..
- Vishio, A. (1991). An Investigation of Structure and Experience in Martino Space. *Perspectives of New Music*, 29(2), 440–76.
- Whittall, A. (2005). The concerto since 1945. In S. P. Keefe (Ed.), *The Cambridge Companion to the Concerto* (pp. 161-173), Cambridge University Press.
- Wuorinen, C. (1991). An 'Appreciation' of Donald Martino. *Perspectives of New Music*, 29(2), 478–478.
- Zeller, M. (2023). Klangfarbenmelodie, Chromophony, and Timbral Function in Arnold Schoenberg's "Farben." *Music Theory Online*, 29 (3), <https://mtosmt.org/issues/mto.23.29.3/mto.23.29.3.zeller.html>
- Zeller, M. (2022). Klangfarbenmelodie in 1911: Timbre's Functional Roles in Webern's Op. 9 and 10, *Music Theory Online* 28(1), <https://mtosmt.org/issues/mto.22.28.1/mto.22.28>.

Timbre as a Pedagogical Tool for Students With Dyslexia in the Music Theory Classroom

Joy Anderson^{1†}

¹ Music Department, Western Michigan University, Kalamazoo, Michigan, United States

[†] Corresponding author:

Students with dyslexia face a range of challenges in the music theory and aural skills classroom, particularly when instruction relies heavily on linguistic processing, rapid visual decoding, and the tight coupling of notation to sound. Dyslexia can hinder a student's ability to track musical lines on the staff or visually follow harmonic progressions and affects the speed of processing across both visual and auditory domains (Parsons, 2015). Dellatolas et al. (2009) and Overy (2003) further note that rhythm perception, auditory sequencing, and phonological processing are connected to broader cognitive systems involving temporal organization and symbolic interpretation. As a result, tasks such as dictation, harmonic analysis, and part-writing often require students to manage multiple cognitive processes simultaneously, increasing the likelihood of cognitive overload. When excessive attention is devoted to decoding symbolic information, fewer cognitive resources remain available for conceptual understanding (Sweller, 1988). Many researchers similarly suggest that multimodal learning can reduce cognitive load while improving student focus and engagement.

This research proposes a timbre-centered pedagogical framework for post-secondary music theory and aural skills instruction as a means of supporting students with dyslexia. Rather than replacing traditional analytical methods, timbral differentiation supplements them by providing additional perceptual entry points into musical structure. Timbral distinctions are often processed rapidly and pre-attentively, which may reduce the cognitive effort required to decode dense musical examples while also supporting attention and memory. The following study explores timbre-based strategies that can be implemented in theory and aural skills classrooms to create more accessible and supportive learning environments for dyslexic students.

Research in auditory cognition suggests that timbre plays a central role in auditory stream segregation, the process through which the brain organizes complex sonic environments into distinct perceptual layers. Albert S. Bregman argues in his work on auditory scene analysis that listeners perceptually separate simultaneous sound sources into coherent streams based on characteristics such as pitch, timing, and timbre. Differences in spectral content, attack, resonance, and acoustic envelope allow listeners to separate simultaneous sound sources and maintain attention on individual musical lines over time. In polyphonic textures, timbral differentiation increases perceptual salience, enabling listeners to organize musical information perceptually before conscious analytical interpretation occurs. By providing salient auditory cues that assist attention and memory, timbre offers an additional pathway through which students can connect musical sound to theoretical representation. Parsons (2015) suggests that for students with visual dyslexia, assigning distinct timbres to individual voices or harmonic functions can support auditory tracking and reduce reliance on continuous visual reference. Figure 1 helps visualize this same idea by showing how timbral cues can shape the way listeners organize sound before they begin analytical processing. It traces how adding timbral differentiation at the stage of notation and listening can support auditory stream segregation and reduce cognitive load, making it easier to move toward clearer analytical understanding.

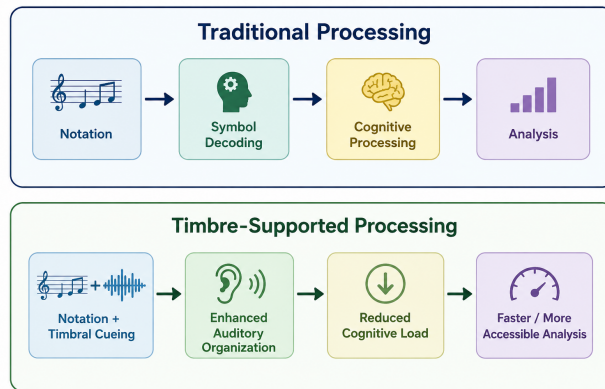


Figure 1: Chart of traditional score-based processing and timbre-supported auditory processing.

In melodic dictation exercises, students must rapidly encode auditory information into symbolic notation while simultaneously retaining sequential order and pitch relationships in working memory. For students with dyslexia, this process may produce substantial cognitive demand because it requires continuous coordination between auditory perception and symbolic transcription (Sweller, 1988; Overy, 2003). A timbre-centered approach to teaching melodic dictation may use a variety of timbral distinctions to differentiate the target melody from accompanying material. In introductory dictation exercises, the target melody may be assigned a distinct timbre while accompanying voices remain comparatively neutral or subdued. Over time, these timbral distinctions can be gradually reduced as students develop stronger internal auditory representations. The timbral contrast assists auditory attention and stabilizes sequential tracking, allowing students to devote greater cognitive resources to notation encoding and intervallic recognition. Using contrasting timbres can be especially beneficial for students with dyslexia, as it allows them to follow the melodic line more easily without relying heavily on visual cues and reduces confusion between musical parts (Parsons, 2015). In a texture performed entirely on one instrument, it can sometimes be difficult to distinguish the melody from the accompaniment. However, when the melody and harmony are assigned to clearly different timbres, the two musical layers become easier to perceive and separate.

Similarly, timbre may also support the perception of contrapuntal independence in part-writing and voice-leading instruction. In four-part textures, students are often required to track multiple simultaneous voices while identifying parallels, tendency-tone resolutions, and contrapuntal motion. Such tasks can produce substantial symbolic density and cognitive load (Sweller, 1988). Assigning distinct timbres to individual voices may strengthen auditory stream segregation and increase the perceptual independence of contrapuntal lines (Bregman, n.d.). In a chorale-style exercise, for example, soprano, alto, tenor, and bass voices might be rendered using contrasting instrumental colours. These timbral distinctions can help students follow individual lines more consistently over time and may reduce reliance on continuous visual parsing of the score.

Figure 2 illustrates this approach using Johann Sebastian Bach Chorale No. 1, “Aus meines Herzens Grunde.” The chorale is first presented on piano alone and then repeated with each voice assigned to a different instrument. In this arrangement, the soprano, alto, tenor, and bass voices are distributed among flute, trumpet, cello, and bassoon, respectively. Comparing these spectrograms suggests how contrasting timbres can make the individual contrapuntal lines easier to perceive and differentiate. In the piano version, the harmonic texture appears visually compressed because the partials of each voice overlap within a relatively unified timbral spectrum. As a result, the individual melodic strands can become more difficult to isolate aurally, particularly for inexperienced listeners. In the orchestrated version, however, the distinct spectral profiles of the flute, trumpet, cello, and bassoon create clearer separation between

voices, allowing the contrapuntal lines to emerge with greater clarity. This separation supports auditory stream segregation, a perceptual process in which the listener organizes simultaneous musical events into distinct layers (Bregman, n.d.). By assigning contrasting timbres to each line, the arrangement reduces the cognitive effort required to follow individual voices and helps listeners more easily attend to the interaction between melodic lines, harmonic motion, and formal structure.

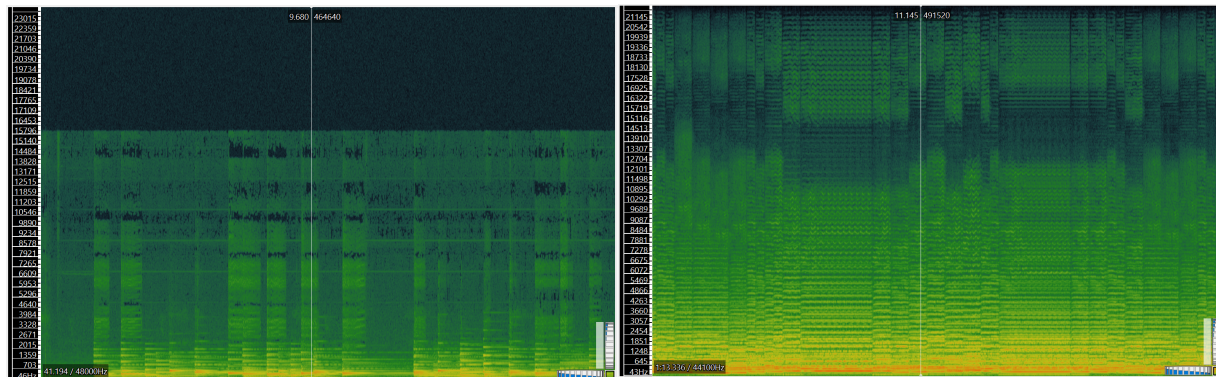


Figure 2: Bach chorale shown in piano reduction, on the left, and with orchestration (flute, trumpet, cello, bassoon), on the right to illustrate timbral separation of voices.

Beyond melodic dictation and the differentiation of contrapuntal voices, timbre may also support the teaching of harmonic function and phrase structure in music theory instruction. When doing analysis or dictation, students are often required to identify cadences, tonicizations, or changes in harmonic direction while simultaneously retaining melodic and rhythmic information in working memory. For students with dyslexia, these tasks may produce substantial cognitive demand (Sweller, 1988). Timbre can help reduce this difficulty by directing auditory attention toward structurally significant harmonic events. For example, an instructor might present a short phrase with dominant-function harmonies reinforced by a brighter timbre such as trumpet or oboe, while tonic harmonies remain in a neutral piano timbre. The contrasting timbral emphasis may make harmonic tension and cadential arrival points more perceptually salient, helping students recognize phrase endings and harmonic motion more efficiently. As students develop stronger internal recognition of harmonic function, these timbral distinctions can gradually be reduced in order to encourage more independent listening.

These examples demonstrate several ways timbre may function as a pedagogical support within music theory and aural-skills instruction, including the differentiation of melodic and contrapuntal lines as well as the reinforcement of harmonic structure and phrase organization. Timbre may also be applied in a variety of additional contexts beyond those explored here. Contrasting timbres can assist interval and chord identification by increasing the perceptual clarity of individual tones within a sonority, support motivic recognition by highlighting recurring thematic material, and aid sequential memory by functioning as auditory landmarks within longer listening exercises. Timbre may furthermore help direct attention toward structurally important musical events such as modulations, suspensions, or formal boundaries. Together, these applications suggest that timbre can serve not only as a means of separating musical textures, but also as a broader cognitive scaffold that supports listening, memory, and musical organization.

This paper argues that timbre is not simply a surface-level sonic element but a substantive pedagogical resource capable of enriching instruction for all learners while offering meaningful support for students with dyslexia. By increasing perceptual salience and supporting auditory organization, timbral differentiation may reduce cognitive load associated with symbolic decoding and strengthen connections between notation and sound (Bregman, n.d.; Sweller, 1988). In this sense, timbre operates as a form of

perceptual scaffolding that supports multimodal musical learning. While this framework is informed by research on dyslexia and symbolic processing, its implications extend more broadly toward inclusive and cognitively flexible models of pedagogy. Integrating timbral variation into music theory and aural skills pedagogy has the potential to reshape how students perceive and conceptualize musical structure, fostering deeper learning, enhanced accessibility, and more inclusive models of musicianship training.

References

- Aldalalah, O. M. A. A., & Fong, S. F. (2010). Music intelligence and music theory learning: A cognitive load theory viewpoint. *International Journal of Psychological Studies*, 2(2), 150–158.
- Birch, S., & Chase, C. (2004). Visual and language processing deficits in compensated and uncompensated college students with dyslexia. *Journal of Learning Disabilities*, 37(5), 389–410.
- Bouchey, B., Castek, J., & Thygeson, J. (2021). Multimodal learning. In J. Ryoo & K. Winkelmann (Eds.), *Innovative learning environments in STEM higher education* (pp. 35–54). Springer.
- Bregman, A. S. (n.d.). Auditory scene analysis. In N. J. Smelser & P. B. Bates (Eds.), *International encyclopedia of the social and behavioral sciences*. Pergamon (Elsevier).
- Cancer, A., Bonacina, S., Lorusso, M. L., Lanzi, P. L., Antonietti, A., & Matic, A. (2016). Rhythmic reading training (RRT): A computer-assisted intervention program for dyslexia. In *Pervasive computing paradigms for mental health*. Springer International Publishing.
- Castilho, F. M., & Pereira, C. (2022). Teacher profile vs student profile: Processes of teaching-learning in the subject of music education. In *Advances in design, music and arts II* (pp. 733–752). Springer.
- Colomaio, O. G. B. (2025). Approaches to teaching music reading to students with dyslexia: A guide for piano teachers (Doctoral dissertation). University of South Carolina.
- Dellatolas, G., Watier, L., Le Normand, M.-T., Lubart, T., & Chevrie-Muller, C. (2009). Rhythm reproduction in kindergarten, reading performance at second grade, and developmental dyslexia theories. *Archives of Clinical Neuropsychology*, 24(6), 555–563.
- Fiveash, A., Schön, D., Canette, L.-H., Morillon, B., Bedoin, N., & Tillmann, B. (2020). A stimulus–brain coupling analysis of regular and irregular rhythms in adults with dyslexia and controls. *Brain and Cognition*, 140, 105531.
- Laes, T. (2017). The (im)possibility of inclusion: Reimagining the potentials of democratic inclusion in and through activist music education. *Sibelius Academy of the University of the Arts Helsinki*.
- Lifshitz-Ben-Basat, A., & Fostick, L. (2019). Music-related abilities among readers with dyslexia. *Annals of Dyslexia*, 69, 318–334.
- Miles, T., Westcombe, J., & Ditchfield, D. (Eds.). (2008). *Music and dyslexia: A positive approach*. John Wiley & Sons.
- Oglethorpe, S. (2002). *Instrumental music for dyslexics: A teaching handbook* (2nd ed.). Whurr.
- Overy, K. (2003). Dyslexia and music: From timing deficits to musical intervention. *Annals of the New York Academy of Sciences*, 999(1), 497–505.
- Parsons, L. (2015). Dyslexia and post-secondary aural skills instruction. *Music Theory Online*, 21(4). <https://mtosmt.org/issues/mto.15.21.4/mto.15.21.4.parsons.html>
- Remillard, A. (n.d.). *Aus meines Herzens Grunde Bach Chorale No. 1 BWV 269 My inmost heart now raises* [Video]. YouTube. <https://www.youtube.com/watch?v=L2cPibJzwbA>
- Rogers, M. R. (2004). *Teaching approaches in music theory: An overview of pedagogical philosophies* (2nd ed.). Southern Illinois University Press.
- Sailer, L. M. (2011). *Dyslexia and music: Toward a comprehensive pedagogy* (Master's thesis). University of Toledo.

- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4
- VanHandel, L. (Ed.). (2020). *The Routledge companion to music theory pedagogy*. Routledge.
- Walker, R. (1978). Perception and music notation. *Psychology of Music*, 6(1), 21–46.
- Witmer, N. S. (2015). *Music lessons from a tablet computer: The effect of incorporating a touchscreen device in teaching music staff notation to students with dyslexia* (Doctoral dissertation). Boston University.
- Wolf, M., & Bowers, P. G. (1999). The double-deficit hypothesis for the developmental dyslexias. *Journal of Educational Psychology*, 91(3), 415–438. <https://doi.org/10.1037/0022-0663.91.3.415>.

Session 13

Synthetic Timbre in Popular Music

The Transmedial Ecology of Robot Timbres: Sonic Metaphors and the Evolution of Dubstep Music

Mark RJ Higgins^{1†}

¹ Department of Music, University of Liverpool, Liverpool, UK

[†] Corresponding author: mark.higgins@liverpool.ac.uk

Dubstep is a form of electronic dance music that began taking shape in London during the early-2000s. The term had begun to circulate by 2001 (Goodman 2001), in reference to a localised clustering of rhythmically broken and bass-heavy music. From these beginnings, over the next ten or so years, an evolutionary trajectory led to the emergence of a texturally abrasive and rhythmically simplistic subvariant, likened to an ‘electronic form of heavy metal’ (Goodwyn 2011). Gaining significant traction in North America, what is most prominent in this later form of dubstep are dynamic bass sounds, with sonic micro-gestures and a timbral tactility that evokes the visual character of animated, cinematic robots (Abebe 2021; Sutton 2012). The output of key North American artists active during the early-2010s, such as Excision and Downlink, has even retroactively attracted the term ‘robostep’. The stylistic gulf between ‘robostep’ and the earlier, avant garde music seems difficult to reconcile. However, there was a crucial period of stylistic transformation in the interim period, and analysis of tracks from this period elucidates both the way the music came to attract interpretation as sonically robotic, and the creative basis on which later artists made these robotic qualities more explicit.

In this paper, I look at the way dubstep’s bass sounds became distinctively charged objects over time. I map the creative influence of some key actors, and examine the way their musical outputs have transmedial resonance. I lend significant depth to what could otherwise be a straightforward, intertextual reading by connecting timbral qualities to those of physical materials and embodied action. To do so, through bringing together their respective scholarship, and by offering tangible case studies, I advance the work of Eric Clarke, Allan Moore, Zachary Wallmark and Maria Perevedentseva.

Intertextually, some of the qualities that emerged in dubstep during the genre’s stylistic evolution can be traced back to trends in regionally antecedent music, such as drum & bass and breakbeat hardcore. Overt references to dystopian sci-fi cinema were the main qualifiers of substyles known as ‘techstep’ and ‘darkcore’ respectively, in the form of both explicit, autsonic quotation, timbre selection and compositional approach, and deployment of non-musical sounds mirroring the ‘sonic ambience’ of the cinematic genre (Schmidt 2023: 169-170). This transmedial aesthetic percolated dubstep from its gestation period onwards, (with one of the stylistic antecedents even referenced in the 2001 article where the term ‘dub step’ first appears in print). A few years later, in the music of dubstep artists Distance and Vex’d in particular, autsonic borrowing from dystopian sci-fi cinema can still be heard. Both acts have also verbally referenced the cinematic genre. While less explicitly intertextual, other artists too were creating cinematic sound worlds – and I note here that the genre figurehead Skream has likewise attested to a love of ‘darker, movie-sounding shit’ (Clark 2005). Within this sonic bed, a particular species of electronic bass timbre emerged that gradually increased in gestural and textural complexity.

I argue that it is dubstep’s transmedial context that informed what would become its robotic sonics. In Skream’s music, notably, the ambience of dystopian sci-fi is routinely conveyed through minor 2nd shuttles. The harmonic structure of his track ‘Rottan’ (2006), for example, pivots between Fm and Gbm triads. What is notable is the way this device maps onto a leitmotif that appears frequently in the *Alien* franchise, which in the inaugural film, through a simple cue ascending from a sustained Em to a decaying Fm, accompanies the first appearance of H.R Giger’s ‘biomechanical’ art vision. Interpenetrating living flesh and synthetic technologies, the gnarled textures and dank colour palette characteristic of Giger’s

biomechanical is a visual apparatus that Skream captures in sound. The aural roughness of the bass timbre in 'Rottan' has a fleshy tactile index, made all the more organic by its throaty, vocable quality, but this is tempered by the artificiality imparted by the electronic technology with which the sound is realised. Within its intertextual bed of minor 2nd intervals, eerie reverberant instrumentation, and evocative sound effects, this kind of bass timbre can be understood as itself 'biomechanical'.

