

Session 1

Contemporary Music Analysis and Dance

Enacting Violence Through Choreography-Orchestration Relationships in Stravinsky's Sacrificial Dance

Rebecca Moranis^{1†}

¹ Department of Music, The Graduate Center, City University of New York (CUNY),
New York, New York, United States

[†] Corresponding author: rebecca.moranis72@gc.cuny.edu

The culmination and shattering climax of Igor Stravinsky's *The Rite of Spring* is its final scene, the Sacrificial Dance, where a young woman (the Chosen One), surrounded and menaced by a vicious crowd, violently dances herself to death to propitiate the sun god. While the music and choreography of the scene have been studied separately (see References), there has been virtually no attention to the relationship between the orchestration and the choreography. In this paper, I examine how three choreographies establish different relationships between the choreography and the orchestration in this scene, thus articulating three specific narrative possibilities of the Chosen One's response to her fate: 1) terror and submissiveness (Millicent Hodson and Kenneth Archer, 1987), 2) madness (Pina Bausch, 1975), and 3) dominance and control (Martha Graham, 1984). Each of these choreographers uniquely designs the music-movement relationship between the Chosen One and the Tribal Elders and Ancestors who surround her. I engage with existing scholarship that debates the definitive narrative of this scene (Levitz, 2004; Taruskin, 2017).

1) In Hodson and Archer's reconstruction of Vaslav Nijinsky's original choreography, the Chosen One's movements are aligned with those of nearly all timbral layers. This extreme coordination between the orchestration and choreography gives the impression that the Chosen One is highly controlled by the Tribal Elders, who limit her freedom narratively as well as spatially, and who force her to dance towards her death.

2) In Bausch's choreography, the Chosen One's movements rapidly switch alignment between various timbral layers. The Elders watch from a large distance, moving at a glacial pace, aligning themselves with the gradual climactic build of the orchestration. The volatile music-movement relationship of the Chosen One makes her appear hysterical.

3) In Graham's choreography, the Chosen One's movements appear to generate those of the Elders, just as the lower register, darker instruments sonically generate those of subsequent higher register, brighter instruments. The Chosen One thus acts as an architect of violence, choosing to dance to her death.

I use three analytical approaches: 1) score-based analysis that describes the introduction and elimination of instruments and their pitch content, 2) audio analysis that accounts for sound intensity and timbral descriptors, and 3) choreographic analysis (performance videos) that adapts existing choreomusical notation systems to transcribe and analyze choreographies. Within each mode (i.e., score, audio, and danced performance), I establish techniques to facilitate the analysis of continuous parameters across the scene.

I build on previous work (Leaman, 2016) to analyze how continuous parameters in the music (e.g., timbre and orchestration) and dance (e.g., motion trajectories and physical expansion and contraction of the body) contribute to the form and narrativity of the scene.

I show that each choreographer shares control with the composer over how the trajectory of the scene is ultimately understood, and therefore that the score and choreography must be analyzed in conjunction.

Within this case study, I show that the portrayal of violence depicted in the conclusion of the narrative is strengthened by the interaction of the choreography and orchestration.

References

1 Musical Elements:

1a) Pitch Organization

- Straus, J. N. (2022a). The melodic organization of *The Rite of Spring*. *Music Theory Online*, 28(4).
- Straus, J. N. (2022b, July 11). *Sacrificial Dance – Stravinsky's Rite of Spring: Analysis by Joseph Straus* [Video]. YouTube. <https://youtu.be/jxJo-tN5KpU?si=3knnqM0NMIlfQgKL5>
- Forte, A. (1978). *The harmonic organization of The Rite of Spring*. Yale University Press.
- Russell, J. (2018). Harmony and voice-leading in *The Rite of Spring* [Doctoral dissertation, Princeton University].
- Taruskin, R. 1980. Russian folk melodies in *The Rite of Spring*. *Journal of the American Musicological Society*, 33(3), 501–543.
- Van den Toorn, P. (1986). Octatonic pitch structure in Stravinsky. In J. Pasler (Ed.), *Confronting Stravinsky* (pp. 130–156). University of California Press.

1b) Rhythmic and Metric Organization

- Boulez, P. (1991). Stravinsky remains (analysis of *The Rite of Spring*). In S. Walsh (Trans.), *Stocktakings from an apprenticeship* (pp. 55–110). Clarendon Press. (Original work published 1953; revised 1966.)
- Van den Toorn, P. (1987). *Stravinsky and The Rite of Spring: The beginnings of a musical language*. University of California Press.
- Van den Toorn, P. (1988). Stravinsky re-barred. *Music Analysis*, 7(2), 165–195.
- Van den Toorn, P. (2017). The physicality of *The Rite*: Remarks on the forces of meter and their disruption. In S. Neff, M. Carr, & G. Horlacher (Eds.), *The Rite of Spring at 100* (pp. 285–303). Indiana University Press.
- Woodruff, E. G. (2006). Metrical phase shifts in Stravinsky's *Rite of Spring*. *Music Theory Online*, 12(1).

1c) Formal Layout

- Boulez, P. (1991). (see above)
- Horlacher, G. (2017). Rethinking blocks and superimposition: Form in the 'Ritual of the Two Rival Tribes.' In S. Neff, M. Carr, & G. Horlacher (Eds.), *The Rite of Spring at 100* (pp. 331–338). Indiana University Press.
- Straus, J. N. (2022a). (see above)

2 Choreographed Elements:

2a) History of the Ballet, Versions, Narrative

- Berg, S. (1988). "Le Sacre du Printemps": Seven productions from Nijinsky to Martha Graham. UMI Research Press.
- Craft, R. (1988). Counterpoint and choreography. *The Musical Times*, 129(1742), 171–176.
- Hodson, M. (2017). Death by dancing in Nijinsky's *Rite*. In S. Neff, M. Carr, & G. Horlacher (Eds.), *The Rite of Spring at 100* (pp. 47–80). Indiana University Press.

- Hodson, M., and Archer, K. 2014. The lost rite: Rediscovery of the 1913 “Rite of Spring.” KMS Press.
- Jordan, S. (2007). Stravinsky dances: Re-visions across a century. Dance Books.
- Jordan, S. (2017). The Rite of Spring as a dance: Recent re-visions. In S. Neff, M. Carr, & G. Horlacher (Eds.), *The Rite of Spring at 100* (pp. 29–38). Indiana University Press.
- Levitz, T. (2004). The Chosen One’s choice. In A. Dell’Antonio (Ed.), *Beyond structural listening? Postmodern modes of hearing* (pp. 70–108). University of California Press.
- Taruskin, R. (2017). Plenary essay: Resisting The Rite. In S. Neff, M. Carr, & G. Horlacher (Eds.), *The Rite of Spring at 100* (pp. 417–446). Indiana University Press.

2b) How Movements Relate to Pitch, Rhythmic, and Formal Structures

- Horlacher, G. (2017). (see above)
- Jordan, S. (2007). (see above)
- Jordan, S. (2015). Machine metaphors in Pina Bausch's *The Rite of Spring*: A choreomusical approach. In G. Brandstetter, G. Klein, & E. Hoops (Eds.), *Methoden der tanzwissenschaft modellanalysen zu Pina Bauschs »Le Sacre du Printemps/Das Frühlingsopfer«* (pp.195–206). transcript Verlag.
- Straus, J. N. (2022a). (see above)

3 Performance Videos and Score:

- Joffrey Ballet. (2010, January 31; originally from 1987). *Rite of Spring (3 of 3)* [Video]. YouTube. https://youtu.be/C_7ndqgwxcM?si=wZKgxgLbht_-MZj&t=263
- Martha Graham Dance Company. (2017, March 14; originally from 1984). *The Rite of Spring*. [Video]. YouTube. https://youtu.be/pZRMiri4c_E?si=amAzF9Tiz8pY0Qzt&t=241
- Stravinsky, I. (1947). *The Rite of Spring*, re-engraved in 1967. Boosey & Hawkes.
- Tanztheater Wuppertal. (2021, October 26; originally from 1975). *Le sacre du printemps - The Rite of Spring - Pina Bausch 1975*. [Video]. YouTube. <https://youtu.be/6Nv9SYhXluI?si=Wjp4IPpiTLR2W38k&t=1830>

4 Methods:

- Bregman, A. S. (1990). Auditory scene analysis: The perceptual organization of sound. MIT Press.
- Dolan, E. I. (2013). The orchestral revolution: Haydn and the technologies of timbre. Cambridge University Press.
- Leaman, K. Y. (2016). Analyzing music and dance: Balanchine’s choreography to Tchaikovsky and the choreomusical score [Doctoral dissertation, Yale University].
- Reymore, L. (2026). Computational approaches to instrumentation analysis with Timbre Trait Profiles. In J. Delisle, R. Hasegawa, J. Noble, and M. Touizrar (Eds.), *The Oxford handbook of orchestration studies* (pp. 356–379). Oxford University Press.

Orchestration and Transformation Procedures in George Benjamin's Dance Figures

Igor Maia^{1†} and Robert Hasegawa²

¹ School of Music, Federal University of Minas Gerais, Belo Horizonte, Minas Gerais, Brazil

² Schulich School of Music, McGill University, Montreal, Quebec, Canada

[†] Corresponding author: imaia@ufmg.br

This paper examines the relationship between George Benjamin's *Piano Figures* (2004) and its orchestral reworking as *Dance Figures* (2004), focusing on the compositional, structural, and timbral transformations shaping the transcription from a piano cycle to an orchestral work. *Piano Figures* consists of ten concise piano pieces, each exploring a single gestural or textural idea. Shortly after composing the set, Benjamin developed an orchestral version of nine movements, *Dance Figures*, planned for a 2006 choreography by Anna Teresa de Keersmaeker. While the orchestral suite preserves much of the ordering and basic expressive character of the original movements, it departs from *Piano Figures* in ways that exceed conventional notions of transcription or orchestration. We present a comparative analysis of both works, asking (1) What role does orchestration play in reconfiguring form, gesture, and texture? (2) What orchestrational and compositional strategies underlie the transformation of *Piano Figures* into *Dance Figures*?

Our approach to the study of these transformational procedures draws on recent scholarship and methods for the analysis of orchestration. We consider the role of orchestral grouping effects in determining the cognitive segmentation of the music as heard (Touizrar & McAdams 2019, McAdams 2022). Lévy's concept of "functional orchestration" (2026) seeks to describe the compositional intent underlying orchestrational decisions: for example, ensuring the separation of two melodic voices or reinforcing upper overtones of a line through parallel couplings. This is particularly valuable for a comparison of the piano and orchestral versions, allowing speculation on the composer's rationale for his orchestration choices. Orchestration impacts not only grouping structure but also timbral associations (Hasegawa 2026). Throughout, we draw on interviews with Benjamin and his own writings to contextualize the study (Benjamin 1997, 2003, 2004) as well as musicological and analytical writings on his other works (Lack 2001, Rupprecht 2005, Whittall 2014).

The comparative analysis of *Piano Figures* and *Dance Figures* reveals a number of essential orchestrational transformations that modify form, gesture, and texture. Processes of superimposition, interruption, and foreground/background reorganisation are especially marked in IV: *Interruptions*. In II: *Recit*, Benjamin uses parallel doublings derived from the overtone series, producing spectral reinforcement that alters the harmonic profile and references organ registration (Benjamin 2004). In VI: *Hammers*, the introduction of repetitive percussion patterns increases rhythmic articulation and drive. The orchestral version often introduces timbral resonances that approximate the piano pedaling, particularly in I: *Spell* and V: *Song*. The orchestration enhances canonic stratification in III: *In the Mirror*, making the contrapuntal layers more distinct than in the piano version. Other important transformations include motivic echoing, superimposed replication (notably in I: *Spell*), and the expansion of gestures (in V: *Song*).

Figures 1 and 2 show a brief excerpt from the versions of V: *Song* in *Piano Figures* and *Dance Figures* respectively. The bullet points that follow quickly summarize the principal transformations between the two versions:

- There are some changes to the overall duration and proportions, with the orchestra version lengthening the initial held note and spreading the gesture over an additional third measure.

- The piano's accented melody line D-B-G is played by the orchestra's E-flat clarinet.
- The piano's flourish in thirty-second notes is entirely recomposed. The idiomatic piano writing with crossing hands (all played *ppp possibile*) is replaced by eight notes on the same pitch classes (though not identical registers) spread across the *divisi* violins and violas, with each instrumentalist alternating between two notes to give a sense of motion. The strings shift from predominantly pizzicato attacks to tremolo *sul ponticello* over the course of a crescendo and decrescendo across the whole gesture.
- The celesta and harp emphasize the end of the rising piano gesture with rapid contrasting scales ascending to a strong attack simultaneous with the E-flat clarinet's arrival on G.
- As a background underlying all of this, the other orchestral strings hold the tritone B-F across multiple octaves (including in held harmonics), while the vibraphone attacks the same tritone at the beginning of the gesture, and a percussionist adds a quiet roll on the ratchet (or fishing-rod reel)

Benjamin's recomposition of this passage shows the subtle creation of a broader and more theatrical gesture (fitting the larger canvas of the orchestra), including a dynamic crescendo and decrescendo not written in the piano. The pizzicato strings, harp, and celesta all crescendo to an accented arrival on the downbeat of the excerpt's third measure, followed by a decrescendo. The two textural layers of the piano become a much more complex texture with multiple layers (at least five), all carefully differentiated by articulation and dynamics.

Particularly significant is the use of orchestral effects to replace the held sustain pedal of the piano, what Lévy 2026 calls a “sound-mastering function” and Touizrar & McAdams 2019 “timbral resonance.” The E-flat clarinet’s melody is presented “dry” (without the virtual pedal), but the blurred resonances of the piano’s flourish are taken up by the held background tritone and the moving eighth notes in the strings. The fadeout of the tremolo *sul ponticello* strings in the third measure imitates the natural dying away of the sound after the last piano attack. As Sibelius famously commented, “The orchestra, you see, is a huge and wonderful instrument that has got everything—except the pedal” (Alesaro 2003). Here in V: *Song*, the piano’s pedal is imitated on a vast scale through a creative and inventive orchestrational solution.

By comparing the piano and orchestra versions of this and other excerpts, this study identifies a set of recurrent transformative procedures—including transforming blends, motivic proliferation, and timbral resonance—that reveal how Benjamin's orchestration is also a process of recomposition.

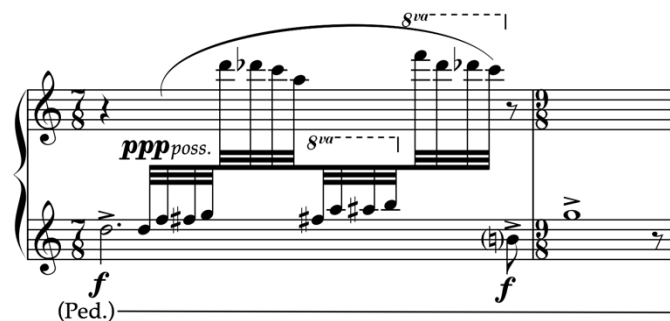


Figure 1: Piano Figures no. 5, bar 32 and 33

The musical score for 'Dance Figures no. 5, bars 188-190' is presented for a full orchestra. The instruments and their parts are as follows:

- Clarinet in E♭:** Plays a melodic line starting with a forte (*f*) dynamic.
- Ratchet:** Provides a rhythmic accompaniment with a tremolo effect, marked *p* (piano).
- Vibraphone:** Plays a melodic line with a *mf* (mezzo-forte) dynamic.
- Celesta:** Features a melodic line with a *pp* (pianissimo) dynamic, marked with a *12* and a *mf* dynamic.
- Harp:** Plays a melodic line with a *p* (piano) dynamic, marked with a *10* and a *f* (forte) dynamic.
- Violins I 1-4:** Play a melodic line with a *pp* (pianissimo) dynamic, marked with a *div. sul pont. (trem.)* and a *mf* (mezzo-forte) dynamic.
- Violas 1-4:** Play a melodic line with a *pp* (pianissimo) dynamic, marked with a *div. sul pont. (trem.)* and a *mf* (mezzo-forte) dynamic.
- Violins II 1. 2.:** Play a melodic line with a *pp* (pianissimo) dynamic, marked with a *div. pizz.* and a *f* (forte) dynamic.
- Violins II 3. 4.:** Play a melodic line with a *pp* (pianissimo) dynamic, marked with a *div. pizz.* and a *f* (forte) dynamic.
- Strings:** Play a melodic line with a *pp* (pianissimo) dynamic.

Figure 2: Dance Figures no. 5, bars 188-190

This study was financed, in part, by the National Council for Scientific and Technological Development (CNPq), Brazil, Process Number 405153/2024-4 and the Social Sciences and Humanities Research Council, Canada.

References

- Alesaro, J. (2003). "Sibelius and his 'Free Moments' at the Piano," in *Sibelius Forum II: Proceedings from the Third International Jean Sibelius Conference*, ed. Matti Huttunen, Veijo Murtomäki, and Kari Kilpeläinen (Sibelius Academy).
- Benjamin, G. (2003). *Canonic codes*. In A. Latham (Ed.), *Sing, Ariel: Essays and thoughts for Alexander Goehr's seventieth birthday* (pp. 21–31). Farnham, England: Ashgate.
- Benjamin, G., & Denut, É. (2004). *Les règles du jeu*. Paris, France: Musica Falsa.
- Benjamin, G., & Nieminen, R. (1997). Interview with George Benjamin. In R. Nieminen (Ed.), *George Benjamin* (pp. 1–38). London, England: Faber & Faber.
- 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 (Eds.), *The Oxford handbook of orchestration studies*. Oxford University Press.

- Lack, G. (2001). Objects of contemplation and artifice of design: Sonic structures in the music of George Benjamin. *Tempo*, 215, 10–14.
- Lévy, F. (2026). Functional orchestration: A research field at the beginning of its history. In J. Delisle, R. Hasegawa, J. Noble, & M. Touizrar (Eds.), *The Oxford handbook of orchestration studies*. Oxford University Press.
- 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>
- Rupprecht, P. (2005). Above and beyond the bass: Harmony and texture in George Benjamin's Viola. *Viola. Tempo*, 232, 28–38.
- Touizrar, M., & McAdams, S. (2019). Aspects perceptifs de l'orchestration dans *The Angel of Death* de Roger Reynolds: Timbre et groupement auditif. In P. Lalitte (Ed.), *Analyse musicale et perception: Actes des Journées d'Analyse Musicale 2016* (pp. 55–78). Dijon, France: Editions universitaires de Dijon.
- Whittall, A. (2014). Rotations and reflections: The musical presence of George Benjamin. *The Musical Times*, 155(1927), 71–84.

Analyzing Orchestration Style: A Case Study of Two Works by Germaine Tailleferre

Jade Roth^{1†}

¹ Schulich School of Music, McGill University, Montreal, Quebec, Canada

[†] Corresponding author: jade.roth@mail.mcgill.ca

Introduction

In *Style and Orchestration*, Read notes an “unconscionable dearth of analysis of the creative orchestrational styles that distinguish one period from another and one composer from another” (1979, p. 2). It is surprising then, that, despite recent interest in timbre and orchestration, orchestration style is so weakly understood nearly half a century later. Indeed, although modern orchestration texts allude to prominent composers’ characteristic orchestration techniques, these analyses are often incomplete pictures of orchestration style. While Adler, for example, notes that Schumann’s orchestrations sound “heavy” and include “many octave doublings,” (2002, p. 292) there is little music theory elucidating how this technique fits into his orchestration style. There is a gap between the aurally identifiable fingerprint of an orchestration style and the systematic theory to support it. This paper, thus, seeks to define a theoretical framework of orchestration style to compare stylistic differences in two works by Germaine Tailleferre (1892–1983): *Concertino for Harp and Orchestra* (1927) and *Concerto for Two Guitars and Orchestra* (ca. 1964).

Orchestration Style in Music Analysis

In writings on musical style, theorists often use the word style in seemingly contradictory ways. For example, Leeuw remarks that style “arises automatically according to the manner of writing in fashion” (2005, p. 97) whereas for Larue, style is the “imaginative escapes from convention” (2011, p. 16). Thus, style embodies two opposing realities simultaneously: it arises from context and yet is a collection of features set apart from their context.

In published analyses, authors often adopt one of these two perspectives of style depending on whether they are interested in broader context or composer individuality. Adler, for example, notes differences in instrumentation between eras, such as the introduction of homophonic brass sections in the Classical era or the increased use and diversity of percussion in the twentieth century (Adler 2002). In her analysis of Haydn, Dolan (2013) highlights the composer’s use of orchestration-based formal processes in his late symphonies. For example, “orchestral growth” is an effect which Haydn often uses in the “slow introduction” openings of his symphonies. Another effect, “enriched recapitulation,” is used by Haydn to bolster the return of the main theme, which is often presented first in a more “anemic” orchestration. She describes Haydn’s repeated use of these orchestral gestures as “collectively...form[ing] a kind of orchestral vocabulary” (2013, p. 16). By identifying recurrent patternings in Haydn’s symphonic forms, Dolan defines characteristic features of Haydn’s orchestration style, without explicitly naming it as such.

Both perspectives, however, are necessary to construct a thorough picture of a style. As Pascall succinctly summarizes: “a work cannot properly be appreciated or studied in isolation, neither can stylistic evolution and trends be distinguished without a thorough understanding of individual examples” (2001, par. 27).

Analysis of Tailleferre’s Orchestration Style

The *Concertino* (1927) and the *Concerto* (ca. 1964) share certain characteristics that offer an appropriate case for comparison. Both works combine traits from contrasting orchestration styles, using both the

vague, hazy textures of Impressionism and the clear, transparent textures of Neoclassicism. The works also feature solo instruments with similar timbral properties; notably, both the harp and the guitar are soft instruments with the shared challenges of balance and audibility alongside sizeable orchestral accompaniment. Composed more than thirty years apart, these two works contrast significantly in the ways of integrating these contrasting styles—a difference that showcases important developments in the composer's orchestration techniques.

The Concertino for Harp and Orchestra combines aspects of Neoclassical and Impressionist orchestration primarily through juxtaposition. These traits are juxtaposed in these two movements to create orchestrational contrast, but the stylistic traits remain in their respective movements. In the first and third movements, the harp fulfills the traditional role of soloist, with tight-knit phrases alternating with a statement by the orchestra. When the orchestra plays the principal melody, the woodwinds, brass, and strings often play in hierarchical, stratified layers without complex blends or busy textures. The second movement, in contrast, features characteristic Impressionist techniques like vague, swirling textures and melodic couplings outlining colourful non-functional harmonies. One such texture is achieved in mm. 22–25, where the strings form an integrated background layer (Figure 1). The transforming foreground melody is carried by the flute, whose colour is augmented by the trumpet, which is then passed to the trumpet and harp. In the background layer, fusion is facilitated by the chordal spacing, which mimics the harmonic series. The parallel motion of the colourful sonority gives the impression of a breathing organism. The harmony of the ebbing string sonority may resemble extended dominants, but with the parallel motion and harmonic series spacing, their use in this context is particularly Impressionistic.

The first movement of the Concerto for Two Guitars plays with contrasting styles by juxtaposing an austere aesthetic with winding, expressive string passages. In Figure 2, for example, the string passage leading up the return of the main theme oozes with late-nineteenth century Romanticism but does not form a coherent tonal progression. The slithering chromatic voice leading and expressive dynamic swells give this passage a marked quality within the movement, which is otherwise stately and detached. This winding descent is destined for the final chord of m. 61, where the strings converge on a French augmented 6th chord of an A minor chord that never materializes.

This paper shows the application of a theory of orchestration style that encompasses both adherence to orchestration stylistic conventions and personal traits that do not conform to a particular orchestration style. Through the study of Tailleferre's orchestration, I show her development from using a montage of established styles to the unconventional and personalized integration of these idioms.

References

- Adler, S. (2002). *The study of orchestration* (3rd ed). W. W. Norton & Company.
- Dolan, E. I. (2013). *The orchestral revolution: Haydn and the technologies of timbre*. Cambridge University Press.
- LaRue, J. (2011). *Guidelines for style analysis* (M. G. LaRue, Ed.; Expanded 2nd ed). Harmonie Park Press.
- Leeuw, T. de. (2005). Timbre. In *Music of the twentieth century: A study of its elements and structure* (pp. 97–133). Amsterdam University Press.
- Lévy, F. (2026). Functional orchestration: A research field at the beginning of its history. In *The Oxford handbook of orchestration studies* (pp. 218–247). Oxford University Press.
<https://doi.org/10.1093/9780197633489.003.0012>
- Pascall, R. (2001). Style. In *Grove music online*.
<https://doi.org/10.1093/gmo/9781561592630.article.27041>

Piston, W. (1955). *Orchestration* (First edition). Norton.

Read, G. (1979). *Style and orchestration*. Schirmer Books.

Roth, J., & McAdams, S. (2026). Analyzing orchestration as a form-defining element in Germaine Tailleferre's *Ballade* (1920). In J. Delisle, R. Hasegawa, J. Noble, & M. Touizarar (Eds.), *The Oxford handbook of orchestration studies* (pp. 549–573). Oxford University Press.
<https://doi.org/10.1093/9780197633489.003.0013>

Satie, E. (1981). *Écrits* (O. Volta, Ed.; New reviewed and expanded ed.). Éditions Champ Libre.

Tailleferre, G. (2024). *Concertino pour harpe et orchestre*. Cmaj7.

Tailleferre, G. (2002). *Concerto pour deux guitares et orchestre*. La Fabrique Musique.

22 4 Très expressif

Picc.

1 Fl.

2 Fl.

Cl.

1 Cor.

2 Cor.

Tpt.

Hpe.

Transforming foreground layer

Partial No.

13 is expressif

12

13

1 Vln. I

2 Vln. I

1 Vln. II

2 Vln. II

1 Alto

2 Alto

1 Vlc.

2 Vlc.

1 Cb.

2 Cb.

Integrated background layer

Identical motion

"F 9 7"

"G add b 6"

"G 13 7"

"F 9 7"

Figure 1: *Concertino for Harp and Orchestra*, ii, mm. 22–25.

Romantic strings

Return of first theme

Tempo

“Counter-melody”

B minor bass line

Amplification of Guitars 1 & 2

B minor bass line

Fr+6 of A

The musical score is for the first movement of the Concerto for Two Guitars and Orchestra, measures 58 to 63. The score is written for a large ensemble including Flutes 1 & 2, Bassoon, Clarinet, Horn, Trombone, Cello, Harp, Guitars 1 & 2, Violins I & II, Viola, Violoncello, and Contrabass. The key signature is B minor (three flats). The time signature is 4/4. The score is divided into two main sections: 'Romantic strings' (measures 58-63) and 'Return of first theme' (measures 64-69). The 'Return of first theme' section is further divided into 'Tempo' (measures 64-69) and 'Counter-melody' (measures 70-75). The 'Counter-melody' section is marked with a blue box. The 'B minor bass line' is marked with a red box. The 'Amplification of Guitars 1 & 2' is marked with a green box. The 'B minor bass line' is also marked with a red box. The 'Fr+6 of A' is marked with a red box. The score includes dynamic markings like pp, mf, and f, and performance instructions like 'expressivo div.' and 'pizz.'.

Figure 2: Concerto for Two Guitars and Orchestra, I, mm. 58–63.

Session 2
Aestheticizing Noise

It May Not Be Noise: Composing Mental Timbre Through Redistribution and Contextual Interference

Rafael Maia^{1†} and José Alberto Gomes¹

¹ Science and Technology of the Arts at the School of Arts, Universidade Católica Portuguesa, Porto, Portugal

[†] Corresponding author: rafamaianet@hotmail.com

This paper presents the sound installation *It May Not Be Noise*, an artistic research project exploring timbral creation through perceptual redistribution and contextual interference. The installation exposes listeners to the same sonic material under contrasting narrative frameworks, encouraging alternative perceptual organizations of identical acoustic cues. The installation investigates how timbral identity may emerge not from fixed acoustic properties alone, but from the intentional interaction between auditory ambiguity, memory, expectation, and contextual framing. (Zhu & McAdams, 2024)

The longstanding tension between objective approaches to timbre as an acoustic phenomenon and subjective approaches to timbre as a perceptual construct—often referred to as the “timbral abyss” (Wallmark, 2022)—continues to challenge both scientific and artistic understandings of sound. Rather than approaching this divide as a limitation, this research treats it as a productive condition for timbral creation. By embracing the instability between what is acoustically present and what is perceptually constructed, timbre is understood here as a contingent relationship instead of a fixed one, emerging through the interaction between acoustic material and listener-dependent perceptual organization.

This practice-based study, situated within artistic research, proposes a creative strategy in which timbre is not fully determined by the acoustic signal, but emerges through the tension between ambiguous acoustic affordances and contextual framing. Rather than composing fixed timbral identities, the approach aims to design conditions under which different listeners may construct different timbral percepts from the same sonic material.