To further deepen the pertinence of the timbral relationship between Skream's tracks and H.R. Giger's biomechanical vision, the original *Alien* movie is seen as having established the core aesthetic of the sci-fi horror movies released during the 1980s and 1990s (Whittington 2007: 129). The example of Skream's 'Rottan' therefore points not just to the pivotal, audiovisual cue in *Alien*, but to the aesthetic of the cinematic subgenre it served to define. Across movies frequently quoted by dubstep artists, such as *Blade Runner* and *The Terminator*, perturbing amalgams of organic (usually human) and technological lifeforms are a central visual and thematic device. The increasingly lifelike bass timbres in dubstep tracks contemporaneous with and following after 'Rottan', therefore, function as transmedial metaphors for the biomechanical antagonists of the sci-fi horror canon.

The 'sonic metaphor', for Zachary Wallmark' is a sound taking on an extrasonic life through a vertical procession of bodily sensation, semantic substitution and symbolic relevance (2022: 78). My descriptive language has begun to outline the first two parts of this process, which connect to the third via dubstep's intertextual relationship with the darker ends of sci-fi cinema. The semantic and symbolic dimensions might give dubstep's bass timbres their context, but it is through embodiment that the sounds really come to life. The semiotic scope of sonic objects is grounded by Eric Clarke in an ecology of environmental constants, reflecting our multi-sensory immersion in a world of solids, liquids, wood, metal, stone, synthetics, and so forth, which forms a crossmodal index of material 'invariants' (Clarke 2005: 32-36). Aural processes traversing to other parts of the body can also be applied to musculoskeletal movement (Wallmark 2022: 64). Allan Moore considers musical sound, in this way, to unfold through a process of embodied cognition (Moore 2012: 238-239). These concepts are nuanced significantly by Maria Perevendentseva, who suggests that the reception of sound draws us into a 'zone of entanglement', extending our selfhood into the ontological fabric of the sound itself (2023: 46). Responding to the qualities of the dubstep bass timbre thus becomes a form of surrogacy: our own bodies map onto the kinds of acts that we associate with vocables, sensuality, and most importantly musculoskeletal gesture.

To give an example of this kind of timbral metaphor at work, in 'Bed Bugz' (2007) by The Others, the pairing of bass timbres is rich with biomechanical qualities. The track opens with a sonic 'micro-ecology' redolent of cinematic dystopian sci-fi ambiance (van Elferen 2021: 47), replete with minor 2nd intervals. Bursting their way out of this bed, the track's bass sounds are hoarsely vocable, and each note in the part is brought to life through microtonal fluctuations in pitch, which transposes onto muscular tension and release. The way in which tense muscles correlate with higher vocal registers, which will naturally sink when muscles are relaxed, is foundational to the embodied cognition through which such sounds are internally indexed. Cutting across other sensory modes, both timbres also have a roughened, fleshy texture, connoting at once the folds of the throat and a rattling metal plate. In Figure 1, the muscular pitch gestures can be seen clearly in the lower frequencies. Also visible are the resonant envelopes that rhythmically exaggerate the upper overtones, giving the timbres their phonetic quality.

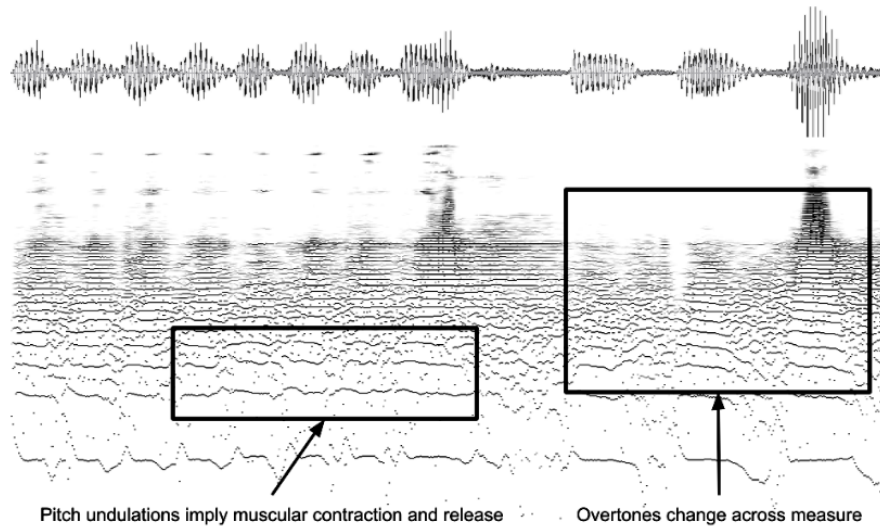


Figure 1: Spectral image of bass part extracted from 'Bed Bugz' (2007)

With its intertextual grounding, through increasingly sophisticated timbral makeup, the biomechanical bass timbre in dubstep developed into an aesthetically charged object. This charge is made visually evident in the cover art for 'If Ya Can't Beat 'Em' (2008) by Reso, which depicts a robotic creature in the style of the 1980s Transformers toy line. This clearly focused, intertextual intent within the genre is emblemised yet further in later tracks by The Others, such as 'Robo Funk' (2009), which in both title and style is a prime example of the way the organic layers of dubstep's sonic biomechanical hybrids were being peeled back in favour of full-fledged artificial lifeforms. The highly complex overtone patterns of the track's bass sounds capture the timbral profile of vintage, computerised 'Speak & Spell' devices. In 'Gravity' (2010), The Others gave this emerging sonic species even greater attention, with a rasping bass part that materially nuances synthetic phoneme shapes with metallic rattling and a chromium sheen.

Crucially, overt references to robots in dubstep music grew in prominence following the release of the first movie in the *Transformers* franchise in 2007. What resulted was an aesthetic convergence, and this was driven by North American artists in particular, who explicitly drew together the motivic devices of the two media. Notably, Datsik and Excision routinely used samples lifted from the *Transformers* films, even using the audio lifted wholesale from an animated battle sequence in the track 'Calypso' (2009).

A paradigmatic example of the aesthetic layers that support the characteristic bass timbres of later dubstep as transmedially charged objects is Datsik and Excision's 'Boom' (2009). The main bass part, supported by an ambiance of crunching metal and whirring machinery, moves through a variety of implied gestures and vocables. In Figure 2, the rasping, metallic quality of the timbre is evident in the nonlinear complexity of the overtones, which nevertheless exhibits greater regularity than the biomechanical timbres depicted previously. Crossmodally, this translates to the tactility of harder materials, resulting in an aural event that pulls the listener closer to associations with robots than biomechanical hybrids. Reinforcing the robot metaphor, in some sections of "Boom" the bass timbre is quite literally suggestive of the movement of a large piece of machinery. Elongated and exaggerated pitch gestures allude to a massive piece of machinery that groans under its own weight as it attempts to manoeuvre itself. This is the sense of inertia that, for Jan Herbst and Mark Mynett, gives quasi-material 'weight' to slow and heavy musical passages (2023: 20-21), which, when coupled with the other qualities exhibited by these kinds of bass timbres brings them to life as aural metaphors for large, animated robots.

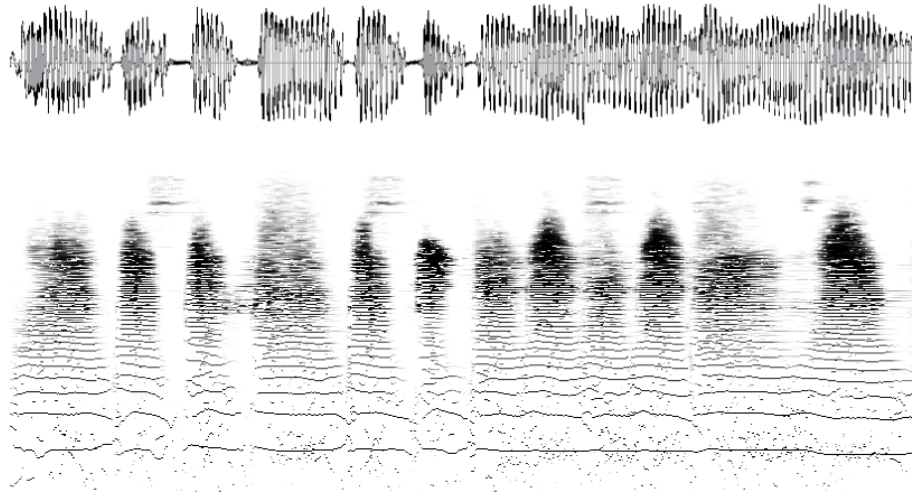


Figure 2: Spectral image of bass part extracted from 'Boom' (2010)

Where intertextuality, tactility and timbral gesture intertwine, dubstep's bass sounds become metaphors for animated robots. The synergies between the work of Clarke, Wallmark, Moore and Perevedentseva shows how the timbres assume a near humanlike vitality, and the intertextual positioning of these timbres aligns this with the same uncanny aura exhibited by animated robots in cinematic CGI spectacles such as *Transformers*. This paper thus demonstrates the processes through which timbre, with uncanny physicality, became a charged object in dubstep music, and this was crucial to the identity that the genre had assumed as it took hold of the popular consciousness in North America during the early-2010s.

References

- Abebe, N. (2012). Why Does America Love Skrillex? *Vulture*.
<https://web.archive.org/web/20211006111730/https://www.vulture.com/2012/01/why-does-america-love-skrillex.html>.
- Clark, M. (2005) Skream. *Blackdown: Dubstep..Grime..UK Funky..Comment..Culture*,
<https://blackdownsoundboy.blogspot.com/search?updated-max=2005-11-27T14:38:00Z>.
- Clarke, E. F. (2005) *Ways of Listening: An Ecological Approach to the Perception of Musical Meaning*. New York: Oxford University Press, 2005.
- Herbst, J., Mynett, M. (2023). Toward a Systematic Understanding of Heaviness in Metal Music Production. *Rock Music Studies*, 10:1, 20-21. <https://doi.org/10.1080/19401159.2022.2109358>.
- Moore, A. F. (2012). *Song Means: Analysing and Interpreting Recorded Popular Song*. Farnham: Ashgate.
- Perevedentseva, M. (2023). Timbre and the 'Zone of Entanglement' in Electronic Dance Music: Rethinking Musico-Social Ontologies with the Mycelial Turn. *Dancecult Journal*, 13:1, 41-57. DOI:10.12801/1947-5403.2023.15.01.03.
- Schmidt, U. (2023). *A Philosophy of Ambient Sound: Materiality, Technology, Art and the Sonic Environment*. Singapore: Palgrave Macmillan.
- van Elferen, I. (2021) *Timbre: Paradox, Materialism Vibrational Aesthetics*. New York: Bloomsbury Academic.
- Wallmark, Z. (2022) *Nothing But Noise: Timbre and Musical Meaning at the Edge*. New York: Oxford University Press.

Deviant Timbre: Sounding Beyond the Human in Pharmakon's *Maggot Mass* (2024)

Bailey Hilgren^{1†}

¹ Music Department, New York University, New York, NY, United States

[†] Corresponding author: bhilgren@nyu.edu

This paper analyzes experimental noise project Pharmakon's 2024 album *Maggot Mass* for the ways it uses timbre to sound bodily and ecological breakdown. These moments exemplify what I theorize as "deviant timbre" (after Tompkins 2024): a mode of sonic materiality that resists normative containment, foregrounds bodily failure, and sounds the entangled destruction of human and nonhuman ecologies. While *Maggot Mass*'s lyrics explicitly articulate ecofeminist critiques of capitalism, anthropocentrism, and mass species death, this paper contends that it is timbre that most forcefully clarifies the album's ethical and political stakes.

Previous work by Pharmakon, the solo project of artist Margaret Chardiet, consistently explored timbral failure as bodily failure, often through a feminist and disability-informed lens. Perhaps best exemplified by her 2014 *Bestial Burden*, these prior albums centered the individual body under sonic duress, with Chardiet's extreme vocals figuring her growling, rasping body as fragile, compromised, and betrayed by its own material limits. *Maggot Mass* returns to these sonic failures but reframes them within a broader, explicitly inhuman and decaying aesthetic that displaces the individual human body as the primary site of agency. Across *Maggot Mass*, timbre operates as an embodied ecological force. The album is saturated with Chardiet's vocals, but also the buzzing of swarming flies; sludgy, flapping bass; distorted, caustic detritus; and the resonant, roaring shudder of a contact-mic'd sheet of metal, performed by Chardiet. These sounds refuse clarity, purity, and resolution or, in my reading, refuse to sound human. Chardiet's vocal timbre in particular decomposes and reconstitutes across tracks, recycling itself into something neither fully human nor fully machinic.

I argue that this timbral excess reveals two central claims embedded in *Maggot Mass*: first, that the bodies at stake – human and nonhuman alike – are real, fleshy, and visceral rather than just abstract or symbolic; and second, that these bodies are fundamentally and irrevocably entangled with one another in rapidly decaying ecologies. Chardiet's voice does not simply represent ecological collapse; it sounds it, staging breakdown, permeability, and contamination as embodied realities rather than metaphors. Through her vocal timbre in particular, Chardiet echoes feminist posthumanist scholar Stacy Alaimo's notion of transcorporeality (2010) by theorizing (sounding) human bodies as fragile, porous, and coextensive with nonhuman environments. Rather than Alaimo's prioritization of slow violence and ambivalence, however, Chardiet's musical theorizing foregrounds rapid destruction, distortion, and disgust.

This paper situates *Maggot Mass* within my larger project to develop a feminist posthumanist theory of timbre. I argue that timbre has been an intensified site of human attempts to discipline unruly musicking human and nonhuman bodies through the development and naturalization of timbral norms that align with ideals of clarity, control, health, and humanity. Yet timbre and its material substrates repeatedly resist such control. Deviant timbre in *Maggot Mass* serves as a sort of painful yet delicious sound from which we might take affirmative ethical and political cues despite its destructive nature. Chardiet's concern with human embeddedness within unruly ecologies provides an important example of how even the human body – that seemingly most knowable, most mythologically "Human" musical material – is still coextensive with the world and exerts unexpected agencies. In doing so, Chardiet performs an ecofeminist posthuman timbral politics that refuses human exceptionalism, builds solidarity across broad notions of difference, and reasserts the embeddedness of all bodies within nature.

References

- Alaimo, S. (2010). *Bodily natures*. Indiana University Press.
- Chen, M. Y. (2012). *Animacies: Biopolitics, racial mattering, and queer affect*. Duke University Press.
- Daughtry, J. M. (2020). Did music cause the end of the world? *Transposition*, Hors-série 2.
- Eidsheim, N. S. (2019). *The race of sound: Listening, timbre, and vocality in African American music*. Duke University Press.
- Lochhead, J. (2023). Timbre realities: A phenomenological study of Liza Lim's *Extinction Events and Dawn Chorus* (2018). In J. De Souza, B. Steege, & J. Wiskus (Eds.), *The Oxford handbook of the phenomenology of music*. Oxford University Press.
- Mundy, R. (2020). Life, death, and humanistic comparison. *Ethnomusicology Review*, 22(2), 85–92.
- Stras, L. (2006). The organ of the soul: Voice, damage, and affect. In N. Lerner & Straus, Joseph N. (Eds.), *Sounding off: Theorizing disability in music*. Taylor & Francis Group.
- Taylor, S. (2024). *Disabled ecologies: Lessons from a wounded desert*. University of California Press.
- Tompkins, K. W. (2024). *Deviant matter: Ferment, intoxicants, jelly, rot*. New York University Press.
- Van Elferen, I. (2020). *Timbre: Paradox, materialism, vibrational aesthetics*. Bloomsbury Academic & Professional.
- Wallmark, Z. (2022). *Nothing but noise: Timbre and musical meaning at the edge*. Oxford University Press.

The Timbre Is the Instrument: The Imagined DX7

Megan Lavengood^{1†}

¹ Dewberry School of Music, George Mason University, Fairfax, VA, USA

[†] Corresponding author: megan@meganlavengood.com

Author's note: An expanded version of this paper will be published as a chapter in the book *Analyzing Musical Instruments: Approaches from Popular Music Studies* (Palgrave Macmillan, forthcoming).

This paper uses “Versace on the Floor” by Bruno Mars (2016) and the Yamaha DX7 E. PIANO 1 sound to illustrate the problems one might encounter when researching synthesizers in recent popular music. I argue that researchers don’t need to think of accurate instrument identification as a prerequisite for analyzing a song’s use of that instrument, because it’s already significant that the timbre suggests the instrument. I support this through my case study of the Imagined DX7: the DX7 as it lives in popular consciousness, derived from memories and imaginings of its most well-known and imitated factory presets. I explore how the E. PIANO 1 preset outgrew the DX7, complicating the sonic identity of the instrument, but simultaneously solidifying a certain timbral signature that captures the essence of the instrument. I also discuss the technological changes—MIDI and the decline of the music trade magazine—that take away sources of truth for identifying instruments, a process that begins in the 1980s and continues to the present day. By considering timbre as the instrument, I show how researchers can analyze synthesizers in recent pop music through intertextuality and listener reception, areas that are unaffected or even enriched by possible inaccuracies in the instrument identification.

The DX7 is a digital synthesizer that uses FM synthesis technology. It was released in late 1983 and quickly became embedded in popular music recording practices. Yamaha shipped the instrument with a robust set of preset sounds (or “patches”) in its internal memory, and it was on these presets that most musicians relied. The most enduringly popular preset was the DX7’s electric piano sound, named “E. PIANO 1” in the internal memory. E. PIANO 1 was ubiquitous in popular music; in 1986, for example, it could be heard on 39% of the Billboard #1 singles on the Hot 100; the preset continued to be a mainstay for power ballads throughout the long 1980s (Lavengood, 2019, 74). The E. PIANO 1 sound was colloquially referred to as the “Rhodes” sound, perhaps more than its official name. The DX7 E. PIANO 1 is easy to distinguish from the Fender Rhodes sound in a side-by-side comparison, however. While it’s clearly evocative of an electric piano like the Rhodes, the DX7 sound has tells—things that make it sound distinct. The easiest to hear is the E. PIANO 1’s gapped overtone structure, which de-emphasizes the harmonics 8–12 and 14 while emphasizing harmonics 13 and 15 (see Figures 1 and 2). The presence and audibility of these high partials in the E. PIANO 1 sound likely contributed to the popular perception of the sound’s “brightness” when compared to older electric pianos.

As a scholar of the Yamaha DX7, the opening of “Versace on the Floor” immediately grabbed my attention on a first hearing: the first full minute of the track is Mars singing against nothing but a DX7 E. PIANO 1 accompaniment. I immediately wanted to know whether Mars did in fact use a DX7 on the recording; if so, I could hold up the song as an exemplar of a twenty-first-century resurgence of the DX7. But definitive evidence one way or the other was hard to find. As a concrete example, let me recount my search for evidence that a DX7 produced the opening sound in “Versace on the Floor.” I checked the liner notes, but specific equipment used for the song is not mentioned. I consulted many interviews with Mars and associated musicians, but found no equipment mentioned in them. I watched concert footage available on YouTube, but the keyboards were visually obscure since the videos’ focus was Mars rather than his touring keyboardist (John Fossit). I was unable to find a way to get in touch with John Fossit. I did have a connection to Greg Phillinganes, who is credited for “keyboards” in the liner notes on the

album. I thought this would clinch my success, but even he didn't know what keyboard had been used for the opening—he wasn't involved in the recording of the song's opening, and overdubbed the synth lead solo that occurs later in the track. The takeaway here is that it's extremely difficult to verify how a song was recorded these days, and while my journey with "Versace" is just one example, I'm certain other similar questions would have similar outcomes. The immense timbral flexibility of synthesizers, the black-boxing effect of MIDI, DAWs obviating hardware synthesizers, and the decline of trade magazines dedicated to music technology seem to create a perfect storm of sound-source illegibility.

The Imagined DX7 is the result of an entrenchment of the E. PIANO 1 sound as its sonic identity becomes blurred through the proliferation of E. PIANO 1 imitators. As the DX7's E. PIANO 1 grew thoroughly entrenched in popular music, some musicians looked for ways to tweak and customize this sound by layering it together with effects or other similar sounds. Synthesizer manufacturers looking to dethrone Yamaha from its market dominance—particularly Roland with the D-50 and Korg with the M1—likewise offered up their own takes on the electric piano patch that musicians had come to expect on their instruments. Today, musicians often use virtual instruments that mimic the signature DX7 sound. Through this process, over the course of more than forty years since its debut, the E. PIANO 1 sound evolves into something larger than just a preset on the DX7. Thus the timbre of the DX7 comes to stand in for the instrument itself.