This strategy is explored through the sound installation *It May Not Be Noise*, conceived as a “narrative duet for a single noise.” Installed in the Purple Room at Casa da Música, the work presents a single six-minute soundscape loop accompanied by two dramaturgically opposing narrative frameworks. Using the Pepper’s Ghost illusion technique, two holographic narratives are projected back-to-back inside glass containers positioned on opposite sides of the room (Figure 1), while a six-channel speaker system spatially distributes the sonic material throughout the space. Although listeners are exposed to the same acoustic material, each narrative framework guides attention toward different sonic associations and perceptual organizations. Through dense sonic clusters, spectral dispersion, and spatial displacement, the installation maintains high levels of auditory ambiguity, destabilizing perceptual grouping processes and encouraging the emergence of alternative timbral identities from the same acoustic material.¹

¹ Audio and video documentation of the installation may be made available upon request to the author.



Figure 1: It May Not Be Noise sound installation—narrative B, viewed from the side B of the room. Narrative A remains invisible on this side and vice-versa.

The proposed strategy articulates two main conceptual axes. First, it draws on Cornelia Fales' concept of *timbral anomaly by redistribution* (Fales, 2002), in which perceptual grouping mechanisms are disrupted, allowing acoustic components to detach from their inferred sources and become available for reorganization into alternative perceptual entities. Second, it employs IRCAM's *SpeaK* lexicon (Carron et al., 2017) as a mechanism of contextual interference, using language-based descriptors to orient attention, expectation, and listening mode. Together, these elements enable the construction of listening situations in which timbre is not imposed by the composer, but perceptually negotiated by the listener under conditions of ambiguity.

From a compositional perspective, the strategy operates through three main parameters: (1) the construction of acoustically rich and ambiguous sound clusters, or perceptual reservoirs (Serra, 1993) capable of supporting multiple perceptual organizations, functioning as reservoirs of latent timbral possibilities; (2) the deliberate destabilization of perceptual grouping mechanisms; and (3) the introduction of contextual cues that guide, but do not determine, perceptual resolution. Within this framework, the composer does not define timbre as a stable outcome, but instead designs conditions for its emergence.



Figure 2: It May Not Be Noise sound installation—narrative A, viewed from the side A of the room. Narrative B remains invisible on this side and vice-versa.

Perceptual responses were gathered through a qualitative methodology consistent with practice-based artistic research. Following the experience of the installation, semi-structured post-experience conversations were conducted with participants, focusing on perceived sound sources, timbral qualities, and interpretative associations (Coutinho et al., 2009). These exchanges were documented through field notes and later examined through thematic analysis. Rather than aiming for statistical validation, this approach sought to identify recurring perceptual tendencies and assess whether the proposed strategy could produce consistent forms of perceptual redistribution across listeners.

The observations indicate significant variability in perceived timbral identities among participants exposed to identical acoustic material. Listeners following narrative A frequently reported distinct and sometimes incompatible sound sources compared to those following narrative B, despite hearing the same underlying soundscape. At the same time, these perceptions often aligned with the contextual interferences presented. This suggests that the interaction between auditory ambiguity and contextual interference can facilitate perceptual redistribution, leading listeners toward different emergent timbral constructions. This supports the viability of the proposed approach as a compositional strategy for timbral creation. At the same time, the installation revealed important constraints. The effectiveness of the strategy appears to depend strongly on the robustness of contextual framing and on the listener's auditory background. Participants with extensive experience in reduced listening practices often resisted source reconstruction, remaining focused on abstract sonic qualities rather than forming virtual source identities. Rather than constituting a failure, this behavior helps delineate the perceptual boundaries of the method and reinforces the situated nature of timbral construction.

It May Not Be Noise demonstrates how perceptual redistribution and narrative framing may function as compositional tools for timbral creation. Rather than defining timbre as a fixed acoustic property, the

installation explores how timbral identities may emerge differently across listeners through the interaction between ambiguous sonic material and contextual interference. Future developments of the project may involve more structured methodologies for perceptual validation and formalization, as well as expanded explorations of perceptual divergence in artistic practice.

This work was developed within the Doctoral Programme in Science and Technology of the Arts at the School of Arts – Universidade Católica Portuguesa, under the supervision of José Alberto Goems, with support from the FCT project 2022.12831.BD, funded by national funds through the Foundation for Science and Technology.

References

- 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. <https://doi.org/10.1504/JDR.2017.086749>
- Coutinho, C. P., Sousa, A., Dias, A., Bessa, F., Ferreira, M. J. R. C., & Vieira, S. R. (2009). *Investigação-ação: Metodologia preferencial nas práticas educativas*. <https://hdl.handle.net/1822/10148>
- Fales, C. (2002). The Paradox of Timbre. *Ethnomusicology*, 46(1), 56–95. <https://doi.org/10.2307/852808>
- Serra, M.-H. (1993). Stochastic Composition and Stochastic Timbre: GENDY3 by Iannis Xenakis. *Perspectives of New Music*, 31(1), 236–257. <https://doi.org/10.2307/833052>
- Wallmark, Z. (2022). *Nothing But Noise: Timbre and Musical Meaning at the Edge*. Oxford University Press.
- Zhu, L., & McAdams, S. (2024). Comparison of Heard and Imagined Blends of Instrumental Dyads. *Music & Science*, 7, 20592043241246751. <https://doi.org/10.1177/20592043241246751>

Phantom Polyphony: Modeling Tinnitus and the Variability of Its Timbre in the Context of an Art Installation

Thomas Augustin^{1†}

¹ Faculty of Music, Université de Montréal, Montréal, Canada

[†] Corresponding author: thomas.augustin@umontreal.ca

Phantom Polyphony is a research-creation project that examines the impact that sound creation can have on the processes of accepting the tinnitus condition, through its musical recontextualization. One of the works that has resulted from the research is a sound and visual immersive installation (*Phantom Polyphony*, 2026) that will be discussed in this artistic report.

Video documentation: <https://vimeo.com/1199769511>

Multidisciplinary Approach

Context in audiology: Tinnitus is a phantom auditory sensation, it's a sound that has no source in the external world, and that can only be heard by the tinnitus sufferer. Its morphology is difficult to access, as it is a highly individual perceptual phenomenon. It can be commonly described as ringing, buzzing, clicking, *wooshing* or hissing sounds, but its verbal description is dependable on individual's background of sensory experiences, and is linked to the variability of individual judgment (Szibor et al., 2021).

The psychoacoustic assessment of the percept is the clinical procedure in audiology which aims to estimate, using measuring devices, the psychoacoustic properties of tinnitus - the dominant frequency and the intensity, and possibly, but in a limited way, a certain consideration for timbre when the tonal or noisy nature of the percept is considered (Henry, 2016).

According to the neurophysiological model of tinnitus (Jastreboff, 2011) these psychoacoustic properties of tinnitus can directly be affected by the emotional reaction of the tinnitus sufferer. Tinnitus can create exponential psychological distress when a negative emotion is associated with it. In the recursive loop of hypervigilance, as more and more attention and emotion are placed towards the phantom sound, it can intensify and transform, becoming impossible to ignore. Some tinnitus therapies seek to break this cycle, and focus on changing the emotion, a process essential for acceptance and management for the person with tinnitus.

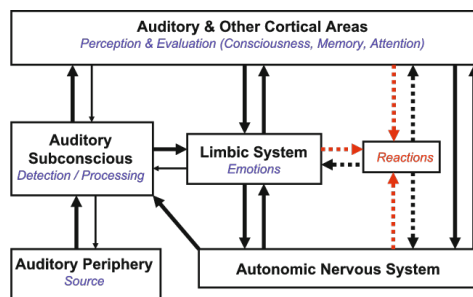


Figure 1: Pawel Jastreboff's neurophysiological model of tinnitus (Jastreboff, 2011)

Hence, tinnitus must be considered as **relational** in its phenomenology, where a key aspect of an acceptance process implies a **transformation of listening**.

Context in Acoustic Ecology: Leading figures in acoustic ecology propose a **conception of listening** that promotes the exploration of relationships between all sounds, real or possible, a perceptual awareness geared towards the productive and non-resigned acceptance of the soundscape, whatever it may be.

- **Hildegard Westerkamp**, a sound artist, pioneer of the acoustic ecology movement that emerged in the 1970s, speaks about the **disruptive nature of listening** to the soundscape. It refers to the experience of sound unfolding in the present moment, and in a potentially difficult but necessary openness to discomfort and the unexpected, allowing us to stay with the sound, regardless of the reactions it may elicit in us (Westerkamp, 2018).
- **Pauline Oliveros**, another sound artist and leading figure of acoustic ecology, explores connection with oneself, others, and the world—a fusion of the external and internal worlds, through **Deep Listening** and the Sonic Meditations. Oliveros proposes being *attuned* to the soundscape, that is, in physical and subjective contact with auditory memories or imagined sounds that have a profound impact on consciousness, that are intimately linked to the psyche, and emotionally charged (Oliveros, 1974).
- **Roberto Barbanti**, a French-Italian philosopher, has deepened the question of the **subject-object relationship** in sound ecology, making it migrate towards what he calls Sound Ecosophy (*Écosophie sonore*). It proposes a paradigm shift: moving from an essentially retinal way of observing the world to a multisensory approach to listening to it. If listening means getting involved, immersing oneself, absorbing, then an **ontological coexistence** induced by listening to the landscape refers to a form of immanence and belonging to the world (Barbanti, 2025).

Research Question

Can a work of sound creation contribute to the processes of acceptance of the tinnitus condition?

Objectives

1. Create an immersive sound and visual installation, in a self-reflective process, aimed at deepening my understanding of my relationship to tinnitus, → through the experimentation with sound and visual synthesis devices that model tinnitus sounds and reveal the richness of these sonic objects.
2. Provoke positive emotional associations that could change the perception of phantom sounds, and bring about a transformation in listening, → through the translation of the individual reality of tinnitus into a shared experience, raising awareness of hearing health issues, and creating emotional and perceptual connections between people with and without tinnitus, scientists and artists.

Questions of Timbre in *Phantom Polyphony*

Synthesis of Tinnitus Portraits (Augustin, 2022): In a preliminary study, interviews were conducted with **eight musicians** who described their tinnitus and engaged in a **sound synthesis protocol** combining quantitative measurements and qualitative **assessments of percepts**, for the creation of tinnitus soundscapes, or portraits (*Portraits acouphéniques*). Sound samples (a **sound collection of typical tinnitus sounds**) oscillators, noise generators, and other audio synthesis tools were used in the development of individual Tinnitus Portraits. The process revealed the relevance of pairing the psychoacoustic properties of frequency and intensity that are typically measured in clinical assessment, but also timbre and spectral content, pitch and amplitude modulations, temporal and spectral envelopes, and for multiple percepts in the same participant.

Tinnitus Portraits: <https://revuemusicaleoicrm.org/rmo-vol9-n2/identification-des-proprietes-psychoacoustiques/>

P7 H 42 ans	« J'entends trois sons : (1) Un sinus très haut qui module lentement en amplitude, (2) Un son plus bruyé qui produit une note claire et qui peut moduler en hauteur entre deux notes, un peu comme le son d'un fax, avec un timbre qui me rappelle la modulation en anneau, donc résonnant, métallique, avec des partiels inharmoniques. (3) Sinus constant à gauche, plus bas. »	Sinus 1 : 9200 Hz. BB : Bruit résonnant, note à 523 Hz (C4), mais harmoniques hautes présentes à 1318 Hz (E5). Bande de fréquence assez large centrée sur 4200 Hz. Sinus 2 : 2900 Hz.	Modulation d'amplitude, lente (moins de 1 Hz) et aléatoire de Sinus 1. Petite Modulation rapide et saccadée de hauteur (4 Hz) de BB, occasionnelle, oscillation entre deux notes.	Unilatéral : Sinus 1 et BB à gauche, Sinus 2 à droite.
-----------------------------------	--	---	--	---

Figure 2: Verbal descriptions, and the corresponding synthesis methodology (Augustin, 2022)

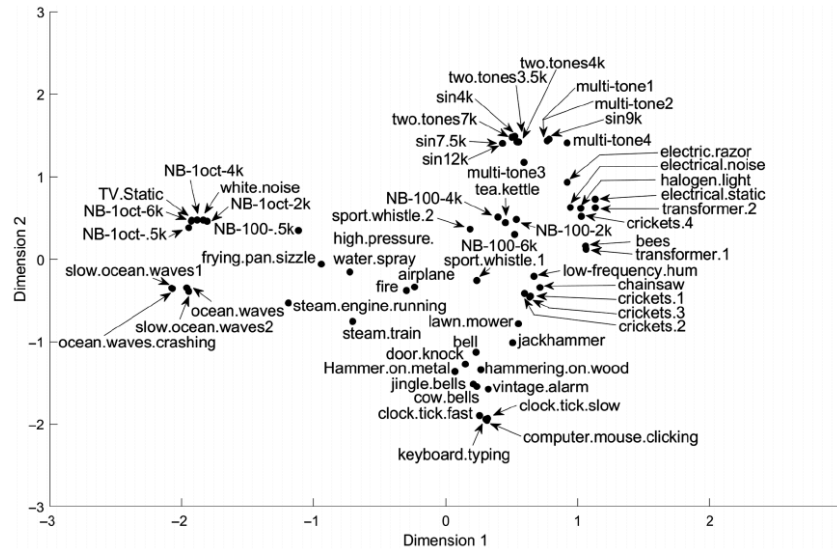


Figure 3: The tinnitus-like sound collection used in the study was modeled on the one used by J. Lentz and Y. He in “Perceptual dimensions underlying tinnitus-like sounds” (Lentz and He, 2020)

Tinnitus generator modules: Using the Tinnitus Portraits as source sound material, a series of algorithmic tinnitus generators render different timbre, and a variety of tinnitus morphology to create musical soundscapes. Implemented in Max/MSP, it uses additive synthesis, where sine waves are combined to create simple, or complex inharmonic tones, or noisy textures.

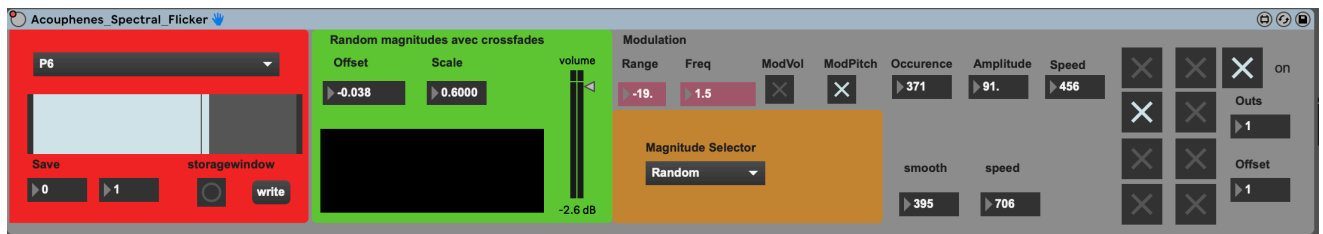


Figure 4: A FFT analysis tinnitus generator module

Sound and light transduction: For sound diffusion, instead of speakers, material sonification (large polystyrene panels) is created with the help of sound transducers. Polystyrene is a flexible material that allows a wide range of timbre, especially responsive in the lower frequency range. Through a chain of successive transductions where the vibration creates an ever-changing sound and audio reactive visuals relationship (sound → transducers → panels → Mylar paper → projection), the relational aspect of tinnitus is illustrated, a phenomenon shaped by multiple external forces, related to space, the sound environment, social and cultural norms and an individual's hearing abilities.

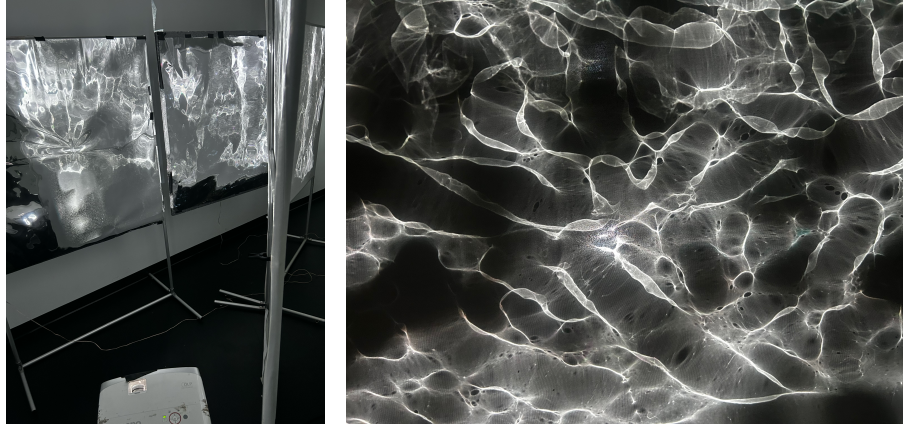


Figure 5: The installation apparatus and a resulting audio reactive visual textures

Conclusion

This artistic report will discuss questions of timbre in *Phantom Polyphony*, a sound and visual immersive installation created in the context of a music research-creation project about tinnitus. For the presentation, a brief sound and video documentation of the work will be presented (1 minute). The public will also be invited to visit the exhibition of the installation, which will take place during the entirety of the conference at B-325 of the Music Faculty of UdeM.

References

- Augustin, T. (2022). Identification des propriétés psychoacoustiques des acouphènes de huit musiciens et synthèse de portraits acouphéniques. *Revue musicale OICRM*, 9(2), 63–82.
- Barbanti, R. (2025). Les sonorités du monde: de l'écologie sonore à l'écophilosophie sonore. Les Presses du réel.
- Henry, J. A. (2016). « Measurement » of Tinnitus. *Otology & Neurotology: Official Publication of the American Otological Society, American Neurotology Society [and] European Academy of Otology and Neurotology*, 37(8), e276-285.
- Jastreboff, P.J. (2011). Tinnitus Retraining Therapy. In: Møller, A.R., Langguth, B., De Ridder, D., Kleinjung, T. (eds) *Textbook of Tinnitus*. Springer, New York, NY.
- Lentz, J. J., & He, Y. (2020). Perceptual dimensions underlying tinnitus-like sounds. *Journal of Speech, Language, and Hearing Research*, 63(10), 3560-3566.
- Oliveros, P. (1974). *Sonic meditations*. Smith publications Baltimore, MD.
- Szibor, A., Hyvärinen, P., Mäkitie, A. et Aarnisalo, A. A. (2021). Low inter-rater consistency in semantic profiles of tinnitus-like sounds rated by tinnitus patients. *Progress in Brain Research*, 262, 93-113.
- Westerkamp, H. (2018). La nature perturbatrice (disruptive) de l'écoute. *Filigrane. Musique, Sons, Esthétique, Société*, (23).

Don't Let Screaming Ruin Your Beauty Sleep, Just Bleach It!

A Study of Frequency Shearing for Noise

Marina Lee

Music Theory, Eastman School of Music, City, NY, United States

Corresponding author: mlee95@u.rochester.edu

Unpleasant, high-intensity sounds—especially human screams and shouts—are a pervasive source of discomfort. These sounds can disrupt everyday life, making their mitigation an important and intriguing challenge. While total silence is often unattainable, transforming these harsh noises into something more tolerable or even calming could have significant benefits (Moniuszko 2023). The central problem addressed here is: How can we convert offensive, high-intensity noises into neutral sounds in real time? This question is not only technically compelling but also socially relevant, as it touches on the intersection of acoustics, psychology, and digital media.

This study draws on several strands of existing research in psychoacoustics and sound analysis, integrating approaches that examine both emotional and psychological dimensions. Foundational work by Plomp and Levelt (1965) and Pressnitzer et al. (2000) examines “roughness” as a key factor in dissonance perception. Bolaños Chamorro’s (2020) analysis of Jonathan Harvey’s *Speakings* provides a framework for understanding speech formants. Research on voice quality (Teixeira et al. 2013; Anikin and Persson 2017; Anikin et al. 2018; Engelberg et al. 2021; and Frühholz et al. 2021) offers tools for evaluating hoarseness through jitter, shimmer, and harmonic-to-noise ratio (HNR). While prior studies often focus on diagnosis or artistic adaptation, this project diverges by proposing a concrete, repeatable method for transforming unpleasant noises into white noise—a process I describe as bleaching.

Based on the above considerations, I proposed a solution concerning a systematic process of frequency shearing—selectively removing certain frequency bands. More specifically, I will show that by identifying and trimming the most active inharmonic frequencies in screams and shouts depending on individual voice characteristics, it is possible to transform these noises into white noise, preserving their beneficial aspects while eliminating their unpleasant qualities. What is new in this approach is the combination of psychoacoustic theory, advanced sound analysis software, and a structured, repeatable methodology applied to real-world audio samples. This method not only addresses the technical challenge of noise transformation but also offers a framework for future research into the acoustic manipulation of other noise types.

The methodology consists of three main components: (1) designing a quadrant (**Figure 1**) to categorize voice qualities (criteria: vocal frequency and hoarseness) across different case studies; (2) employing three specialized sound analysis tools—Praat for formant detection, Spear for frequency editing, and Sonic Visualiser for hoarseness analysis; and (3) executing a sequence of nine tasks to systematically “bleach” the target sounds. The quadrant helps differentiate the unique characteristics of each voice, while the software tools provide both visual and auditory feedback for precise frequency manipulation. The nine tasks include identifying formants, locating and cutting the busiest areas, using auditory checks to ensure sufficient trimming, removing both high and low frequency bands, and finally, visualizing the results to confirm the transformation into white noise. This methodology is applied to four case studies, each representing a different combination of vocal frequency and hoarseness, drawn from well-known sitcom characters.

The results demonstrate the effectiveness of the proposed method across four distinct voices. In the attached charts, I will only show the related analysis for Michael Scott’s case in detail for explanatory

purposes. Following the nine-step task list, we can see through the four examples Michael’s case. **Figure 2** shows the formants, and **Figure 3** compares the Spear spectrograms before and after according to frequency shearing. **Figure 4** is the Sonic Visualiser spectrogram, and **Figure 5** displays the final result. For Michael Scott’s shout, the process involved removing frequencies up to 7000 Hz, resulting in a sound resembling static hiss—an auditory equivalent of white noise. Amy Santiago’s panicked outbursts required trimming up to 8000 Hz, with the final product sounding like a pencil tip scratching paper. Ron Swanson’s hoarse bellow necessitated cutting up to 9000 Hz, producing a tinkling, bell-like noise, while Donna Meagle’s high-frequency, hoarse scream was transformed into a “crickets and frogs” effect. In each case, the methodology successfully eliminated the unpleasant qualities of the original sound, as confirmed by spectrogram analysis and auditory testing.

In summary, this proposal outlines a novel approach to mitigating unpleasant noises by converting them into white noise through targeted frequency shearing. I will show that this method, grounded in psychoacoustic theory and implemented with advanced analytical tools, can effectively transform disruptive screams and shouts into calming or neutral sounds. The findings suggest that the key variables are vocal frequency and hoarseness, which determine the range and intensity of frequencies to be trimmed. This work opens new avenues for research in sound design, noise reduction, and even therapeutic applications, such as sleep aids or focus enhancers. Future directions include extending the methodology to crowd noises, non-human sources, and incorporating amplitude as an additional dimension of analysis. Ultimately, “bleaching noise” offers a practical and metaphorically apt solution to the age-old problem of sonic discomfort.

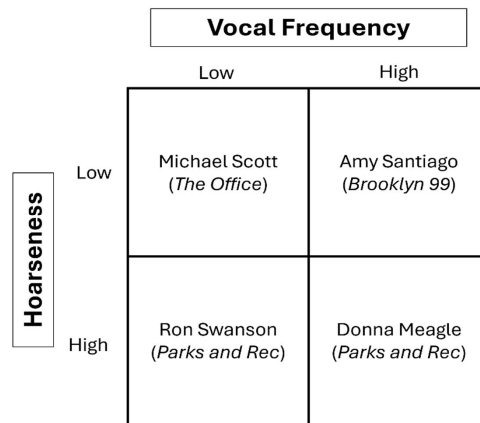


Figure 1: Quadrant of vocal qualities for four different characters in the case studies

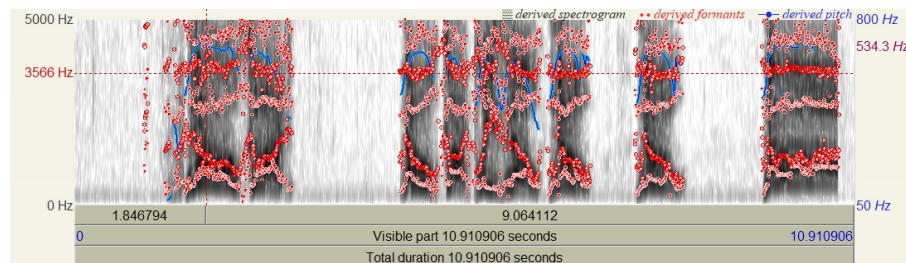


Figure 2: Praat spectrogram of Michael’s shout, showing the positions of formants (red dots and blue lines)

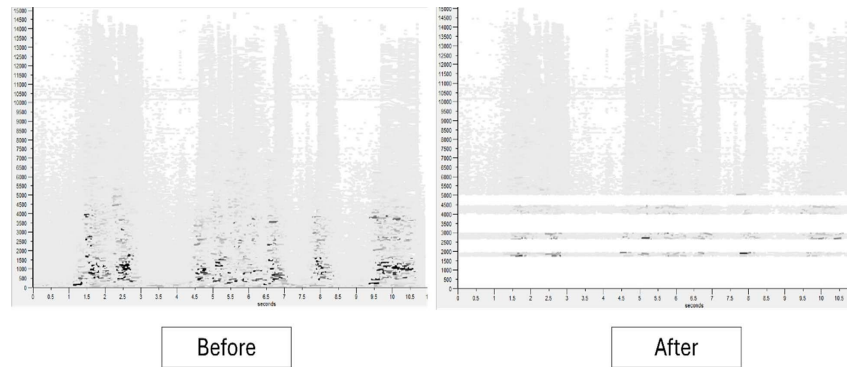


Figure 3: Original (before) and partially sheared (after) Spear spectrograms

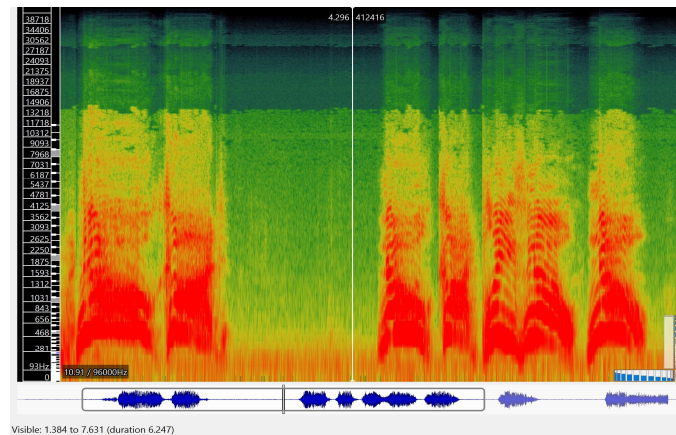


Figure 4: Sonic Visualiser representation of Michael's scream

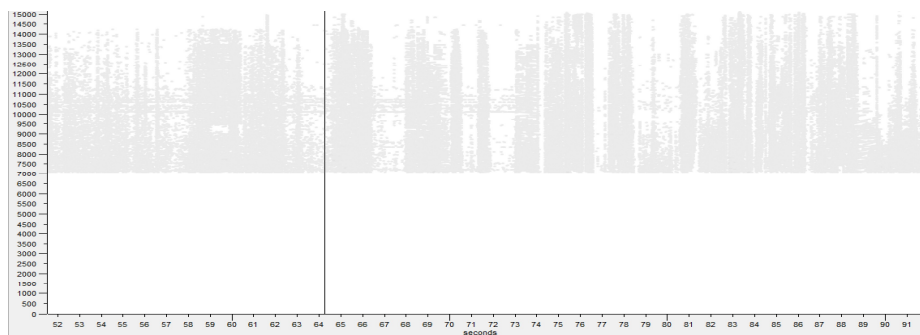


Figure 5: Finished, trimmed spectrogram of Michael's shout, showing all frequencies below 7000 Hz

References

- Anikin, Andrey, and Tomas Persson. “Nonlinguistic vocalizations from online amateur videos for emotion research: A validated corpus.” *Behavior research methods* 49.2 (2017): 758-771.
- Anikin, Andrey, Rasmus Bååth, and Tomas Persson. “Human non-linguistic vocal repertoire: call types and their meaning.” *Journal of nonverbal behavior* 42.1 (2018): 53-80.
- Bolaños Chamorro, G. J. (2020). *An analysis of Jonathan Harvey’s Speakings for orchestra and electronics*. *Ricercare*, 13, 72–107.
- Engelberg, Jonathan WM, Jay W. Schwartz, and Harold Gouzoules. “The emotional canvas of human screams: patterns and acoustic cues in the perceptual categorization of a basic call type.” *PeerJ* 9 (2021): e10990.
- Franzoni, Valentina, Giulio Biondi, and Alfredo Milani. “Emotional sounds of crowds: spectrogram-based analysis using deep learning.” *Multimedia tools and applications* 79.47 (2020): 36063-36075.
- Frühholz, Sascha, et al. “Neurocognitive processing efficiency for discriminating human non-alarm rather than alarm scream calls.” *PLoS biology* 19.4 (2021): e3000751.
- von Helmholtz, Hermann. *On the Sensations of Tone as a Physiological Basis for the Theory of Music*. Longmans, Green, 1912.
- McAdams, Stephen, and Kaija Saariaho. “Qualities and functions of musical timbre.” 1985 international computer music conference. 1985.
- Moniuszko, S. (2023, August 25). *White noise to help you sleep? A neurologist explains why people are turning to “sound masking”*. CBS News – HealthWatch. (Accessed October 31, 2025.)
- Plomp, Reinier, and Willem Johannes Maria Levelt. “Tonal consonance and critical bandwidth.” *The journal of the Acoustical Society of America* 38.4 (1965): 548-560.
- Pressnitzer, D., McAdams, S., Winsberg, S., & Fineberg, J. (2000). Perception of musical tension for nontonal orchestral timbres and its relation to psychoacoustic roughness. *Perception & Psychophysics*, 62(1), 66–80.
- Schwartz, Jay W., Jonathan WM Engelberg, and Harold Gouzoules. “Was that a scream? Listener agreement and major distinguishing acoustic features.” *Journal of Nonverbal Behavior* 44.2 (2020): 233-252.
- Teixeira, J. P., Oliveira, C., & Lopes, C. (2013). Vocal acoustic analysis: Jitter, shimmer and HNR parameters. *Procedia Technology*, 9, 1112–1122.

Session 3
Extended Techniques

Classifying String Instrument Preparation Techniques: Analysis, Taxonomies, and Applications in Computer-Aided Orchestration

Maurizio Azzan^{1†} and Luca Guidarini¹

¹Department of Theory, Analysis and Composition, Conservatory of Mantua, Italy

[†] Corresponding author: maurizio.azzan@conservatoriomantova.com

Among the new orchestration practices that have emerged in recent decades within the field of experimental contemporary music, the preparation of string instruments has gained increasing significance and has dramatically expanded their timbral palette. Patafix, paper clips, hairpins, clothespins, and aluminium foils have encouraged composers to explore complex sounds, often produced with different materials that yield seemingly similar results, while still allowing for highly specific choices by individual composers. However, despite the growing use of these techniques in contemporary repertoire, we still lack an operational organological model which can adequately describe them and enable more informed orchestration strategies, particularly in the context of computer-aided orchestration.

This paper proposes a theoretical model and an operational taxonomy aimed at understanding how commonly used preparation techniques act on strings and influence their spectromorphology. Building on this framework, we discuss the specific case of strings preparation with patafix and clothespins, and introduce a dedicated sound database designed for integration into the computer-aided orchestration software Orchidea.

1. Theoretical Model and Taxonomy

Musical instruments and the bodies that play them can be understood as adaptive (Reybrouck, 2006) and dynamic systems (Heyde, 1975), characterized by their ability to “construct meaning [...] out of the perceptual (sounding) flux” through “adaptations at each level of their functioning,” within a continuous feedback process (Reybrouck, 2006: 64). According to Heyde’s organological model (1975), each human–instrument system consists of interconnected elements endowed with specific functions, which together establish a relation of spectromorphological co-determination and contributes, in a distinct way, to defining both the spectral content of the produced sound and its temporal evolution. By integrating Heyde’s model (1975: 77) with subsequent developments in acoustics (Rossing, 2010), bowed string instruments can be described as systems composed of five main components as follows:

Table 1: Organological functions in bowed string instruments.

Component	Function	Action
<i>Muscular system</i>	<i>Exciter</i>	<i>provide energy</i>
<i>Bow</i>	<i>Mediator</i>	<i>channel energy into the string and determine its excitation regime (stick–slip)</i>
<i>String</i>	<i>Transducer</i>	<i>transform the energy delivered by the mediator into mechanical oscillations</i>

<i>Bridge</i>	<i>Channel</i>	<i>act as an acoustical lever, transform the motion of the string into forces applied to the soundboard, thereby modifying the instrument's frequency response</i>
<i>Body of the instr.</i>	<i>Resonator</i>	<i>selectively filter, amplify, and radiate the output of the transducer</i>

Since the system described here is not perfectly linear (Fletcher & Rossing, 1998), identifying the role of each of its components in spectromorphological co-determination requires considering an idealized performance situation, in which a single sound emission is produced under conditions of constant bow speed, pressure, and position, with an open string and a motionless instrument. In such a situation, the transducer defines the frequency space within which energy will be distributed, the mediator determines the modalities of this distribution and its temporal articulation, while channel and resonator selectively filter some of its components. Within this framework, a preparation consists in altering the operating conditions of these elements in order to modify their spectromorphological contribution without, however, changing their corresponding organological function. The application of external objects to the strings, bow, bridge, or body of the instrument, as well as the partial or total replacement of their constitutive elements, reconfigures their geometry, material, mass, impedance, and conditions of contact and friction, thus turning these aspects into modifiable parameters within a standardized instrumental model and, at times, adding further organologically functionalized elements to it.

In the case of the bow, since its preparation entails a modification of the stick–slip mechanism, it alters both the modalities of energy transmission within the frequency space and its temporal articulation. Likewise, the use of mutes or additional resonators applied to the bridge modifies its lever action and frequency response, while the placement of external objects on the body of the instrument results in a selective attenuation of specific resonance frequencies (Rossing, 2010: 221). As for the strings, preparation reconfigures their organological arrangement by introducing an impedance discontinuity. It acts both as a constraint, partially dissipating and reflecting vibrational energy, and, when not in contact with the nut or the bridge, also as a channel transmitting part of it to the portion of string not directly excited by the bow. The resulting frequency space is therefore influenced by parameters such as acoustic conductivity, inertia, density, shape, weight, and position of the preparation, which co-determine the mechanical impedance of the system. The more closely the impedance of the preparation matches that of the string to which it is applied, the more efficient the transfer of energy it enables and, consequently, the signal radiated by the portion of the string not directly excited by the bow may reach an amplitude sufficient to exceed the masking threshold imposed by the signal produced by the bowed one.

From an acoustical-perceptual perspective, the coexistence of multiple spectra gives rise to a particular class of complex inharmonic sounds, which we define as multispectral, distinct from both noisy complex sounds (e.g., *al tasto* overpressure) and complex harmonic sounds (multiphonics), both deriving from the organological model initially described. This distinction between noisy and multispectral sounds makes it possible to relate the organological model to its perceptual outcomes, making explicit the correlation between the topology of the vibrating system and the resulting spectromorphological organization (Table 2).

Table 2: Acoustical-Perceptual Taxonomy of Bowed String Instrument Sounds.

Balanced Sounds	<i>optimal balance of the harmonic spectrum favoring the fundamental</i>
Filtered/Equalized Sounds	<i>certain components of the harmonic spectrum are favored over others, behaving as a multiband equalizer</i>
<i>Band pass</i> (harmonics)	
<i>Low-shelf cut</i> (rubber mute)	

Resonant peaks filter (formant mutes)
High-shelf cut (heavy metal mute)

Complex Sounds

Harmonic sounds (multiphonics)

Inharmonic sounds

- Multi-spectral (patafix)
- Noisy (overpressure 'al tasto')

Energy does not decrease linearly from low to high frequencies across the spectrum, but is instead redistributed or amplified in upper frequency bands to compensate for a loss of energy in the fundamental

2. Case Study: Patafix and Clothespins

The corpus was recorded by cellist Alessandro Mazzacane, using IRCAM's SOL (Cella et al., 2020) as a methodological reference. Analyses were performed using audio descriptors from *Essentia*, IRCAM's *pm2*, *pYin* and *supervp* algorithms, and STFT to identify the prominent partials and their amplitude (Peeters et al. 2011). Data are then organized into a database and corresponding dataset, all implemented with the bach, ears, and dada libraries in Max8. The dataset is compatible with the Orchidea architecture for computer-aided orchestration.

The analyses visualization and navigation interface provide all the information needed to read and interpret the databases from different dimensions. (Figure 1)

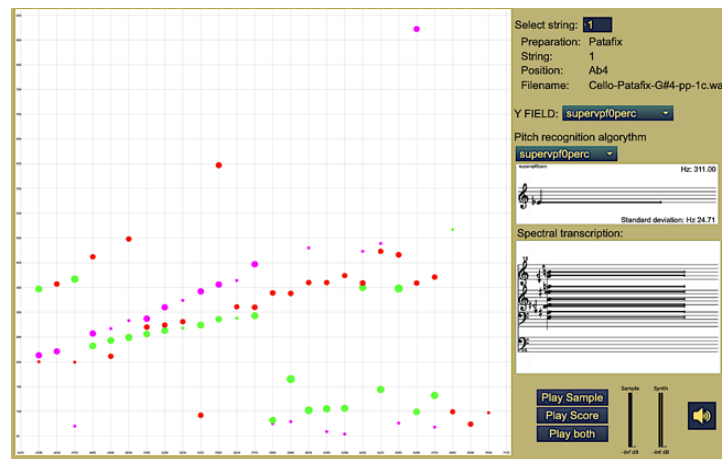


Figure 1: Interface developed for the navigation and analysis of the sound corpus.

The three colours determine the preparation: Patafix in red, big clothespin in green, small clothespin in magenta. The dataset is organized to map, for each sample, the location in the timbral space (centroid and spread), the nine STFT spectral peaks with the corresponding musical notation transcription, the pitch estimates returned by five algorithms (*pYIN*, *supervp maximum peak*, *supervp:f0percussion*, *predominant pitch melodia*, *pm2 inharmonic partial detections*) together with their standard deviation, and the position on the fingerboard.

The standard deviation across algorithms operates both as an indicator of measurement error and as the operational moniker of the complex sound of the preparation itself. Since a single signal contains multiple and simultaneously plausible *f0* candidates, each algorithm locks onto a different one, and the resulting disagreement quantifies the machine-listening ambiguity of pitch that defines this sound category.

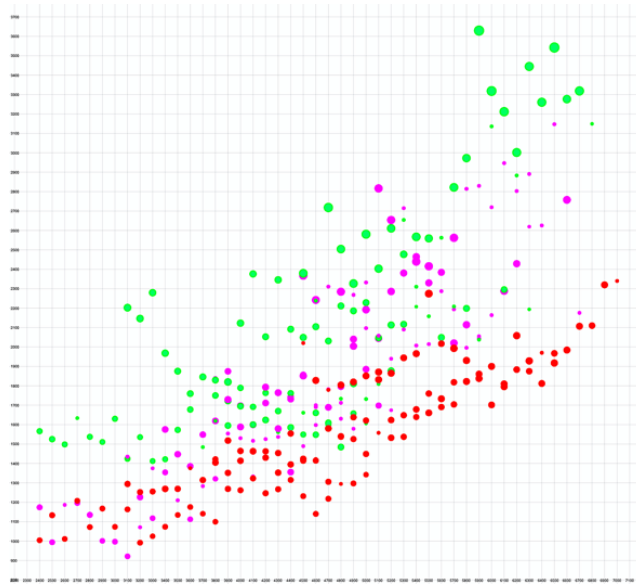


Figure 2: Centroids on Y axis, chromatic position on the instrument on X axis.

In this first case (Figure 2) the spectral centroid of the whole dataset is ordered by the pitch of the played note, and it shows how the three preparations occupy different regions of the timbral space. The Patafix forms a coherent and relatively linear and narrow trajectory in which brightness scales with pitch, showing that the mechanical (and then spectral) behavior of perturbation of the string function in a similar way across the different strings, and overall is the preparation with the darkest timbral characteristics. The two different clothespins, namely ClipBig and ClipSmall are positioned systematically higher and with greater dispersion: the centroid decouples from the played note, indicating that the preparation reconfigures its overall physical and spectromorphological behaviour.

The distance among the three populations in timbral space is therefore correlated with the relative mechanical impedance. The greater the impedance of the preparation with respect to the local impedance of the string, the larger the deviation from the perturbed-string regime.

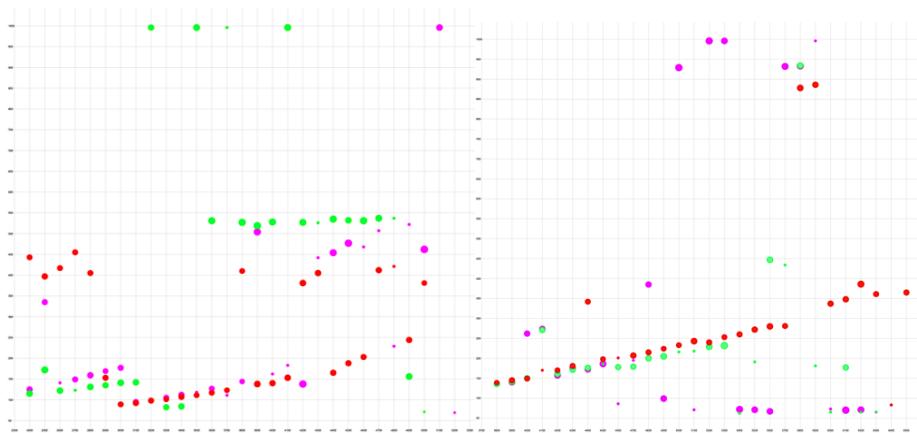


Figure 3: Strings 2 and 4. *Supervp:f0percussion* on Y axis, chromatic position on the instrument on X axis.

The two examples in figure 3 show the output of *supervp:f0percussion*, the most accurate among the algorithms employed, along the chromatic scale, for the second and the fourth strings, which differ in characteristic impedance. On both strings the statistical dispersion of the clothespin clouds becomes particularly evident from the first octave of the fingerboard onwards. This threshold is geometric. As the stopping point approaches the position where the vibrating length is halved, the preparation is fixed at one quarter of the open string from the bridge, maximizing its mass-loading effect on that mode. Above this threshold, the comparison between the two strings reveals the impedance gradient itself. On the second string, with lower characteristic impedance, the dispersion is broadly distributed and chaotic. On the fourth string of higher impedance, the same preparations produce more orderly clusters running parallel to the Patafix curve.

References

- Cella, C. E., Ghisi, D., Lostanlen, V., Lévy, F., Fineberg, J., & Maresz, Y. (2020). OrchideaSOL: A dataset of extended instrumental techniques for computer-aided orchestration. *ArXiv*. <https://doi.org/10.48550/arXiv.2007.00763>
- Fletcher, N. H., & Rossing, T. D. (1998). *The Physics of Musical Instruments* (2nd ed.). New York, USA: Springer.
- Heyde, H. (1975). *Grundlagen des natürlichen Systems der Musikinstrumente*. Leipzig: Deutscher Verlag für Musik.
- Peeters, G., Giordano, B. L., Susini, P., Misdariis, N., & McAdams, S. (2011). The Timbre Toolbox: Extracting audio descriptors from musical signals. *Journal of the Acoustical Society of America*, 130, 2902–2916.
- Reybrouck, M. (2006). Music Cognition and the Bodily approach: Musical Instruments as Tools for Musical Semantics. *Contemporary Music Review*, 25(1–2), pp. 59–68.
- Rossing, T. D. (Ed.). (2010). *The Science of String Instruments*. New York, USA: Springer.
- Thelin, H. (2011). *Multiphonics on the double bass. An investigation on the development and use of multiphonics on the double bass in contemporary music*. Norwegian Academy of Music.

Analyzing String Instrument Preparation Techniques in Context: Timbral Space, Syntax and Form

Ingrid Pustijanac^{1†} and Michele Leggieri¹

¹ Department of Musicology and Cultural Heritage, University of Pavia, Italy

[†] Corresponding author: ingrid.pustijanac@unipv.it

1. Situating the Research

In his seminal article *The Revolution of Complex Sounds*, Tristan Murail acknowledges the expansion of musical material to include complex sounds, emphasizing that “any attempt to integrate these new sounds, which are above all [...] sounds of a ‘complex’ character, necessitates a profound revision of traditional compositional techniques” (Murail, 1982, 2005, p. 121). He further states that “if these sounds [...] are studied with some scrutiny, more rational methods could be discovered that could well give rise to a new musical logic” (Murail, 1982, 2005, p. 133). Murail records a well-established practice that, since the 1980s, has received new stimuli from research in psychoacoustics and timbre (Lerdahl 1987; Smalley 1997, Pressnitzer & McAdams, 2000), providing accurate data to support the exploration and creation of new materials—a situation that, with increasingly sophisticated computer-assisted composition and orchestration techniques, has led to an extraordinary development both technically and imaginatively (Cella et al. 2020). In this context, the pivotal research on the extraction of spectral descriptors from musical signals presented in 2011 (Peeters et al.) opened the way to a more systematic and scientifically grounded investigation of the characteristics and relationships among different sounds and sonic objects.

This research inaugurated a new phase in contemporary composition of complex sounds, making explicit the data that had long informed compositional practice on an empirical basis. The spectromorphological complexity of individual sounds produced through the extensive use of extended techniques (such as multiphonics, scrapes, and other effects), direct intervention on the instrument’s body (ranging from minimal, easily reversible preparations to radical transformations), and the construction of self-made (DIY) musical instruments has opened up numerous new possibilities. As the range of sonic materials has expanded, the question that Murail first raised becomes even more pressing, prompting a central inquiry: what principles and rules enable the organization of this new type of material in a way that is coherent, intelligible, and expressive?

The present study seeks to identify and examine key parameters involved in a selected subset of complex sounds produced on prepared string instruments, concentrating in particular on sounds generated by preparations that directly modify the mechanics of string vibration such as Patafix, paper clips, hairpins, plastic and wooden clothespins, and aluminium foil.

2. Methodology and Theoretical Framework

Due to the practical, often tacit nature of the data embedded in the composer-performer/ensemble relationship—which constitutes the primary space in which the foundational elements of this research manifest—the methodology integrates three complementary approaches: (1) research on prepared string sounds based on recording and analysis protocols, guided by String Instrument Preparation Techniques Taxonomy (SIPTT), developed by the CCoRP Group (see Azzan, Guidarini 2026); (2) an ethnographic perspective on performers’ experience and composers’ practices, based on questionnaires and interviews conducted between 2019 and 2026 within the *Musica del XXI secolo* Project at the University of Pavia; and (3) systematic analysis of scores and performance notes. This integrated approach allowed for the

combination of data pertaining to the “grammar” of the studied sounds from a broader perspective. First, spectromorphological data were observed in their embodied dimension by directly examining the methods of sound production; second, the SIPTT served as the lens of observation, enabling an understanding of the data also in their gestural and energetic dimension and adding layers of meaning to the acoustic and psychoacoustic aspects of discretized phenomena.

The information obtained was subsequently cross-referenced with data gathered from composers’ questionnaires and interviews, allowing for the consolidation of insights pertaining to the spectromorphological domain—for example, the selection of sonic qualities based on direct exploration of the instrument, either individually or in collaboration with a performer. These data also enabled the research to expand toward a more semantic level, reflecting the ways in which composers verbally describe complex sounds, their behavior, and the compositional projects in which they are employed. The final level, where all these data converged, was the analysis of compositions—through score study and recording analysis—which provided a context-specific framework for the materials under investigation. From the integration of these different sources, several preliminary results were derived, which are presented in the following paragraph.

3. Data Analysis

The composers and works examined represent only a portion of a much larger repertoire, yet they are sufficiently representative of certain general trends presented in this study. For each composer, works for solo strings, string quartets or small ensembles (up to five performers), as well as pieces for larger ensembles and orchestra, were considered. This classification, as will be discussed later, proved crucial for structuring the analytical data. In addition to instrumentation, information was recorded for each work on the year of composition, duration, dedicatee performer (when present, often collaborating in the exploratory phase), and the use of amplification or electronics. All SIPTs active in each work were also recorded, along with their descriptions in the performance notes, including those that went beyond the previously mentioned categories, which primarily concern actions on the transducer (= string). This approach allowed for the collection of broader data on the entire ecosystem, supporting the hypothesis regarding the dynamic interactions and correlations between different domains, as demonstrated in Azzan & Guidarini (2026). The final element included in the systematic survey concerned spectromorphological aspects. This allowed for a comparative analysis of the use of the same preparation across different contexts and compositional projects, enabling the formulation of preliminary results.

As illustrative examples, several specific case studies are provided. For instance, in Mauro Lanza’s work, preparations are progressively integrated into complex timbral contexts. In *La bataille de Caresme et de Charnage* (2012), the prepared sound serves a contrastive function with respect to unprepared strings and multiphonics. In *John Conway in Gondola* (2016), for a prepared trio, the preparation loses its oppositional character and instead emphasizes the spectral content of melodic modules and chords. In *Gretchen and the Fragment on Machines* (2022), the prepared strings are merged with wind multiphonics and prepared piano, creating cohesive yet unstable textures: the prepared sound no longer emerges in isolation, but rather as part of an instrumental synthesis.

A similar distinction between contrastive and integrated use of preparations can be observed in Giovanni Verrando’s works. In *Fourth Born Unicorn, Angular Version* (2016), for prepared viola, the tuning fork replaces the traditional bow (mediator), and articulation arises from spectromorphological micro-differences—such as variations in contact points of the tuning fork, degrees of pressure on the strings, and bowing or striking speed. In *Instrumental Freak Show* (2017–2020), by contrast, the same preparation is embedded within a complex timbre alongside electric guitar, distorted voice, and thunder tube; here, the micro-differences are no longer perceptible as autonomous structural elements.

In the *Dead Wasps in the Jam Jar* cycle (2014–2018), Clara Iannotta develops three works based on the same sound, produced by applying circular clips to the strings. In the solo violin piece, spectromorphological variations maintain a contrastive and syntactic function, differentiating gestures through varying bow pressures. In the string quartet, the sound exhibits textural and temporal expansion, characteristic of a timbre-based composition. In the string orchestra, the prepared sound becomes a true generative core: the initial sound is expanded using sinusoids, birdcall, polystyrene, crystal glasses, and plexiglass. Iannotta conceives the sound as a combination of layers and perspectives (“zoom in/out”): by manipulating spectromorphology, register, and perceptual distance, she constructs depth and formal transformation without relying primarily on dynamics.

4. Preliminary Results and Discussion

The comparative analysis of the considered corpus shows how in a solo context, the syntactic dimension—that is, the temporal articulation of individual spectromorphological morphemes—can be shaped even by subtle differences between sounds (dynamics, roughness, brilliance, level of inharmonicity or multispectral content, stability/instability), as such distinctions are clearly perceptible to the listener. Preparations can thus assume contrastive functions, which are significant at the grammatical and structural level and can be valorised by a focalised amplification and/or electronics. Prepared sound can assume a contrastive function, for example, in relation to overpressure and multiphonics on unprepared strings (Mauro Lanza, *La bataille de Caresme et de Charnage*), or it can support entire sections through micro-spectromorphological variations achieved by manipulating bow pressure, contact points, and the speed of bowing and/or striking (Clara Iannotta, *Dead Wasps in the Jam Jar I*), or by using a substitute mediator (Heyde 1975)—for instance, the tuning fork in Giovanni Verrando, *Fourth Born Unicorn, Angular Version*. The dimension most strongly correlated with the properties of this type of prepared sound is duration. In all cases examined, the sounds do not appear in isolation or as discretized events, but rather as sustained (or reiterated) gestures, long enough to allow the perception of all characteristic components—described by the composers using terms such as inharmonic, unstable, residual, imperfect, parasitic, or degraded, blending technical descriptions with a more abstract semantic references (Noble et al. 2020).

In a chamber or orchestral context, on the contrary, the same micro-differences are absorbed into the overall timbre and are not perceptually robust enough to sustain the same contrastive function. In the face of a composite ensemble, within a post-spectral context, the sound of a preparation tends instead to assume the role of a model, from which both materials (model as a core) and compositional processes (model as an internal logic) can be derived.

An illustrative example of a prepared sound functioning as a model-as-core is Verrando’s already mentioned *Instrumental Freak Show* (2017–2020), in which the same preparation is integrated into a complex timbre. In terms of model-as-internal-logic, several works highlight how preparations reinforce or expand the underlying spectral structure of the material. In Mauro Lanza’s *John Conway in Gondola*, the preparation emphasizes the spectral content of the melodic modules, while in *Gretchen and the Fragment on Machines*, it contributes to an expansion of the model by reinforcing specific spectromorphological qualities. Similarly, in Clara Iannotta’s *Dead Wasps in the Jam Jar II and III*, the model is expanded through the reinforcement of particular spectral features, such as artificial harmonics and the use of crystal glasses as an echo of the model sound. A similar function of mirroring and expanding the spectral qualities of the prepared string sounds can be found in Maurizio Azzan’s *Wasteland_beyond*. Here, the percussionists and pianists employ a wide array of objects and preparations capable of coloring and highlighting, in each instance, specific properties of the prepared strings, exploring different levels of noisiness.

The study of prepared sounds within larger ensemble contexts is thus situated within the broader field of contemporary orchestration (Lévy 2026). The expansion of resources—namely, the variety of sound sources used to optimally render model or target sounds (as in the case of the OAC)—becomes an integral part of the research, contributing to the development of new means for formal articulation. These methods are inevitably shaped by the demands of highly diverse individual compositional poetics. Across different dimensions—perception, sonic material, processes, gestures, concepts, and temporal structure—clear organizational logics emerge, yet these remain in need of more systematic study, a gap to which the present research seeks to contribute.

References

- Azzan, M., Guidarini, L. (2026). Classifying String Instrument Preparation Techniques: Analysis, Taxonomies, and Applications in Computer-Aided Orchestration. *Proceedings*.
- Cella, C. E., Ghisi, D., Lostanlen, V., Lévy, F., Fineberg, J., & Maresz, Y. (2020). OrchideaSOL: A dataset of extended instrumental techniques for computer-aided orchestration. *ArXiv*. <https://arxiv.org/abs/2007.00763>
- Heyde, H. (1975). *Grundlagen des natürlichen Systems der Musikinstrumente*. Leipzig: Deutscher Verlag für Musik.
- Lerdahl, F. (1987). Timbral hierarchies. *Contemporary Music Review*, 2(1), 135–160.
- Lévy, F. (forthcoming 2026). Functional orchestration: A research field at the beginning of its history. In R. Hasegawa, J. Delisle, J. Noble, & M. Touizrar (Eds.), *The Oxford handbook of orchestration studies*. Oxford University Press.
- Murail, T. (2005). The revolution of complex sounds (J. Cody, Trans.). *Contemporary Music Review*, 24(2–3), 121–135.
- Noble, J., Thoret, E., Henry, M., & McAdams, S. (2020). Semantic dimensions of sound mass music: Mappings between perceptual and acoustic domains. *Music Perception: An Interdisciplinary Journal*, 38(2), 214–242.
- Peeters, G., Giordano, B. L., Susini, P., Misdariis, N., & McAdams, S. (2011). The Timbre Toolbox: Extracting audio descriptors from musical signals. *Journal of New Music Research*, 40(3), 201–231.
- Pressnitzer, D., & McAdams, S. (2000). Acoustics, psychoacoustics and spectral music. *Contemporary Music Review*, 19(2), 33–59.
- Smalley, D. (1997). Spectromorphology: Explaining sound-shapes. *Organised Sound*, 2(2), 107–126.

Hearing the Unconventional: A Study of Perceptual Fusion of Extended Techniques in Contemporary Classical Music

Jia Yi Lee^{1†} and Jenine Brown²

¹ Department of Music, Davidson College, Davidson, North Carolina, USA

² Peabody Institute of The Johns Hopkins University, Baltimore, Maryland, USA

† Corresponding author: jilee@davidson.edu

The study of orchestration in Western classical music has traditionally been focused on the effective combination of instrumental sounds to create a blended timbre. In contemporary classical music from the mid-20th century and beyond, this idea of blend is still relevant, but composers have become more interested in experimenting with extended instrumental techniques to create more novel and interesting sounds (Adler, 2016). Where extended techniques are discussed in orchestration treatises, they are often noted as “special effects” and not examined from a perceptual standpoint (Adler, 2016; Blatter, 1997; Piston, 1955).