To illustrate this, Figure 3 and Table 1 visualize the relationships among the electric piano presets on the Yamaha DX7, the Roland D-50, and the Korg M1, as well as the Yamaha Clavinova and a VST called Dexed. In the leftmost column Table 1 are opposed pairs of timbre descriptors, defined in Lavengood (2020). For each instrument discussed, I have indicated which of the two opposed terms better describes the sound of the instrument. To make the differences among instruments readily visible, the negative sign (–) represents the term on the left of the slash, and the positive sign (+) the term on the right. Since the DX7 E. PIANO 1 sound is the basis for comparison in this analysis, I have assigned all the E. PIANO 1 qualities to the negative side of the opposition. This allows the differences between E. PIANO 1 and any other sound to be immediately visually apparent: each positive sign corresponds to an audible timbral distinction from E. PIANO 1. The columns of this table organize the instruments on a continuum, from the most DX7-like (left) to the least (right). The footer row shows the total count of differences from E. PIANO 1, which roughly reflects this progression. Examining the rows that track each timbre descriptor shows these sounds are all bright, steady, and harmonic. Figure 3 is a spectrogram of all the sounds captured in the Table 1. This analysis illustrates the characteristic brightness that many people associate immediately with the sound of the DX7, present across a variety of sounds from various synthesizers.

The Imagined DX7 functions in part due to what Cornelia Fales (2002) identifies as the “paradox of timbre”—even while timbre is the strongest aural clue we have for identifying a sound source, timbre is most susceptible to being warped by our minds when compared to rhythm or pitch. Human perception shapes our hearing of a timbre profoundly. When assessing the likely source of a timbre, we are informed in part by the physical acoustic signal, but just as important are cultural and contextual clues: where we are, who else is around us, why sounds are being made, and so forth. Why would an observer, hearing a stereotypical “bright” and “clean” digitally synthesized sound in a 1980s performance and seeing the musician play a DX7, not assume that the sound is generated by the DX7 (despite the possibility that the DX7 is merely a controller for the true sound source)? In a situation where the sound source would be surprising, contextual clues might instead convince us of an inaccurate, but more likely-seeming, source. And without any evidence to contradict that sound-source identification, it persists—“we need no other” (Fales, 2002, 58).

We've reached a point in music technology where the study of instruments must take on a new, perhaps less obvious, form. The advent of MIDI, digital sampling, and the virtualization of music-making renders

a study of materiality nearly obsolete. The Imagined DX7 illustrates that analysis of instruments doesn't need to rely on concrete factual identification of the instrument in order to lead to meaningful insights. Indeed, for recent pop music, the question "how was this sound made?" is probably inevitably answered "with a DAW and VST." Today's musicians, especially those with the resources of Bruno Mars, can create any sound imaginable; more interesting than "how was it made?" is "why was it made?" What does it mean to be a hugely successful Top-40 artist and choose a sound for your record that's been widely panned as dated, overused, and grating? What's under examination in that question is the sound itself, the timbre, rather than the object that produced it. This isn't to say that anything goes, or to encourage intellectual sloppiness or laziness in sound-source identification, but rather to suggest that if the identification works and is productive, perhaps we need no other, to borrow Fales's phrasing. Accurately identifying the gear used in the recording is one avenue of research, but how many listeners can discern the difference between Dexed and a DX7? How many care? Does it impact the most interesting parts of my research?

Ultimately, what I am interested in is what it means to use a DX7 E. PIANO 1 sound—how listeners in 2016 experience such a distinctly '80s sound, how the sound might prompt every review I found to mention the track's retro sound, how the sound connects Bruno Mars to a specific lineage of R&B balladeers. When "Versace" starts, the timbre of the opening synthesizer immediately calls up all these associations of that familiar, even cheesy sound that "must die," but in a sleek, revitalized, 2016 package. Not just anyone could pull this off, but Mars has the panache and the musical background to make the DX7 sound cool—or if not cool, then at least charming.

It isn't particularly revolutionary to urge analysts to attend to what the sound means culturally; sound studies has done this for decades. The novelty is in thinking of this as a methodology for the analysis of musical instruments, an area that traditionally (and still) is preoccupied with gear—physical instruments and their material presentations—as sound sources. The critical idea here is that the timbre is the instrument. Timbre can carry the essence of an instrument even when the timbre is produced by something else, as in the case of the Imagined DX7. Timbral similarities, especially when combined with cultural plausibility, encourage the misattribution of a timbre to a sound source. Misattributed or not, the various instances of the (Imagined) DX7 sound all add new nodes to the vast web of intertextual meanings that the DX7 calls up. The practical obsolescence of hardware synthesizers means that traditional organological methods may fail, but even if we can't know for certain what instrument is used in a recording, we can always learn what listeners believe the instrument to be.



Figure 1: A comparison between a) an ideal harmonic series, b) the partials in a Rhodes sound, and c) the partials in a DX7 sound. Pitches marked with an asterisk (*) are inharmonic. Representation of pitches is approximate; the actual harmonic series does not adhere to the twelve-tone equal temperament implied by staff notation.

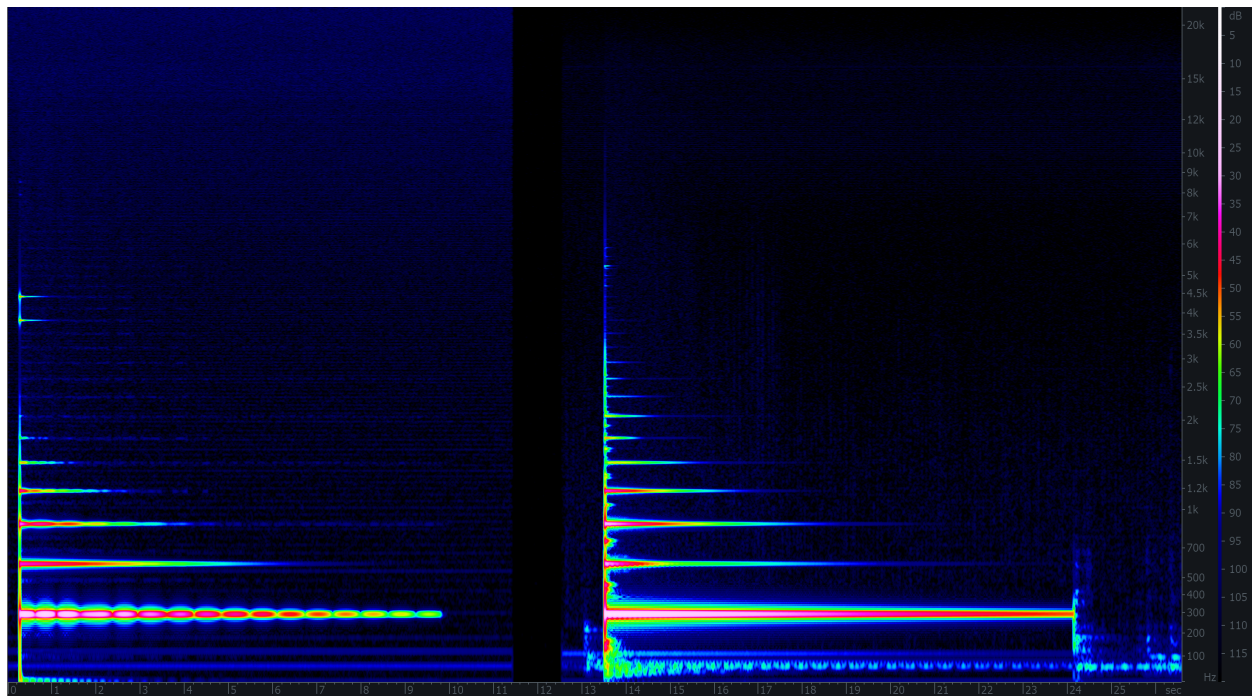


Figure 2: A spectrogram of the Fender Rhodes sound and the DX7 E. PIANO 1 sound.

Table 1: An opposition table comparing the DX7 E. PIANO 1 to other electric piano sounds.

– / + OPPOSITION – indicates DX7	DX7 E. PIANO 1	VST Dexed	Clavinova EP 1	M1 E. Piano 1	M1 E. Piano 4	Roland D- 50 Synth. Elec.
<i>bright / dark</i>	–	–	–	–	–	–
<i>pure / noisy</i>	–	–	–	+	+	+
<i>hollow / full</i>	–	–	–	–	–	+
<i>sparse / rich</i>	–	–	–	–	–	+
<i>beating / beatless</i>	–	–	+	–	–	–
<i>steady / wavering</i>	–	–	–	–	–	–
<i>harmonic / inharmonic</i>	–	–	–	–	–	–
total differences (+) from DX7		0	1	1	1	3

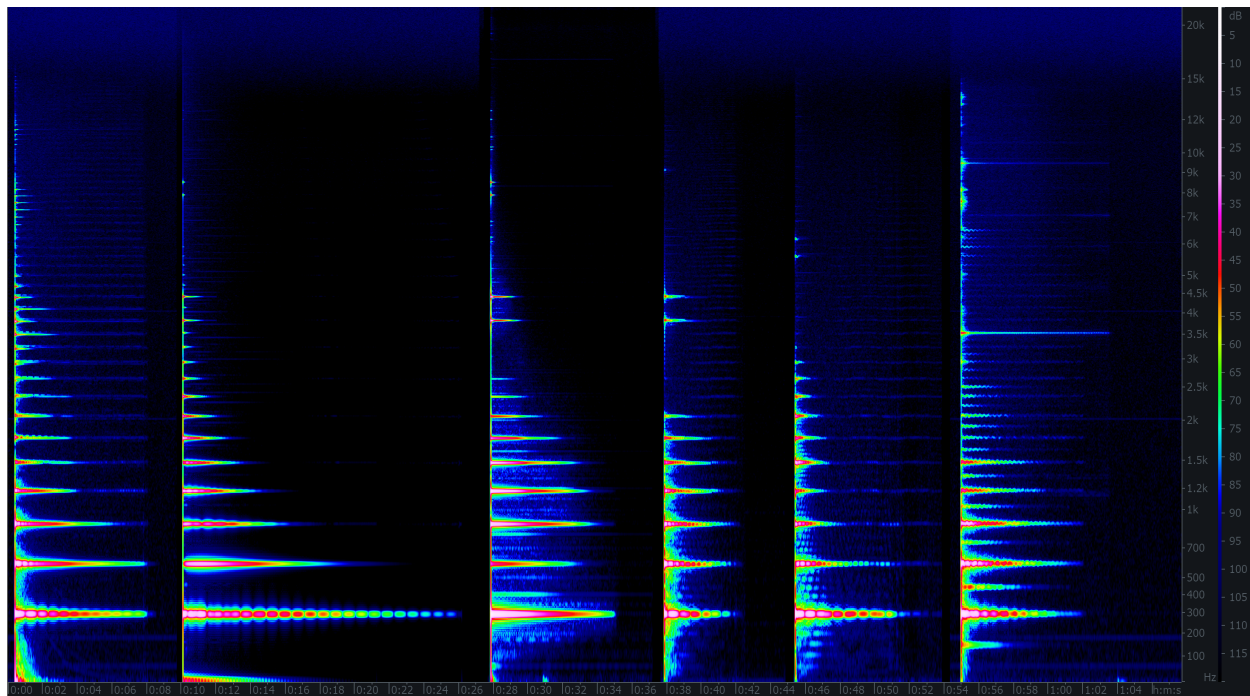


Figure 3: Spectrogram images of the electric piano sounds in Table 1.

References

- Fales, C. (2002). The Paradox of Timbre. *Ethnomusicology*, 46(1), 56–95.
<https://doi.org/10.2307/852808>
- Lavengood, M. L. (2019). "What Makes It Sound '80s?": The Yamaha DX7 Electric Piano Sound. *Journal of Popular Music Studies*, 31(3), 73–94. <https://doi.org/10.1525/jpms.2019.313009>
- Lavengood, M. L. (2020). The Cultural Significance of Timbre Analysis: A Case Study in 1980s Pop Music, Texture, and Narrative. *Music Theory Online*, 26(3). <https://doi.org/10.30535/mt.26.3.3>

Synthetic Timbres, Cinematic Worlds: Queen's *Flash Gordon* Soundtrack as a Paradigm for Sound Synthesis and Processing

Nicola Bizzo^{1†}

¹ CESEM, University of Lisbon, Portugal

[†] Corresponding author: wiz.eutropio@gmail.com

Timbre is not merely a secondary attribute of sound but one of the principal materials through which musical meaning is produced, perceived, and culturally encoded. This claim is especially productive in the study of film music and popular music, where sound color often mediates between compositional structure, technological mediation, and systems of representation. Queen's soundtrack for *Flash Gordon* (Hodges, 1980) offers an exemplary case in this respect. Released at a crucial moment in the band's development and in the wider history of analog synthesis, the soundtrack foregrounds synthetic timbre not as a decorative supplement but as the primary medium through which narrative space, affective intensity, and audiovisual identity are articulated. This article argues that *Flash Gordon* should be understood as a paradigmatic instance of timbral world-building: a work in which sound synthesis and studio processing function simultaneously as compositional resources, technological mediators, and cultural signifiers.

Such an approach also contributes to Queen scholarship. Much analytical work on the band has understandably privileged formal expansiveness, stylistic plurality, vocal layering, and the dialectic between operatic ambition and rock energy. Braae (2021), in particular, has shown how Queen's music is best understood through the idea of an idiolect constituted by recurring strategies of contrast, synthesis, and sonic excess. Yet *Flash Gordon* reveals another dimension of that idiolect: the capacity to reconfigure Queen's established theatricality through electronically mediated timbre. The soundtrack does not simply transpose the band's style into a cinematic setting. Rather, it intensifies the technological and spectral aspects of Queen's musical language, making timbre itself the principal bearer of futurity, spectacle, and estrangement.

The theoretical premise of this study is that timbre in popular audiovisual music must be read both structurally and semiotically. Middleton (1990) argues that the analysis of popular music cannot isolate musical detail from its social and cultural signification; Tagg (1999) similarly insists that musical parameters generate "paramusical" associations, linking sound to bodies, spaces, gestures, and ideological frames. From this perspective, synthetic timbre in *Flash Gordon* is not simply a matter of sonority in the narrow acoustical sense. It is a semiotic interface through which listeners identify technological otherness, imperial menace, erotic ambiguity, and heroic propulsion. Benjamin's (2008) reflections on technological reproducibility provide a further lens for understanding this phenomenon. In electronically generated and studio-processed music, the sound object is not merely reproduced; it is constituted as technologically mediated presence. In *Flash Gordon*, this condition is not concealed but aestheticized. The soundtrack foregrounds its own synthetic ontology, turning technical mediation into a core component of cinematic meaning.

At the same time, Queen's work on *Flash Gordon* needs to be placed within the band's broader aesthetics of performance and spectacle. Brooks (1995) documents the monumental scale and dramaturgical ambition of Queen's live performances, while Gunn and Jenkins (1992) emphasize the group's sustained investment in media-conscious self-fashioning. These aspects are highly relevant to the soundtrack. Although *Flash Gordon* belongs to the domain of film scoring rather than live performance, it extends the band's spectacular logic into the studio. The result is not a retreat from Queen's established identity but a rearticulation of it. Synthetic timbres, electronic layers, and studio effects become equivalent to costumes,

staging, and lighting in the band's concert work: devices for the construction of heightened presence. In this sense, the soundtrack is fully consistent with Queen's longstanding interest in the theatrical transformation of sound.

An organological perspective clarifies the originality of this procedure. Schaeffner (2020) reminds us that the history of instruments is inseparable from the history of human conceptions of sound, gesture, and cultural function. Read in these terms, the analog synthesizer (in this case, an Oberheim OB-X) is not simply a new tool added to an existing palette; it is an instrument that redefines the relation between performance, design, and musical authorship. Duxbury's (1985) account of the interchange between rock and classical traditions is also useful here, because Queen's soundtrack inhabits a territory where orchestral rhetoric, rock instrumentation, and electronic design intersect. *Flash Gordon* does not imitate the symphonic idiom typical of mainstream blockbuster scoring, nor does it reduce itself to a conventional rock soundtrack. Instead, it constructs a hybrid environment in which filters, envelope shapes, detuning, oscillatory instability, reverberant depth, and stereo placement operate as genuine compositional parameters. One might say that Queen are orchestrating with timbre rather than merely adding synthesizers to pre-existing songs.

Electronic timbre has long played a central role in science-fiction cinema, from the theremin in *The Day the Earth Stood Still* (1951) to the fully electronic soundtrack of *Forbidden Planet* (1956). Across these early examples, electronically generated sound became increasingly associated with alterity, technology, and the otherworldly. By the time of *Blade Runner* (1982), this association had become definitive: Vangelis's synthesizer-based score was so closely integrated with the film's atmosphere that electronic timbre itself came to signify futurity.

Methodologically, the article treats analysis as a mode of understanding, combining close listening, audiovisual cue analysis, organological reflection, and production-oriented interpretation. This approach is further supported by selected spectral, temporal, and dynamic descriptors applied comparatively to cues from the soundtrack and their corresponding film scenes.

Queen became involved with *Flash Gordon* in 1980, when they were commissioned to write the soundtrack for a new film adaptation. In doing so, the band combined synthesizers with orchestral forces, giving both the film and the score a scale and dramaturgical profile that at times recall grand opera. The use of **leitmotif** is equally significant: like Wagner's practice of associating recurring musical ideas with characters or dramatic situations, the score employs recognizable thematic cells to articulate narrative tension and identity. This approach was, of course, already well established in film music, and one of its most prominent twentieth-century examples is John Williams's work for the original *Star Wars* trilogy.

In *Flash Gordon*, the themes are often deliberately simple, but that very simplicity facilitates recognition and memorability, enabling viewers to associate them readily with particular characters and situations. The use of electronic instruments—still relatively new for Queen, who had only begun to employ synthesizers more extensively on *The Game*, released earlier the same year—gives the film a distinctive sonic orientation. In this sense, the score decisively shapes the viewer's perception of the film.

Queen's contribution to the project is predominantly instrumental, and an orchestral component associated with Howard Blake can be heard throughout the film. Yet the orchestral layer often remains in the background, while the synthetic and electronically processed elements emerge as the score's most distinctive and memorable features.

The principal excerpts considered are "Flash", "Ming's Theme (In the Court of Ming the Merciless)", "Football Fight", "The Ring (Hypnotic Seduction of Dale)", "The Kiss (Aura Resurrects Flash)" and

“Battle Theme”. In these cues, timbral behavior can be tracked through parameters such as brightness, textural density, attack profile, decay contour, modulation rate, and dynamic shaping. The purpose of such measurements is not positivistic quantification for its own sake, but the clarification of how timbral design supports narrative differentiation and cinematic pacing.

“Flash” establishes the soundtrack’s core aesthetic by combining emphatic declamation, rhythmic drive, and sharply defined synthetic timbres. Rather than simply introducing the protagonist, the cue frames a sonic world in which action, technology, and spectacle are inseparable, merging song form with sound design. A different strategy is audible in “Ming’s Theme” (see Figure 1).



Figure 1: Excerpt from “Ming’s Theme” (Freddie Mercury).

In this cue, villainy is not conveyed primarily through melodic complexity but through the darkened and pressurized quality of the timbral field. Low-register emphasis, sustained resonance, and unstable electronic color produce an effect of concentrated authority. The cue exemplifies how synthetic timbre can operate as political iconography. Ming’s sonic identity is not simply “evil” in an abstract sense; it is coded as technologically enhanced domination, an acoustic correlate of imperial power. Tagg’s semiotic framework is particularly useful here, since the cue mobilizes a network of paramusical associations—machinery, remoteness, ritualized menace, superhuman control—without requiring explicit verbal explanation. Timbre functions as narrative compression.