More recently, music cognition scholars such as McAdams et al. (2004) have coined the term “perceptual fusion” (p. 184) to understand how we perceive blended timbres. Existing studies have largely addressed the perceptual fusion of instrumental sounds, often focusing on conventional techniques (i.e., pitched sounds naturally produced by instruments) (Kendall & Carterette, 1993; Lembke & McAdams, 2015; Lembke et al., 2019; Peynircioğlu et al., 2015; Sandell, 1995; Tardieu & McAdams, 2012). However, the perception of novel timbral fusion involving extended techniques remains underexplored. As composers experiment with extended techniques in combination with conventional sounds, listeners are presented with new combinations of timbres to hear and internalize.

This paper utilizes a two-pronged approach in studying the perceptual fusion of extended techniques. First, an analytical approach studies trends of timbral combinations in selected contemporary compositions after 1970, cataloging dyadic blends involving extended techniques in works by composers such as Lachenmann, Murail, and Saariaho. The analysis illustrates that composers use combinations of sounds to augment a primary timbre or create a novel timbre, the latter of which often uses extended techniques with a higher noise content to create dissonance and musical tension. Hence, a new method of categorization of extended techniques is proposed based on timbral characteristics (degree of pitch vs. noise content), which then informs the creation of stimuli for the subsequent experiment.

Second, we describe an experiment that examines the perceptual fusion of timbre dyads, replicating elements of the experimental design of Tardieu and McAdams (2012) but replacing the stimuli with frequent timbral combinations resulting from the prior analysis. The experiment tests the effect of instrumental combination and type of performance technique on rating of blend. Findings suggest that the total spectral centroid as well as the total noisiness of both sounds together influenced the perception of blend, specifically, the higher the total spectral centroid and noisiness, the lower the participants’ rating of blend, resulting in a sound that is perceived as not well-blended. Moreover, participants, who were non-musicians, were able to perceive these various categories of sounds as perceptually different, despite the unfamiliarity of such novel timbres.

This study has numerous implications for composition and future experimental studies in timbre perception. This study can motivate a new way of analyzing music that focuses primarily on extended techniques. By integrating the methodology of extended technique categorization into existing methods of timbre analysis, we can examine how the perception of timbre relates to the perception of gesture and form in music, for example, through the use of timbral combinations with different characteristics to

indicate structural moments in music. Furthermore, since listeners can perceive different levels of timbral blend involving both conventional sounds and extended techniques, it may be that they can also differentiate between complex timbral changes in contemporary music and are able to use this knowledge in the understanding of formal design in these works. This work also has implications for composers, given that study findings suggest that the perception of blend is dependent on instrumental combination and spectral characteristics; this enables a better justification for using combinations of sounds that are perceived to be well-blended or not well-blended for specific musical and formal effects. Future work can involve replicating this study with musician participants, comparing responses between musicians and non-musicians, as well as build upon this current methodology of categorizing extended techniques to analyze music by a wider range of composers, and other genres/styles of music.

Table 1: List of individual sounds.

Instruments / Techniques	Conventional (C)	Extended technique 1 (E1)	Extended technique 2 (E2)
<i>Strings:</i> <i>Violin (Vn) /</i> <i>Cello (Vc)</i>	<i>ord. (arco)</i>	<i>sul ponticello</i>	<i>overpressure</i>
<i>Woodwinds:</i> <i>Flute (Fl) /</i> <i>Clarinet (Cl)</i>	<i>ord.</i>	<i>aeolian*</i>	<i>fluttertongue</i>

*Note: half-aeolian & half-pitched sounds were selected so as to retain a somewhat pitched quality

Table 2: Example of one stimulus in each category (Instrument & Technique), according to a 2x2 ANOVA design.

Instrument	Technique			
	same vs. same (C vs. C, E1 vs. E1, E2 vs. E2)	C vs. E1	C vs. E2	E1 vs. E2
<i>Same instrument</i>	<i>e.g. Vn ord. + Vn ord.</i>	<i>e.g. Vn ord. + Vn sul pont.</i>	<i>e.g. Vn ord. + Vn overpressure</i>	<i>e.g. Vn sul pont. + Vn overpressure</i>
<i>Same family (strings)</i>	<i>e.g. Vn sul pont. + Vc sul pont.</i>	<i>e.g. Vn ord. + Vc. sul pont.</i>	<i>e.g. Vn ord. + Vc overpressure</i>	<i>e.g. Vn sul pont. + Vc overpressure</i>
<i>Same family (woodwinds)</i>	<i>e.g. Fl fluttertongue + Cl fluttertongue</i>	<i>e.g. Fl ord. + Cl aeolian</i>	<i>e.g. Fl ord. + Cl fluttertongue</i>	<i>e.g. Fl aeolian + Cl fluttertongue</i>
<i>Different family (string + woodwind)</i>	<i>e.g. Vn ord. + Fl ord.</i>	<i>e.g. Vn ord. + Fl aeolian</i>	<i>e.g. Fl ord. + Vc overpressure</i>	<i>e.g. Fl aeolian + Vc overpressure</i>

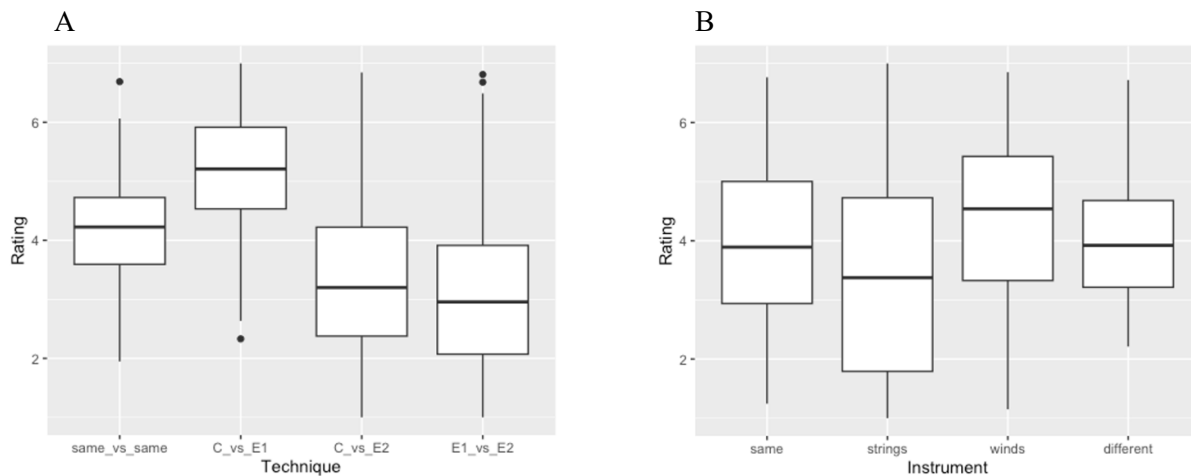


Figure 1: Box plots showing effect of Technique (A) and Instrument (B) on participants' mean blend ratings.

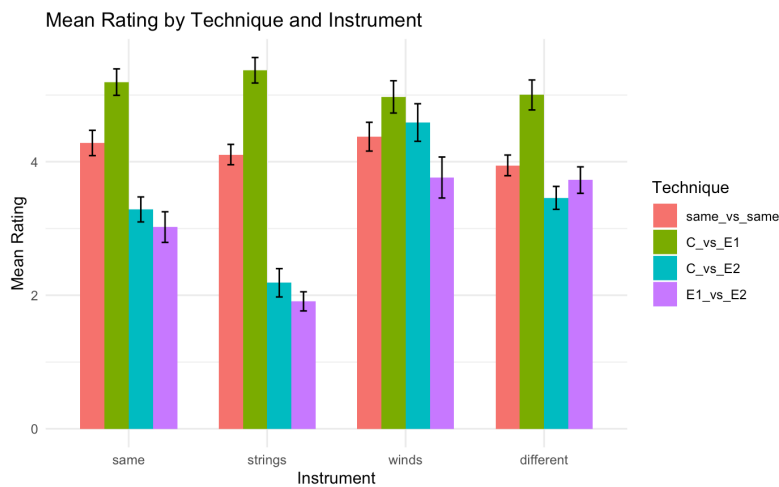


Figure 2: Grouped bar plot showing the interaction between Technique and Instrument on participants' mean blend ratings.

References

- Adler, S. (2016). *The study of orchestration: Fourth edition*. W.W. Norton.
- Belkin, A. (1988). Orchestration, perception, and musical time: A composer's view. *Computer Music Journal*, 12(2), 47–53. <https://doi.org/10.2307/3679941>
- Blake, D. (2018). Timbre. In *The Oxford Handbook of Critical Concepts in Music Theory* (pp. 136–159). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780190454746.013.21>
- Czernowin, C. (2010). *The Quiet* [Musical Score]. Schott Music.
- Czernowin, C. (2015). *Ayre* [Musical Score]. Schott Music.
- Davies, H. (2020). Instrument modifications and extended techniques. *Grove Music Online*. Retrieved 5 May 2025, from

<https://www.oxfordmusiconline.com/grovemusic/view/10.1093/gmo/9781561592630.001.0001/omo-9781561592630-e-90000329463>

- Fales, C., & McAdams, S. (1994). The fusion and layering of noise and tone: implications for timbre in african instruments. *Leonardo Music Journal*, 69–77.
- Fischer, M., Soden, K., Thoret, E., Montrey, M., & McAdams, S. (2021). Instrument timbre enhances perceptual segregation in orchestral music. *Music Perception: An Interdisciplinary Journal*, 38(5), 473–498.
- Gervasoni, S. (2020). *Eufonique* [Musical Score]. Schott Music.
- Kendall, R. A., & Carterette, E. C. (1993). Identification and blend of timbres as a basis for orchestration. *Contemporary Music Review*, 9(1-2), 51–67.
- Krumhansl, C. L., & Iverson, P. (1992). Perceptual interactions between musical pitch and timbre. *Journal of Experimental Psychology: Human Perception and Performance*, 18(3), 739.
- Lachenmann, H. (1971). *Kontrakadenz* [Musical Score]. Breitkopf & Härtel.
- Lachenmann, H. (1992). „...zwei Gefühle...“, *Musik mit Leonardo* [Musical Score]. Breitkopf & Härtel.
- Lembke, S. A., & McAdams, S. (2015). The role of spectral-envelope characteristics in perceptual blending of wind-instrument sounds. *Acta Acustica*, 101(5), 1039–1051.
- Lembke, S. A., Parker, K., Narmour, E., & McAdams, S. (2019). Acoustical correlates of perceptual blend in timbre dyads and triads. *Musicae Scientiae*, 23(2), 250–274.
- McAdams, S. (2019). Timbre as a structuring force in music. In *Timbre: Acoustics, perception, and cognition* (pp. 211–243). Cham: Springer International Publishing.
- McAdams, S., Depalle, P., & Clarke, E. (2004). Analyzing musical sound. *Empirical musicology: Aims, methods, prospects*, 157–196.
- 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>
- McAdams, S., & Giordano, B. L. (2014). The perception of musical timbre. In *The Oxford Handbook of Music Psychology* (pp. 113–124). Oxford University Press.
<https://doi.org/10.1093/oxfordhb/9780198722946.013.12>
- Murail, T. (1976). *Mémoire/Érosion* [Musical Score]. Editions Musicales Transatlantiques.
- Murail, T. (1978). *Ethers* [Musical Score]. Editions Musicales Transatlantiques.
- Peeters, G., Giordano, B. L., Susini, P., Misdariis, N., & McAdams, S. (2011). The timbre toolbox: Extracting audio descriptors from musical signals. *The Journal of the Acoustical Society of America*, 130(5), 2902–2916.
- Peynircioğlu, Z. F., Brent, W., & Falco, D. E. (2016). Perception of blended timbres in music. *Psychology of Music*, 44(4), 625–639.
- Piston, W. (1955). *Orchestration*. W.W. Norton.
- Rimsky-Korsakov, N. (1964). *Principles of Orchestration*. Dover Publications.
- Saariaho, K. (2002). *Orion* [Musical Score]. Chester Music.
- Sandell, G. J. (1995). Roles for spectral centroid and other factors in determining "blended" instrument pairings in orchestration. *Music Perception*, 13(2), 209–246.
- Saunders, R. (2011). *Still* [Musical Score]. Edition Peters.
- Tardieu, D., & McAdams, S. (2012). Perception of dyads of impulsive and sustained instrument sounds. *Music Perception*, 30(2), 117–128.
- Wallmark, Z. (2019). A corpus analysis of timbre semantics in orchestration treatises. *Psychology of Music*, 47(4), 585–605.

Applying Timbre Analysis in the Performance of Chamber Music With Timpani and Cello

Manuel López Tovar^{1†}

¹ Schulich School of Music & Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT),
McGill University, Montréal, Québec, Canada

[†] Corresponding author: manuel.lopeztovar@mail.mcgill.ca

Pieces written for timpani and cello are more common than one would expect. Even though this pairing seems odd due to their different sonic characteristics, it often works very well. “Fantasy for Cello and Percussion” (1986) by David Maves (b. 1937) is one clear example of a successful composition using both instruments. Nonetheless, special attention must be given to certain sections in which the timbre of the timpani may not match the cello’s, overpower its leading line and create an ensemble imbalance. Previous research (Rossing, 1982; Benade, 1990; Rossing et al., 2000; Fleischer, Fastl, 2008) has explored timpani acoustics extensively, yet little attention has been given to its interaction with other instruments in chamber settings. With the instrument’s particular acoustic properties, how can we achieve the right blending in challenging passages? A practical solution would involve altering the instrument’s timbre by changing the playing spot (centre, normal playing spot, edge), changing mallets (medium, soft, hard, different cores, etc.), the type of stroke, and many more. However, the music in the first movement, mm. 23-40 (**Erreur! Source du renvoi introuvable.**), presents a sudden change in character (articulated, fast rhythms vs legato, continuous roll) in a continuous musical line that prevents what would be the performer’s first choice: a change of mallets. The remaining possibilities are then a change in the roll speed, roll type (open vs buzz), beating position (centre, normal, edge), or spacing of the hands (mallets close together vs separated).

FANTASY
I
David Maves

The image displays a musical score for 'Fantasy I' by David Maves, specifically measures 23 through 42. The score is written for two instruments: Cello and Timpani. The Cello part is in treble clef, and the Timpani part is in bass clef. The score shows a complex interplay between the two instruments. In measure 23, the Cello has a long rest, while the Timpani plays a fast, articulated roll. In measure 24, the Cello enters with a series of eighth notes. In measure 25, the Cello has a long rest, and the Timpani plays a fast, articulated roll. In measure 26, the Cello enters with a series of eighth notes. In measure 27, the Cello has a long rest, and the Timpani plays a fast, articulated roll. In measure 28, the Cello enters with a series of eighth notes. In measure 29, the Cello has a long rest, and the Timpani plays a fast, articulated roll. In measure 30, the Cello enters with a series of eighth notes. In measure 31, the Cello has a long rest, and the Timpani plays a fast, articulated roll. In measure 32, the Cello enters with a series of eighth notes. In measure 33, the Cello has a long rest, and the Timpani plays a fast, articulated roll. In measure 34, the Cello enters with a series of eighth notes. In measure 35, the Cello has a long rest, and the Timpani plays a fast, articulated roll. In measure 36, the Cello enters with a series of eighth notes. In measure 37, the Cello has a long rest, and the Timpani plays a fast, articulated roll. In measure 38, the Cello enters with a series of eighth notes, and the Timpani plays a fast, articulated roll. In measure 39, the Cello has a long rest, and the Timpani plays a fast, articulated roll. In measure 40, the Cello enters with a series of eighth notes, and the Timpani plays a fast, articulated roll. In measure 41, the Cello has a long rest, and the Timpani plays a fast, articulated roll. In measure 42, the Cello enters with a series of eighth notes, and the Timpani plays a fast, articulated roll. A purple box highlights the section in measure 38 where the timpani plays a fast, articulated roll while the cello plays a slower, legato line. The score includes various musical notations such as notes, rests, and dynamic markings like mf, p, and cresc.

Figure 1: *Fantasy for Cello and Percussion* by David Maves, mm. 23-42.

Materials and Methods

To determine how these changes altered the timbre of the timpani and the blending with cello, multiple audio samples were recorded from the most problematic excerpt in the first movement (mm. 36-38), including timpano (timpani= plural, timpano=singular) playing solo, and simultaneous cello+timpano playing for blending. The timpano used for the recording was a 24" Yamaha TP-7224H equipped with a year-old Evans Orchestral head. The drum was tuned in Ab₃ (A₄=442Hz). The selected mallets were a pair of ultra staccato Freer Percussion US2H – STACCATO Hickory Extra Hard. The microphones used for the timpano recording were a pair of MKH20 (omni-directional condenser) (**Erreur! Source du renvoi introuvable.**, left panel), placed over 2m high on the opposite side of the performer. For the cello, a single microphone of the same reference was placed 1m above the ground, 80cm from the performer (Figure 2, right panel). The recordings were done in the CIRMMT hemi-anechoic room to avoid any sort of room interference. For the timpano-solo recordings, 6 different playing positions were proposed: center, normal, and near the rim positions, each with 2 variants—hands close together and separated by 10cm (See bottom of **Erreur! Source du renvoi introuvable.**3).

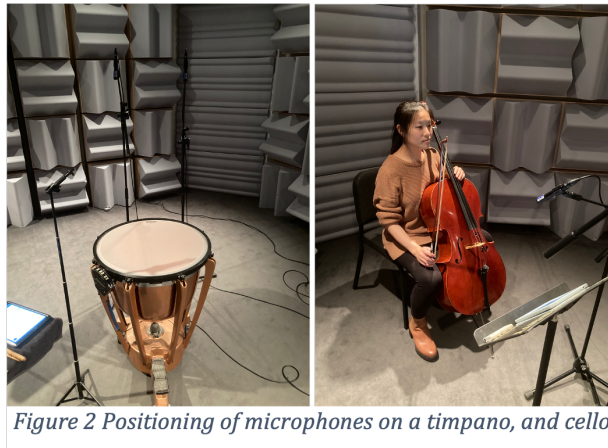


Figure 2 Positioning of microphones on a timpano, and cello

Results

A simple timbral analysis was performed with the help of spectrograms, and “The Timbre Toolbox” (Kazazis et al., 2022), updated from the original version by Peeters et al., 2011. The timpano-solo samples were represented with the help of a spectrogram for subsequent visual comparison (Figure 3). The findings were that playing in the centre excites the fundamental (first vibration mode (0,1)), and not enough higher frequencies. Because the perceived pitch on timpani comes from the first partial (vibration mode (1,1)), by playing in the centre the perceived pitch is E₃~F₃ instead of the previously tuned Ab₃ (see Rossing 2005). Playing at the normal beating spot excites vibration mode (1, 1) providing a strong sense of pitch. Playing near the rim excites higher frequencies, giving a brighter sound, with a less intense preferred vibration mode (1, 1). A buzz roll gives a smoother waveform, compared to an open roll. This translates into a continuous sound vs. a peakier, rougher sound.

The ‘timpani+cello’ samples were analysed with the toolbox, classifying the recorded sound signals within different descriptors: spectral centroid which is related to auditory brightness, spectral spread which indicates how rich the spectrum is, spectral kurtosis which indicates how peaky the distribution is around the centroid, and spectral crest which indicates (Kazazis et al., 2022). These descriptors, represented as values, served to establish a comparison (Figure 4). The conclusions were that the spectral centroid rises when playing closer to the rim. The spectrum is richer when playing closer to the rim with open rolls, contributing to a better blending. However, open rolls near the centre create a peaky

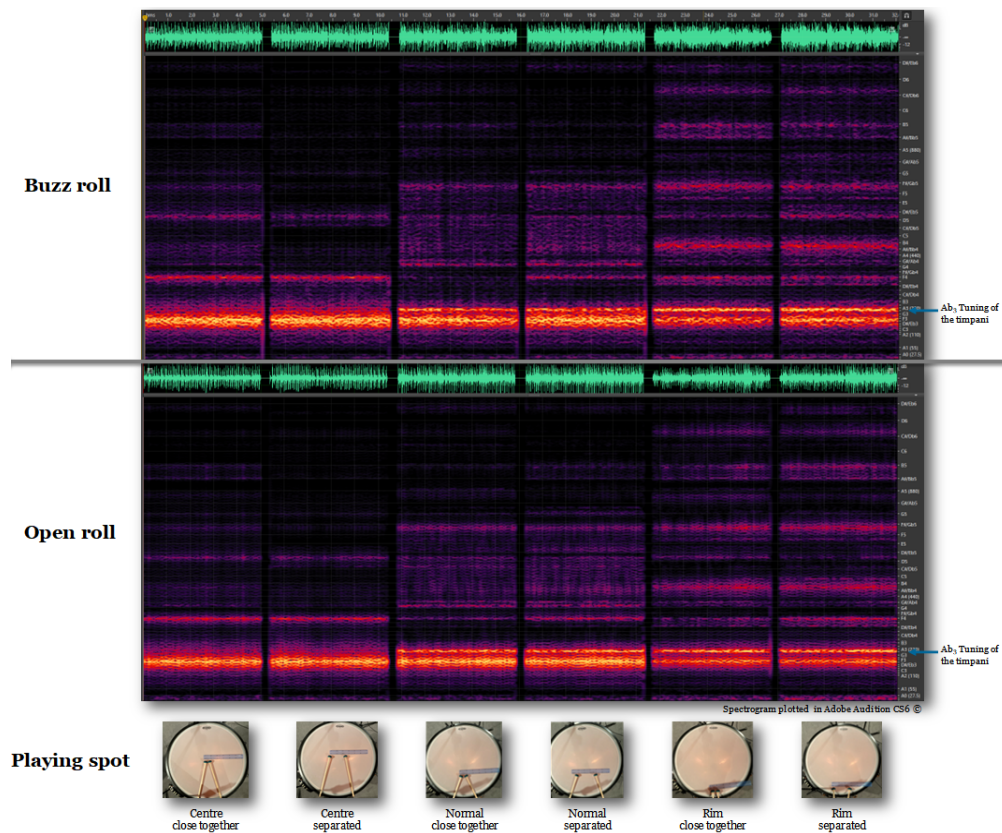


Figure 3: Spectrogram of timpano-solo recordings comparing 2 types of rolls, with 6 different playing spots

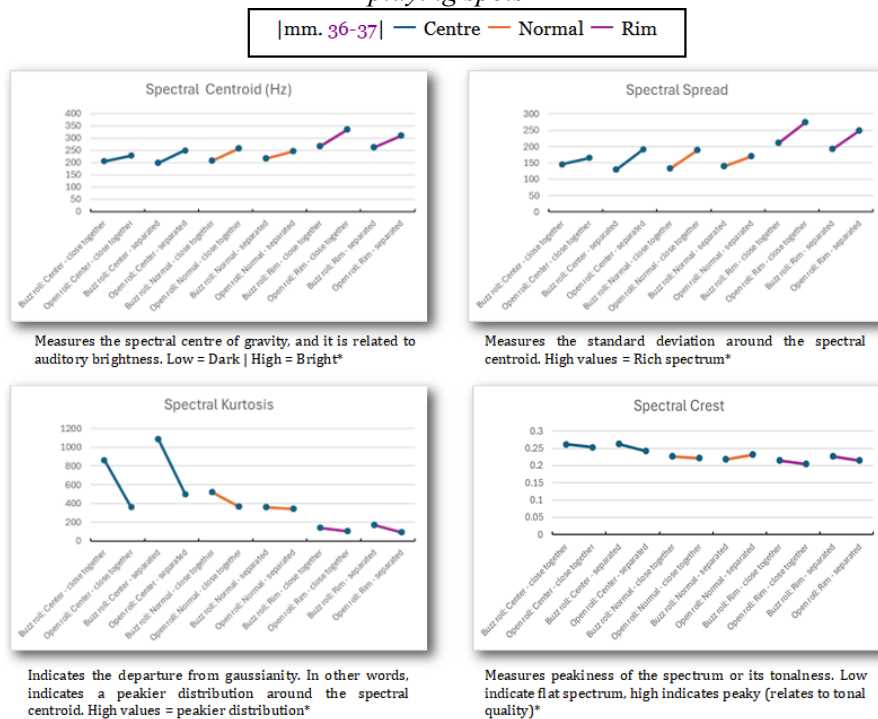


Figure 4: Spectral centroid, spectral spread, spectral kurtosis, and spectral crest of samples involving timpani and cello

spectrum. In contrast, near the rim is more uniform, but still peakier than a buzz roll: closer to the rim = stronger higher frequencies = brighter sound. The results allowed the performer to make an informed interpretation choice for the specific section. Moreover, other timbre balance issues in different sections of the music were discovered and analysed, proposing solutions that could be considered when performing with similar instrumentation.

Conclusion

Timpani timbral capabilities can be exploited not only by changing mallets, but also by altering the playing technique and beating spot. These can result in a darker or brighter sound, uniform or uneven sound, and/or an alteration of perceived pitch, which can help (or not) in blending with other instruments, in this specific case, cello. For the purpose of this excerpt, we were looking for a continuous, pedal-like dark roll sound that could blend enough with the cello line. After analysing the data and trying all the different options, it was decided that a buzz roll, normal beating spot with separated hands would be the best option. This preliminary study opens the door for further timpani timbre blending research, which can benefit performers by enhancing their tonal mastery over the instrument, useful in any ensemble playing settings.

References

- Benade, A. H. (with Internet Archive). (1976). *Fundamentals of musical acoustics*. New York: Oxford University Press. <http://archive.org/details/fundamentalsofmu0000bena>
- Fleischer, H., & Fastl, H. (2008). *Fell, Kessel und Gestell der Orchesterpauke*. Inst. für Mechanik.
- Kazazis, S., Depalle, P., & McAdams, S. (2021). *The Timbre Toolbox* [Computer software]. Retrieved <https://github.com/MPCL-McGill/TimbreToolbox-R2021a>
- López, M. (2024, October 18). *Preliminary study on timpani timbre-blending with cello* [Research poster]. The Space Between III. <https://percnet.ca/wp-content/uploads/2025/08/Acadia-program-FINAL.pdf>
- Peeters, G., Giordano, B., Susini, P., Misdariis, N., & McAdams, S. (2011). The Timbre Toolbox: Extracting audio descriptors from musical signals. *The Journal of the Acoustical Society of America*, 130(5), 2902–2916.
- Rossing, T. (2005). *Science of Percussion Instruments* (Vol. 3). World Scientific. (pp. 5-15).
- Rossing, T. (1982). The Physics of Kettledrums. *Scientific American*, 247(5), 172–179.

Acknowledgments

The audio recordings were done by Michael Borsellino, the cello was played by Crystal Kim, and the timpano was played by the author

Ethics certification: McGill University Research Ethics Board certificate 24-06-034.

Session 4
Crossmodality

How Transmodal Performative Settings Shape Timbral Perception

Ajtony Csaba^{1†}

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

[†] Corresponding author: acsaba@uvic.ca

Music may be understood not exclusively as organized sound, but as a modality-independent syntax of energetic density transformations capable of migrating across sensory domains while retaining perceptual identity. Conducting operates as a particularly revealing transmodal practice in this regard, translating visual and kinetic information into auditory structure while unfolding within a broader multisensory performance environment. Building on work in musical gesture studies (Godøy and Leman 2010), gesture–sound causality (Emerson and Egermann 2018), timbre semantics (Saitis and Weinzierl 2019), and auditory perception in orchestration (McAdams, Goodchild, and Soden 2022; Goodchild and McAdams 2021), this presentation examines how visible performative gestures connected to music shape the perception of timbral and energetic processes.

While relationships between musical loudness and the scale of conducting gestures have been comparatively well documented, the interaction between visible gesture and timbral perception remains less clearly theorized. Here, performative settings designate the combined visual, spatial, kinetic, and ritual conditions of performance, while timbral processes refer to musically relevant transformations occurring predominantly within the domains of timbre, texture, density, and perceptual organization.

The presentation proposes a unified framework in which gesture and timbral transformation may be approached as consubstantial expressions of a shared energetic syntax. The framework distinguishes between coordinating dimensions of conducting gesture—linked to measurable parameters such as tempo, inertia, and dynamic coordination—and qualitative dimensions associated with phrasing, energetic shaping, and narrative continuity. The qualitative layer enables the migration of musical identity across modalities, including in extended conducting practices and in the adoption of extra-musical gesture vocabularies functioning in conducting-like roles.

Rather than proposing a fully empirical model, the framework operates through descriptors signifying perceptual categories meaningful in both movement and timbre, building on research into timbre semantics, crossmodal correspondence, and embodied musical perception (McAdams 2019; Saitis and Weinzierl 2019; Wallmark 2019). Drawing on concepts of virtual agency (Hatten 2018), embodied musical gesture, and crossmodal perceptual organization, the presentation explores movement not merely as accompaniment to sound, but as an extended timbral and dramaturgical domain capable of sustaining musical continuity even beyond the acoustical duration of sound itself.

Case studies include Giacinto Scelsi's *Quattro pezzi*, Karlheinz Stockhausen's *Inori*, Hans Zender's concept of *Umschrift*, and practices such as Soundpainting, alongside examples from my own creative work involving movement as compositional material. Particular attention is given to ritualized and syncretic gesture vocabularies that, while not functioning as literal conducting techniques, nevertheless acquire musically operative identities within performative contexts. The analyses further demonstrate the potential use of the framework for performance dramaturgy planning and for the development of dramaturgical listening lenses.

The presentation concludes by proposing anchor points for an expanded framework in which static and dynamic spatialization, alongside movement and orchestration, also function as perceptually active components in timbral processes.

References

- Donin, N. (2012). Empirical and historical musicologies of compositional processes: Towards a cross-fertilisation. In D. Collins (Ed.), *The act of musical composition* (1st ed.). Routledge. <https://doi.org/10.4324/9781315612256-1>
- Donin, N. (Ed.). (2018). *The Oxford handbook of the creative process in music*. Oxford University Press.
- Emerson, G., & Egermann, H. (2018). Gesture–sound causality from the audience’s perspective: Investigating the multisensory experience of performances with digital musical instruments. *Psychology of Aesthetics, Creativity, and the Arts*, 12(1), 96–109.
- Godøy, R. I., & Leman, M. (2010). *Musical gestures: Sound, movement, and meaning* (1st ed.). Routledge. <https://doi.org/10.4324/9780203863411>
- Goehr, L. (1992). *The imaginary museum of musical works: An essay in the philosophy of music* (1st ed.). Oxford University Press. <https://doi.org/10.1093/0198235410.001.0001>
- Goodchild, M., & McAdams, S. (2021). Perceptual processes in orchestration. In E. I. Dolan & A. Rehding (Eds.), *The Oxford handbook of timbre*. Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780190637224.013.10>
- Hatten, R. S. (2018). *A theory of virtual agency for Western art music*. Indiana University Press.
- Lewin, D. (2001). Music theory, phenomenology, and modes of perception. *Music Theory Online*, 7(4).
- 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>
- McAdams, S. (2019). *The perceptual representation of timbre*. In K. Siedenburg, C. Saitis, S. McAdams, A. N. Popper, & R. R. Fay (Eds.), *Timbre: Acoustics, perception, and cognition* (pp. 23–57). Springer.
- Orning, T. (2017). Music as performance – gestures, sound and energy. *Journal for Research in Arts and Sports Education*, 1, 79–94. <https://doi.org/10.23865/jased.v1.946>
- 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. 119–149). Springer. https://doi.org/10.1007/978-3-030-14832-4_5
- Scelsi, G. (1959/1987). *Quattro pezzi su una nota sola* [Score]. Salabert.
- Stockhausen, K. (1974/1995). *Inori: Adorations for one or two soloists and orchestra* [Score]. Stockhausen-Verlag.
- Zender, H. (2004). *Die Sinne denken: Texte zur Musik 1975–2003* (J. P. Hiekel, Ed.). Breitkopf & Härtel.

Crossmodal Semantic Priming Between Timbre and Olfaction: Does Timbre Activate Semantic Associations?

Asterios Zacharakis^{1†}

¹ School of Music Studies, Aristotle University of Thessaloniki, Greece

[†] Corresponding author: aszachar@mus.auth.gr

Introduction

Recent work on the still underexplored domain of audio–olfactory correspondences has highlighted a prominent role for semantic mediation in linking timbres with aromas (Zacharakis, 2025). Specifically, it has been shown that portions of the variance in the strength of association between timbres and aromas—interpretable as degrees of perceived similarity or dissimilarity—are reflected by their relative positions within a shared semantic space. While these findings establish a robust descriptive structure for organising crossmodal associations, an open question remains as to whether such crossmodal semantic correspondences are functionally active, that is, capable of biasing semantic evaluations in real time.

The present contribution proposes a follow-up experiment designed to address this question using a crossmodal priming paradigm, commonly employed in the study of multisensory integration with a focus on timbre (Wallmark et al., 2021; Saitis & Wallmark, 2024). The core aim is to test whether brief exposure to a timbre or an aroma can systematically influence subsequent semantic ratings of stimuli in the other modality, in ways that reflect the semantic structure previously identified.

The primary hypothesis of this study is that semantically (in)congruent sensory and/or lexical primes may bias semantic ratings of target stimuli relative to control conditions on the semantic dimensions corresponding to the primes. This approach is adopted to further investigate the role of meaning in multisensory integration, assessing potential asymmetries between auditory-to-olfactory priming and vice versa.

Method

The experiment employed a within-subjects, bidirectional crossmodal priming design with an incomplete set of prime–target combinations per participant. A subset of five synthetic tones named after intended aromatic targets (Lemon Blossom.wav, Melon.wav, Vanilla.wav, Black Pepper.wav, and Truffle.wav) and five aromas (Lemon Blossom, Melon, Vanilla, Black Pepper, and Tobacco) was selected from a prior, larger stimulus set (Zacharakis, 2024, 2025) to represent distinct regions of the shared semantic space. Distances within this semantic space and similarities derived from direct associations were used to quantitatively define semantically congruent and incongruent sound–scent pairings (see Fig.1).

A convenience sample of 43 students from the School of Music Studies at the Aristotle University of Thessaloniki (29 females, mean age = 21.7 years, age range = 18–50 years, mean years practising music = 11.9) participated in the study. Participants were presented with a prime stimulus followed by a target stimulus that they then evaluated along a fixed battery of six semantic rating scales (i.e., brightness, sharpness, fullness, warmth, sweetness, and freshness) shown to be salient in prior work. The salience of these six semantic variables (drawn from an initial set of 12) was determined using a combination of criteria. These included weighting coefficients derived from modelling direct crossmodal similarity judgments for the same stimulus set (Zacharakis, 2024; 2025), inter-variable structure revealed by cluster analysis, and inspection of individual semantic profiles of timbres and aromas. The rating scales were presented on a computer screen after the target stimulus had been experienced, as horizontal sliders with

the cursor positioned at the midpoint by default. The endpoints of each scale were anchored by the terms “very” (right) and “not at all” (left). Participants were instructed to use at least one scale to describe each target, while being free to use the full range of the rating battery.

Primes belonged in three categories: (i) neutral control primes, (ii) lexical primes consisting of written nouns presented on a screen that were either congruent or incongruent with the target, and (iii) sensory primes. Auditory stimuli were presented through headphones, whereas olfactory stimuli were delivered using 5 ml glass bottles sealed with plastic screw caps, each containing a piece of cotton moistened with the corresponding aromatic oil. Tobacco was presented in the form of leaves placed inside the bottle. The auditory neutral control prime consisted of low-level pink noise, whereas the olfactory control prime consisted of odourless distilled water. Lexical primes consisted of the following six Greek terms, each selected for its conceptual affinity with one of the six semantic evaluation scales: Λιακάδα (Sunshine), Αιχμή (Point/Spike), Μάζα (Mass/Bulk), Θαλωρή (Coziness/Glow), Νέκταρ (Nectar), and Αύρα (Breeze/Aura). Table 1 presents the combinations of congruent, incongruent and control primes for each target.

Each participant completed successive sessions that included the five aromatic and five timbral targets, under lexical and sensory priming categories, representing the three experimental conditions (congruent, incongruent, and control). The lexical priming and olfactory stimuli presentation lasted 7 seconds, while the auditory stimuli ranged from 6 to 7 seconds. Each session comprised one-fifth of all possible trials and was separated from the previous session by at least 24 hours to minimise memory and carryover effects. Trial order was counterbalanced both across and within participants using an algorithm that dynamically generated trials for each session while maintaining balanced target–prime frequencies and preventing trial repetition within participants. Most participants completed three sessions, yielding a total of 100 sessions and 1000 observations per target–prime pair. The first session lasted approximately 12–14 minutes, including instructions, for most participants, whereas subsequent sessions were generally shorter. The study was approved by the Research Ethics Committee of the Aristotle University of Thessaloniki. Participants received course credit and snacks as compensation for their time.

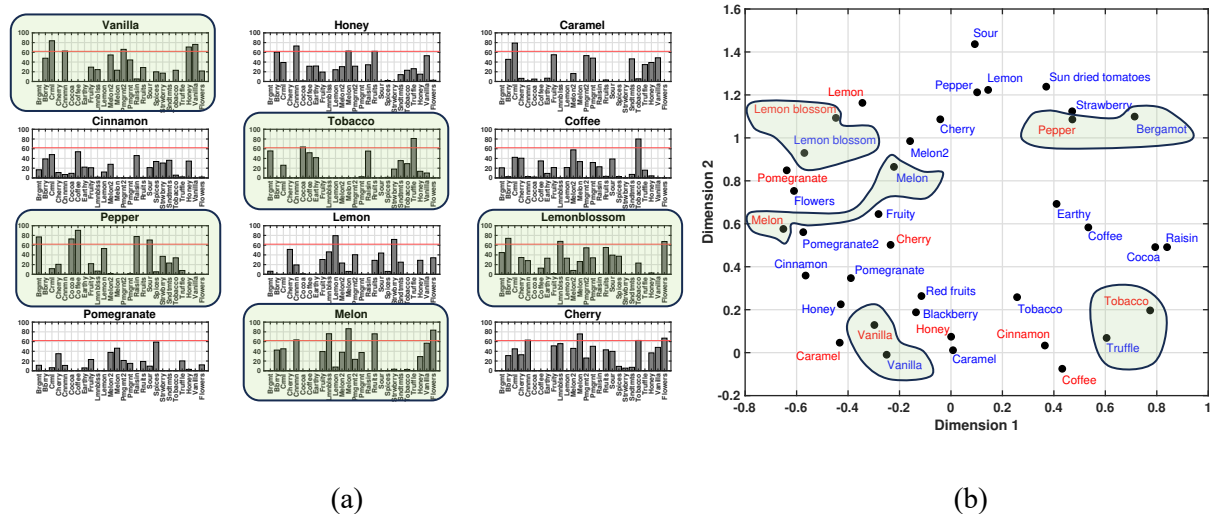


Figure 1: (a) Direct correspondences between the 12 aromas and the 26 timbres from Zacharakis (2024). The five aromas selected for the present study are highlighted. Red lines indicate the threshold above which correspondences become statistically significant based on bootstrapping. (b) Shared semantic space for aromas (in red) and timbre (in blue) from Zacharakis (2025). Highlighted regions indicate the five selected congruent aroma–timbre pairs. Synthetic timbres are labeled according to their intended aromatic targets. Aroma selection aimed to provide representative coverage of the

semantic space, whereas the selection of congruent and incongruent timbres was informed by both their relative proximity within the semantic space (b) and the strength of direct aroma–timbre correspondences (a) (see Table 1 for the congruent-incongruent pairs).

Table 1: Olfactory and auditory targets (1st column) with their corresponding prime pairs for each category and congruency condition.

Target		Condition				
		Control	Congruent		Incongruent	
		Olfactory	Olfactory	Lexical	Olfactory	Lexical
Auditory	Lemnbloss.wav	Distilled water	Lemon blossom	Λιακάδα (Sunshine)	Tobacco	Μάζα (Mass)
	Melon.wav	Distilled water	Melon	Λιακάδα (Sunshine)	Tobacco	Μάζα (Mass)
	Vanilla.wav	Distilled water	Vanilla	Νέκταρ (Nectar)	Black Pepper	Αιχμή (Spike)
	Bergamot.wav	Distilled water	Black Pepper	Αιχμή (Spike)	Melon	Θαλπωρή (Coziness)
	Truffle.wav	Distilled water	Tobacco	Μάζα (Mass)	Lemon blossom	Αιχμή (Spike)
Olfactory		Auditory	Auditory	Lexical	Auditory	Lexical
	Lemon blossom	Pink noise	Lemnbloss.wav	Αύρα (Breeze)	Truffle.wav	Θαλπωρή (Coziness)
	Melon	Pink noise	Melon.wav	Λιακάδα (Sunshine)	Truffle.wav	Μάζα (Mass)
	Vanilla	Pink noise	Vanilla.wav	Νέκταρ (Nectar)	Bergamot.wav	Νέκταρ (Nectar)
	Black Pepper	Pink noise	Bergamot.wav	Αιχμή (Spike)	Vanilla.wav	Θαλπωρή (Coziness)
	Tobacco	Pink noise	Truffle.wav	Μάζα (Mass)	Lemnbloss.wav	Αύρα (Breeze)

Results

Semantic ratings on the six semantic scales (bright, sharp, full, warm, sweet, fresh; range: −50 to +50, where 0 = neutral) were analysed using linear mixed-effects models (LMMs) fitted separately per target stimulus in R (R Core Team, 2026) using lme4 (Bates et al., 2015). Each model included Congruency (Control, Congruent, Incongruent) and prime type (sensory: olfactory or auditory; lexical) as fixed effects and participant as a random intercept. Three planned contrasts were evaluated per model using emmeans: Congruent vs Control, Incongruent vs Control, and Congruent vs Incongruent, with false discovery rate correction applied within each target model. Figure 2 summarises congruency effects per target, semantic scale, and prime type, shown with and without FDR correction.

Auditory targets (timbres primed by olfactory or lexical stimuli). Three out of five timbres showed significant differences between congruent and incongruent priming conditions. *Vanilla.wav* and *Melon.wav* were rated as sweeter in congruent compared to incongruent conditions, while *Melon.wav* was also rated as fresher and less sharp under olfactory priming (Figure 2a). For lexical priming, congruent trials led to higher warmth ratings for *Melon.wav* and lower warmth ratings for *Lemon blossom.wav* relative to incongruent trials.

Olfactory targets (scents primed by auditory or lexical stimuli). As shown in Figure 2b, lexical priming did not produce any observable effects on olfactory targets. In contrast, auditory primes significantly affected the perceptual profiles of three out of five olfactory targets. More specifically, congruent timbres increased the perceived fullness of *Tobacco* and the perceived brightness and sweetness of *Lemon Blossom*. Congruent auditory priming also reduced perceived sharpness for *Melon*.

Similar effects were observed when both congruent and incongruent conditions were compared to the control across both target types.

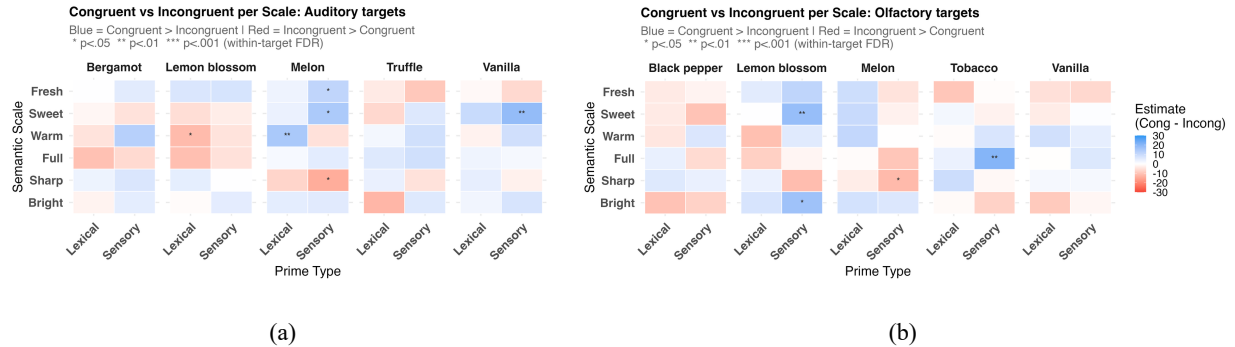


Figure 2: Significant effects following false discovery rate (FDR) correction per target stimulus (auditory left, olfactory right), across priming types and semantic scales.

Discussion

This abstract presents a fraction of the analyses possible on this rich dataset. The core finding is that both sensory and lexical primes have demonstrated potential to bias the semantic profiles of timbres and scents in theoretically coherent ways. Although significant effects may appear sparse, this is expected: congruency effects should not be assumed to operate uniformly across all six semantic dimensions, nor should prior assumptions about what constitutes congruent or incongruent priming for specific target–prime pairs be taken as given. The relevant dimensions, their sensitivity to priming, and the magnitude of expected effects remain empirical questions at this stage. Viewed in this light, the fact that six out of ten target stimuli showed semantic profile alterations as a result of priming constitutes meaningful complementary evidence of crossmodal semantic crosstalk between auditory and olfactory perception. Notably, for olfactory targets, auditory priming appeared more influential than lexical priming, suggesting that direct crossmodal sensory activation may be a more potent channel for olfactory semantic modulation than verbal mediation.

Although congruent priming generally enhanced expected semantic dimensions, occasional reversals suggest that priming effects may also arise through contextual contrast mechanisms. This possibility warrants further investigation.

Acknowledgements

The research project is implemented in the framework of H.F.R.I call “4th Call for H.F.R.I.’s Research Projects to Support Postdoctoral Researchers” (H.F.R.I. Project Number: 28975).

References

- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- R Core Team (2026). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Saitis, C., & Wallmark, Z. (2024). Timbral brightness perception investigated through multimodal interference. *Attention, Perception, & Psychophysics*, 86(6), 1835-1845.

- Wallmark, Z., Nghiem, L., & Marks, L. E. (2021). Does timbre modulate visual perception? Exploring crossmodal interactions. *Music Perception*, 39(1), 1-20.
- Zacharakis A. (2024). Sonic Bouquet: Decoding Cross-Modal Correspondences Between Timbre and Scent, *Music & Science*, 7, doi:10.1177/20592043241274258
- Zacharakis A. (2025). The poetry of senses: exploring semantic mediation in timbre-aroma correspondences, *Frontiers in Psychology*, 16, doi.org/10.3389/fpsyg.2025.1520046

Microsensorium: Timbral Crossmodal Worlds in Orchestral Music and AI Video

Chelsea Komschlies^{1†}



Microsensorium is a five-minute orchestral composition with projected AI video, commissioned by the Louisville Orchestra for its March 2026 children’s programming. Designed to engage young audiences with imaginative, multi-sensory worlds, the work explores how timbre can convey texture, temperature, and taste through sound. Composed as five micro-scenes—*Feral Furs & Teeth*, *Cold*, *Warm*, *Rusty Clockwork*, and *Bubblegum Unicorn Rainbows*—interspersed with dronelike interludes, the piece uses crossmodal correspondences, onomatopoeic gestures, and cultural associations to transform experimental orchestral textures into vivid sensory experiences.

The accompanying video, created in Midjourney, shows a picture frame on a gallery wall that opens into fantastical miniature worlds, visually reflecting the sensory elements of each scene.



In *Feral Furs & Teeth*, coarse, dark, bristling textures are evoked with low, biting brass tones, trombone singing with growl effects, flutter tongue, and short percussive accents, suggesting sharp, brittle surfaces and musty, animalistic textures.

Cold relies on high inharmonic metals, delicate string harmonies, sparkling glockenspiel, and upper-range harp, paired with minor-tinged modes, to conjure icy, crystalline sensations.





Warm contrasts by sinking into slow, rich, harmonic-influenced chords in strings and brass, combined with plush, diffuse flutter-tongue effects to evoke soft, plushy, enveloping warmth and peachy, comforting colors.

In *Rusty Clockwork*, harmonies across multiple instruments mimic the inharmonic spectra of metal objects and steam whistles, generating chimera-like timbres that suggest mechanical movement, junkyard textures, and imaginary machinery.



Bubblegum Unicorn Rainbows blends bright, high-pitched timbres, pure low-centroid tones, and harmonic spectral blends to evoke the taste, scent, and playful energy of candy and citrus, while drawing on the cultural associations of 1990s synthwave soundscapes.

The presentation will examine the timbral strategies in each micro-scene, highlighting

how spectral shaping, orchestration, and playing techniques contribute to crossmodal perception. Perceptual studies linking timbre, texture, temperature, and color will be discussed, illustrating how these principles make experimental sound design accessible and engaging for children.



The session concludes with a full audiovisual presentation of *Microsensorium*, showcasing the integration of orchestral timbre and AI-generated visual worlds. The work demonstrates how timbre functions not only as foundational musical material but also as a bridge between sound, vision, and multi-sensory perception. By exploring crossmodal correspondences in orchestration, *Microsensorium* offers a model for music that is simultaneously experimental, expressive, and appealing to young audiences, suggesting new possibilities for research and practice at the intersection of timbre, cognition, and multimedia art.

door opens

Play 4 times

A Title

hands

Flute 1

Flute 2

Oboe

English Horn

Clarinet in B♭ 1

Clarinet in B♭ 2

Bassoon 1

Bassoon 2

Horn in F 1

Horn in F 2

Trumpet in C 1

Trumpet in C 2

Trombone 1

Trombone 2

Euphonium

Tuba

Vincent

Vibraphone

Musical Tree

Harmonica

Violin I

Violin II

Viola

Violoncello

Double Bass

f

mf

p

pp

ff

random chatter effect

(F# or G-sharp)

Duo

All Speak any phrase or random syllables.

Speak any phrase or random syllables.

MSIP → ord MSIP → ord

MSIP → ord MSIP → ord

1 2 3 (4) (5) (6) 7 8 9 10 11

References

- Adeli, M., Rouat, J., & Molotchnikoff, S. (2014). Audiovisual correspondence between musical timbre and visual shapes. *Frontiers in Human Neuroscience*, 8, 352. <https://doi.org/10.3389/fnhum.2014.00352>
- Crisinel, A.-S., & Spence, C. (2010). As bitter as a trombone: Synesthetic correspondences in nonsynesthetes between tastes/flavors and musical notes. *Attention, Perception, & Psychophysics*, 72(7), 1994–2002. <https://doi.org/10.3758/APP.72.7.1994>
- Hasegawa, R. (2019). Timbre as harmony—harmony as timbre. In *The Oxford handbook of timbre*. Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780190637224.001.0001>
- McAdams, S. (2013). Musical timbre perception. In *The Oxford handbook of music psychology*. Oxford University Press.
- Palmer, S. E., Schloss, K. B., Xu, Z., & Prado-León, L. R. (2013). Music–color associations are mediated by emotion. *Proceedings of the National Academy of Sciences of the United States of America*, 110(22), 8836–8841. <https://doi.org/10.1073/pnas.1212562110>
- Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics*, 73(4), 971–995. <https://doi.org/10.3758/s13414-010-0073-7>
- Traube, C. (2013). De la thermoception à la perception auditive : en quête de l'identité du son « froid ». *Les Cahiers de la Société québécoise de recherche en musique*, 14(1), 9–15. <https://doi.org/10.7202/1016193ar>
- Wang, Q., Keller, S., & Spence, C. (2017). Sounds spicy: Enhancing the evaluation of piquancy by means of a customized crossmodally congruent soundtrack. *Food Quality and Preference*, 58, 1–9. <https://doi.org/10.1016/j.foodqual.2016.12.014>
- Wang, Q. J., & Spence, C. (2017). The role of pitch and tempo in sound–temperature crossmodal correspondences. *Multisensory Research*, 30(3–5), 307–320. <https://doi.org/10.1163/22134808-00002564>
- Ward, J., Huckstep, B., & Tsakanikos, E. (2006). Sound–colour synaesthesia: To what extent does it use cross-modal mechanisms common to us all? *Cortex*, 42(2), 264–280. [https://doi.org/10.1016/S0010-9452\(08\)70352-6](https://doi.org/10.1016/S0010-9452(08)70352-6)
- Photos of the premiere courtesy of Oneil Arnold and the Louisville Orchestra.

Poster Session: Day 1

The Contours of Timbre in Rebecca Saunders's *Fury II*

Ash Mach^{1†}

¹ Department of Music Theory, Eastman School of Music, Rochester, New York, U.S.A.

[†] Corresponding author: amach2@u.rochester.edu

Introduction

Rebecca Saunders (1967-) is a Berlin-based British composer whose music is self-described as having a “distinctive and intensely striking sonic language.” Timbre, and by extension orchestration, is the main parameter from which Saunders works. Each instrument is carefully chosen and given a timbral palette, which when combined with other instruments in the ensemble creates a single palette of sounds (Saunders 2016). Analyses of her music therefore necessitate methods that place timbre and orchestration at the forefront. Previous analyses of her works draw upon cross-disciplinary semantic connections to describe timbral characteristics in an accessible manner (McMullan-Glossop 2017), theorize perceptual groupings based on instrumental technique (Abram 2021), or develop piece-specific sound descriptors based on orchestral grouping effects (Hasegawa 2026). However, questions remain regarding the analysis of timbre and orchestration as *dynamic*, rather than static entities. Using Saunders's double bass concerto *Fury II* as a case study, this paper introduces a methodology that puts orchestration in conversation with acoustic correlates of timbre, allowing us to hear the relationships between differing sound components.

Methods

This paper adopts tools from *contour theory*, namely contour segments, reduction algorithms, and matrices, as the common language across musical dimensions (Morris 1993, Marvin and Laprade 1987). The use of contour analysis to compare timbre and orchestration aids in determining the effect of changing instrumental combinations on global trends in timbre. Orchestration is analyzed through piece-specific descriptors and onset density graphs. The paper proposes nine piece-specific sound categories based on type of attack, presence of pitch, or break in sound, expanding upon Hasegawa's (2026) four sound descriptors for the same piece. The paper then utilizes these descriptors to create onset density graphs (Goodchild 2018) mapping “pitch saturation” and “length of attack” based on the number of onsets of each sound type. These graphs demonstrate the timbral prominence of each descriptor in a given section. Contour segment vectors (CSEGs) and comparison matrices (Morris 1993) are then used to compare contour, and therefore timbral prominence, of each sound type. Two spectral descriptors, spectral crest and spectral centroid, build upon the orchestration analysis to describe changes in timbre over time. These can then be mapped onto the onset density graphs to show how the nine sound categories contribute to global trends in noisiness and brightness.

Results and Discussion

Piece-Specific Descriptors

Hasegawa's analysis of *Fury II* combines Touizrar and McAdams's orchestral grouping effects with Thoresen's symbol system to describe the segmentation and association of sound objects. He argues that *Fury II* can be understood through four sound categories: clearly pitched sounds, low and mostly unpitched sounds, high inharmonic sustain or resonance, and percussive sounds. This approach encourages the analyst to move away from the score and towards the perception of instrumental combinations. Through close listening of the emerging timbres from the instruments' interactions, I propose nine categories that can be sorted according to pitch or length of attack.

Table 1: Pitch and length-of-attack descriptors

Pitch Saturation	Length of Attack	Silence
Initiated pitch	Sustained	Break in sound
Resultant pitch	Repeated percussive attacks	
Low, rumbling	Sharp attack	
Unpitched noise		

“Pitch saturation” refers to the level of pitch present and how they are created, distinguishing between pitches that Saunders intentionally wrote versus pitches that emerge from an extended technique. “Type of attack” refers to the presence of an attack point as well as the number of hits. “Silence” is also added as a sound category, for her music always “emerges from silence” (Saunders 2016). These categories are not meant to be clear-cut delineations, but rather fluid guidelines based upon aural analysis.

Onset Density

In measures 1-7, the “low, rumbling” and “sharp percussive attacks” are introduced. These descriptors arise from a complex set of emergent phenomena and vary in length, density and complexity. The details of an orchestration analysis can get lost among the shuffle, making the overall trajectory difficult to envision. To mitigate this, I create onset density graphs (Goodchild 2018) to analyze pitch saturation and length of attack. The x-axes represent time in beats per measure, whereas the y-axes represent the number of onsets of a sound per beat. In this paper, “onset density” is defined as *the number of onsets of each sound type as a measure of timbral prominence*. The purpose of these graphs is to quantify the most prominent timbres in each section.

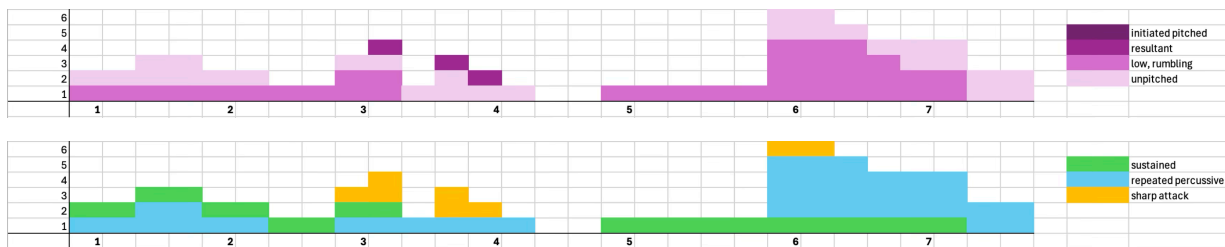


Figure 1: pitch and length of attack onset density graphs, measures 1-7

The low, rumbling timbre, paired with unpitched noise, is the most prominent timbre in the opening seven measures. The rumbling timbre is accomplished through repeated percussive attacks, whereas the unpitched noise through sustained gestures. The pitch and attack length graphs do not create one-to-one mappings; repeated percussive attacks account for most of the articulation between mm. 6-7 (25 onsets within 8 beats) but are heard through both low and rumbling and unpitched sounds. Unpitched sounds also take the form of both sharp and repeated percussive attacks within these same two measures.