“Football Fight” and “Battle Theme” deploy timbre to convey propulsion, impact, and collective energy, showing how analog synthesis can construct action as a sensory event while sustaining epic scale within a rock idiom. By contrast, “The Ring” and “The Kiss” explore timbre’s more ambiguous and transformative potential, using filtered motion, softened contours, and reverberant diffusion to evoke seduction, suspension, and threshold states (see Figure 2).



Figure 2: Excerpt from “The Kiss” (Freddie Mercury).

The cue presents timbre as a medium of transformation, making resurrection sensorially credible through a technologically mediated sonic aura. In this way, its emotional intensity arises not despite technical mediation, but through it.

Taken together, these cues demonstrate that *Flash Gordon* is not simply an episode in Queen's discography or a curious detour into film work. It is a major statement about timbre as an artistic, technological, and cultural problem. In *Flash Gordon*, sound synthesis is not reducible to equipment, novelty, or genre marker. It becomes a form of composition in its own right, one that organizes space, encodes character, and gives audible shape to futuristic imagination.

The broader implication is that timbre should occupy a more central place in the study of popular music and media. If analysis is to account fully for how music means, it must attend not only to pitch, form, and text, but to the concrete materiality of sound and to the cultural histories embedded in instrumental and studio technologies. Queen's *Flash Gordon* soundtrack provides an especially rich site for such work because it unites organology, semiotics, production, and audiovisual narration within a single corpus. Its enduring value lies in showing that the futuristic in music is not only a matter of subject matter or iconography. It is something composed in and through timbre.

References

- Benjamin, W. (2008). *The work of art in the age of its technological reproducibility* (Second version). In M. W. Jennings, B. Doherty, & T. Y. Levin (Eds.), *The work of art in the age of its technological reproducibility, and other writings on media* (pp. 19–55). Harvard University Press.
- Braae, N. (2021). *Rock and rhapsodies: The music of Queen*. Oxford University Press.
- Brooks, G. (1995). *Queen live: A concert documentary*. Omnibus Press.
- Duxbury, J. R. (1985). *Rockin' the classics and classicizin' the rock: A selectively annotated discography*. Greenwood Press.
- Eggebrecht, H. H. (1997). *Comprendere attraverso l'analisi. Il saggiatore musicale*, 4(2), 373–384.
- Gunn, J., & Jenkins, J. (1992). *Queen: As it began: The authorized biography*. Sidgwick & Jackson.
- Hodges, M. (Director). (1980). *Flash Gordon* [Film]. Dino De Laurentiis Corporation.
- Middleton, R. (1990). *Studying popular music*. Open University Press.
- Queen. (1980). *Flash Gordon* [Album]. EMI.
- Schaeffner, A. (2020). *The origin of musical instruments: An ethnological introduction to the history of instrumental music* (R. Taylor, A. Lih, & E. Lih, Trans.). Routledge.
- Tagg, P. (1999, July). *Introductory notes to the semiotics of music* (Version 3) [Manuscript]. <https://www.tagg.org/xpdfs/semiotug.pdf>

Session 14

Communicating About Timbre

A Multi-Space Hierarchical Timbre Model: Perceptually Grounded Prediction of Timbral Similarity

Siavosh Banihashemi^{1†}

¹ Music Psychology and Brain Research, Department of Psychology, University of Graz, Austria

[†] Corresponding author: siavosh@live.at

Abstract

This paper presents a computational model of musical timbre based on perceptually interpretable audio descriptors and multisensory-inspired integration principles. The proposed Multi-Space Hierarchical Timbre Model integrates harmonic, spectral, and temporal information across three perceptual spaces: Relative Spectral Energy Distribution, Absolute Spectral Energy Distribution, and Temporal Energy Distribution. Using isolated instrumental tones, the model was evaluated against behavioral similarity ratings and compared with MFCC-based baselines. Cross-validated analyses showed strong correlations with perceptual data ($r = 0.78\text{--}0.82$), consistently outperforming MFCC models while maintaining full interpretability. Results suggest that hierarchically combining perceptually grounded descriptors provides an effective and transparent framework for timbre similarity prediction.

Introduction

Timbre perception arises from interactions among spectral structure, energy distribution, and temporal evolution (McAdams & Giordano, 2008; McAdams, 2019). Although many computational models achieve high predictive accuracy, they often rely on abstract or high-dimensional representations with limited perceptual interpretability. This restricts their applicability in composition, orchestration, and music pedagogy, where understanding the relationship between acoustic features and perception is essential.

The proposed Multi-Space Hierarchical Timbre Model (MSHTM) addresses this limitation by integrating psychoacoustically meaningful descriptors within a structured framework. Timbre is represented across three perceptual domains, each comprising dimensions derived from timbre descriptors (Peeters et al., 2011). First, the conceptual framework of the MSHTM is introduced. Next, Experiment 1 identifies perceptually relevant descriptors for the selected stimuli. Based on these descriptors, similarity predictions are generated by the model. Finally, these predictions are compared with perceptual similarity ratings obtained in Experiment 2 to evaluate the predictive validity of the model.

Model Structure

A total of 42 timbre descriptors were extracted using the Timbre Toolbox (Peeters et al., 2011), covering spectral, harmonic, and temporal properties. Additionally, a formant-based descriptor using a modified double-Bark scale (DB) was introduced. The DB descriptors are based on psychoacoustic evidence suggesting that formants are perceived categorically within overlapping critical bands rather than as precise frequency values (Zwicker, 1961; Xu et al., 2004). Spectral peaks were therefore mapped onto a reduced double-Bark representation, providing a perceptually relevant description of dominant and secondary formant regions while reducing dimensional complexity.

For each descriptor, nine temporal summary measures were computed, including the median, maximum, and minimum of the descriptor trajectory, as well as the median, maximum, and minimum of its first and

second derivatives. Principal component analysis (PCA) reduced these measures to a single component per descriptor, yielding a 42-dimensional representation.

Descriptors were grouped into three perceptual spaces: Relative Spectral Energy Distribution (RSED), Absolute Spectral Energy Distribution (ASED), and Temporal Energy Distribution (TED). Pairwise similarities were computed independently within each space using Euclidean distance and subsequently combined using equal-weighted linear integration. Figure 1 illustrates the proposed model, which is structured around three acoustically and perceptually salient domains:

RSED captures the distribution of energy across partials, reflecting perceptual attributes such as harmonic balance (e.g., odd–even ratios, harmonicity) and temporal

ASED describes energy across absolute frequency regions, representing features such as spectral centroid, low–high frequency balance, formant structure, and their temporal evolution.

TED models the temporal evolution of overall energy, including attack, sustain, modulation, and decay characteristics relevant to timbre perception.

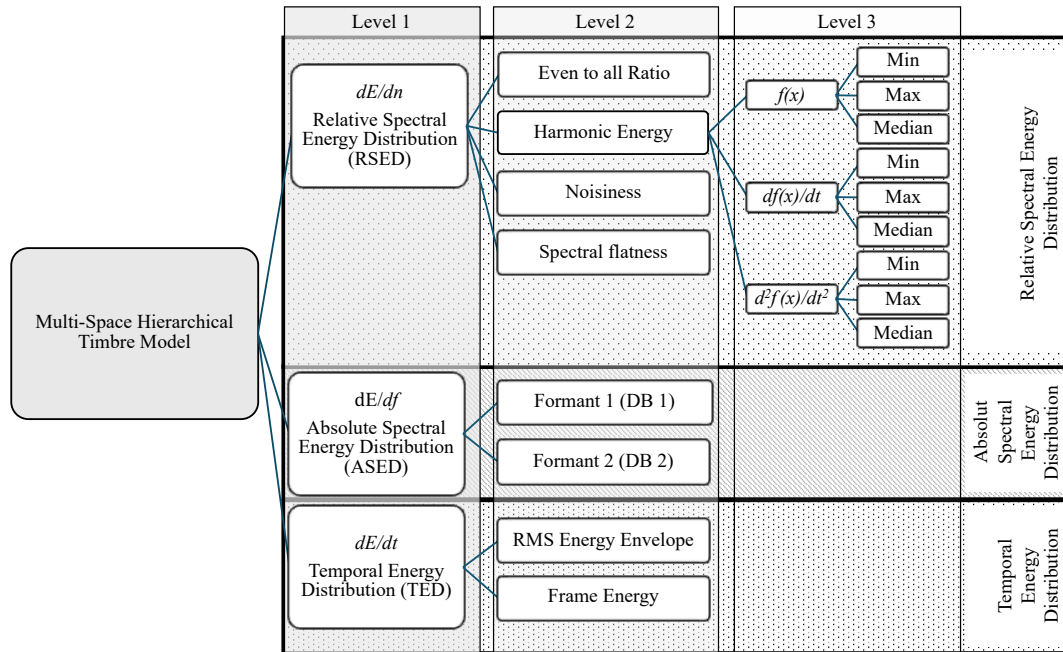


Figure 1: Hierarchical structure of the Multi-Space Hierarchical Timbre Model.

The three domains are represented as distinct layers, while hierarchical processing is organized into three levels. At the first level, each domain is defined by a set of descriptors. For each descriptor, nine temporal measures are computed and reduced using Principal Component Analysis (PCA), retaining only the first principal component (PC1) as a compact representation of perceptually relevant variance.

At the second level, pairwise similarities between stimuli are computed independently within each domain. PC1 values define the dimensions of the corresponding timbre spaces, and similarity is measured using Euclidean distance, yielding three domain-specific similarity estimates per stimulus pair.

At the third level, these estimates are integrated using a weighted linear combination inspired by multi-cue perceptual integration. Each domain contributes proportionally to its reliability, consistent with probabilistic models of perception (Ma, Beck, Latham, & Pouget, 2006; Trommershäuser, Kording, & Landy, 2011). In the present study, equal weighting is assumed for simplicity, although the framework allows adaptive weighting depending on stimulus characteristics and listener factors.

Experiment 1: Similarity of Musical Tones to Prototypes, Descriptor Selection

Forty participants ($M = 24.5$ years, $SD = 8.85$) with normal hearing participated in the study. Ninety-five percent reported experience playing a musical instrument. Stimuli were selected from IRCAM's TinySOL dataset (Cella, 2020) and consisted of 45 recordings (400 ms) of five wind instruments (clarinet, flute, oboe, saxophone, and trumpet) across three pitch levels (C4, A4, F5) and three dynamic levels: fortissimo (ff), mezzo forte (mf), and pianissimo (pp). All tones were steady-state, normalized for perceived loudness, and presented via headphones in a controlled environment.

Participants rated the similarity of each stimulus to a reference instrument. Results indicated that instrument identity was the strongest factor influencing similarity judgments, followed by dynamic level and pitch. Gender showed no significant effect ($p = 0.850$), whereas age demonstrated a weak trend ($p = 0.094$).

Correlation analyses (Pearson and Spearman) between acoustic features and perceptual ratings identified descriptors consistently contributing to similarity judgments. In the RSED space, energy of even to all harmonic ration, Harmonic Energy, Noisiness, and Spectral flatness were significant. In the ASED space, the proposed double-Bark formant descriptor (DB) showed strong relevance. In the TED space, RMS-Energy Envelope and Frame Energy were consistently significant across instruments.

Model Implementation

The MSHTM was implemented in MATLAB using the same 45 stimuli. Only descriptors identified as perceptually relevant in Experiment 1 were retained. Descriptor measures were normalized to the range $[0, 1]$, reduced using PCA, grouped into the three perceptual spaces, and compared via Euclidean distance. Final similarity predictions were obtained through equal-weighted summation of the three similarity matrices.

Experiment 2: Perceptual Similarity and Model Validation

The same participants rated pairwise similarity for all stimulus pairs from Experiment 1 ($N = 675$) across three pitch-specific sessions using a continuous rating scale. Results revealed non-normal rating distributions but consistent perceptual structures. Instruments such as oboe and trumpet exhibited stable timbral identities, whereas saxophone showed greater variability across conditions.

Predictive validity was assessed by correlating model predictions with mean perceptual similarity ratings from Experiment 2. The MSHTM showed strong correlations across pitch conditions (C4: $r = 0.80$, A4: $r = 0.82$, F5: $r = 0.78$). Among the three spaces, ASED contributed most strongly to prediction accuracy, followed by RSED, while TED had a smaller influence.

Comparison With MFCC Models

To benchmark the model, MSHTM predictions were compared with MFCC-based similarity models using 13, 24, and 40 coefficients. MFCCs were extracted from short-time frames and compared using

Euclidean distance. MFCC models showed significantly lower correlations with perceptual data ($r = 0.28 - 0.40$), with the best performance observed at $r = 0.40$. Increasing the number of coefficients did not improve accuracy. In contrast, the MSHTM consistently achieved correlations above $r = 0.78$. Statistical tests confirmed that the MSHTM significantly outperformed all MFCC configurations ($p < 0.001$).

Conclusion and Discussion

The results indicate that spectral shape and formant-related features play a dominant role in timbre similarity perception under steady-state conditions for the present stimulus set. Future studies using stimuli with greater temporal variation may clarify the role of temporal descriptors and the relative contribution of each perceptual space.

The proposed MSHTM provides a perceptually interpretable framework for timbre similarity prediction by integrating spectral and temporal information across hierarchical perceptual spaces. Results suggest that perceptual relevance of descriptors is more critical than feature dimensionality alone, as demonstrated by substantial improvements over MFCC-based models. Future work should investigate adaptive weighting strategies and applications to more temporally dynamic stimuli and complex musical contexts.

References

- Cella, C.-E. (2020). TinySOL [Data set]. IRCAM. <https://forum.ircam.fr/projects/detail/tinysol/>
- Ma, W. J., Beck, J. M., Latham, P. E., & Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nature Neuroscience*, 9(11), 1432-1438.
- McAdams, S. (2019). The Perceptual Representation of Timbre. In K. Siedenburg, C. Saitis, S. McAdams, A. Popper, & R. Fay, *Timbre: Acoustics, Perception, and Cognition*.
- McAdams, S., & Giordano, B. L. (2008). The perception of musical timbre. In S. Hallam, I. Cross, & M. Thaut, *The Oxford handbook of music psychology* (pp. 72-80). Oxford University Press.
- Peeters, G., Giordano, B. L., Susini, P., Misdariis, N., & McAdams, S. (2011, November). The Timbre Toolbox: Extracting audio descriptors from musical signals. *J. Acoust. Soc. Am.*, 130(5), 2902–2916.
- Trommershäuser, J., Kording, K. P., & Landy, M. S. (2011). *Sensory cue integration*. Oxford University Press.
- Xu, Q., Jacewicz, E., Feth, L. L., & Krishnamurthy, A. K. (2004, April). Bandwidth of spectral resolution for two-formant synthetic vowels and two-tone complex signals. *J. Acoust. Soc. Am.*, 115(4), 1653-1664.
- Zwicker, E. (1961). Subdivision of the audible frequency range into critical bands. *The Journal of the Acoustical Society of America*, 33(2), 248.

**From Terms to Concepts:
Towards an Epistemology of North-American Timbre Theory and Analysis
of the Mid-1970s**

Nathalie Hérold^{1†}

¹ Faculty of Arts and Humanities/IReMus (UMR 8223), Sorbonne University, Paris, France

[†] Corresponding author: nathalie.herold@sorbonne-universite.fr

In the field of music theory and analysis, timbre-based studies can be considered as relatively new in comparison to approaches related to other music parameters – in particular pitch and duration. Meyer (1989) still assigned timbre a status of secondary parameter in terms of syntax and Christensen (2002) considered timbre in music theory principally in an indirect way, in relation to acoustics, psychology, and twelve-tone theory. In contrast, more recent research argued in favor of a structural role of timbre in several musical repertoires, from the 18th to the 20th centuries and from piano to orchestra, as evidenced – among others – by Hérold (2011) and Dolan (2013), as well as Guigue (2009). Contrary to the first two, this latter work – although clearly part of a timbral approach – used however the term “timbre” with great caution, preferring the expression “unité sonore composée” (composite sound unit) (Guigue, 2009, p. 37). A similar terminological issue is also raised by the idea of a Taxonomy of Orchestral Grouping Effects (TOGE) designed for analytical purposes by McAdams, Goodchild, and Soden (2022), which relates timbre to orchestration and declines it within several categories and sub-categories referred to by specific terms such as “timbral augmentation”, “timbral emergence”, “surface texture”, “stratification”, “timbral contrasts”, etc.

Based on this observation, the following question arises: how is it possible to consider the diversity of terms related to the general concept of timbre in music analysis and theory, and the different systems and conceptions that underlie them, beyond apparent terminological differences? The aim of this paper is not to answer this question in general terms, but rather to focus on a specific case study. As highlighted by Hérold (2020), the genesis of timbre-based theory and analysis in the mid-1970s was a particularly fruitful period in North America for the development of new concepts and theoretical systems incorporating timbre aspects as an essential dimension. This paper thus proposes to examine a corpus of three pioneering works in this field: *Sound Structure in Music* by Robert Erickson (1975), which gives the term “timbre” an essential role, *Sonic Design: The Nature of Sound and Music* by Robert Cogan and Pozzi Escot (1976), which bases its reflection on the expression “sound color”, and *Structural Functions in Music* by Wallace Berry (1976), which employs the term “texture” to refer to a primary musical and sound dimension. Due to their strong conceptual convergence despite surface differences in terminology, it will be interesting to consider these three works individually and in relation to each other.

This study, which takes the form of an analysis of analyses, in other terms a meta-analysis, is based on a number of methods developed in relation to the field of timbre semantics (Wallmark, 2018; Wallmark & Kendall, 2021; Saitis & Weinzierl, 2019), applying them to the corpus of three books mentioned above from a specifically theoretical and analytical perspective, in line with research such as Reymore (2020). First, it will be necessary to identify and extract from the corpus textual data such as terms and lexical fields related to the general concept of timbre, in order to examine their definitions – whether explicit or implicit –, their recurrence patterns, and their relationships – in particular through the study of co-occurrences. Data statistics, structuring and modeling, in particular through categorization and the establishment of diagrams highlighting conceptual organizations and hierarchies, will be particularly useful, along with the application of computational and digital methods – partly based on artificial intelligence – in order to automate some analytical procedures given the amount of data to be managed (Guillotel-Nothmann, 2026). Then, it will also be important to consider the internal contexts of the terms’

usage: the musical repertoires (periods, genres, cultures) to which these terms and concepts relate, the analytical representations they give rise to, and above all their role within the general theoretical systems set out in each of the corpus books, namely in relation to other dimensions such as harmony, melody, rhythm, dynamics, etc. Finally, it will be essential to compare the use of theoretical and analytical terms and concepts pertaining to timbre in the three works of the corpus, in order to highlight convergences and divergences between the different systems proposed by the authors, while contextualizing them within the scientific and technological environment of the time – both from a disciplinary and interdisciplinary perspective.

Through this case study, this paper aims to outline an epistemology of North-American timbre theory and analysis of the mid-1970s from a synchronic perspective. Based on an articulation between qualitative and quantitative approaches applied to textual sources through the creation of a specific database, this research can be considered as a pilot study, with a view to extending it in the future to other textual corpus, in order to gain a better understanding of timbre-based theory and analysis from a diachronic perspective, from its beginnings to current developments.