CSEG analysis of the individual components of each graph reveal which sonic aspects inform or conflict with each other. Using Morris’s (1993) contour reduction algorithm, each CSEG was normalized to 8

integers. Subsections of each CSEG, (csubsegs) can also be analyzed for equivalence or similarity. “Unpitched” and “repeated percussive attacks” share a csubseg of <120>. This contour occurs during the same stretch of time, suggesting that unpitched timbres are directly influenced by the repeated percussive attacks. The “low, rumbling” sound shares the same initial contour, <120>, but at double the length. This suggests that the “low, rumbling” timbre is an augmentation of the “unpitched” and “repeated percussive” timbres. The second half of “low, rumbling” and “repeated percussive attacks” share similar csubsegs (<4320> and <0432>, respectively), suggesting another manipulation of time.

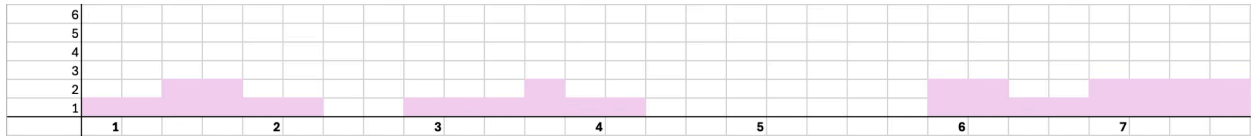


Figure 2: sustained unpitched noise onset density graph; CSEG <12020212>

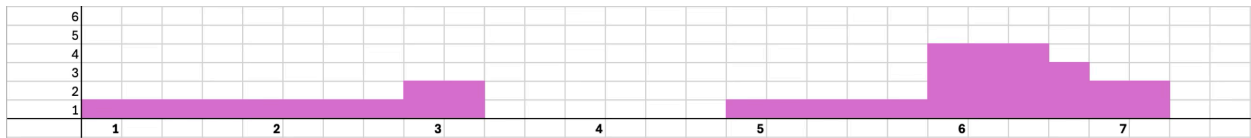


Figure 3: sustained and low, rumbling onset density graph; CSEG <1201430>

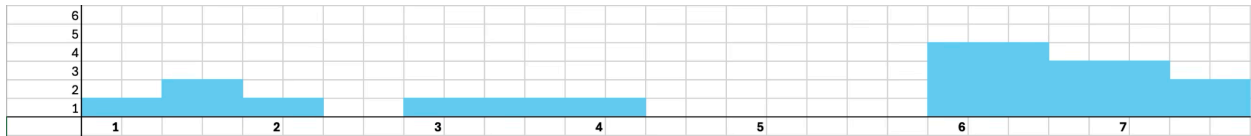


Figure 4: semi-pitched repeated percussive attack onset density graph; CSEG <12010432>

Spectral Descriptors for Analysis

In measures 30–40, the “sustained, initiated pitch” descriptor takes prominence. To capture the variation within this descriptor, I extract the spectral crest and spectral centroid from a recording of *Fury II* by Ensemble Linea. The data was extracted using the Timbre Toolbox in the MATLAB environment and was smoothed through a moving average with a window of 175 data points. The local minima and maxima were then found using the “argelextrema” function in the scipy.signal library. Increases in spectral crest correspond to where initiated pitch is the primary timbre, whereas decreases correspond to the onset of resultant and/or unpitched sounds. Spectral centroid is dependent on the initiated and emergent pitch in the solo bass. Based on a cursory overview of the graphs, spectral crest and centroid seem to move in opposition to each other. While the double bass *sul ponticello* glissando causes the lowest dip in spectral crest, it simultaneously causes the first major peak in spectral centroid. Likewise, the global maxima in spectral crest (0:34) occurs at the peak of the bass clarinet and accordion’s swell before m. 36, yet only causes a slight rise in spectral centroid. The contour relationships of spectral crest and spectral centroid, normalized to 13 data points, are shown in Table 2. Based on the opposing contours, it can be argued that Saunders creates the perception of a brighter timbre through the addition of high frequencies and unpitched noise.

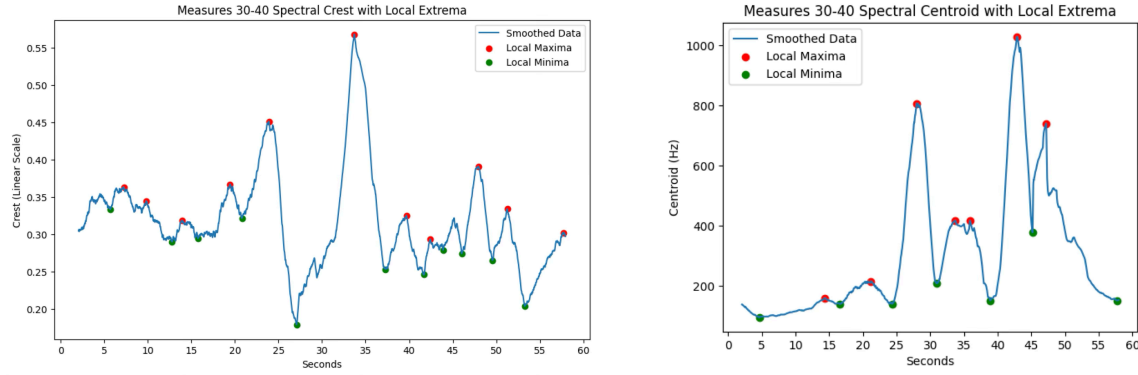


Figure 5: spectral crest and spectral centroid graphs with plotted local extrema, mm. 30-40

Table 2: INT_1 contour relationships for spectral crest and centroid, mm. 30-40

Spectral Crest	-	+	-	+	-	+	-	+	-	+	-	+
Spectral Centroid	-	-	+	-	+	-	+	-	+	-	+	-

Conclusion

Rebecca Saunders’s music is known for its extreme contrasts, inventive extended techniques, and evocative timbres. This paper offers a new perspective on listening to and analyzing *Fury II* by using the shared vocabulary of contour analysis to show how different timbral parameters inform or conflict with each other over time, allowing the analyst to formulate dynamic groupings that inform listening perception. Questions remain regarding the scale of this analytical method and how analyses may differ per recordings.

References

- Caetano, M., Saitis, C., and Siedenburg, K. (2019). Audio content descriptors in timbre. In K. Siedenburg C. Saitis, S. McAdams, A.N. Popper, and R.R. Fray (Eds.), *Timbre: Acoustics, perception, and cognition* (pp. 297-334). Cham, Switzerland: Springer Nature Switzerland.
- Goodchild, M. (2016). *Orchestral gestures: music-theoretical perspectives and emotional responses* [Doctoral dissertation, McGill University].
- Hasegawa, R. (2026). Timbre and orchestration in two works by Rebecca Saunders for soloist and ensemble. In J. Delisle, R. Hasegawa, J. Noble, and M. Touizrar (Eds.), *The Oxford handbook of orchestration studies* (pp. 574-598). New York, USA: Oxford University Press.
- Marvin, E.W., and Laprade, P. (1987). Relating musical contours: Extensions of a theory for contour. *Journal of Music Theory* 31(2), 225-267.
- Morris, R.D. (1993). New directions in the theory and analysis of musical contour. *Music Theory Spectrum* 15(2), 205-228.
- Saunders, R. (2016). *Analysis: Fury II by Rebecca Saunders – English* [video]. IRCAM. <https://medias.ircam.fr/fr/xa3ce1d#information>.

Coloring Shapes and Forms: Timbral and Orchestration Strategies in the Serial Works of Elisabeth Lutyens

Aidan McGartland^{1†}

¹ Schulich School of Music, McGill University, Montreal, Quebec, Canada

[†] Corresponding author: aidan.mcgartland@mail.mcgill.ca

Elisabeth Lutyens was the first British composer to employ serial techniques with her *Chamber Concerto* (1939). Yet in her own writings she consistently downplays serialism, instead emphasizing musical “shapes and forms” and the resulting “character and gesture” (Lutyens, 1972, pp. 267–268). While this position has been linked to the unpopularity of serialism in mid-century Britain, it also reflects the growing role of timbre within her compositional thinking. It is important to note that she does not serialize timbre, only pitch; this leads to an interesting relationship between serialism and timbre in many of her works, with timbre sometimes even superseding serialism as a structural parameter. Despite a growing body of recent scholarship on Lutyens’ music (Parsons, 2016; Cole, 2023; Forkert, 2023), the aspect of timbre remains mostly unexplored. Specifically, in this study the concept of timbre is conceived as the abstract musical parameter that is the effect of a more concrete technique, orchestration.

Through sketch and score analysis, this paper identifies four techniques through which timbre shapes Lutyens’ music: (1) orchestrational blocks as formal demarcations, (2) a segmented textural approach resembling *Klangfarbenmelodie*, (3) the manipulation of register, and (4) the timbral evolution (of usually a single line). The first technique, which I term orchestrational blocks, involves the use of distinct, non-overlapping instrumental groups in succession to articulate form through timbral contrast. In *And Suddenly It’s Evening* (1966), the ensemble is divided into three contrasting groups—in a percussive *ritornello* 1, a strings-dominated *ritornello* 2, and a brassy *coro di strumento* (Table 1). These groups alternate symmetrically around the poems of each movement, and seldom overlap, so that the overall form of the work emerges through shifts in timbre. This timbral-demarcated form is derived from the Baroque concept of “ritornello,” noting the use of the term, where a solo part is interspersed with interludes from a larger ensemble. The second technique occurs more locally. Lutyens employs techniques related to *Klangfarbenmelodie*, where linear pitch material is distributed in small fragments across the ensemble with changing instrumental colors (see Zeller 2022). At the opening of *The Valley of Hatsu-Se* (1965), the row is segmented into short fragments that circulate rapidly among instruments, producing a highly disjunct texture that introduces the ensemble through timbral rotation (Figure 1). This opening example has an exposition-like function as it introduces all the instruments – and their main timbres – that foreshadows the rest of the work. The third technique is the manipulation of register. Lutyens often fixes material within a specific registral spacing and later shifts or flips it, preserving the pitch-classes while changing the timbre. In the row chart for *Présages* (1963), a major sixth appears in place of a registrally conjunct minor third (Figure 2). This major sixth is then preserved throughout much of the opening “Recitative,” implying that this was motivated by the distinct oboe registers (cf. the nasal, loud lower register and the shrill, soft higher register) and this suggests a direct link between pitch and timbre from the outset (Figure 3). The final technique, timbral evolution, progressively transforms the sound (often a single pitch or solo line) through articulation, dynamics, and extended techniques. In *Echo of the Wind* (1981) for solo viola, gradual changes between *sul ponticello*, *flautando*, and varying degrees of tremolo create a continuous transformation of timbre that shapes the work’s expressive trajectory (Figure 4) and is somewhat disconnected from the serial structure.

While timbre remains an understudied aspect of Lutyens’s oeuvre, it emerges as a central compositional parameter that increases in importance throughout her career. Recognizing these timbral strategies enables a richer understanding of her music and reframes the emphasis on “shapes and forms” that

complements serial readings of Lutyens' works. Beyond Lutyens' music, this study has implications for (1) later composers taught or influenced by Lutyens, including Richard Rodney Bennett and Harrison Birtwistle; (2) the structural role of timbre in other works defined by pitch serialism; and (3) the role of timbre in post-tonal music more generally.

Figure 1: *The Valley of Hatsu-Se* (1965) opening (mm. 1–6) featuring the parsing of rows through the ensemble in the manner of *Klangfarbenmelodie*.

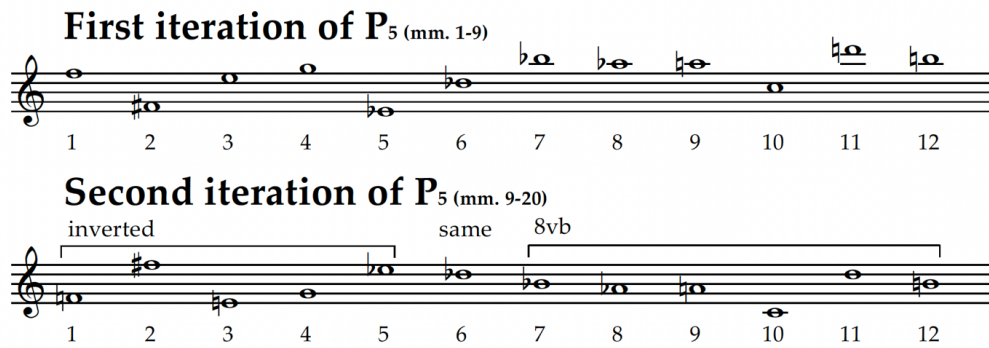


Figure 3: Fixed register in the opening movement, “Recitative,” of *Présages*.

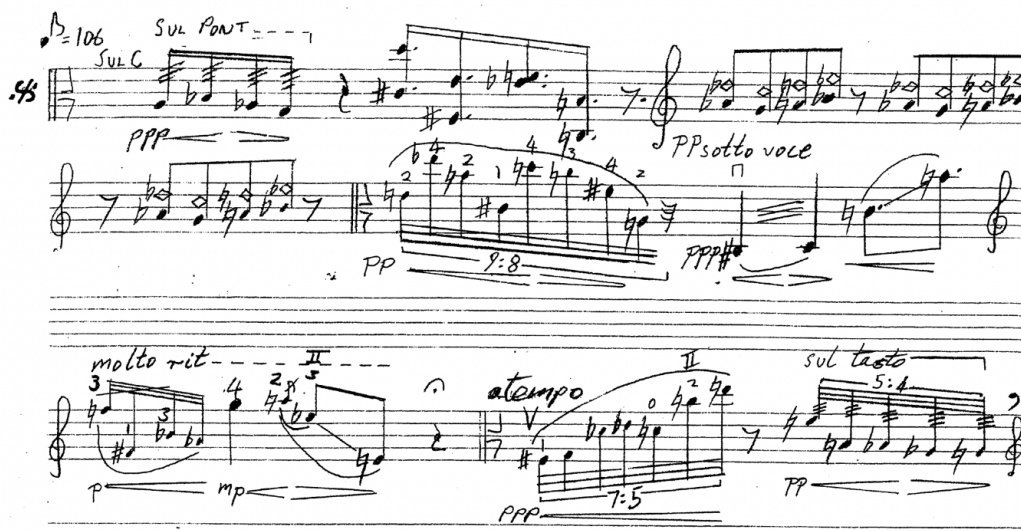


Figure 4: *Echo of the Wind* (1981), opening, published holograph (Lutyens, 1999, p. 1)

Table 1: Instruments groups in *And Suddenly It's Evening* (1966) with timbral descriptors based on Reymore (2022).

Instrument group (as labelled by the composer)	Instruments within each group	Main instrument family of each group	Overall timbre
Tenore	Voice (tenor)	Voice	Breathy, nasal, airy
Coro di Strumenti	2 Trumpets, 2 Trombones, Double-Bass	Brass	Warm, rich, brassy, thick, “choral,” blended, bright
Ritornello 1	Harp, Celesta, Percussion (lots)	Percussion	Percussive, struck, lack of sustain, bell-like
Ritornello 2	Violin, Horn, Cello	Strings	Resonant, warm, rich

References

- Adorno, T. W. (2018). The function of timbre in music (1966). In E. Dolan & A. Rehding (Eds.), *The Oxford handbook of timbre* (pp. 124–170). New York, USA: Oxford University Press.
- Alegant, B. (2010). *The twelve-tone music of Luigi Dallapiccola*. Rochester, USA: University of Rochester Press.
- Bradshaw, S. (1971). Music of Elisabeth Lutyens. *The Musical Times*, 112(1541), 653, 655–656.
- Cole, C. (2023). *The work of the performer: Enlivening performative effort through the music of Elisabeth Lutyens* (Doctoral dissertation, Indiana University).
- Forkert, A. (2023). *Elisabeth Lutyens and Edward Clark: The orchestration of progress in British twentieth-century music*. Cambridge, UK: Cambridge University Press.
- Harries, M., & Harries, S. (1989). *A pilgrim soul: The life and work of Elisabeth Lutyens*. London, UK: Michael Joseph.
- Hasegawa, R. (2018). Timbre as harmony—harmony as timbre. In E. Dolan & A. Rehding (Eds.), *The Oxford handbook of timbre* (pp. 525–551). New York, USA: Oxford University Press.
- Lewin, D. (2007). *Generalized musical intervals and transformations* (2nd ed.). New York, USA: Oxford University Press. (Original work published 1987)
- Lochhead, J. (2016). *Reconceiving structure in contemporary music: New tools in music theory and analysis*. Abingdon, UK: Routledge.
- Lutyens, E. (c. 1963). *Présages* row chart [Unpublished manuscript]. British Library, Lutyens Collection, Add MS 64789.
- Lutyens, E. (1999). *Echo of the wind* [Published holograph musical score]. York, UK: University of York Musical Press. (Original work composed 1981)
- Lutyens, E. (1972). *A goldfish bowl*. London, UK: Cassell.
- McAdams, S., Goodchild, M., & Soden, K. (2022). A taxonomy of orchestral grouping effects derived from principles of auditory perception. *Music Theory Online*, 28(3).
- McGartland, A. (forthcoming 2027). *Twelve-Tone Lizzie: An Analytical and Critical Study of the Serialism of Elisabeth Lutyens* (Doctoral dissertation, McGill University).
- Manning, J. (1986). *New vocal repertory: An introduction*. London, UK: Macmillan.
- Parsons, L. (2016). Elisabeth Lutyens, *Essence of Our Happinesses* (1968). In L. Parsons & B. Ravenscroft (Eds.), *Analytical essays on music by women composers: Concert music from 1960–2000* (pp. 196–219). New York, USA: Oxford University Press.
- Payne, A., & Calam, T. (2001). Lutyens, (Agnes) Elisabeth. *Grove Music Online*. Oxford Music Online.
- Reymore, L. (2022). Characterizing prototypical musical instrument timbres with timbre trait profiles. *Musicae Scientiae*, 26(3), 648–674.
- Schoenberg, A. (2010). *Theory of harmony* (R. E. Carter, Trans.). University of California Press.
- Straus, J. N. (2009). *Twelve-tone music in America*. Cambridge, UK: Cambridge University Press.
- Straus, J. N. (2022). *The art of post-tonal analysis: Thirty-three graphic music analyses*. New York, USA: Oxford University Press.
- Zeller, M. Klangfarbenmelodie in 1911: Timbre’s functional roles in Webern’s opp. 9 and 10. *Music Theory Online*, 28(1).

Hearing the Form, Seeing the Material: Global and Local Feature Correspondences Across Vision and Audition

Asterios Zacharakis^{1†} and Amit Gur²

¹ School of Music Studies, Aristotle University of Thessaloniki, Greece

² Department of Philosophy, University of Antwerp, Belgium

† Corresponding author: aszachar@mus.auth.gr

Introduction

Crossmodal correspondences between hearing and sight have been widely studied, linking fundamental auditory attributes such as pitch and loudness, as well as higher-level musical features such as tempo and mode, to a range of visual properties including elevation, size, angularity, colour hue, saturation, and brightness (see Liu & Lupyan, 2026, for a recent review). When it comes to timbre, however, reported visual associations are comparatively scarce and less well-defined. One of the most frequently cited attempts is that of Adeli et al. (2014), who reported that softer timbres tend to correspond to rounded shapes and cooler colours (e.g., blue or green), whereas harsher timbres are more often associated with angular shapes and warmer colours (e.g., red or yellow). In that work, however, timbre was defined primarily in terms of musical instrument source. A subsequent body of work by Wallmark and colleagues (Wallmark et al., 2018; Wallmark, 2019; Wallmark & Allen, 2020; Wallmark et al., 2021; Saitis & Wallmark, 2024) supports the existence of crosstalk between timbral characteristics and visual attributes. Reymore & Lindsey (2025) showed that the associations of colours with musical timbres covary with crossmodal semantic ratings, suggesting that timbre–colour mappings are partly mediated by shared semantic structure.

This study builds on recent work by Zacharakis et al. (2025) on crossmodal correspondences between timbre and visual contour. Their findings demonstrated systematic associations of round, rough, and sharp timbres (as determined through expert-panel tagging and acoustic modelling) with a circle whose outline varied from smooth to increasingly jagged and ultimately spiky. In this experiment, the timbral associations were made with the outline rather than the form, as the underlying shape (a circle) was held constant across stimuli. This suggests that crossmodal correspondences may operate at both shape and outline levels, which may be alternatively framed in terms of the global form of an object and its local material properties.

Research on global/local features in audition is relatively limited and typically addresses the hierarchy between global melodic shape (often called melodic contour) and the local pitch intervals that constitute it (Susini et al., 2020). In vision, the distinction is often cast more broadly, separating a global form from the local elements, or material, that comprise it (Kimchi, 1982, 1988, 1992). Gur (2024) has recently suggested a similar distinction for audition, treating a melodic phrase as the global form and timbre as the local, material-like feature that constitutes it, drawing an analogy to visual global/local hierarchical structures.

Through this study, we sought to explore whether the global (form) and local (material) dimensions of visual and auditory objects interact to give rise to systematic crossmodal correspondences.

Method

Aiming to examine potential audio-visual associations when local and global structures are jointly varied, we designed a crossmodal matching experiment including four auditory and four visual stimuli (Fig. 1). In this design, global structure refers to geometrical shape in vision and melodic contour in audition,

while local structure refers to shape outline (potentially conveying object material) and timbre (conveying sonic material). The visual stimuli consist of Kiki- and Bouba-style forms (Ramachandran & Hubbard, 2003) with varied local features (smooth versus jagged line contours), yielding a 2×2 combination of shape and outline. Similarly, the auditory stimuli consist of melodic phrases that are either ‘smooth’ or ‘spiky’. Smooth describes a gentle, flowing melodic contour dominated by stepwise motions, whereas spiky describes a jagged melodic contour dominated by jumps and abrupt changes. Each melodic phrase is played back in either a soft clarinet timbre or a rough timbre produced by distorting the clarinet sound, resulting in a parallel 2×2 combination of melodic shape and timbral quality.

A two-alternative forced-choice experiment was conducted online, in which participants matched one of two auditory stimuli to a visual reference and vice versa. Forty participants (18 female, mean age: 30.8 years, age range: 18-56 years) completed the task with a visual reference, and thirty-nine (16 female, mean age: 32.1 years, age range: 19-60 years) with an auditory reference. In both conditions, the total number of trials was 24 (4 references × 6 pairs), and the test lasted 6-7 minutes on average.

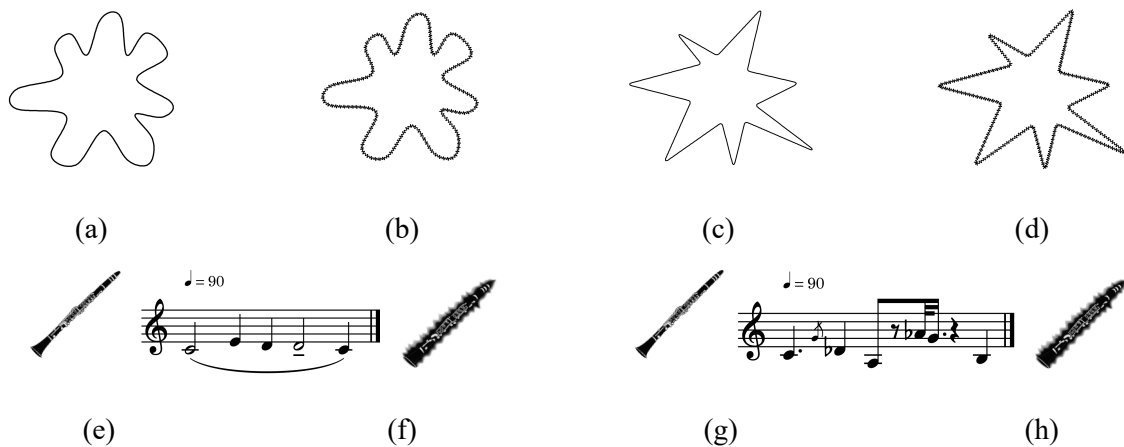


Figure 1: The four visual (top) and the four auditory stimuli (bottom). (a): Bouba shape – smooth outline, (b) Bouba shape – jagged outline, (c): Kiki shape – smooth outline, (d): Kiki shape – jagged outline, (e): ‘smooth’ melody – clarinet timbre, (f): ‘smooth’ melody – distorted clarinet timbre, (g): ‘spiky’ melody – clarinet timbre and (h): ‘spiky’ melody – distorted clarinet timbre.

Results

Two-tailed exact binomial tests were conducted against a chance baseline of .50 to assess whether participants systematically selected one of the options provided for each trial type. The results are presented separately for internally congruent (see Fig 1. (a), (d), (e) and (h)) and incongruent (see Fig 1. (b), (c), (f) and (g)) reference conditions across both experiments. An internally congruent reference is defined as one whereby both form and material point towards spikiness-roughness or roundness-smoothness, whereas an internally incongruent reference is characterised by conflict between these two hypothetical dimensions. False discovery rate correction (Benjamini-Hochberg) was applied within each experiment × congruency block to account for multiple comparisons. Table 1 presents the results of this general analysis.

The most notable observation is that, irrespective of reference modality, when the reference stimulus is internally congruent (e.g., Kiki form with jagged outline), matching exceeds chance for most trial types. The opposite is true for internally incongruent references, for which no trial type reaches significance. Among congruent reference trials, participants reliably selected the option aligning with the reference on both dimensions over one matching on form alone or neither dimension. Crucially, however, alignment

on material only versus both dimensions yielded chance-level responses, pointing to material as the primary driver of crossmodal matching. This is further corroborated by the material vs neither and form vs material trial types, where in both cases it is material congruency that steers selection. Form vs neither, however, also yielded above-chance selection – albeit weaker than material vs neither – indicating that form retains a distinct influence on crossmodal matching.

Table 1. Proportion of choices by pair type, reference modality and congruency condition. Proportions refer to the first stimulus within a given pair as presented in column 1; those in bold indicate above-chance matching. Visual Reference: $n = 80$ per cell (pooled across both references, i.e., congruent/incongruent); Auditory Reference: $n = 78$ per cell. All tests were two-tailed binomial vs. chance ($p = .50$); significance markers reflect FDR correction (Benjamini-Hochberg) applied within each experiment $\times$ congruency block.

Trial type in terms of crossmodal congruency	Visual Reference				Auditory Reference			
	Intern. Congruent		Intern. Incongruent		Intern. Congruent		Intern. Incongruent	
	Prop. %	Sig.	Prop. %	Sig.	Prop. %	Sig.	Prop. %	Sig.
BOTH vs NEITHER	75	***	57.5	n.s.	73.1	***	52.6	n.s.
BOTH vs FORM	71.3	***	53.7	n.s.	74.4	***	60.3	n.s.
BOTH vs MATERIAL	58.8	n.s.	60	n.s.	60.3	n.s.	38.5	n.s.
FORM vs NEITHER	65	*	50	n.s.	65.4	*	42.3	n.s.
MATERIAL vs NEITHER	76.2	***	48.7	n.s.	80.8	***	52.6	n.s.
FORM vs MATERIAL	25	***	47.5	n.s.	46.2	n.s.	35.9	n.s.

* $p < .05$; ** $p < .01$; *** $p < .001$.

Discussion

This study investigated whether global (visual shape; melodic contour) and local (object material; timbre) features independently contribute to auditory–visual correspondences. This initial analysis supports the hypothesis that form and material, as operationalised here, function as independent cues driving such associations. Whereas internally congruent references — where shape and material align — seem to constitute a clear cue for matching stimuli in the other modality, references in which these two dimensions conflict fail to produce above-chance associations. This latter finding suggests the existence of dimensional interference that gives rise to perceptual incoherence in the reference stimulus, thereby preventing systematic matching.

It is possible that form and material jointly contribute to a latent Kikiness–Boubaness dimension, such that stimuli combining conflicting cues — for example, a Bouba shape with jagged outline contour, or a ‘spiky’ melody performed on a clarinet with a smooth timbre — are less effective as crossmodal anchors. Thus, the absence of significant associations in the internally incongruent conditions should not necessarily be interpreted as evidence that shape and material lack distinct perceptual salience. Rather, it may reflect the fact that both dimensions integrate into a unified crossmodal percept, yielding stimuli that occupy an ambiguous intermediate position along the proposed Bouba–Kiki continuum. Future work will employ a 2 x 2 forced choice design with a focus on the four incongruent stimuli, similar to the classic Bouba/Kiki paradigm. Although such paradigms have been criticised for potentially exaggerating correspondence effects (e.g., Nielsen & Rendall, 2001; Chen et al., 2016), they may nevertheless draw participants’ attention to finer details in the stimuli and could potentially result in the dissociation of the form and material dimensions even in the incongruent reference conditions.

Finally, a particularly noteworthy finding from this preliminary analysis is that material appears to be weighted more strongly than form in the construction of the latent Bouba–Kiki continuum. This

asymmetry may reflect the salient and rapidly accessible nature of timbral information in audition, which may bias crossmodal integration toward material-related cues.

References

- Adeli, M., Rouat, J., & Molotchnikoff, S. (2014). Audiovisual correspondence between musical timbre and visual shapes. *Frontiers in Human Neuroscience*, 8, 352.
- Chen, Y. C., Huang, P. C., Woods, A., & Spence, C. (2016). When “Bouba” equals “Kiki”: Cultural commonalities and cultural differences in sound-shape correspondences. *Scientific reports*, 6(1), 26681.
- Gur, A. (2024). *Form and material in visual and auditory perception* (Doctoral dissertation, University of Antwerp).
- Liu, Q., & Lupyan, G. (2026). The unity of sense and mind: A review of cross-domain mapping. *Psychonomic Bulletin & Review*, 33(1), 25.
- Ramachandran, V. S., & Hubbard, E. M. (2003). Hearing colors, tasting shapes. *Scientific American*, 288, 52–59.
- Reymore, L., & Lindsey, D. T. (2025). Color and tone color: audiovisual crossmodal correspondences with musical instrument timbre. *Frontiers in Psychology*, 15, 1520131.
- Kimchi, R., & Palmer, S. E. (1982). Form and texture in hierarchically constructed patterns. *Journal of Experimental Psychology: Human Perception and Performance*, 8(4), 521.
- Kimchi, R. (1988). Selective attention to global and local levels in the comparison of hierarchical patterns. *Perception & Psychophysics*, 43(2), 189-198.
- Kimchi, R. (1992). Primacy of wholistic processing and global/local paradigm: a critical review. *Psychological Bulletin*, 112(1), 24.
- Nielsen, A., & Rendall, D. (2011). The sound of round: evaluating the sound-symbolic role of consonants in the classic Takete-Maluma phenomenon. *Canadian Journal of Experimental Psychology*, 65(2), 115-124.
- Saitis, C., & Wallmark, Z. (2024). Timbral brightness perception investigated through multimodal interference. *Attention, Perception, & Psychophysics*, 86(6), 1835-1845.
- Susini, P., Jiaouan, S. J., Brunet, E., Houix, O., & Ponsot, E. (2020). Auditory local–global temporal processing: Evidence for perceptual reorganization with musical expertise. *Scientific Reports*, 10(1), 16390.
- Wallmark, Z. (2019). Semantic crosstalk in timbre perception. *Music & Science*, 2.
- Wallmark, Z., & Allen, S. E. (2020). Preschoolers’ crossmodal mappings of timbre. *Attention, Perception, & Psychophysics*, 82, 2230-2236.
- Wallmark, Z., Iacoboni, M., Deblieck, C., & Kendall, R. A. (2018). Embodied listening and timbre: Perceptual, acoustical, and neural correlates. *Music Perception*, 35(3), 332-363.
- Wallmark, Z., Nghiem, L., & Marks, L. E. (2021). Does timbre modulate visual perception? Exploring crossmodal interactions. *Music Perception*, 39, 1-20.
- Zacharakis A., Velenis, K., & Cambouropoulos, E. (2025). Between Bouba and Kiki lies Grougi: Modelling the Rough Middle Ground of Sound–Shape Correspondences, In *Proceedings of the 17th International Symposium on Computer Music Multidisciplinary Research (CMMR2025)* (pp. 93-104), London, UK.
- Nielsen, A., & Rendall, D. (2011). The sound of round: evaluating the sound-symbolic role of consonants in the classic Takete-Maluma phenomenon. *Canadian Journal of Experimental Psychology*, 65(2), 115-124.

MusicVis: Leveraging Cross-Modal Correspondences for Fine-Grained Visual Representations of Music

Kelly Raines^{1†} and Xin Wei Sha²

¹ School of Computing and Augmented Intelligence, Arizona State University, Tempe, Arizona, United States

² The Games, Arts, Media, and Engineering School, Arizona State University, Tempe, Arizona, United States

[†] Corresponding author: kraines5@asu.edu

Introduction

When musicians describe particular characteristics or qualities of music, they often choose descriptors such as “bright”, “warm”, “rough”, and other visually derived attributes. This sense of terminology emerges early in child development, independent of specialized training or cultural knowledge, reflecting the deeply rooted connection between sound and visuals (Wallmark & Allen, 2020). Also known as cross-modal correspondences, this occurs when two seemingly distinct sensory modalities are systematically associated (Spence, 2011). This paper presents MusicVis, a system that leverages cross-modal correspondences to transform music into non-figurative, artistic representations.

MusicVis introduces an alternate approach to existing solutions with a three-step pipeline: stemming, feature extraction, and visual mapping. Stemming isolates the individual parts of the music’s composition in which each are analyzed for contributing features such as pitch, tempo, or instrument sources through pre-trained deep learning models and classifiers. These features are then mapped to corresponding visual textural masks and colors, created by procedural algorithms. This methodology enables a fully transparent system with fine-grained control over the configurable parameters for generated visuals.

The purpose of MusicVis is to provide a tool for performers to enhance practice and performance. By visualizing the sounds, the performer can analyze the colors, textures, and shapes produced and evaluate whether it aligns with their intended expression. Based on the results from a study conducted by Heyen et al. (2026), visualizations allow musicians to reflect and learn from patterns that may not be immediately apparent. As for live performance, visualizations create an additional expressive dimension for the performer that may support audience understanding under conditions that allow focused engagement, such as controlled or intimate settings (Lartillot et al., 2025). In these two aspects, performers benefit from MusicVis’ explainability and fine-grained control for self-evaluation and expression.

Related Work

A review of existing literature on existing music-to-image systems informed MusicVis on current approaches’ strengths and areas of improvements. One example formalizes a set of methods: direct associations, random associations, and genetic algorithms (GAs) to generate abstract images from MIDI files (Aleixo et al., 2021). Direct associations between visuals and audio were guided by established cross-modal correspondences and author preferences, random associations removed instrument-based assumptions of shape and texture, and GA combines these methods to iterate on colors, location of elements in the canvas, and shape distributions. Majority of the participants reported positively that the audiovisual connections are understandable, favoring the GA methodology. However, the system was not as robust since it relies on MIDI files in which compositional details of the music may be lost.

Another approach takes advantage of generative AI models (GPT-4) to interpret a textual description of music to generate corresponding images in real-time (Yang et al., 2024). Textual description captures details such as tempo, key signature, note length, and instrument in which GPT-4 is then tasked with

detecting the emotion and providing a description to Stable Diffusion XL (SDXL) Turbo to create the related imagery. Participants' feedback revealed that although it was a feasible system, it still suffered from the aforementioned MIDI limitations and wished for more fine control and influence over specific parts of the generated images as their playing developed on certain motifs.

Lastly, Soundsketcher uses a mix of low-level and high-level features from audio to perform line sketches and create aural scores (Zaman et al., 2025). While existing results are compelling, it mentions future integration of deep learning models could improve accuracy and adaptability of the system across diverse musical contexts. Specifically, stream separation techniques would help Soundsketcher distinguish overlapping sound sources which inform the overall sound.

Based on these works, a common issue was capturing subtleties in music either due to restrictions in system input or information loss throughout the process. This is where MusicVis aims to improve with refined control on each of the contributing parts for fine-grained visual generation, direct from the music.

Implementation

MusicVis is agnostic to any musical genre and instruments as all it requires for input is a wav file which will be separated by the `htdemucs_6s` stemmer, provided by Demucs (Rouard et al., 2023). This splits the music into 6 parts: bass, drums, guitar, piano, vocals, and any other unidentified part. The music is then cut into 4-second clips for feature extraction by classifiers, described in the following subsection, and saved into a local database. If a stem is completely silent, below -35 dB, then feature extraction is skipped and it is not added to the database. To begin visualization, a request is sent to the database with the song title to retrieve the features which will be mapped to the corresponding visual textures. Each clip can have anywhere from 1 to 6 stems stacked at once, each with a varying number of visual textures concatenated together. Thus, to get a full picture of the features at each time step, an animation blends each of the stacked visual textures with varying cycle of weights. The finalized visual animation is displayed in time to the music.

Feature Extraction and Classifiers

To perform feature extraction essential for MusicVis, a group of classifiers and pre-trained deep learning models gather both high-level, human-understandable features alongside low-level features from sound analysis. For pitch detection, MusicVis utilizes the small CREPE pre-trained model (Kim et al., 2018). Librosa, a python package for audio analysis, is responsible for tempo estimation and evaluating onset strength envelope, spectral centroid, and spectral flatness (McFee et al., 2025). Loudness was estimated using a separate python package, `pyloudnorm` (Steinmetz & Reiss, 2021). As for detecting instrument source and note quality, which contributes to the tone and timbre, custom classifiers were developed using supervised learning under the NSynth dataset developed by Engel et al (2017). The source classifier had three classes: acoustic (0), electronic (1), synthetic (2). The note quality classifier has four classes: bright, dark, reverb, and percussive which assigns to each either 0 (False) or 1 (True) to indicate whether the feature is present. Although there were additional quality labels in this dataset, MusicVis focuses on these four which are more prevalent across the literature. Both source and quality classifier models achieved an overall accuracy of about 74% on the test set of which the architecture and features input to the classifiers were inspired by a survey by Zaman et al. (2023) which compiled existing models that pertain to timbre-related tasks such as environmental sound classification or accent detection. Often using a CNN architecture, the most performant features were log mel-spectrograms and Mel Frequency Cepstral Coefficient (MFCC) combined with mel-spectrograms. However, for the source classification, it reveals a slight negative impact on performance, so only log mel-spectrograms were used.

Visual Generation

To generate the corresponding visuals, procedural algorithms utilize custom textural functions: scratchy, wavy, blobs, or splatter. Each function contains a set of controllable parameters determined by the music qualities detailed in Table 1. Generation criteria represent playing techniques which affect the shape perceived. According to Schubert et al. (2019), “staccato articulation...and rapid changes in register, harmony, dynamics (Bizet and Chopin in this study) [are] being associated with sharp, jagged objects ..., while slower, legato music (Brahms and Mozart) is characterized by the smooth, curvaceous shapes”. To capture this, low-level features such as spectral flatness or reverb classification determine the style.

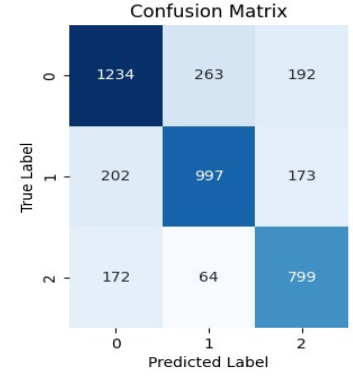


Figure 1: Source Classification Confusion Matrix

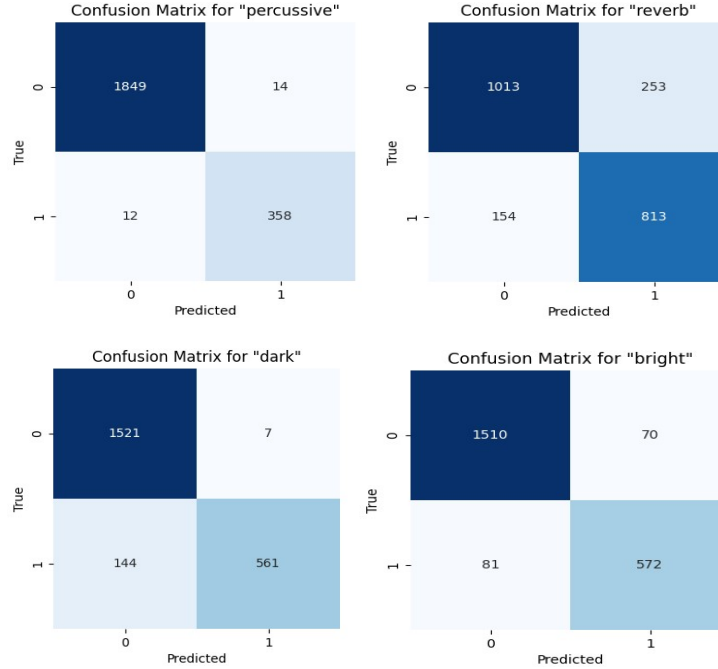
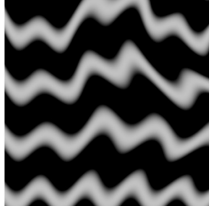
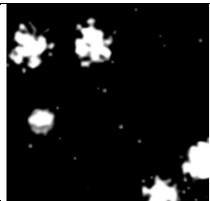


Figure 2: Quality classification confusion matrices

Table 1: Aesthetic Decisions for MusicVis

Texture	Generation Criteria	Parameters	Image Example
Scratchy	Rapid change in spectral flatness	Tempo → Number of Lines Spectral Centroid Deviation → Minimum Separation Line Length → $[\frac{50}{N}, \frac{500}{N}]$, where N is the number of unique notes	

Wavy	Reverb quality present	Onset Strength Deviation → Thickness Acoustic, Electric, and Synthetic Source → Distortion assigned values 1, 3, and 5, respectively	
Blob	Non-acoustic source and low mean velocity OR if no texture	Number of unique notes → Number of blobs	
Splatter	Non-acoustic source and high mean velocity OR if no texture	Number of unique notes → Number of splatters	

As for color, it was separated into three parts: hue, value, and saturation. Since “higher pitches are associated with lighter colors, while lower pitches are associated with darker colors”, the values provided to the image reflect this information (Reymore & Lindsey, 2025). Additionally, louder music is given a higher saturation value (Anikin & Johansson, 2019). Hues are mainly based on warmth or coolness where high spectral centroids are generally cooler and low spectral centroids are warmer (Komschlies, 2022). The split determined was green, blue, and purple fit in the cool category (hue range: [201, 330] ∪ [110, 169]), whereas pinks, reds, orange, yellow were in the warm category (hue range: [241, 330]).

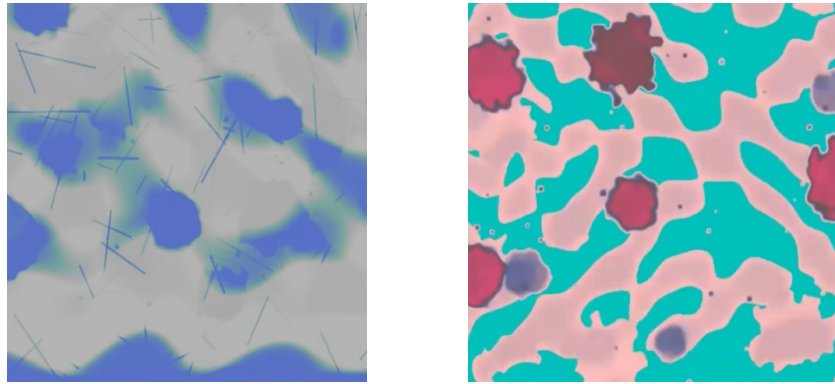


Figure 3: Example visual generations for contrasting music

Limitations & Future Work

An unintended behavior of MusicVis is that it disproportionately generates visual textures for music with more production layers. This occurs since more stems are generated for more distinct instruments, thus causing a loss of complexity from orchestral or classical pieces. To solve this, possible solutions may be explored further in future work:

- Alternate workflow which accepts scores as input to determine the number of textural layers.
- Specialized music source separation/stemmers to handle each voice of each instrument.

- Instrument classifier which replaces the stemmer and determines the number of textural layers.
- Alternative blending algorithms that are less dependent upon the number of textural layers.

Instrument classification may be an interesting first step since it could prove useful to refine color and textural information as the literature suggests a direct tie between tone and color (Reymore and Lindsey, 2025). This would also include an expansion of available procedural algorithms which generate textures would allow visuals to capture larger varieties and shapes that may be described by music.

A secondary goal for future work would involve the source classifier. To make results more reliable and accurate since it was trained on monophonic sources, it may not be as performant for realistic music sources. Based on the research done by Baelde et al. (2019), they suggest that monophonic may be transformed into polyphonic classifiers using the normalized power spectra (NPS). Additionally, expanded datasets with polyphonic examples or architectural improvements can also enhance accuracy.

The long-term vision for MusicVis is its integration into both live and pre-recorded uses. While the system already supports pre-recorded music, its current limitation lies in feature extraction, which is performed offline. Enabling live audio streams to be analyzed by MusicVis will require further research focused on optimizing and refining models to support efficient, real-time analysis.

References

- Aleixo, L., Pinto H. S., & Correia, N. (2021). From Music to Image a Computational Creativity Approach. In J. Romero, T. Martins, & N. Rodríguez-Fernández (Eds), *Artificial Intelligence in Music, Sound, Art and Design* (pp. 379–395). doi:10.1007/978-3-030-72914-1_25
- Anikin, A., & Johansson, N. (2019). Implicit associations between individual properties of color and sound. *Attention, Perception & Psychophysics*, 81, 764–777. doi:10.3758/s13414-018-01639-7
- Baelde, M., Biernacki, C., & Greff, R. (2019). Real-Time monophonic and polyphonic audio classification from power spectra. *Pattern Recognition*, 92, 82–92. doi:10.1016/j.patcog.2019.03.017
- Heyen, F., Gleicher, M., & Sedlmair, M. (2026). Make the Unhearable Visible: Exploring Visualization for Musical Instrument Practice. *IEEE Transactions on Visualization and Computer Graphics*, 32(4), 2927–2943. doi:10.1109/TVCG.2026.3658216
- Kim, J. W., Salamon, J., Li, P., & Bello, J. P. (2018). Crepe: A Convolutional Representation for Pitch Estimation. *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 161–165. doi:10.1109/ICASSP.2018.8461329
- Komschlies, C. (2022). *Crossmodal correspondences in composing and listening to music*. Retrieved from <https://youtu.be/qpmJTasiGKE>
- Lartillot, O., Swarbrick, D., Upham, F., & Cancino-Chacón, C. E. (2025). Video Visualization of a String Quartet Performance of a Bach Fugue: Design and Subjective Evaluation. *Music & Science*, 8.
- McFee, B., McVicar, M., Faronbi, D., Roman, I., Gover, M., Balke, S., ... Südholt, D. (2025). *librosa/librosa: 0.11.0*. doi:10.5281/zenodo.15006942
- Engel, J., Resnick, C., Roberts, A., Dieleman, S., Eck, D., Simonyan, K., & Norouzi, M. (2017). *Neural Audio Synthesis of Musical Notes with WaveNet Autoencoders*.
- Steinmetz, C. J., & Reiss, J. D. (2021). pyloudnorm: A simple yet flexible loudness meter in Python. *150th AES Convention*.
- Reymore, L., & Lindsey, D. T. (2025). Color and tone color: audiovisual crossmodal correspondences with musical instrument timbre. *Frontiers in Psychology*, 15–2024. doi:10.3389/fpsyg.2024.1520131

- Rouard, S., Massa, F., & Défossez, A. (2023). Hybrid Transformers for Music Source Separation. 2023 *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1–5. doi:10.1109/ICASSP49357.2023.10096956
- Schubert, E., Murari, M., Rodà, A., Canazza, S., Da Pos, O., & De Poli, G. (05 2019). Verbal and cross-modal ratings of music: Validation and application of an icon-based rating scale. *Iperception*, 10(3).
- Spence, C. (2011). Crossmodal correspondences: a tutorial review. *Attention, Perception & Psychophysics*, 73(4), 971–995. doi:10.3758/s13414-010-0073-7
- Velenis, K., Zacharakis A., & Cambouropoulos, E. (2025). Soundsketcher: graphical representation of sound based on Crossmodal audiovisual correspondences, *22nd Sound and Music Computing Conference (SMC2025)*, 342-349.
- Wallmark, Z., & Allen, S. E. (2020). Preschoolers’ crossmodal mappings of timbre. *Attention, Perception, & Psychophysics*, 82(5), 2230–2236. doi:10.3758/s13414-020-02015-0
- Yang, M., Llano, M. T., & McCormack, J. (2024). Exploring Real-Time Music-to-Image Systems for Creative Inspiration in Music Creation. In K. Grace, M. T. Llano, P. Martins, & M. Hedblom (Eds), *Proceedings of the 15th International Conference on Computational Creativity, ICCC’24*. Association for Computational Creativity (ACC).
- Zaman, K., Sah, M., Direkoglu, C., & Unoki, M. (2023). A Survey of Audio Classification Using Deep Learning. *IEEE Access*, 11, 106620–106649. doi:10.1109/ACCESS.2023.3318015

Toward a Conceptual Framework for AI-Driven Timbre Transformation: Exploring Ethics, Occlusion, Transparency, and Interaction in Algorithmic Design

Steven Lewis^{1†}

¹ College of Design, Art & Performance, University of South Florida, Tampa, Florida, USA

[†] Corresponding author: smlewis5@usf.edu

AI-enhanced musical models are a means for the generation of sound, computational performance analysis, and the development of a new collaborative co-creator paradigms in emergent musical ecosystems (Thelle & Waerstad, 2023; Wang et al., 2025). These ecosystems of co-creativity represent an ontological transformation for digitally mediated timbral manipulation and human-computer interaction, altering not only how sounds are made, but how musicians relate to and interact with their instruments. This relational shift shapes musical motivations and the interdisciplinary nature of creative identities during a time when the once-distinct roles in music—the performer, composer, engineer, instrument designer, to name only but a few—are converging into a new model of musicianship.

AI-mediated timbre analysis and transformation software has introduced profound yet equally unresolved questions for music technology, computational aesthetics, and the cultural politics of computer-mediated composition and performance. As these technologies become further integrated into real-time or asynchronous workflows for hybrid and fully cybernetic instruments, new forms of bias, shifting notions of authenticity, and increasingly layered and obfuscated networks of authorship, meaning, and mediation will continue to emerge (Born, 2005; Brown, 2021; McCormack et al., 2019). These questions warrant investigation into the ways in which audio and gestural analysis, corpora creation, and interface design collectively contribute to the infrastructural and cultural dynamics that bound the sonic possibilities and creative affordances of AI-mediated timbre transformation.

Open toolkits and proprietary commercialized software present divergent models for user experience and interaction. Toolkits for machine learning, analysis, and corpus-based concatenative synthesis (Schwarz 2006, 2007) such as FluCoMa (Tremblay et al., 2022; Moore et al., 2021) and the Data Knot² packages for Cycling74's Max/Msp~ emphasize education, transparency and customization, while commercialized tools such as Synplant 2 and its Genopatch technology³, center systematized operation and opaque function in lieu of a granular examination and intervention of its underlying algorithmic processes. These distinctions call for a comparative framework that examines the fundamental differences between these approaches while building on existing research into how artists engage with computer-human interaction and improvisation (McCormack & d'Inverno, 2016), materiality (Magnusson, 2019; Born, 2011) musicking with machine learning assemblages (Dahlstedt, 2021; Fiebrink, 2016; Gioti et al., 2021; Jourdan & Caramiaux, 2023), and timbre transformation in AI-mediated digital and electroacoustic music practice (Stefánsdóttir & Magnusson, 2025). These differing digital music interfaces provide equally contrasting user experiences, altering the way musicians understand and engage with these emergent technologies. A way to organize the panoply of AI-mediated tools, especially as it relates to algorithmic

² Made by programmer and percussionist Rodrigo Costanza, Data Knot is set of tools made for Max/Msp~ that leverages machine learning for descriptor analysis, classification and clustering, regression and prediction, corpus analysis, among other techniques for real-time performance. <https://rodrigoconstanzo.com/2025/10/data-knot-machine-learning-tools-for-low-latency-real-time-performance/>.

³ Made by Sonic Charge and engineer Magnus Lindström, Synplant 2's Genopatch uses proprietary AI to create individual synth settings—referred to as patches—based on user-loaded audio file. As it analyzes the input source, Genopatch generate patches that increasingly match the chosen audio. <https://soniccharge.com/synplant>.

transparency, is imperative for understanding how particular user interaction paradigms contextualize the creative choices related to timbral transformation.

Framework

The framework presented here is comprised of four interrelated and contingent dimensions. *Transparency vs. Occlusion* stresses the degree to which access to and visibility of algorithmic design influences user intervention and their analytical capacity. *Ethics of Access* examines how algorithm design encodes bias and cultural ideas of creativity in performance. *Interaction Modalities* probes how the presentation of the interface, be it through skeuomorphism, flat design, or otherwise, greatly shapes notions of agency and control related to performance outcomes. *Authorship of Process* explores how systemic transparency impacts beliefs about creative authenticity in data-driven musical processes. These four dimensions provide a conceptual framework for designers, researchers and musicians to connect critique to the practical considerations that guide creative practices involving AI-mediated timbre manipulation.

To elucidate on this framework, the user experience of performing an original electroacoustic piece — *Fracture and Flow*—will be detailed to explore this conceptual model’s intersectional and contingent components. Using the same acoustic percussion performance, one version of the piece was processed through Synplant 2’s highly visualized, familiar, yet algorithmically opaque interface, while the other was done so through the comparatively transparent, iterative workflow of the FluCoMa and Data Knot toolkits in Max/Msp~. This juxtaposition effectively demonstrates how differing algorithmic opacities and interface designs influence the range of perceptual sonic possibilities, gestural interactions, and creative agencies across the framework’s four dimensions.

Table 1: Framework Applied to Commercial and AI-mediate Open Toolkits

<i>Dimensions</i>	Genopatch (Synplant2)	Data Knot (Max Msp~/FluCoMa)
	Description	Description
<i>Transparency vs. Occlusion</i>	<i>Occluded entirely; user cannot see algorithmic functions or processes.</i>	<i>Semi-transparent; users can examine abstractions, and the functions of foundational FluCoMa infrastructure.</i>
<i>Ethics of Access</i>	<i>Opaque and fixed; software design decides what users access and change at an abstracted level.</i>	<i>Open and fluid; users can both intervene on existing algorithms and iteratively configure them to be customized instruments of analysis and timbral transformation.</i>
<i>Interaction Modalities</i>	<i>Unidimensional; computer mouse movement on main interface further modulates sounds.</i>	<i>Multimodal; users can control parameters in real-time with any number of gestural interactions, controllers, and networked technologies.</i>
<i>Authorship of Process</i>	<i>Input source selection is the main aspect of authorship; users can select the sounds Genopatch creates, repeat the process, or make changes to the input source.</i>	<i>Users can create custom interactive systems, combining programming with modular AI and ML-driven algorithms.</i>

Analysis

Through analyzing the conceptual framework and applying it to the case study of *Fracture and Flow*, this research will illuminate how algorithmic transparency shapes the cultural situatedness of AI-mediated timbral transformation. The first dimension—transparency vs occlusion—has a downstream effect on the framework’s other considerations, as well as on the perception of creative affordances in an interface’s user experience. The degree to which fixed digital products (as opposed to nearly fully customized, interactive programming environments) affect creative decisions is greatly determined by skeuomorphic design choices, which itself is a result of a lack of transparency into any underlying algorithmic function.

Skeuomorphism is a design approach that uses material metaphors to present a range of options, referred to by J.J. Gibson as “affordances” (1979) — to the user in virtual spaces (Chandler & Munday, 2020). In commercial digital software, skeuomorphic design (or any hybrid thereof) choices limit the range of gestural affordances so that a virtual object’s behavior roughly mirrors its real-world equivalent. Indubitably, the combination of a skeuomorphic design methods, unidimensional interactions, and a lack of access and agency over algorithmic processes can offer a demystifying heuristic approach to AI-mediated timbral transformation. Yet, grafting familiar acoustic metaphors into virtual spaces invites the musician to expect a familiar sonic skeuomorphism to accompany the virtual simulacrum of real-world interactions and metaphors. Transforming timbres in ways novel to AI’s creative affordances may necessitate interface design choices that are contemporaneous to this moment of technological development and integrated musical mediation. As this presentation will demonstrate, the real-time processes and sonic outcomes resulting from AI-driven timbre transformations are not merely determined by algorithms, but through the ways in which these visible and interactive virtual elements are presented to and perceived by the user.