References

- Berry, W. (1976). *Structural Functions in Music*. Englewood Cliffs (NJ): Prentice-Hall.
- Christensen, T. (Ed.). (2002). *The Cambridge History of Western Music Theory*. Cambridge: Cambridge University Press.
- Cogan, R., & Escot, P. (1976). *Sonic Design: The Nature of Sound and Music*. Englewood Cliffs (NJ): Prentice-Hall.
- Dolan, E. I. (2013). *The Orchestral Revolution: Haydn and the Technologies of Timbre*. Cambridge (MA): Cambridge University Press.
- Erickson, R. (1975). *Sound Structure in Music*. Berkeley (CA), Los Angeles (CA), London: University of California Press.
- Guigue, D. (2009). *Esthétique de la sonorité: l'héritage debussyste dans la musique pour piano du XX^e siècle*. Preface by Makis Solomos. Paris: L'Harmattan.
- Guillotel-Nothmann, C. (2026, June 4–5). Beyond Words: Team-Based AI Approaches to Unlocking Conceptual Structures in Early Modern Music Theory at the IReMus [Paper presentation]. 2nd Annual Conference of the RMA French Music Study Group, Paris, France.
- Hérol, N. (2011). *Timbre et forme: la dimension timbrique de la forme dans la musique pour piano de la première moitié du dix-neuvième siècle* [Doctoral dissertation, University of Strasbourg]. theses.fr. https://publication-theses.unistra.fr/public/theses_doctorat/2011/HEROLD_Nathalie_2011.pdf
- Hérol, N. (2020). *Le timbre face à l'hétérogénéité musicale: conceptualisation, enjeux, pertinence*. In M.-N. Masson (Ed.), *Entre théorie et analyse musicale: corpus et méthodes* (pp. 189–211). Sampzon: Delatour.
- McAdams, S., Goodchild, M., & Soden, K. (2022). A Taxonomy of Orchestral Grouping Effects Derived from Principles of Auditory Perception. *Music Theory Online*, 28(3). <https://mtosmt.org/issues/mto.22.28.3/mto.22.28.3.mcadams.html>
- Meyer, L. (1989). *Style and Music: Theory, History, and Ideology*. Philadelphia (PA): University of Pennsylvania Press.
- Reymore, L. (2020). *Empirical Approaches to Timbre Semantics as a Foundation for Musical Analysis* [Doctoral dissertation, Ohio State University]. OhioLINK Electronic Theses and Dissertations Center. http://rave.ohiolink.edu/etdc/view?acc_num=osu1586792450387823

- Saitis, C., & Weinzierl, S. (2019). The Semantics of Timbre. In K. Siedenburg, C. Saitis, S. McAdams, A. N. Popper & R. R. Fay (Eds.), *Timbre: Acoustics, Perception, and Cognition* (pp. 119–49). Cham: Springer.
- Wallmark, Z. (2018). A Corpus Analysis of Timbre Semantics in Orchestration Treatises. *Psychology of Music*, 47(4), 585–605.
- Wallmark, Z., & Kendall, R. A. (2021). Describing Sound: The Cognitive Linguistics of Timbre. In E. I. Dolan & A. Rehding (Eds.), *The Oxford Handbook of Timbre* (pp. 579–608). New York (NY): Oxford University Press.

The Classroom Case for Timbre

Holly Bergeron-Dumaine^{1†}

¹ School of Music, University of British Columbia, Vancouver, British Columbia, Canada

[†] Corresponding author: hollyber@student.ubc.ca

Recent critiques from within the field of music theory (e.g. Ewell, 2023; Hisama, 2021) have sought to redress persistent disciplinary inequities and biases through calling for significant epistemological shifts. Pedagogical critiques similarly highlight the necessity of welcoming a broader variety of contexts while deemphasizing western classical repertoires in favour of broader pluralities (Attas, 2019; Campbell et al., 2016; Cunningham et al., 2020; Hess, 2016). Pedagogical work continues to underscore the importance of moving beyond merely diversifying repertoire, to reconsidering the way in which music theory is taught (Rush & Acevedo, 2025).

I argue that granting timbre a core place within the undergraduate music theory core curriculum directly responds to these prompts for reparative action. Now solidly established, timbre studies exhibit considerable methodological richness which instructors can distill and leverage in reconceiving both curricular content and teaching approaches. I discuss three key rationales for making timbre an integral part of the theory core, highlighting connections to literature and suggesting examples suitable for the classroom. Equipped with these examples, I propose activities suitable for first- or second-year theory that introduce learners to new tools and ways of listening.

Timbre impels instructors and learners to engage with a variety of representations of sound, helping the curriculum move away from a monolithic music literacy and toward a plurality of music *literacies* encompassing a wider spectrum of skills relevant to contemporary musicians (Abrahams, 2021). Now mainstays of timbre scholarship, spectrograms and textural-formal diagrams (as used by e.g. Dolan, 2013; Malawey, 2020; Peres, 2016) each foreground dimensions of sonic experience largely abstracted away by western staff notation. Working with these modern “instruments of music theory” (Rehding, 2016) explicitly legitimizes new epistemic frames; these visualizations effectively disrupt the assumption that pitch relationships primarily shape musical experience and conversely foreground the structural role of timbre (McAdams, 2019). Paralleling scholarship, instructors can leverage spectrograms to teach formal analysis using learners’ experience of texture as a platform without recourse to a score. I outline such an analysis of the opening of Sergei Rachmaninov’s 1909 tone poem *Isle of the Dead* (Figure 1). Instructors can combine formal analysis with an intuitive primer on spectrograms and their epistemic affordances, validating alternatives to pitch-based analysis.

When used in conjunction with comparative analysis, this experience-and-spectrogram formula helps to build a more pluralistic curriculum which admits many different repertoires from the outset. Megan Lavengood (2021) has argued that paying greater attention to rhythm and texture as structuring parameters valorizes oft-marginalized repertoires such as EDM; similarly, comparing and contrasting textural processes in works drawn from a variety of styles and periods places popular, contemporary art, and western classical genres on a level playing field. Informed by Figures 2 and 3, I discuss a three-way comparison of *Isle of the Dead* with short pieces of a similar length—electronic producer Laxcity’s “run/fly” (2026), and Anna Clyne’s “River” from the orchestral suite *Abstractions* (2016). Putting these pieces in dialogue defuses the skewed “hidden curriculum” of classical repertoire (Palfy & Gilson, 2018). Moreover, it productively challenges the implicit notion that increased pitch complexity and proximity to current day means higher aural and analytic barriers to entry—a narrative that traditional core curricula reproduce by “locking” contemporary repertoires and popular musics behind end-of-sequence courses or electives (Clercq, 2019; Endicott, 2020).

More broadly, the immediacy and subjectivity of timbral experience invites productive engagement with the *affective*—as opposed to strictly *cognitive*—dimensions of learning and practicing music theory. Theorized by Kratwohl, Bloom, and Masia (1964), the affective domain of learning encompasses outcomes often relegated to the hidden curriculum. These outcomes can range in complexity from capacity to focus on a particular dimension of musical experience, to cultivating a broader analytic attitude toward the music one encounters. Instructors can leverage the tension between timbre’s semiotic power and its difficulty of description to foreground explicitly affective work, prompting learners to describe timbral experiences and consider their interpretations of musical meaning. I propose an activity that prompts learners to compare Queen’s “I Want to Break Free” (1984) with a translated and rearranged cover by Inuk artist Elisapie (from the 2023 album *Inuktitut*). Learners listen to both versions of the song, then describe their emotive response to each. In groups, they generate descriptions of timbres and textures they hear while discussing which changes from version to version might connect to changes in interpretation and response. The instructor summarizes points of convergence and divergence in interpretation and description, connecting analysis to questions of instrumental associations, vocal proxemics, and social context. Through this activity, the “materialist/idealist” divide (Van Elferen, 2020) becomes a platform for valorizing individual learners’ musicianship and affective capacities, framing music theory as an active, subjective practice of interpretation informed by a keen analytic ear.

Taken together, these avenues of pedagogical change promote a reconceived curriculum that transforms learner’s conceptions of what music theory can “be,” “do,” and “be about.” Timbre facilitates classroom practices that espouses the growing epistemic and stylistic plurality championed in the field at large, enabling theory in the classroom to become an active relational practice (see Luong, 2017) in which undergraduates critically appraise notation, genre, and their own listening experiences.

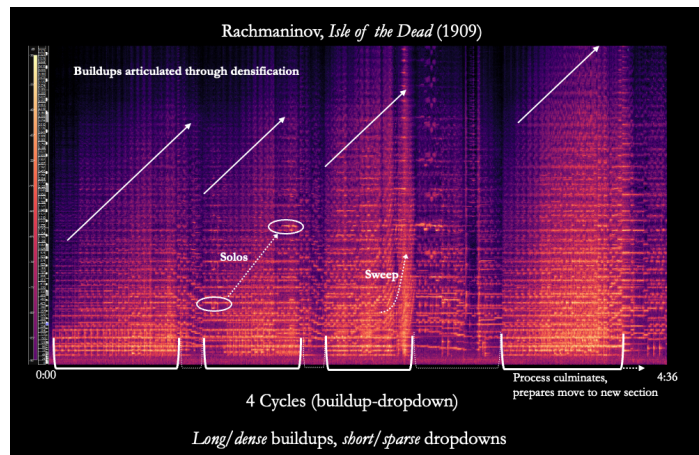


Figure 1: Spectrogram of excerpt from Sergei Rachmaninov’s Isle of the Dead, annotated with descriptions of textural processes.

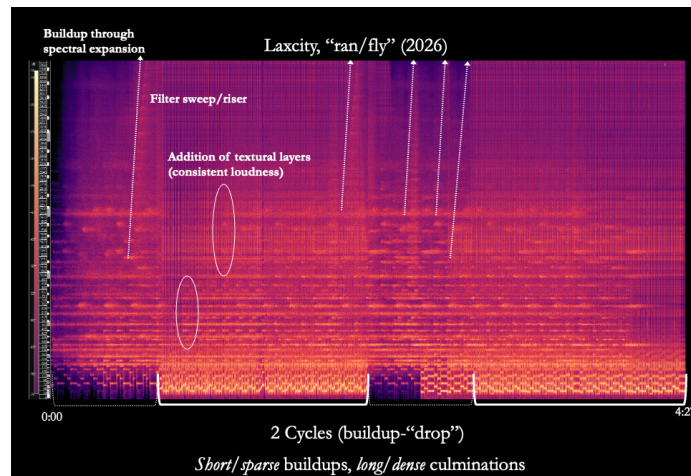


Figure 2. Spectrogram of Laxcity's "ran/fly," annotated with descriptions of timbral processes.

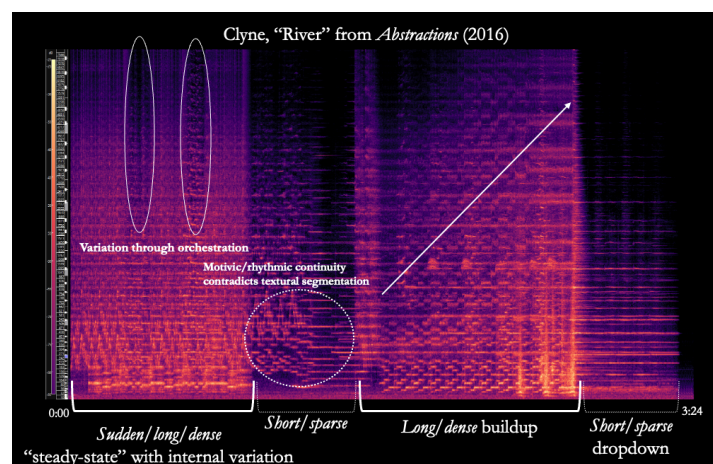


Figure 3. Spectrogram of Anna Clyne's "River" from *Abstractions* (2016), annotated with descriptions of timbral processes.

References

- Abrahams, R. (2021). Rethinking Music Literacy in the Undergraduate Theory Core. *Journal of Music Theory Pedagogy*, 35 (Article 2). <https://doi.org/https://doi.org/10.71156/2994-7073.1223>
- Attas, R. (2019). Strategies for Settler Decolonization: Decolonial Pedagogies in a Popular Music Analysis Course. *Canadian Journal of Higher Education*, 49, 125–139.
- Campbell, P. S., Myers, D., & Sarath, E. (2016). *Transforming Music Study from its Foundations: A Manifesto for Progressive Change in the Undergraduate Preparation of Music Majors*. College Music Society. <https://www.music.org/pdf/pubs/tfumm/TFUMM.pdf>

- Clercq, T. de. (2019). A Music Theory Curriculum for the 99%. *Engaging Students: Essays in Music Pedagogy*, 7. <https://doi.org/10.18061/es.v7i0.7359>
- Cunningham, M., Robinson, D., Stover, C., Tilley, L., & Wang, A. Y. (2020). Introduction: Beyond Western Musicalities. *Engaging Students: Essays in Music Pedagogy*, 8. <https://doi.org/10.18061/es.v8i0.8653>
- Dolan, E. I. (2013). *The Orchestral Revolution: Haydn and the Technologies of Timbre*. Cambridge University Press.
- Endicott, N. (2020). Mamlok on Day One: Strategies for Incorporating Recent Music Into the Early Undergraduate Music Theory Curriculum. *Journal of Music Theory Pedagogy*, 34 (Article 8). <https://doi.org/10.71156/2994-7073.1219>
- Ewell, P. (2023). *On Music Theory, and Making Music More Welcoming for Everyone*. University of Michigan Press.
- Hess, J. (2016). “How Does That Apply to Me?” The Gross Injustice of Having to Translate. *Bulletin of the Council for Research in Music Education*, (207–208), 81–100.
- Hisama, E. M. (2021). Getting to Count. *Music Theory Spectrum*, 43(2), 349–363.
- Krathwohl, D. R., Bloom, B. S., & Masia, B. B. (1964). *Taxonomy of Educational Objectives, Handbook II: Affective Domain*. David McKay Co.
- Lavengood, M. L. (2021). Timbre, Rhythm, and Texture within Music Theory’s White Racial Frame. In L. M. Garcia-Mispireta & R. James (Eds.), *The Oxford Handbook of Electronic Dance Music* (Online edn). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780190093723.013.17>
- Luong, V. (2017). Rethinking Music Loving. *Music Theory Online*, 23(2). <https://doi.org/10.30535/mto.23.2.4>
- Malaway, V. (2020). *A Blaze of Light in Every Word: Analyzing the Popular Singing Voice*. Oxford University Press.
- McAdams, S. (2019). Timbre as a Structuring Force in Music. In K. Siedenburger, C. Saitis, S. McAdams, A. N. Popper, & R. R. Fay (Eds.), *Timbre: Acoustics, Perception, and Cognition* (pp. 211–243). Springer International Publishing.
- Palfy, C. S., & Gilson, E. (2018). The Hidden Curriculum in the Music Theory Classroom. *Journal of Music Theory Pedagogy*, 32 (Article 5), 79–110. <https://doi.org/10.71156/2994-7073.1194>
- Peres, A. (2016). *The Sonic Dimension as Dramatic Driver in 21st-Century Pop Music* (Ph.D. dissertation, University of Michigan).
- Rehding, A. (2016). Instruments of Music Theory. *Music Theory Online*, 22(4). <https://doi.org/10.30535/mto.22.4.4>
- Rush, T. W., & Acevedo, S. (2025). Deceptive Gestures: Lies Our “Diverse” Curriculum Is Telling. *Engaging Students: Essays in Music Pedagogy*, 9. <https://doi.org/10.18061/esm.6981>
- Van Elferen, I. (2020). *Timbre: Paradox, Materialism, Vibrational Aesthetics*. Bloomsbury.

Toward a Cross-Disciplinary Framework for Strategic Communication About Timbre

Lindsey Reymore^{1†}

¹ School of Music, Dance and Theatre, Arizona State University, Tempe, AZ, USA

[†] Corresponding author: lreymore@asu.edu

The complexities of timbre description are rich subjects for inquiry, offering insight into how listeners construct meaning from sound. Yet, these same complexities also pose persistent challenges. Timbre-semantic clarity is critical for effective pedagogy, for advancing theories of music, and for developing new analytical techniques. In creative contexts, shared language is essential for successful communication between composers and performers. More broadly, developing standard—but flexible—approaches to timbre description can support more frequent and more innovative collaborations among academic researchers, practitioners, and industry partners.

Fales (2002) introduced “the paradox of timbre,” observing that while hearing timbre is paramount to identifying sound sources, timbre perception is not equivalent to the acoustic signal. I identify a related phenomenon—the paradox of timbre semantics. On one hand, timbre is arguably the most accessible musical parameter. With minimal guidance, listeners can describe their phenomenal experience of a sound without need for formal training or specialized, technical vocabulary. On the other hand, timbre has historically been treated as the most elusive, inaccessible parameter, eschewed for its supposed resistance to language. As a result, timbre has little status in traditional music pedagogy, often appearing only briefly, if at all, in undergraduate curricula—despite the central importance of tone quality and expressive timbral nuance in musical practice.

I review four sources of timbre-related vocabulary: audio features (e.g., Caetano, Saitis, & Siedenburt, 2019; Kazazis, Depalle, & McAdams, 2021), Aural Sonology (Thoresen, 2015), qualitative descriptions (Carron, et al., 2017; Reymore, 2022; Reymore & Huron, 2020; Reymore et al., forthcoming; Wallmark, 2019), and the Taxonomy of Orchestral Grouping Effects (McAdams, Goodchild, & Soden, 2022). I argue that a balanced, productive framework must incorporate multiple strategies spanning both objective and subjective description. No single lexicon is sufficient; the “best” vocabulary depends on context. Accordingly, the framework is conceived not as a prescriptive method but as a toolkit, in which multiple lexical approaches act as complementary resources that can be strategically leveraged to improve communication.

At the highest level, the framework distinguishes between physical characteristics of sound and perceptual qualities of timbre; next, it differentiates among functional references to “sound,” “timbre,” and “orchestration.” For quantitative, measurable aspects of sound, a bespoke subset of audio features provides a shared technical vocabulary, including descriptors such as spectral centroid, inharmonicity, and noisiness. For timbre description, the framework integrates common qualitative terms, drawn from empirical studies, with the envelope-based descriptive approach of Aural Sonology, facilitating differentiated characterization of onset, sustain, and release. Vocabulary drawn from the Taxonomy of Orchestral Grouping Effects supports descriptions of the perceptual effects of the combination or juxtaposition of sound sources.

The framework distills these four lexicons to a limited number of terms manageable for entry-level learning, such as in the undergraduate music classroom. Additional terms from the original sources can be introduced with more advanced levels of study and adapted according to particular musical goals. In this way, the framework supports both accessibility and depth, offering a scalable approach to timbre

semantics that facilitates communication across disciplinary boundaries while respecting the richness and complexity of timbral experience.

References

- Caetano, M., Saitis, C., & Siedenburg, K. (2019). Audio content descriptors of timbre. In *Timbre: Acoustics, Perception, and Cognition* (pp. 297–333). Springer.
- Carron, M., Rotureau, T., Dubois, F., Misdariis, N., & Susini, P. (2017). Speaking about sounds: a tool for communication on sound features. *Journal of Design Research*, 15(2), 85–109.
- Fales, C. (2002). The paradox of timbre. *Ethnomusicology*, 46(1), 56–95. <https://doi.org/10.2307/852808>
- Kazazis, S., Depalle, P., and McAdams, S. (2021). The Timbre Toolbox Version R2021a, User's Manual. Available online at: <https://github.com/MPCL-McGill/TimbreToolbox-R2021a>
- McAdams, S., Goodchild, M., & Soden, K. (2022). A taxonomy of orchestral grouping effects derived from principles of auditory perception. *Music Theory Online*, 28(3). <https://doi.org/10.30535/mto.28.3.6>
- Reymore, L. (2022). Characterizing prototypical musical instrument timbres with Timbre Trait Profiles. *Musicae Scientiae*, 26(3), 648–674. <https://doi.org/10.1177/10298649211001523>
- Reymore, L., & Huron, D. (2020). Using auditory imagery tasks to map the cognitive linguistic dimensions of musical instrument timbre qualia. *Psychomusicology: Music, Mind, and Brain*, 30(3), 124–144. <https://doi.org/10.1037/pmu0000263>
- Reymore, L., Nestorova, T., Soden, K., & Marchand Knight, J. (forthcoming). Timbre semantics of the singing voice. *Music Perception: An Interdisciplinary Journal*.
- Thoresen, L. (2015). *Emergent Musical Forms: Aural Explorations*. Department of Music Research and Composition, University of Western Ontario.
- Wallmark, Z. (2019). A corpus analysis of timbre semantics in orchestration treatises. *Psychology of Music*, 47(4), 585–605. <https://doi.org/10.1177/0305735618768102>

Session 15
Emotions and Intimacy

Emotions in song and speech: Acoustic features underlying the multidimensional space of perceived emotions

Lena Heng^{1†} and Annabel Cohen²

¹ Psychology, St. Francis Xavier University, Antigonish, NS, Canada

² Psychology, University of Prince Edward Island, Charlottetown, PE, Canada

† Corresponding author: lheng@stfx.ca

Effective vocal communication is crucial for survival because it allows the transmission of important information to others, thereby influencing behaviours and allowing for the inferences of intentions (Juslin & Scherer, 2005). Humans normally process emotion information with ease and speed. In verbal communication, even when direct semantic connotations of words are removed, it is still possible to discern the underlying emotional meanings (Coutinho & Dibben, 2013), although accuracy can vary with different emotions (Banse & Scherer, 1996). Studying this communication, however, is complicated because it involves a multitude of interwoven, biological, psychological, social, and cultural aspects that also interacts with the context within which such communication is conceived and understood. Scherer (1982) proposes basing theory and research on the modified Brunswikian lens model with a tripartite process of encoding, transmission, and decoding. Because the lens model frames communication into components of expression, transmission, and reception, it is useful for exploring the communicative process by clearly demarcating contributions from each of these stages of the communicative process and allowing a systematic study of each stage.

Studies of emotion communication inevitably assume particular conceptual frameworks of emotion, whether it is attempting either to explain how each emotion is expressed or to compare how cues differ between various emotions. These include the categorical model of basic emotions (e.g., Ekman, 1992), the dimensional model of affect (e.g., Russell, 1980), emotion prototypes (Scherer, 2009), and constructionist accounts of affect (e.g., Barrett, 2017). Barrett (2006) also questions the validity of assuming emotions to be natural kinds. Kinds are particulars of scientific inquiry; when a kind is natural, it is implied that it reflects the structure of the natural world (Bird & Tobin, 2024). If emotions are not natural kinds, such an assumption could erroneously guide scientific inquiry.

Characteristics of the sound provide important cues in decoding expressed emotions. Scherer (1982) discusses various objective measurements of acoustic parameters of the voice and considers parameters that are based on time, amplitude, and frequency. These parameters could involve higher level features of structure and phrasing, melodic characteristics, and changes in dynamic levels. They could also involve low level acoustic features that characterize timbre. Timbre carries important information for emotion communication, and various studies have explored the timbre features that carry this perceptually salient information (e.g. Eerola et al., 2012; Heng & McAdams, 2024; Juslin & Laukka, 2004). Based on a meta review of emotion communication in vocal expression and music performance by Juslin and Laukka (2003), speech prosody and music share psychoacoustic cues. A difficulty in exploring acoustic cues that could be salient in emotion communication is the numerosity of such cues, and traditional modelling methods have difficulty considering contributions from a large number of cues without overfitting.

To this end, the present study has the following aims:

- 1) To examine the multidimensional space of emotions communicated through song and speech, with the purpose of determining whether emotions are cognitively structured into categories or dimensions, and what this structure is like.
- 2) To investigate the acoustic features that play a role in influencing perception of emotions in both song and speech using interpretable machine learning methods.

Method

There were 29 participants recruited from the university community for a Song affect rating task (mean age = 22.03 years, S.D. = 5.25) and 26 for a Speech affect rating task (mean age = 23:00 years, S.D. = 6.03). The study was certified for ethics compliance by the UPEI Research and Ethics Board. Stimuli were selected from the Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS) (Livingstone & Russo, 2018) with four each of Angry, Calm, Fearful, Happy, Sad, and Neutral excerpts for Song and three each of Angry, Calm, Fearful, Happy, Sad, Disgust, Surprise, and Neutral excerpts for Speech. Each clip was normalized to -0.9 dBFS and silences at the beginning and ending of each clip were removed with the MIRTtoolbox (Lartillot, 2022) to ensure that the peak amplitude levels were controlled for, and judged differences in emotions were directly related to dynamic variations within the clip. In each experimental trial, participants heard two excerpts and were asked to compare their perceived emotion content expressed by the voice actors in Song or in Speech, and to rate the level of dissimilarity on a scale of 1 (very similar) to 9 (very dissimilar). The Song and Speech stimuli were also analyzed with the revised Timbre Toolbox in the MATLAB environment, used to computing various temporal, spectral, and spectrotemporal acoustic descriptors that characterizes the timbre of sounds (Kazazis et al., 2021). Medians and interquartile ranges of 24 time-varying descriptors: spectral centroid, spectral spread, spectral skewness, spectral kurtosis, spectral flatness, spectral crest, spectral slope, spectral decrease, spectral rolloff, spectral variation, spectral flux, pitch, harmonic spectral deviation, tristimulus 1 to 3, inharmonicity, harmonic energy, noise energy, noisiness, harmonic to noise energy, partials to noise energy, zero crossing rate, and root mean square energy, and three global descriptors: effective duration, and frequency and amplitude of energy modulation were obtained. For Song, median pitch was removed because the actors had to sing a prescribed melody which means that pitch variation was inconsequential. The interquartile range of pitch was retained for Song, however, because slight variations within the prescribed pitch sequence could contribute to emotion expression and therefore the range in which the pitch of an excerpt varied could be of interest. This gave a total of 50 features for the Song and 51 features for Speech.

Results

MDS of Song and Speech

A pairwise dissimilarity matrix was obtained from the 288 pairs of Song and 288 pairs of Speech excerpts. Relatively low AIC and plateauing stress values suggested that four dimensions were sufficient to capture the underlying structure of both perceived emotions in Song and Speech. In Song, the first dimension has angry, fearful, and happy on opposite sides of calm, neutral, and sad. This therefore appears to correspond to the arousal dimension of affect. The second dimension in Song has calm, happy, and neutral lying on one side, angry in the middle, and fearful and sad lying on the other side. This approximates the valence dimension. Although dimensions 3 and 4 do not show clear clustering of emotions, they likely contribute to greater nuances for differentiating a variety of emotions. For example, angry and fearful which are less differentiated in dimensions 1 and 2 are more clearly differentiated by dimension 3. Calm and neutral are also differentiated more clearly on dimensions 3 and 4.

In Speech, the first dimension has angry, disgust, fearful, and surprised lie on one side; calm, happy, neutral, and sad on the other, again corresponding to the arousal dimension, similarly to Song. On the second dimension, happy, neutral, and surprised were on one end, angry, calm, disgust in the middle, and fearful and sad on the opposite end. This dimension reflects the valence dimension. Dimensions 3 and 4 do not appear to relate clearly to any established emotion dimensions. However, they could contribute to differentiating some of the emotions which are clustered closely in dimensions 1 and 2. Fearful and disgust, for instance, were not significantly different from each other in dimensions 1 and 2, but they are

differentiated in dimensions 3 and 4. These two dimensions therefore again appear important in differentiating a wide variety and nuances of emotions.

Acoustic features of communicated emotions

Acoustic features were used as explanatory predictors to model the loadings on each of the four dimensions of the MDS space using the decision-tree based XGBoost algorithm (Chen & Guestrin, 2016). Separate models were trained for each dimension of Song and Speech. Tables 1 and 2 lists the top five features for each dimension. As XGBoost models for dimensions 2 and 3 of Speech did not perform above chance, it suggests that cues other than these low-level timbre features could be informing listeners about emotion content on these dimensions.

Table 1: Top five features of each dimension in Song and Speech.

	<i>Song</i>				<i>Speech</i>	
	Dimension 1	Dimension 2	Dimension 3	Dimension 4	Dimension 1	Dimension 4
1	Spectral centroid (med)	Spectral centroid (med)	Effective duration	Harmonic spectral deviation (IQR)	Spectral decrease (med)	Zero-crossing rate (IQR)
2	Spectral spread (med)	Spectral centroid (IQR)	Spectral roll-off (IQR)	Spectral roll-off (IQR)	Spectral flatness (med)	Spectral slope (IQR)
3	Spectral flatness (IQR)	Spectral crest (IQR)	Spectral centroid (med)	Harmonic to noise energy (med)	Harmonic to noise energy (med)	Spectral flatness (med)
4	Zero-crossing rate (med)	Pitch (IQR)	Spectral flatness (med)	Effective duration	Spectral centroid (med)	Harmonic spectral deviation (med)
5	Tristimulus 2 (IQR)	Frequency modulation	Zero-crossing rate (IQR)	Spectral decrease (med)	Tristimulus 2 (IQR)	Spectral crest (IQR)

Discussion

MDS results reveal comparable first two dimensions for emotions communicated through both Song and Speech, the first corresponding to arousal and the second to valence. However, even though participants appear to cognitively structure emotions similarly in both modalities, the acoustic features that contribute to their communication appear to be somewhat different. A reason could be due to different constraints in speaking versus in singing, such as in the control of how harmonic or inharmonic the voice sounds. Other parameters not captured in the low-level timbral features could also be different across the modalities of Song versus Speech. Rhythm, for example, would be highly structured in singing, and less so in speaking. Communicating valence information is also different in Song and Speech. Valence information could be carried through prosodic aspects of speech or through time-varying relationships which were not captured in the low-level acoustic features that were analyzed.

As with previous research (e.g., Eerola et al., 2012; Laukka et al., 2005; Heng & McAdams, 2024), arousal appears to be a dimension that is consistently agreed on by participants, both in perceiving the arousal levels, and in the acoustic cues that inform it. Dimension 1 (arousal) explains more than 40% of

variance in the MDS models of both Song and Speech whereas dimension 2 (valence) explains only 31% of variance in Song and 24% in Speech. There are also overlapping low level acoustic cues that contribute to communicating arousal in both Song and Speech but not for communicating valence. Even though there are no low-level acoustic features that appear to contribute to communicating valence in Speech, listeners are still accurate and consistent, as demonstrated in the MDS models. This implies that valence might be carried by higher-level cues such as phrasing, accents, rhythm, dynamic variations, and in the time-varying components of sound.

This study provides support for the dimensional model of emotion (Russell, 1980) and clarifies certain low level acoustic features that contribute to auditory communication of emotions in Song and Speech. The use of interpretable machine learning techniques (Molnar et al., 2018) allows for a better understanding of how psychoacoustic cues play a role in the communication of emotions. It also demonstrates similarities in perceived emotions from both Song and Speech, and highlights differences in acoustic cues for each modality.

References

- Banase, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, 70(3), 614–636. <https://doi.org/10.1037/0022-3514.70.3.614>
- Barrett, L. F. (2006). Are emotions natural kinds? *Perspectives on Psychological Science*, 1(1), 28–58. <https://doi.org/10.1111/j.1745-6916.2006.00003.x>
- Barrett, L. F. (2017). The theory of constructed emotion: An active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience*, 12(11), 1833–1833. <https://doi.org/10.1093/scan/nsx060>
- Bird, A., & Tobin, E. (2024). Natural Kinds. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy* (Spring 2024). Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/spr2024/entries/natural-kinds/>
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16*, 785–794. <https://doi.org/10.1145/2939672.2939785>
- Coutinho, E., & Dikken, N. (2013). Psychoacoustic cues to emotion in speech prosody and music. *Cognition & Emotion*, 27(4), 658–684. <https://doi.org/10.1080/02699931.2012.732559>
- Eerola, T., Ferrer, R., & Alluri, V. (2012). Timbre and affect dimensions: Evidence from affect and similarity ratings and acoustic correlates of isolated instrument sounds. *Music Perception*, 30(1), 49–70. <https://doi.org/10.1525/mp.2012.30.1.49>
- Ekman, P. (1992). Are there basic emotions? *Psychological Review*, 99(3), 550–553.
- Heng, L., & McAdams, S. (2024). The function of timbre in the perception of affective intentions: Effect of enculturation in different musical traditions. *Musicae Scientiae*, 28(4), 675–702. <https://doi.org/10.1177/10298649241237775>
- Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*, 129(5), 770–814. <https://doi.org/10.1037/0033-2909.129.5.770>
- Juslin, P. N., & Laukka, P. (2004). Expression, perception, and induction of musical emotions: A review and a questionnaire study of everyday listening. *Journal of New Music Research*, 33(3), 217–238. <https://doi.org/10.1080/0929821042000317813>
- Juslin, P. N., & Scherer, K. R. (2005). Vocal expression of affect. In *The new handbook of methods in nonverbal behavior research* (pp. 65–136). Oxford University Press. <https://doi.org/10.1093/oso/9780198529613.003.0003>

- Kazazis, S., Depalle, P., & McAdams, S. (2021). *The Timbre Toolbox version R2021a, user's manual*.
<https://github.com/MPCL-McGill/TimbreToolbox-R2021a>
- Lartillot, O. (2022). *MIRtoolbox*. *MATLAB Central File Exchange*. [Computer software].
<https://www.mathworks.com/matlabcentral/fileexchange/24583-mirtoolbox>
- Laukka, P., Juslin, P., & Bresin, R. (2005). A dimensional approach to vocal expression of emotion.
Cognition & Emotion, 19(5), 633–653. <https://doi.org/10.1080/02699930441000445>
- Livingstone, S. R., & Russo, F. A. (2018). The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English. *PLOS ONE*, 13(5), e0196391. <https://doi.org/10.1371/journal.pone.0196391>
- Molnar, C., Bischl, B., & Casalicchio, G. (2018). iml: An R package for Interpretable Machine Learning. *JOSS*, 3(26), 786. <https://doi.org/10.21105/joss.00786>
- Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6), 1161–1178. <https://doi.org/10.1037/h0077714>
- Scherer, K. R. (1982). Methods of research on vocal communication: Paradigms and parameters. In K. R. Scherer & P. Ekman (Eds.) *Handbook of methods in nonverbal behavior research* (pp. 136-198). Cambridge University Press.

Timbre as a Compositional and Relational Tool in an Immersive Contemporary Music Performance for Infants (0–2 Years)

Anne Chabot-Bucchi^{1†} and Joshua Bucchi^{2†}

¹ Faculté de musique, Université de Montréal, Montréal, Québec, Canada

² Independent artist, Montréal, Québec, Canada

[†] Corresponding authors: anne-clementine.chabot@umontreal.ca, contact@joshuabucchi.com

Over the last decade, an increasing number of live music performances have been created specifically for very young audiences, offering infants opportunities to engage with music in non-formal and embodied ways (Dionysiou & Fytika, 2017; Creech et al., 2020). Infants are now widely recognized as competent listeners from birth, endowed with strong musical memory and early auditory skills (Trehub & Degé, 2015). Recent studies further indicate that infants can remain engaged during extended live performances and show heightened attention in concert settings compared to recorded formats (Barbosa et al., 2023; Kragness et al., 2023).

While existing research has primarily examined infants' engagement with tonal classical music or repertoire specifically composed for children, little is known about how they respond to contemporary music practices centered on timbre, texture, and sound transformation. Addressing this gap, this artistic report presents a research-creation project that explores timbre as a primary compositional and relational material in an immersive contemporary music performance for infants aged 0 to 2 years. Excerpts from the work (whether audiovisual recordings or documentation from working sessions) will illustrate how timbre-based contrasts, gradual transformations, and embodied gesture contribute to engagement and listening continuity.

The project involves the creation of a 45-minute performance for percussion and live electronics, structured as a sequence of five short tableaux designed to renew attention through contrast and continuity. Each tableau foregrounds a distinct timbral category (stone, wood, metal, skin, and then all together) with electronics derived from natural sounds and voices in Motherese serving as a unifying thread while maintaining an underlying coherence through recurring sonic gestures and spatial configurations. The compositional process relies on *comprovisation* (Bhagwati, 2013), a method in which improvisation serves as the foundational site for composition. Rather than treating timbre as a secondary parameter, this approach places timbre at the core of musical form, guiding both instrumental writing and electronic sound processing.

A central aspect of the work is the construction of a shared timbral space between acoustic percussion and electronics derived exclusively from transformed real-world sounds. These include natural environments (lake water, forest ambiances), urban soundscapes, everyday sounds (footsteps, crackling fire), and vocal material inspired by motherese, a form of infant-directed speech characterized by exaggerated prosody and expressive contours. Familiarity and recognizability of timbre are treated as essential compositional constraints, in line with developmental neuroscience emphasizing infants' sensitivity to timbral surface features in long-term memory (Trainor, Wu & Tsang, 2004).

The electronic component is developed through a collaborative *comprovisational* process with composer and computer music designer Joshua Bucchi. Sound transformations (such as spatialization, source-filter synthesis, and spectral emphasis) are deliberately subtle, aiming to preserve the perceptual identity of the original sound sources. This process revealed that certain materials (e.g., natural ambiances) support timbral transformation without losing affective clarity, while others (notably motherese) resist heavy processing, as excessive transformation can undermine their communicative function. Improvisation

sessions serve as exploratory laboratories where timbral interactions between percussion and electronics are tested, refined, and gradually formalized into semi-open notated structures, maintaining flexibility in performance while preserving a clear dramaturgical trajectory.

Through this iterative process, timbral affordances emerge as compositional guides: rhythmic cells arise organically from environmental sounds (e.g., footsteps or birdsong), and instrumental gestures evolve in response to electronic textures. Timbre thus becomes both a structural agent and a communicative medium, shaping musical form while fostering engagement between performers, sound, and audience.

Observations From the Creative Process

Two presentations of the work were conducted with infant audiences, yielding observations across two age groups. For the 0–1 age group, engagement was consistently high: 6 out of 7 infants displayed sustained visual attention during the second tableau (wood), where the woodblock repeatedly drew them physically toward the performance space. Timbral contrasts between tableaux proved effective in renewing attention throughout the full duration of the performance. Infants expressed enthusiasm through vocalizations (babbling, joyful cries) and bodily movement, with some actively accompanying musical passages, most notably during the vibraphone tableau, where one infant moved their arm in synchrony with an *accelerando/ritardando* passage. Negative affect was rare and consistently linked to external factors (hunger, discomfort); once these needs were met, infants spontaneously re-engaged. Sign language-inspired gesture generated strong interest, particularly in tableaux 3 and 4, and direct performer–infant interaction (eye contact, mirroring, responsive play) proved consistently effective in sustaining engagement.

For the 1–2 age group, the observations confirm that, just like the 0–1 age group, sustained attention over 40 minutes is achievable. Recognition of sound sources emerged as a key engagement mechanism: one child who identified water sounds in the first tableau pointed excitedly toward the performance space and remained actively engaged throughout. Active participation (touching instruments, being included in the performance) visibly deepened immersion. As in the younger group, parental engagement proved to be a significant mediating factor: attentive, responsive caregivers consistently supported their child's involvement, while disengaged parents sometimes hindered it. Spatial organization also emerged as central: a more clearly delimited performance space, with explicit guidance for parents, would likely strengthen the quality of collective attention.

By foregrounding timbre as both a perceptual and compositional tool, this project contributes to interdisciplinary discussions on timbre, embodiment, and early auditory experience, while offering concrete artistic methodologies for contemporary creation aimed at very young audiences. These observations further confirm that contemporary music does not constitute a barrier to infant engagement, provided that structural, timbral, and relational conditions are met.

References

- Barbosa, M., Vences, M., Rodrigues, P. M., & Rodrigues, H. (2023). Babies' engagement in music theater performances. A microanalytical study of the aesthetic experiences in early childhood. *Psychology of Aesthetics, Creativity, and the Arts*. <https://doi.org/10.1037/aca0000379>
- Bhagwati, Sandeep (2013a), «Comprovisation – Concepts and Techniques», in Henrik Frisk et Stefan Östersjö (dir.), *(Re) Thinking Improvisation: Artistic explorations and conceptual writing*, Lund UP

- Malmö, Université de Lund, 1-9, <https://concordia.academia.edu/sandeepBhagwati> (consulté le 3 juin 2024).
- Bhagwati, Sandeep (2013b), «Notational Perspective and Comprovisation», in Paulo de Assis (dir.), *Sound & Score. Essays on Sound, Score and Notation*, Louvain, Presses universitaires de Louvain, p. 165-177, <https://concordia.academia.edu/sandeepBhagwati> (consulté le 3 juin 2024).
- Ben Moshe, H., & Gluchankof, C. (2021). The story of “Beat” A concert program in kindergartens. *International Journal of Music Education*, 40(1), 149-160. <https://doi.org/10.1177/02557614211029091>
- Creech, A., Varvarigou, M., & Hallam, S. (2020). Contexts for music learning and participating
Developing and sustaining musical possible selves. Palgrave Macmillan.
- Dionyssiou, Z., & Fytika, A. (2017). Musical concerts for young children in Greece. *International Journal of Community Music*, 10, 317-326. https://doi.org/10.1386/ijcm.10.3.317_1
- Geoffroy, J. (2006). Le geste dans l’oeuvre musicale, la musique et le mouvement. Rencontres musicales pluridisciplinaires: Le feedback dans la création musicale, Lyon, France.
- Ilari, B. (2020). Musical engagement in the early years: A summary and synthesis of research findings. *Annals of the New York Academy of Sciences*, 1467(1), 1–16. <https://doi.org/10.1111/nyas.14292>
- Kragness, H. E., Berezowska, B., & Cirelli, L. K. (2023). A concert for babies: Attentional, affective, and motor responses in an infant audience. *Developmental Science*, 26(2). <https://doi.org/10.1111/desc.13297>
- Kragness, H. E., Eitel, M. J., Anantharajan, F., Gaudette-Leblanc, A., Berezowska, B. et Cirelli, L. K. (2023). An itsy bitsy audience: Live performance facilitates infants’ attention and heart rate synchronization. *Psychology Of Aesthetics, Creativity, And The Arts*. <https://doi.org/10.1037/aca0000597>
- Lizère, E. (2010). Tigouli: La création à inventer. *Circuit*, 16(2), 57-68. <https://doi.org/10.7202/902397ar>
- Muhonen, S. (2016). Students’ experiences of collaborative creation through songcrafting in primary school: Supporting creative agency in ‘school music’ programmes. *British Journal of Music Education*, 33(3), (pp 263-281). <https://doi.org/10.1017/S0265051716000176>
- Trainor, L. J., Wu, L., & Tsang, C. D. (2004). Long-term memory for music: Infants remember tempo and timbre. *Developmental Science*, 7(3), 289–296.
- Trehub, S. E., & Degé, F. (2015). Reflections on infants as musical connoisseurs. In G.E. McPherson (Éd.), *The Child as Musician* (p. 31-51). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780198744443.003.0002>
- Young, S. (2016). Early childhood music education research: An overview. *Research Studies in Music Education*, 38(1), 9–21. <https://doi.org/10.1177/1321103X16645631>

The Intimate Timbres of the Kantele: Materiality, Touch and Cultural Memory

Arja Kastinen^{1†}

¹ Folk Music Department, Sibelius Academy, The University of the Arts Helsinki, Finland

[†] Corresponding author: arja@temps.fi

The Finnish–Karelian kantele instruments are part of the Baltic psaltery family – a group of instruments found among Baltic, Finnic, and Slavic peoples of Northeastern Europe. The instrument exhibits regional specificities that reflect the history and culture of the areas in which it flourished. The unknown origin of this presumably at least thousand-year-old instrument has long puzzled and fascinated scholars, as its unique shape and distinctive acoustic qualities set it apart from the psalteries of medieval Europe and Asia. Since the nineteenth century, the Baltic psaltery has been developed into numerous models – with national characteristics – to meet the needs of changing musical cultures. (Temkin 2004; Haas 2001.)

This presentation focuses on the acoustic properties and cultural features of the nineteenth-century Finnish–Karelian carved kantele, as well as one of its successors at the turning point in the instrument’s development at the beginning of the twentieth century. From a musician’s perspective, the intimate, multidimensional timbre of early kantele models emerges as an entanglement of materiality, embodied performance, and cultural memory, produced through the interaction of structural design, wood species, string material, specific tunings, and historically informed playing techniques.

The example of a carved kantele is a replica of a Finnish–Karelian southern model, hollowed out from the top with a separate top plate fastened with nails. (The northern model is hollowed out from below and has no bottom plate.) It has twelve brass strings. The body is made of pine, and the top plate is made of spruce. The replica was made by Rauno Nieminen. The original instrument was built in the nineteenth century in Suistamo, Border Karelia, and was purchased for the Perniö Museum in South-West Finland in 1906 from the Karelian runosinger and kantele player Iivana Šemeikka.



The other instrument is a slightly modified version of a steel-string kantele that was purchased for the National Museum of Finland in 1914 in Korpiselkä, Border Karelia. It is made of several pieces of wood, has bracing under the top plate, features a rounded end instead of the *ponsi* found in carved kanteles, has its strings attached to pins instead of a metal bar, and is equipped with metal tuning pegs. The sides and bottom are made of pine, and the top plate is made of spruce. Support plates under the tuning pegs and

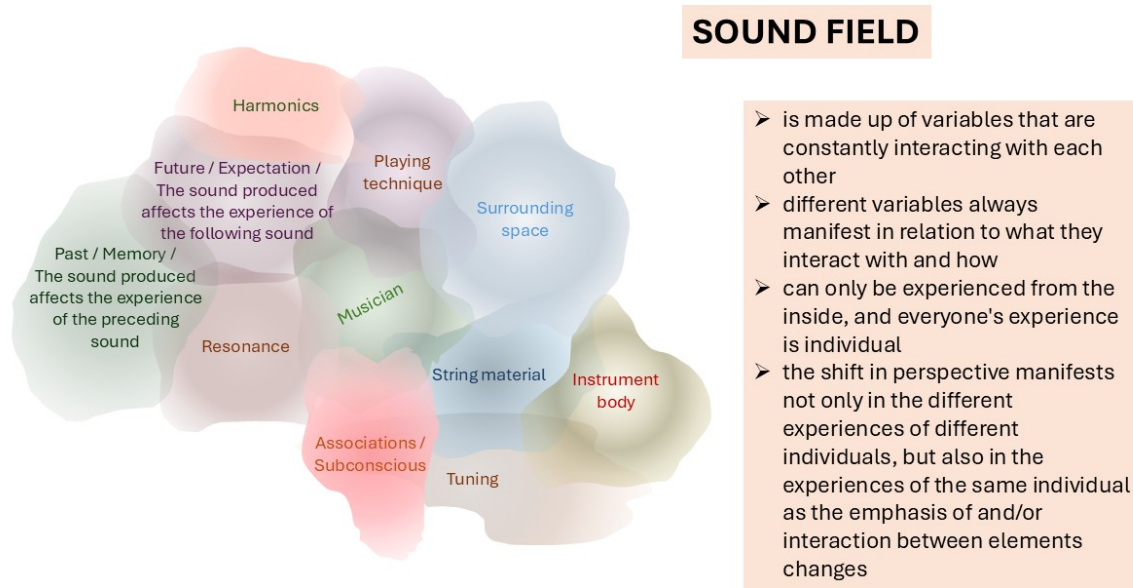
the pins are made of maple. It represents the change in musical culture at the turn of the twentieth century. The kantele was made by Keijo Säteri in 1992.



The acoustic investigations of the kantele have demonstrated several nonlinear behaviours of the instrument (Erkut et al., 2002; Karjalainen et al., 1993; Tahvanainen 2023; Tahvanainen & Hochmuth 2024). Henna Tahvanainen and Mathis Hochmuth (2024) list five of them as follows:

1. The inharmonicity of the strings, related to their tension and material – the inharmonicity of the string partials varies and does not appear to follow a clear trend.
2. Because there is no bridge beneath the strings, and because of their knotted termination at the other end, the different polarisations of the transverse string vibration result in different speaking lengths and consequently in beating – the beating effects of the string partials increase with the vertical plucking angle.
3. The pluck causes the string to be slightly longer in vibration than at rest, which leads to pitch glides.
4. The transverse, longitudinal, and torsional vibrations can couple at the string termination points, producing a strong second harmonic in the kantele – contrary to linear theory, in traditional small kanteles a pluck at the node of a vibrational mode does not prevent the corresponding harmonic from sounding.
5. The string vibration exhibits losses mainly due to air viscosity and internal friction, as well as terminal losses, resulting in different decay times for different partials.

To ensure repeatability in acoustic investigations, the string can be excited mechanically with a wire while the other strings are damped. From a musician's perspective, however, the timbre of a Finnish–Karelian carved kantele is approached differently: the behaviour of an individual string and its coupling with the soundbox becomes meaningful only through its interaction with the full set of variables that create the sound field – the foundation of ancient Finnish–Karelian kantele music. The musician's view of the sound field can be illustrated, for example, as follows:



The historical plucking technique and tunings documented in ethnographic sources and early recordings are closely tied to the musical aesthetics of the Finnish–Karelian runosong culture (– more information at Kastinen 2026: <https://www.temps.fi/en/research/kantele-of-the-runosong-culture-and-the-dialogue-of-creativity/>). Because the strings are plucked upward and rarely stopped, they resonate freely with the sound box and structural components, producing a multidimensional sound field. By altering the plucking technique, the musician varies the timbre of this sound field, and the music unfolds as a long continuum of constant variation that can lead the musician into a trance-like state. This phenomenon was later described by the folk music researcher Armas Otto Väisänen as *quiet exaltation*, while the musicians themselves referred to this improvised music as *playing their inner power*.

The human touch, with its conscious and unconscious modifications in strength, choice of plucking finger, finger angle, and plucking point, for example, is one of the variables within the sound field. In other words, human fingers are seen as part of the instrument and thus the human touch is one of the components constructing the timbre. Even small changes in these parameters can produce distinct timbral shifts. In acoustic research this can be seen as a hindrance to reliable analysis, but from the perspective of the instrument's musical background, embedded in its cultural history, it is regarded as an essential element of the timbre.

Because of the instrument's intimate, close-range sonic character, such nuances become especially meaningful for the musician. The resonance, with its varying harmonics, slowly grows through the sympathetic vibration of the open strings. The structural components of the instrument's body contribute their own vibrational modes to the sound, and thus the resulting sound field is far more complex than merely the notes produced by the individual strings. The old plucking technique, the tactile negotiation of strings, and the instrument's resonant behaviour all act as agents in the unfolding of the music. In addition, resonance functions both as an acoustic phenomenon and as a conceptual tool, revealing how each sound carries traces of previous gestures and how improvisation becomes a dialogue between intention, materiality, and memory.

During the late nineteenth and early twentieth centuries, the string material in carved kanteles changed from the previously common copper alloys to steel, which altered the timbre dramatically. Brass-string kanteles were tuned lower, and their string tension was lower as well, which also results in a different touch. Because kantele players often made their own instruments, the individual variation in shape, size, and structural solutions is wide among the museum instruments. This means that each instrument had its unique timbre. In addition, by varying the tuning, musicians could adjust the sound field to match their mood.

Although these two example instruments can offer only a narrow glimpse into the sonic world of the old Finnish–Karelian kanteles, they nevertheless allow the dramatic shift from brass-string carved kanteles to larger steel-string kanteles to be demonstrated. The sounding qualities of the historical kantele replicas carry traces of past musical worlds. Combining these with historical information on playing technique, tuning solutions, musical aesthetics, and practice-based experimentation has led me to the obvious conclusion that there is no exact answer to the question: “What did the old kantele sound like?” Just as the kantele’s sound field exists in a state of constant variation, the music itself has changed along with the personalities of the players and their unique instruments in different places and times. Music is the result of a complex interplay between the human body and mind, the instrument, the acoustic environment, the surrounding space, history, and resonance. Thus, to find a musical connection with the ancient Finnish–Karelian kantele, one must find one’s own way into the deeply personal *inner-power improvisation*.

With these two instruments, I will demonstrate how the old plucking technique, combined with the tuning method based on fifths and octaves, creates the sound field, and how its timbral structure varies with changes in touch. The remarkable timbral contrast between the carved brass-string kantele and the more modern steel-string instrument also becomes evident.

References

- Erkut, C., Karjalainen, M., Huang, P. & Välimäki, V. (2002). Acoustical analysis and model-based sound synthesis of the kantele. *The Journal of the Acoustical Society of America*, vol. 112, no. 4, 1681–1691.
- Haas, A. (2001). Intercultural Contact and the Evolution of the Baltic Psaltery. *Journal of Baltic Studies*, Volume XXXII (3), 209–250.
- Karjalainen, M., Backman, J. & Pölkki, J. (1993). Analysis, Modeling, and Real-time Sound Synthesis of the Kantele, A Traditional Finnish String Instrument. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing ICASSP’93*, Vol. 1 (pp. 229–232).
- Kastinen, A. (2026). *Kantele of the Runosong Culture and the Dialogue of Creativity*. Accessed 14 May 2026 at <https://www.temps.fi/en/research/kantele-of-the-runosong-culture-and-the-dialogue-of-creativity/>
- Tahvanainen, H. (2023). *On the acoustics of the concert kantele*. Conference paper. Stockholm Musical Acoustics Conference 2023.
- Tahvanainen, H. & Hochmuth, M. (2024). Measurements on the effect of vertical plucking angle of the kantele string. In *INTER-NOISE and NOISE-CON Congress and Conference Proceedings 270(2)* (pp. 9473–9478).
- Tëmkin, I. (2004). Evolution of the Baltic psaltery: a case for phyloorganology? *Galpin Society Journal*. 57, 219–230.

Session 16

Appropriating Sounds of Nature

Lithophonica: Timbral Explorations of Stones and Sticks

James Harley^{1†} and Gayle Young²

¹ School of Fine Art and Music, University of Guelph, Guelph, Ontario, Canada

² Independent, Grimsby, Ontario, Canada

[†] Corresponding author: jharley@uoguelph.ca

Lithophone is the term for a musical instrument created from stone. There is a long history of music being created from rocks, throughout the world (Adcock 2025, 2017). The Lithophonica project dates back several years, and builds from previous work Gayle Young has pursued, building instruments from strings, wood, and metal. These instruments were primarily designed to support explorations of alternate tuning systems (Young 1992).

Stones, and other natural elements, such as dried wooden sticks gathered and sized by beavers in forest locations (for example in wilderness areas in Ontario, Canada) have the potential to become instruments for making music. One of the core elements of this process is listening carefully, including seeking out each item's balance point where the natural resonances can be heard and activated.

In the case of our project, Lithophonica, a set of slate stones from the shores of Newfoundland, Canada, were collected, then gathered to form what might be considered a (non-standard) scale of pitches. With these stones, the next process was to explore different ways of activating them: sticks, mallets, washers, screws, etc. These stones can be struck, rubbed, scraped. The main concern is to be mindful of the resonance of each. The beaver sticks were treated in similar ways, although the activation explorations are more restricted.

The next step in our process was to explore ways to capture these sounds with microphones, especially to clearly capture the high frequencies and balanced range of these instruments. From there, the recordings were brought into the computer, where they were subjected to a range of signal processing techniques. These include: resampling, looping, transpositions, granulations, delays. Finally, spatialization was applied to layers of the audio using an 8-channel software diffusion system. All of the elements of the DSP processing are controlled in real time.

The stones and sticks are performed together with the computer processing based on improvisation and highly responsive listening. Harley and Young have performed live on a number of occasions.

The recording *Lithophonica* was the result of several days of recording, listening, and editing. Long-duration takes were honed to 35 short(er) tracks (Young, Harley, 2024).



Figure 1: Lithophone with various resonators and activators (photo: Reinhard Reitzenstein).

References

- Adcock, M. (2025, October 8). Music Stones: The Rediscovery of Ringing Rock. The Archaeopress Blog. <https://archaeopress.wordpress.com/2025/10/08/music-stones-the-rediscovery-of-ringing-rock/>
- Adcock, M. (2017). The Ancient Stone Instruments of Vietnam. *Time & Mind, The Journal of Archaeology, Consciousness and Culture* 10:23-37
- Young, G. (1992). Pitch Probe: Musical Instruments for Unusual Tuning Systems. *Contemporary Music Review* 7 (1) 51-64
- Young, G., Harley, J (2024) *Lithophonica*. Farpoint Recordings fp099

Timbre across Species: Appropriating Bird Vocalizations in Contemporary Music

Matthieu Galliker^{1†}, Jimmie LeBlanc¹, Caroline Traube¹

¹ Faculté de musique, Université de Montréal, Montréal, Québec, Canada

[†] Corresponding author: matthieu.galliker@umontreal.ca

1. Context

When humans appropriate non-human sound material, such as animal vocalizations, both its acoustic properties and its associated meaning are inevitably transformed, from subtle modifications to profound recontextualizations. In music, such appropriations can take many forms, including imitation, recomposition, and direct reproduction. The closer an appropriation remains to its source material, the more difficult it becomes to distinguish the original from its artistic reinterpretation. Non-human vocalizations and their human appropriations are produced, perceived and interpreted through different mechanisms and within different contexts, although certain sound parameters are shared across species, such as a repertoire of idiomatic structural elements, context-dependent and culturally shaped behavioral codes, and meanings inferred from individuals' experiences. This complexity becomes particularly apparent when considering timbre which emerges from the interaction of acoustic, perceptual and contextual factors. Examining the appropriation of bird vocalizations through the lens of timbre therefore provides an opportunity to explore how acoustic features, behavioral meanings, and cultural interpretations are transformed as sounds move from one ecological and communicative context to another. As with timbre and its multifactorial nature, the appropriations of bird vocalizations in human music take highly diverse forms. The timbre of bird vocalizations varies according to context and inferred meaning, including territorial defense, courtship, time of day and season. It may also be shaped by social configuration, from the song of a solitary individual to the collective vocal activity of a group. Bird calls are generally higher-pitched than the sounds we make (Catchpole & Slater 2018; Marler & Slabbekoorn 2004), and cognitive information processing occurs over a shorter time frame than our own (Hermann 2016). Finally, it is important to consider that bird vocalizations exhibit both interspecific (between species) and intraspecific (between individuals within the same species) variations. In short, these variations are perceived differently by human listeners—who associate them with new meanings—than by birds themselves, and are further transformed through artistic appropriation.

From the perspective of anthropozoomusicology, these considerations call for a distinction between the immanent (or intrinsic) structures of sounds, the social behaviors associated with them, and the subjective experiences they engender (Martinelli 2009, 9–10). This will raise the following question: how do contemporary composers negotiate these dimensions, and what aesthetic relationships with bird vocalizations emerge from these negotiations? This paper suggests a preliminary typology of appropriations of bird vocalizations in contemporary music by identifying distinct approaches related to timbre, its perception, and its representation.

2. Methodology and selection of works/vocalizations

Even though timbre—a concept specific to human culture—cannot fundamentally be applied using an emic approach to non-human cultures such as those of birds (Hui 2018), it nonetheless remains a useful tool for reflection in this study. In this context, an emic perspective refers to the meanings, functions, and categories that emerge from within a given communicative system, whereas an etic perspective designates externally constructed analytical frameworks. Rather than claiming analytical equivalence between human and non-human sonic systems, this study treats timbre as a heuristic and translational concept. It is not understood as a property that would exist in the same terms within avian

communication, but as an externally applied category that enables the articulation of relationships between acoustic structures, perceptual processes, and culturally mediated forms of listening. Its analytical value therefore lies in its capacity to operate as a bridge between different regimes of sound production and interpretation. As a multifactorial attribute of sound, timbre has thus been understood as the intersection of the acoustic, psychological, and symbolic dimensions of sound.¹² The aim will therefore be to account for these dimensions and to evaluate their transformation from an avian culture to a human culture.

Inspired by Silvers' proposal for multispecies ethnomusicology (2020), the methodology used here combines acoustic, ethological, cognitive, and ecological perspectives, in three stages of analysis:

1. Acoustic analysis of selected bird vocalizations, contextualizing sound parameters in relation to timbre and inferred meaning.
2. Perceptual analysis based on aural sonology (Thoresen & Hedman 2015), examining how these vocalizations are experienced by human listeners.
3. Musical analysis of appropriations, taking into account the contexts of production and the composers' intentions.

The three stages follow one another in order to trace the successive transformations: 1) beginning with the sound object that is a bird's song—an object situated within its context and interpreted according to meanings specific to the bird's native culture, 2) continuing with the psychoacoustic reception by a human subject, 3) and culminating in the analysis of the final result as it appears in the musical work where the songs have been incorporated. The perceptual analysis in the second stage constitutes an inductive aesthetic approach (Nattiez 1990, 140–142) that is, one based on the analyst's—in this case, my own—reception of the work itself. The results of this analysis will serve as a tool for comparison and reflection between stages 1 and 3. Aural sonology serves here as an alternative means of representation to the spectrogram and its objective physical dimension, in order to incorporate the subjective dimension of such a process. The spectromorphological nature of sound will thus be highlighted through the graphic forms of sonology, to which categories have been assigned.

3. A case study: Messiaen's *Le Merle noir* (1952)

In order to show how this methodology can be applied, the analysis of Messiaen's *Le Merle noir* (1952) will be detailed as a case study. This work, as its name suggests, is based on the vocalizations of blackbirds (*Turdus merula*). Their songs are characterized by strophes separated by pauses of 3 to 5 seconds. Each strophe has a general two-part structure lasting approximately 3 to 5 seconds as well. They sing a first motivic section with a fundamental frequency usually located between 1,500 and 3,000 Hz, while the second section, known as the chirping section, can reach up to 9,500 Hz (Dabelsteen 1984). Furthermore, the repertoire includes an average of 34 to 63 different motifs, distributed variably among individuals (Rasmussen 2002). Since no recordings or working transcriptions were found in connection with Messiaen's piece, we have chosen two strophes from a recording by Fernand Deroussen made in 1991, one morning in May in the Île-de-France region (Deroussen 1992), which are representative of some vocalization possibilities of blackbirds. The discussion will therefore be generalized, but will nonetheless provide an interesting comparison. The spectrograms of these two strophes, realized with the Acousmographie¹³ software, are represented on Figure 1. The timbral elements of these strophes, consisting of two to three parts, are identified with symbols from aural sonology.

¹² This refers to the process of sound production, reception, conceptualization and symbolization as described by Zachary Wallmark in his book *Nothing but Noise: Timbre and Music at the Edge* (2022).

¹³ See <https://inagrm.com/en/showcase/news/203/acousmographie>.

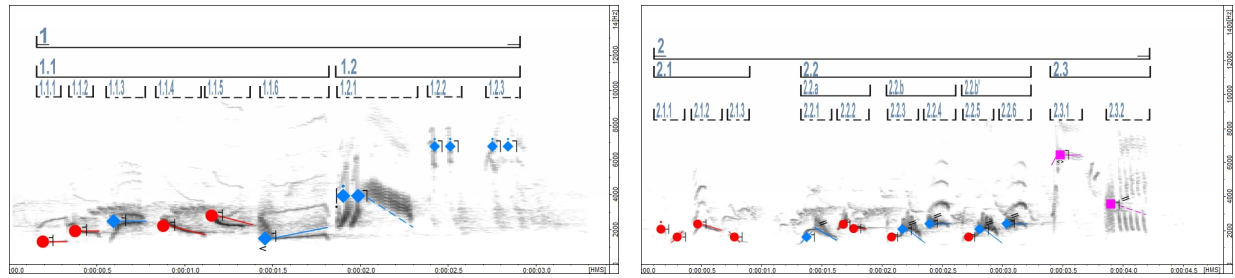


Figure 1: Two verses of a blackbird song analyzed with Aural Sonology on the Acousmographie software.

They include pitched sounds (circles in red), dystonic sounds (diamonds in blue) and complex or unpitched sounds (squares in magenta). The second strophe is richer in timbral variations, although the central motifs are repeated. Blackbirds play with these different variations to construct their songs.

In Messiaen's work, the spectromorphological diversity of blackbird strophes, here performed on the flute by Anne Cartel (Cartel & Vermeulin 2022), is less rich and more pitch-tempered (tuned to a 12 semitones scale) than the original song. There is a near-omnipresence of tonic sounds, with the occasional exception of a dystonic sound produced by a *flatterzunge* on the flute (Figure 2).

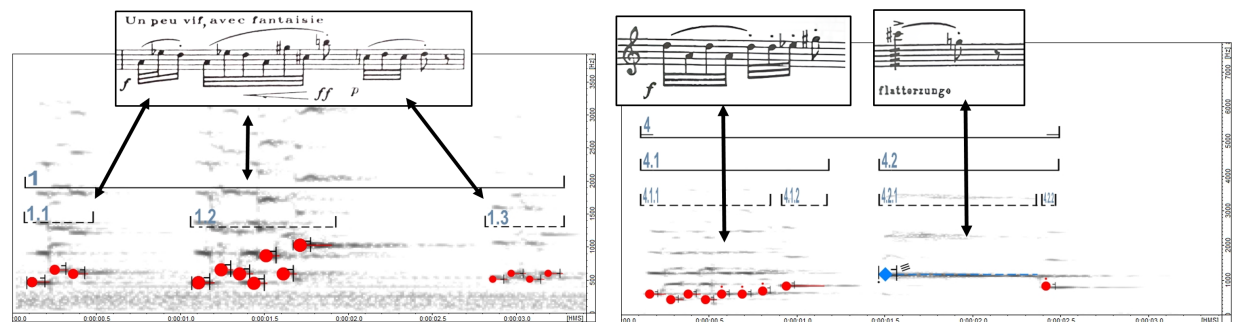


Figure 2: Measures 3 (left) and 6 (right) from Messiaen's *Le Merle noir* (1952).

In addition to this comparison of blackbird/flute phrases, we will list here a few elements related to the behavior of blackbirds and their environment.¹⁴ First, the flute part is enriched with notes, rhythms, and articulations, from its first solo section (mm. 3–8) to its second solo section (mm. 46–53): this enrichment can be interpreted as the development of the blackbirds' sonic and timbral repertoire over their lifetime (since they develop it through learning). Second, during the duet sections (mm. 9–26 and 54–71) that follow the flute solos, the instruments play a canon motif based on a quotation from another work by Messiaen symbolizing love, and the second duet section features a greater number of homorhythms than the first section: this rhythmic evolution in the relationship between the instruments represents the complicity between two birds after a male has sung with a richer repertoire to attract a mate (a variety of elements in the song being more likely to attract a mate). Finally, in the transitional subsection (mm. 76–82), the flute raises its melody by a fourth and plays louder to be heard in the final section (mm. 91–125) above a piano that plays a cloud of sounds based on a series of 11 pitches, with the pedal: this situation can be interpreted as a representation of a territorial struggle in which noise pollution in urban areas—already a reality in Messiaen's time (Loubes 2021; Granger 2014)—forces the birds to sing louder, higher

¹⁴ The analysis and interpretation of these elements were presented at the Journées d'analyse musicale, which took place in Paris in November 2024 (Galliker 2024 for the conference video; forthcoming conference proceedings).

and brighter (Nemeth & Brumm 2008; Nemeth et al. 2013), a phenomenon also found in human vocalization, known as the Lombard effect. Although Messiaen did not explicitly state that he was depicting these aspects of the blackbird's life, the choice of sounds, their timbre, and the relationship between the instruments suggest such a representation of the sounds in their context.

4. Discussion

Two key points emerge from the perceptual analysis: 1) Messiaen reproduces the blackbird's timbre not by replicating its spectromorphology identically—indeed, the limited number of extended techniques employed does not allow the transverse flute to approximate the bird's acoustic parameters more closely, either from an objective or a perceptual perspective; 2) instead, the focus is on representing the timbre according to the blackbird's behavior and the context in which its song is produced. For these reasons, we have identified two types of correspondences between the avian vocalization model and human appropriation: 1) the sound itself, and 2) the sound in context. The first, which concerns the intrinsic properties of sound, allows us to estimate the degree to which the acoustic parameters and perceptual qualities of the songs correspond to those of the appropriation themselves. The second, for extrinsic properties, allows us to compare the para-elements of these sound productions—namely, the birds' behaviors as musical representations, including their experience and appreciation of the songs, the reconstruction of the sound context, and the relationship between bird songs and their environment.

This method thus makes it possible to assess the degree to which bird vocalizations are appropriated by humans (from “very distant” to “very close”), based on the concept of timbre as a vector of multiple dimensions associated with sound: acoustic parameters and perceptual qualities on the one hand, and behaviors, appreciation, and the environment in relation to these sounds on the other. In Messiaen's *Le Merle noir*, the timbral correspondence of the sound in itself is distant, whereas that of the sound in context is very close. By applying this method to other works—such as *Cantus Arcticus* (1972) by Einojuhani Rautavaara, *Birdcalls* (1972) by Louise Lawler, *Sopiana* (1980) by François-Bernard Mâche, *Night Black Bird Song* (1999) by Emily Doolittle, *Drôles d'oiseaux* (1985-86/2001) by Francis Dhomont—and by considering the composers' intentions and the creative media (instrumental, mixed, electroacoustic), we can observe that two specific trends emerge regarding the correspondences between sound in itself and sound in context: 1) the instrumental works (Messiaen, Doolittle) and Rautavaara's mixed-media work show greater attention to timbre-related behaviors, whereas the spectromorphological reproduction of the timbre corresponds less closely to the original; 2) electroacoustic works (Dhomont) or works with an electroacoustic tendency (voices with electroacoustic processing in Lawler's case; mixed in Mâche's work, in a unique instance where instrumental voices are superimposed on a soundtrack of vocalizations) show greater correspondence regarding the spectromorphology of timbre, whereas that of the timbre associated with its context corresponds less. Regarding their intentions, Messiaen and Doolittle focus on the birds themselves, thereby signaling a *comprehensive naturalistic* approach that seeks to incorporate extrinsic dimensions of bird vocalizations. In contrast, the approaches of Dhomont are more *technical* and *idealistic* (involving the decontextualization of vocalizations). Other types of appropriation may also be more technical and naturalistic, and others more comprehensive and idealistic. This preliminary typology will be further refined following additional analyses in order to identify clear trends.

References

- Cartel, A., & Vermeulin, M. (2022). Birds of a Feather [Album]. Paraty.
- Catchpole, C. K., & Slater, P. J. B. ([1995]2018). Bird Song: Biological Themes and Variations. Cambridge: Cambridge University Press.
- Dabelsteen, T. (1984). An Analysis of the Full Song of the Blackbird *Turdus merula* with Respect to

- Message Coding and Adaptations for Acoustic Communication. *Ornis Scandinavica*, 15(4), 227–239.
- Deroussen, F. (1992). Chants d'oiseaux [Album]. Nashvert.
- Galliker, M. (2024, November 22-23). *Le merle noir* d'Olivier Messiaen : enjeux anthropo-zoomusicologiques dans l'analyse comparative de productions sonores humaine et non humaine [Conference]. Journées d'analyse musicale 2024, Paris (France).
<https://sfam.org/2024/11/jam-2024-journees-danalyse-musicale-2024/> (consulted on May 15, 2026).
- Granger, C. (2014). Le coq et le klaxon, ou la France à la découverte du bruit (1945-1975). *Vingtième siècle, revue d'histoire*, 123(3), 85–100. <https://doi.org/10.3917/vin.123.0085>.
- Herrmann, D. (2016). Avian Cognition: Exploring the Intelligence, Behavior, and Individuality of Birds. Boca Raton, FL: CRC Press.
- Hui, A. (2018). The Naturalization of Timbre: Two Case Studies. In E. I. Dolan & A. Rehding (Eds), *The Oxford Handbook of Timbre* (pp. 464–483). New York, USA: Oxford University Press.
- Loubes, T. (2021). Le bruit de la circulation et l'invention de la pollution sonore dans les villes occidentales des années 1930, entre tournant matériel et évolution des sensibilités. *Histoire politique*, 43. <https://doi.org/10.4000/histoirepolitique>.
- Marler, P. & Slabbekoorn, H. (Eds) (2004). *Nature's Music: The Science of Birdsong*. San Diego: Academic Press.
- Martinelli, D. (2009). *Of Birds, Whales, and Others Musicians*. Scranton: University of Scranton Press.
- Martinelli, D. (2008). Introduction (to the Issue and to Zoomusicology). *TRANS-Transcultural Music Review*, 12. <https://www.sibetrans.com/trans/article/93/introduction-to-the-issue-and-to-zoomusicology> (consulted on May 15, 2026).
- Nattiez, J.-J. (1990). *Music and Discourse: Toward a Semiology of Music*. Princeton, NJ: Princeton University Press.
- Nemeth, E., & Brumm, H. (2008). Blackbirds Sing Higher-Pitched Songs in Cities: Adaptation to Habitat Acoustics or Side-Effect of Urbanization? *Animal Behaviour*, 78(3), 637–641.
<https://doi.org/10.1016/j.anbehav.2009.06.016>.
- Nemeth, E., Pieretti, N., Zollinger, S. A., Geberzahn, N., Partecke, J., Miranda, A. C., & Brumm, H. (2013). Bird Song and Anthropogenic Noise: Vocal Constraints May Explain why Birds Sing Higher-Frequency Songs in Cities. *Proceedings of the Royal Society B*, 280(1754).
<https://doi.org/10.1098/rspb.2012.2798>.
- Rasmussen, R. & Dabelsteen, T. (2002). Song Repertoires and Repertoire Sharing in a Local Group of Blackbirds. *Bioacoustics*, 13(1), 63–76.
- Silvers, M. (2020). Attending to the Nightingale: On a Multispecies Ethnomusicology. *Ethnomusicology*, 64(2), 199–224.
- Thoresen, L., & Hedman, M. (2015). *Emergent Musical Forms: Aural Explorations*. London, ON: University of Western Ontario.
- Wallmark, Z. (2022). *Nothing but Noise: Timbre and Music at the Edge*. New York: Oxford University Press.

Ethereality in Jennifer Higdon's Early Works

Laura Dallman^{1†}

¹ Musicology, School of Music, University of Florida, Gainesville, FL, USA

[†] Corresponding author: ldallman.rorick@ufl.edu

Jennifer Higdon, a Pulitzer and Grammy-winning composer, is perhaps best known for her orchestral work *blue cathedral*. Premiered in 2000, the piece has received over 600 performances in the past 25 years. In *blue cathedral*, Higdon introduces an ethereal soundscape that resurfaces in a selection of her orchestral works from the early 2000s. In this paper, I argue that Higdon's early compositional style is timbrally marked by this ethereal soundscape and that the soundscape functions not only as a sonic and stylistic marker, but also a structural tool. More specifically, I use a threefold methodology that includes 1) describing Higdon's ethereal soundscape, 2) interrogating the semantics of my timbral descriptions, and 3) analyzing how the ethereal segments help delineate form. In doing this work, I offer not only a critical examination of Higdon's timbre, but I also help fill a lacuna in substantial scholarly studies of Higdon's music, which are few in number (e.g. McKinney, 2011; Reitz, 2018).

In *blue cathedral* and “river sings a song to trees,” from *City Scape* (2002), Higdon begins and ends each symphonic work by featuring a silvery, crystalline collection of percussion instruments. She includes vibraphone, crotales, and triangles in both works, adding sizzle cymbal and crotales on the timpani in “rivers sings a song to trees” and chimes, glockenspiel, celesta, prepared piano (with screws), crystal classes, and Chinese reflex balls in *blue cathedral*. Yet, the significance is not just the instrumentation, but a sense of mystery surrounding how the sounds are produced. As Dolan and Patteson note, aesthetic pleasure, paired with instrumental obfuscation, is a touchstone of ethereal timbres (Dolan & Patteson, 2018). Audience members cannot easily see percussionists situated at the back of the orchestra or Chinese reflex balls in the hands of other orchestral musicians. Furthermore, listeners of CD or digital recordings will have no visual references at their disposal. Dolan and Patteson write, “the less we know about where these tones are coming from and how they are made, the stronger the spell they cast” (Dolan & Patteson, 2018). Higdon's ethereal soundscapes are captivating in large part because of their inexplicability—or at least their inexplicability until one examines the score, recordings, and performance techniques in detail.

But what exactly makes the timbre “silvery” and “crystalline”—or, to draw on my own earlier descriptive notes regarding Higdon's ethereal soundscape, “bright, crisp, sparkling, twinkling, shining, and metallic”? Scholars note multisensory (Wallmark, 2019b), metonymic (van Elferen, 2018), and metaphorical (Saitis & Weinzierl, 2019) approaches to discussing timbre, and van Elferen gives credence to the aesthetic role of timbre. Yet, as the work of these scholars suggests, perhaps the best approach is to utilize semantic examination alongside studies of physical sound production. As such, I sort my eight aforementioned adjectives using Wallmark's seven categories of timbre descriptors as a guide (Wallmark, 2018a), then I posit how visualizations of the physical attributes of Higdon's ethereal soundscapes (e.g. spectrograms, spectral centroid) may help clarify my perceptual frame.

Structurally, Higdon's ethereal soundscapes appear at the beginning and end of *blue cathedral* and “river sings a song to trees.” In these early works, the soundscapes are bookends; they provide transitions into and out of the main material. To better understand the boundaries of these soundscapes, I analyze “secondary parameters” (Bauer, 2001) and consider timbral discontinuity (McAdams, 2019; McAdams & Goodchild, 2017). Bauer explains that secondary parameters, as defined by Leonard B. Meyer, include timbre, articulation, dynamics, and tempo. She argues that secondary parameters can delineate formal structure when primary, and analytically more traditional, parameters of pitch and rhythm are static.

Building on Bauer, I argue that Higdon's ethereal soundscapes are defined by the secondary parameters of timbre and dynamics; there are typically no strong pitch or rhythmic elements in these soundscapes. Furthermore, McAdams and Goodchild assert that segmentation can occur due to timbral discontinuity. Higdon's shifts from her ethereal soundscapes into denser instrumentation, or vice versa, also help mark these ethereal soundscapes as separate sections.

While this paper focuses on two of Higdon's earliest symphonic works, she does use her ethereal soundscape in later symphonic works, specifically "1726" from her Violin Concerto (2008) and "Cathedrals" from *All Things Majestic* (2011). In these instances, there are clear modifications; although the soundscape remains similar, the structural function begins to change as Higdon experiments with moving ethereal soundscapes inside of a piece. Thus we have begun to explore semantics surrounding Higdon's ethereal soundscapes, as well as how the soundscapes function formally, providing expedient analysis of a major living composer's early oeuvre while foregrounding timbre. But there is room for more work to be done.

Jennifer Higdon's [*blue cathedral*](#) and [*"river sings a song to trees"*](#)

References

- Bauer, A. (2001). "Composing the sound itself": Secondary parameters and structure in the music of Ligeti. *Indiana Theory Review* 22(1), 37-64.
- Dolan, E. I., and Patteson, T. (2018). Ethereal timbres. In E. I. Dolan & A. Rehding (Eds.), *The Oxford handbook of timbre* (pp. 183-204). Oxford University Press.
- McAdams, S. (2019). Timbre as a structuring force in music. In K. Siedenburg, C. Saitis, S. McAdams, A. N. Popper, & R. R. Fey (Eds.), *Timbre: Acoustics, perception, and cognition* (pp. 255-92). Springer International.
- McAdams, S., and Goodchild, M. (2017). Musical structure: Sound and timbre. In R. Ashley & R. Timmers (Eds.), *The Routledge companion to music cognition* (pp. 129-39). Routledge.
- McKinney, D. (2011). Jennifer Higdon. In M.K. Slayton (Ed.), *Women of influence in contemporary music: Nine American composers* (pp. 141-89). Scarecrow Press.
- Reitz, C. L. (2018). *Composing in color*. McFarland & Company.
- Saitis, C. & Weinzierl, S. (2019). The semantics of timbre. In K. Siedenburg, C. Saitis, S. McAdams, A. N. Popper, & R. R. Fey (Eds.), *Timbre: Acoustics, perception, and cognition* (pp. 148-87). Springer International.
- van Elferen, I. (2018). Timbrality: The vibrant aesthetics of tone color. In E. I. Dolan & A. Rehding (Eds.), *The Oxford handbook of timbre* (pp. 105-34). Oxford University Press.
- Wallmark, Z. (2019a). A corpus analysis of timbre semantics in orchestration treatises. *Psychology of Music* 47(4), 585-605.
- Wallmark, Z. (2019b). Semantic crosstalk in timbre perception. *Music and Science* 2, 1-18.