References

- Born, G. (2005). On Musical Mediation: Ontology, Technology and Creativity. *Twentieth-century music*, 2(1), 7-36.
- Born, G. (2011). Music and the materialization of identities. *Journal of Material Culture*, 16(4), 376-388.
- Bown, O. (2021). Sociocultural and design perspectives on AI-based music production: why do we make music and what changes if ai makes it for us?. In *Handbook of artificial intelligence for music: Foundations, advanced approaches, and developments for creativity* (pp. 1-20). Cham: Springer International Publishing.
- Chandler, D., & Munday, R. (2011). *A dictionary of media and communication*. OUP Oxford.
- Dahlstedt, P. (2021). Musicking with algorithms: Thoughts on artificial intelligence, creativity, and agency. In *Handbook of artificial intelligence for music: Foundations, advanced approaches, and developments for creativity* (pp. 873-914). Cham: Springer International Publishing.
- Fiebrink, R. (2016). Machine learning as meta-instrument: Human-machine partnerships shaping expressive instrumental creation. In *Musical Instruments in the 21st Century: Identities, Configurations, Practices* (pp. 137-151). Springer Singapore.
- Gibson, J. J. (1977). The theory of affordances. *Hilldale, USA*, 1(2), 67-82.
- Gioti, A. M., Einbond, A., & Born, G. (2022). Composing the assemblage: Probing aesthetic and technical dimensions of artistic creation with machine learning. *Computer Music Journal*, 46(4), 62-80.
- Jourdan, T., and Caramiaux, B. (2023). “Machine learning for musical expression: a systematic literature review,” in *Proceedings of the International Conference on New Interfaces for Musical Expression*. Mexico City: NIME. 319–331.
- McCormack, J. & d’Inverno, M. (2016). Designing improvisational interfaces. In *International*

- Conference on Computational Creativity. Paris, France, 27 June - 1 July 2016.
- Magnusson, T. (2019). *Sonic Writing: Technologies of Material, Symbolic, and Signal Inscriptions*. Bloomsbury Academic.
- McCormack, J., Gifford, T., & Hutchings, P. (2019, April). Autonomy, authenticity, authorship and intention in computer generated art. In *International conference on computational intelligence in music, sound, art and design (part of EvoStar)* (pp. 35-50). Cham: Springer International Publishing.
- Moore, T., Bradbury, J., Tremblay, P. A., & Green, O. (2021). Making Machine Learning Musical: Reflections on a Year of Teaching FluCoMa. *Journal SEAMUS*, 32(1-2), 24.
- Schwarz, D. (2006). Concatenative sound synthesis: The early years. *Journal of New Music Research*, 35(1), 3–22.
- Schwarz, D. (2007). Corpus-based concatenative synthesis. *IEEE signal processing magazine*, 24(2), 92-104.
- Stefánsdóttir, H. S., & Magnusson, T. (2025). Of altered instrumental relations: a practice-led inquiry into agency through musical performance with neural audio synthesis and violin. *Frontiers in Computer Science*, 7, 1578595.
- Thelle, N. J. W., & Wærstad, B. I. G. (2023). Co-Creatives Spaces: The machine as a collaborator. In *Proceedings of the International Conference on New Interfaces for Musical Expression*.
- Tremblay, P. A., Roma, G., & Green, O. (2021). Enabling programmatic data mining as musicking: The fluid corpus manipulation toolkit. *Computer Music Journal*, 45(2), 9-23.
- Wang, Y., Zuo, Y., & Zhang, X. (2025, May). The Practice and Challenges of “Human-AI Co-creation” in Music Composition. In *International Conference on Human-Computer Interaction* (pp. 219-233). Cham: Springer Nature Switzerland.

Concrete Listening in Latent Space: Reduced Semantic Conditioning of Diffusion Transformers as a Compositional Strategy

David Piazza^{1†}

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

[†] Corresponding author: david.piazza@umontreal.ca

This poster presents a compositional project where I explore reduced semantic conditioning in a text-based latent audio diffusion model as a method for composing timbre (Evans et al., 2024; Liu et al., 2023). Rather than treating generated sounds as the successful or unsuccessful realizations of prompts, the project approaches the model’s outputs as concrete materials to be auditioned, transformed, and organized after their generation. The central question is how reductions of semantic control produce sounds whose source identity, morphology, density, and instrumental affordances remain unstable enough to become compositional material.

The project is situated within electroacoustic and musique concrète practices in which timbre is a primary bearer of musical form (Schaeffer, 1966; Smalley, 1997). In this context, listening is the principal method through which the musical potential of the generated material is discovered. The proposed notion of concrete listening in latent space names a studio practice in which diffusion processes are deliberately under-specified, then evaluated through repeated audition according to their spectral, morphological, and formal potential (Smalley, 1997). This orientation shifts attention from prompt adherence toward the musical consequences of ambiguity, incompleteness, and hybridization.

The work is developed using the Stable Audio Open model in ComfyUI (ComfyUI Contributors, 2026; Evans et al., 2024). Four generation regimes are compared. The first uses relatively explicit prompts naming recognizable sources, actions, or timbral qualities, along with associated model sampling parameters. The second uses underspecified or abstract prompts, which reduces the likelihood of a stable source category in the output. The third uses unconditioned generation, producing sounds with minimal explicit semantic anchoring. The fourth applies partial denoising to encoded recordings, allowing recorded and model-generated material to occupy a shared intermediate space. Across these conditions, seed, denoising amount, sampling parameters, and prompt text are logged so that listening comparisons can be repeated.

Generated and hybrid materials are auditioned in multiple passes and annotated according to perceived source ambiguity, noise-pitch balance, spectral envelope, attack profile, temporal continuity, and capacity to function as figure, background, transition, or harmonic material (Smalley, 1997). Sounds are selected when they suggest possible formal roles or invite further timbral transformation. This protocol responds to the need for experimental specification while remaining appropriate to artistic research: the comparison concerns modes of listening and compositional affordance, not the accuracy of text-to-audio retrieval (Ben-Tal et al., 2021; Gioti et al., 2022).

Selected materials are then developed in a digital audio workstation alongside other non-generated recordings. Some outputs remain close to their generated form and are heard as such; others are edited, layered, filtered, granulated, or used as sources for harmonic extraction. When pitched content emerges from the diffusion process, salient harmonic components are selectively tracked and re-instrumentalized through MIDI-based resynthesis, virtual instruments, and hardware synthesizers. These post-generative procedures operate as a form of orchestration: they respond to the timbral behavior of the model’s outputs rather than predetermining those outputs through increasingly precise prompts.

The poster presents the comparative generation protocol, annotated listening categories, and excerpts from the resulting composition. By bringing latent audio diffusion into dialogue with electroacoustic approaches to timbre, the project argues that reduced semantic conditioning can be understood as a deliberate compositional condition through which timbre is encountered in text-to-audio models as emergent, unstable, and musically actionable.

References

- Ben-Tal, O., Harris, M. T., & Sturm, B. L. T. (2021). How Music AI Is Useful: Engagements with Composers, Performers and Audiences. *Leonardo*, 54(5), 510–516.
https://doi.org/10.1162/leon_a_01959
- ComfyUI Contributors. (2026). *ComfyUI* [Computer software].
<https://github.com/comfyanonymous/ComfyUI>
- Evans, Z., Parker, J. D., Carr, C., Zukowski, Z., Taylor, J., & Pons, J. (2024). *Stable Audio Open* (Version 2). arXiv. <https://doi.org/10.48550/ARXIV.2407.14358>
- Gioti, A.-M., Einbond, A., & Born, G. (2022). Composing the Assemblage: Probing Aesthetic and Technical Dimensions of Artistic Creation with Machine Learning. *Computer Music Journal*, 46(4), 62–80. https://doi.org/10.1162/comj_a_00658
- Liu, H., Chen, Z., Yuan, Y., Mei, X., Liu, X., Mandic, D., Wang, W., & Plumbley, M. D. (2023). AudioLDM: Text-to-Audio Generation with Latent Diffusion Models. *Proceedings of the 40th International Conference on Machine Learning*, 21450–21474.
<https://proceedings.mlr.press/v202/liu23f.html>
- Schaeffer, P. (1966). *Traité des objets musicaux*. Éditions du seuil.
- Smalley, D. (1997). Spectromorphology: Explaining sound-shapes. *Organised Sound*, 2(2), 107–126.
<https://doi.org/10.1017/S1355771897009059>

Prony Informed Percussion Orchestration: A Case Study for *Elements of Adventure*

Joseph Chang^{1†} and Philippe Depalle¹

¹ Schulich School of Music, McGill University, Montreal, QC, Canada

[†] Corresponding author: joseph.chang@mail.mcgill.ca

Abstract

Elements of Adventure (Chang, 2026), a composition for percussion sextet, is shaped by the Taxonomy of Orchestral Grouping Effects (TOGE; McAdams, Soden, and Goodchild, 2022), informed by the author’s participation in the third edition of the Composer-Performer Orchestration Research Ensemble (CORE 3) as a vibraphonist. This dual analytical-performative perspective supports a compositional approach in which timbre functions as a primary structural parameter. The perceptual principles underlying TOGE are complemented through Prony analysis (Laroche, 1989), which models percussion signals as sums of exponentially damped sinusoids. This method enables the estimation of prominent modes, inharmonicity, and decay characteristics beyond the limits of Fourier-based analyses. Prony’s method provides a quantitative basis for perceptual attributes of timbral fusion and segregation. The relationship informs the compositional design of *Elements of Adventure*, particularly in the movement “Foreign Landscapes,” where orchestration is informed by projected spectral and temporal behaviours. By exploiting differences in modal structure and decay, the work articulates perceptual grouping across registral layers and over time. The study demonstrates how parametric analysis can inform perceptually grounded orchestration strategies in contemporary percussion writing.

Analysis of Percussion Sounds Using Prony’s Algorithm

Percussion sounds present significant challenges for analysis using FFT-based methods due to their inharmonic spectra, pronounced transients, and rapid decay. Early work by Jean Laroche (1989) demonstrated that Prony’s method can effectively estimate parameters of short audio signals, including frequency, damping factor, amplitude, and phase. The analysis is understood in three steps. First, a structured data matrix (Hankel or Toeplitz matrix) is constructed from the signal, enabling separation of signal and noise subspaces via eigen decomposition. Second, the signal subspace is used to estimate a prediction polynomial, whose roots encode the modal frequencies and damping factors of the signal. Finally, by forming the analytic signal (Hilbert transform), the remaining parameters—amplitude and phase—are estimated through a least-squares solution that involves a Vandermonde matrix (Kay and Marple 1981).

A key advantage of Prony’s method lies in its ability to extract highly accurate data pertaining to the dominant modal components from short signal durations. By examining the eigenvalue distribution for a given model order p , one can distinguish the number of significant modes and their parameters from percussion spectra. For example, the eigenvalue spectrum of wind chime at F6 (Figure 1) displays a sharp decay ending at -30dB, followed by dense clusters below—the boundary between signal and noise subspaces. The number of modes (L) approximates to half the number of poles (p) above the threshold, since sinusoids are represented as complex conjugate pairs. In this case, three dominant modes are identified, corresponding to the principal resonant components of the Sabian wind chimes.

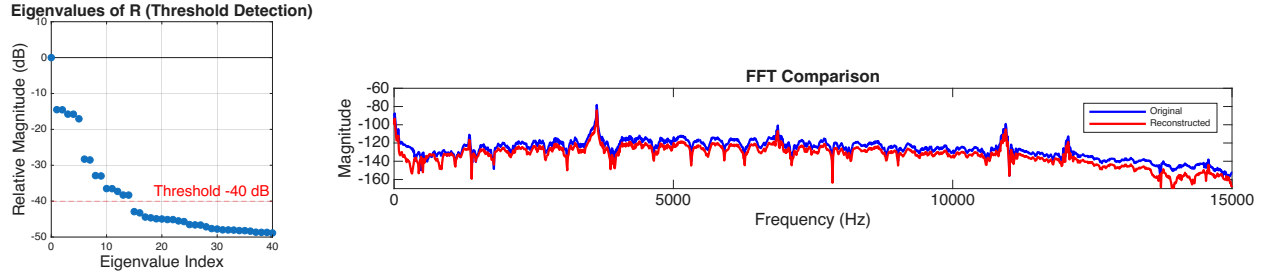


Figure 1: Eigenvalue and FFT Plots of F6 Wind Chime Note.

The Fourier transform plot of two comparable pitches from the wind chime and the glockenspiel illustrate clear differences in modal structure (Figure 2). The glockenspiel note exhibits two dominant spectral peaks, whereas the wind chime displays a more complex, inharmonic distribution with approximately five prominent modes. However, Prony analysis estimates the model order as $L = 7$, revealing an additional closely spaced pair of modes at 5511 and 5531 Hz that are not clearly resolved in this magnitude spectrum. This highlights a key strength of Prony’s method: its ability to resolve closely spaced components beyond the frequency resolution limits of FFT-based analysis. A similar effect is observed in the Eb6 wind chime note, where five dominant modes in the 5-11 kHz range (Figure 3) are identified through Prony modeling (6.35, 6.39, 10.13, 10.14, 10.27 kHz, indicated in blue) despite appearing partially merged in the Fourier representation (Figure 3).

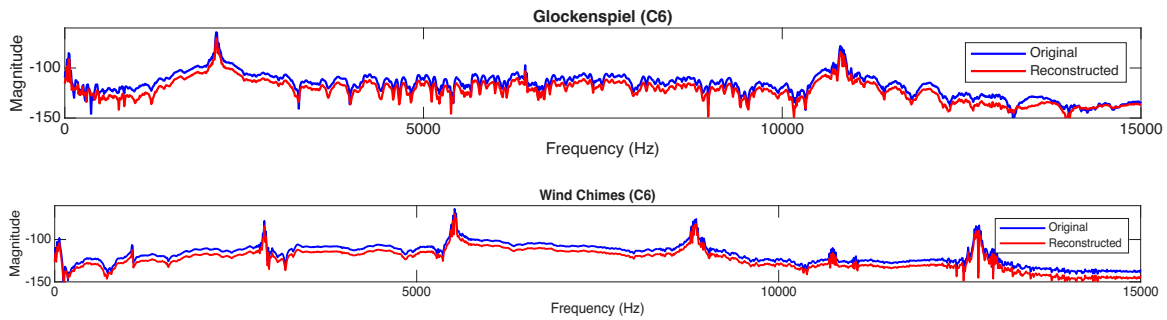


Figure 2: FFT Plots for C6 (approx.) Note from Glockenspiel and Wind Chimes

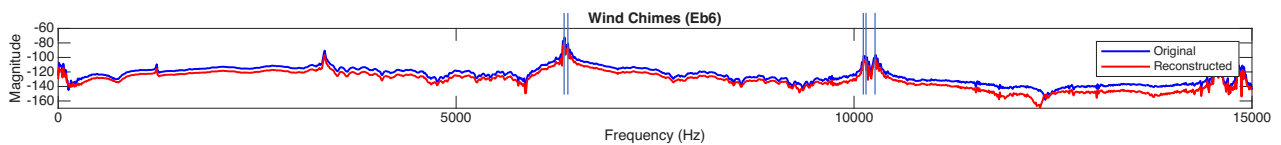


Figure 3: FFT Plot for Eb6 from Wind Chimes.

Prony Informed Percussion Orchestration

Comparison of spectral characteristics—especially modes, degrees of harmonicity, and decay rates—provides a quantitative basis for perceptual attributes of timbral blend as described in the Taxonomy of Orchestral Grouping Effects (TOGE; McAdams, Soden, and Goodchild 2022). For the C6 tones of the glockenspiel and wind chimes, Prony analysis reveals contrasting modal structures in terms of the number of prominent modes and their ratios, as well as differences in the distribution of noise components within its decay. The wind chime exhibits a greater number of inharmonic modes,

contributing to a perceptually “richer” timbre, whereas the glockenspiel is characterized by fewer, and harmonically stabler modes, yielding a comparatively “cleaner” or “hollower” quality. In both cases, noise is concentrated around the most prominent spectral components—near the fundamental in the glockenspiel and the third lowest mode in the wind chime—shaping the initial attack. Their temporal behaviors further differentiate the two timbres. Whereas the glockenspiel exhibits a more uniform and sustained decay profile, the wind chime displays a faster, asymmetrical decay across its modes, reflecting differences in its resonant structure. When combined at normalized amplitudes, these characteristics produce an emerging timbre: the initial onset is shaped by the glockenspiel’s concentrated noise and stable modes, while the wind chime’s broader inharmonic spectrum becomes increasingly perceptible as the sound decays.

This relationship forms the basis of the compositional approach in *Elements of Adventure*, particularly in the movement “Foreign Landscapes,” where orchestration is closely aligned with projected spectral and temporal behaviours. The work systematically exploits differences in modal structure, inharmonicity, and decay to shape perceptual grouping across registral layers and over time. Before turning to the analysis of the movement, a brief clarification of the TOGE framework is necessary.

Taxonomy of Orchestral Grouping Effects (TOGE)

TOGE, developed by McAdams, Goodchild, and Soden (2022), provides a framework for understanding orchestral grouping based on principles of auditory scene analysis. In other words, TOGE proposes a taxonomy for understanding orchestral grouping in relation to instrumentation, primarily within acoustic orchestral works. This framework can be extended to chamber music—mixed ensembles and electronic works, where its descriptors for textural differentiation are especially effective.

TOGE identifies three primary categories of grouping: concurrent, sequential, and segmental. These categories are better understood as a two-dimensional analytical system. The concurrent grouping principles concern the formation of timbral events at the first order (small scale) level, while sequential grouping principles take these unitary auditory events and observe their interactions with other textures through time. Segmental grouping principles concern interactions between textures within shorter bounded segments. Musically, concurrent grouping concerns relationships among individual instruments and their timbral fusion or segregation, sequential grouping takes multiple of such groups and relate their interaction over time, and segmental grouping concerns specific textural effects in a moment of time, such as echoes and antiphonal contrasts.

Analysis of Foreign Landscapes With TOGE

“Foreign Landscapes” is the fourth movement of *Elements of Adventure*, a work written for a percussion sextet using musical idioms drawn from 21st century video games. “Foreign Landscapes” draws on idioms derived from Gérard Grisey’s spectralism (Grisey, 1987) and the registral motion characteristic of percussion works by Iannis Xenakis (Xenakis, 1969). The movement evolves in time through registral shift from the lowest to the highest spectral strata, complemented with extended techniques on drums and keyboard instruments. These gestures ultimately create growling sounds, breaths, and hums which interact with one another to deliver narratives of danger and the discomfort from exploring a hostile environment. Two sections are analyzed using the TOGE framework: mm. 1-7 and 88-90. These segments represent the introduction and key transitional passages linking distinct timbral regions. While mm. 1-7 introduce the primary timbral materials, mm. 88-90 marks a transition from noise-based sounds to more clearly defined pitched materials with higher spectral centroids.

In the analysis from mm. 1-7, highlighted colours represent timbral streams based on concurrent grouping principles (Fig. 4). In the upper registral strata, the crotales and the mark tree are integrated as a single stream, with mark tree as the prominent voice while the crotales function to augment the mark tree. A second stream emerges among cymbals, tam-tam, and bass drum. The cymbal bowing and the scraping of the tam-tam excite overtones, producing sharp transients and sustained spectra, while their decays are blended seamlessly with the lowest spectra of the tam-tam and cymbals. The decay of the bass drum similarly blends with the lower spectrum of the tam-tam. While their onsets remain heterogenous in timbre (mm. 1, 3, 5), the arrival of the sustain and their prominent decay blends well with the 19" cymbal, ultimately producing an emergent, evolving composite timbre.

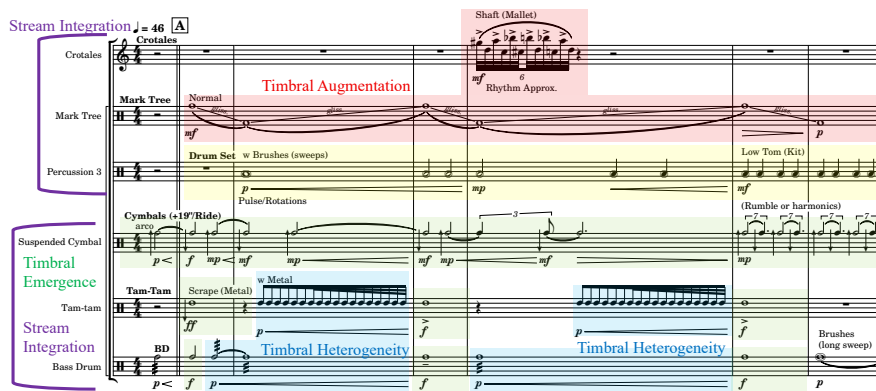


Figure 4: Analysis of “Foreign Landscapes” mm. 1-7.

The final analytical segment from mm. 88-90 concerns the blend between the higher metallic keyboard and percussion instruments contrasting with the lower tam-tam and bass drum textures (Figure 5). The blend among crotales, mark tree, and triangle is particularly salient, with the mark tree being the most prominent timbre, supported by the crotales and triangle. This produces a continuous stream of bell-like “twinkling” attacks, while the vibraphone is the only texture which remains partially segregated with crotales, mark tree, and the triangle. The other two streams of textures, tam-tam and bass drums are recurring periodically yet distinct in timbre to be perceived as a single auditory stream. This presents three hierarchically differentiated streams of textures: the first in the highest spectrum led by crotales, the second, lower vibraphones, and finally the combination of tam-tam and the bass drums.

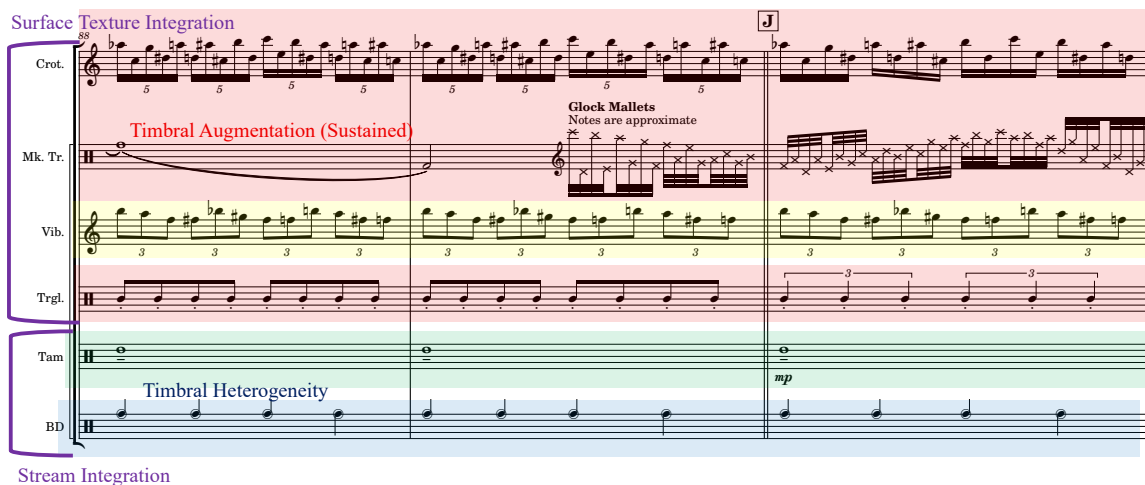


Figure 5: Analysis of “Foreign Landscapes” mm. 88-90.

The analyses presented for “Foreign Landscapes” suggest that the framework is effective in the analysis of works written for chamber percussion ensembles given the extensive timbral variability of the percussion ensemble. The final analysis of the crotales lines for mm. 88-90 especially demonstrates how timbral groupings can play a crucial role not only analytically but also in the conception of the work. Extending this approach in future compositions—through the systematic deployment of unconventional techniques (e.g. mark tree played with glockenspiel mallets)—enables the generation of novel timbral configurations that remain underexplored analytically. Analyses that account for timbral blending and registral movement using the TOGE framework clarify perceptual organization, as demonstrated in “Foreign Landscapes” and informed by CORE 3.

References

- Besada, L. (2021). Xenakis’s sieve theory: a remnant of serial music? *Music Theory Online*, 28(2).
<https://www.mtosmt.org/issues/mto.22.28.2/mto.22.28.2.besada.html>.
- Grisey, G. (1987). Tempus ex Machina: A composer’s reflections on musical time. *Contemporary Music Review*, 2(1), 239-275.
- Laroche, J. (1989) “A new analysis/synthesis system of musical signals using Prony’s method: Application to heavily damped percussive sounds,” *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, Glasgow, U.K., pp. 2053-2056.
- 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>.
- Steven M. Kay, S.M., and Marple, S. L. (1981) “Spectrum analysis – a modern perspective”, *Proceedings of the IEEE*, Vol 69, pp 1380-1419, Nov.
- Xenakis, I. (1969). *Persephassa* [Score]. Salabert.
- Xenakis, I. (1992). *Formalized music: thought and mathematics in composition*. Pendragon Press.

Timbre Discrimination in Drum Tom-Toms: An Approach Based on Audio Descriptors and Timbre Space

Carlos Eduardo Leal de Castro^{1,2,†} and Marcelo Gomes de Queiroz²

¹ Ibmec São Paulo, São Paulo, SP, Brazil

² Institute of Mathematics and Statistics, University of São Paulo, São Paulo, Brazil

[†] Corresponding author: carlos.castro@professores.ibmec.edu.br

Abstract

Drum timbre emerges from the interaction of shell material, shell thickness, reinforcement, drumheads, hoops, tuning, and excitation conditions. This study investigates whether controlled micro-variations in real tom-tom construction can be discriminated through audio descriptors and represented in micro-timbre spaces. We built a controlled dataset of recorded 12" × 10" double-headed tom-toms, varying one construction factor at a time: wood species, shell thickness, reinforcement ring, drumhead type, and hoop type. Recordings were obtained under controlled close-miking, tuning, and striking conditions. Temporal and spectral descriptors were extracted from the recordings, and four complementary methods—Support Vector Machine, Random Forest, Analysis of Variance, and Mutual Information—were integrated to identify robust descriptors for timbral discrimination. The descriptors most consistently selected across methods were then used to construct three-dimensional micro-timbre spaces for each construction factor. The results show that subtle physical modifications in tom-tom construction produce structured differences in descriptor space, especially in dimensions related to spectral energy distribution, spectral complexity, sustain, damping, and harmonic control. These findings support the use of descriptor-based micro-timbre spaces as compact representations of perceptually meaningful timbral variation across tom-tom configurations.

1. Introduction

The sonic identity of a drum is shaped by a complex network of physical and mechanical variables that dictate its acoustic behavior. In tom-tom drums, small construction changes may produce perceptually relevant differences, making timbre analysis important for both acoustics research and instrument design. This study investigates whether micro-variations in real tom-tom construction can be discriminated through audio descriptors and represented in micro-timbre spaces under controlled experimental conditions (Ibáñez-Arnal et al., 2020; La Cerra, 2020; McAdams et al., 1995; Siedenburg & McAdams, 2017).

2. Methodology

The experiment comprised controlled dataset construction, recording of all tom-tom configurations, feature extraction, multi-method feature selection, and micro-timbre space construction. ANOVA, Mutual Information, Support Vector Machine, and Random Forest were combined to identify robust descriptors for timbral discrimination.

2.1 Building the Dataset

The dataset was designed to isolate one construction factor at a time in real tom-tom recordings, allowing controlled comparison of wood type, shell thickness, reinforcement ring, drumhead type, and hoop type (La Cerra, 2020).

2.1.1 Tom-Tom Shell, Drumheads, and Hardware

Custom 12" × 10" double-headed tom-toms were fabricated through a partnership with Odery Drums to enable controlled comparisons across construction factors. All shells were built with the same 45° bearing edge and consistent manufacturing conditions, while varying one factor at a time. The dataset included shells in Araucaria (*Araucaria angustifolia*), Copaiba (*Copaifera* spp.), and Teak (*Tectona grandis*); shell thicknesses of 5.5 mm and 8.5 mm; the presence or absence of a reinforcement ring; two 12-inch Evans batter drumhead types (clear and coated); and two hoop types (double-flanged steel and die-cast aluminum). This design allowed controlled analysis of wood, thickness, reinforcement ring, drumhead, and hoop effects on timbre.

2.1.2 Recording, Tuning, and Striking Control

To avoid confounding construction effects with size-related pitch differences, all comparisons were carried out on 12" × 10" double-headed tom-toms. Recordings were obtained in a dedicated room using a close-miking setup with an omnidirectional Behringer ECM8000 microphone positioned at approximately 60° and 5 cm from the drumhead. Drumhead tension was controlled with a Tama Tension Watch to promote even tuning and reduce tuning-related variability. Four tension levels (60, 70, 80, and 85 t.m.u.) were combined across batter and resonant heads, yielding 16 tuning conditions. Excitation was standardized with a pendulum-based striking apparatus: shell recordings used a silicone-tipped 2B drumstick to reduce hard-impact artifacts, whereas assembled tom-toms were struck with a timpani mallet to ensure stable pendular excitation. A repeatability test based on repeated strikes supported the consistency of the excitation procedure across recordings.

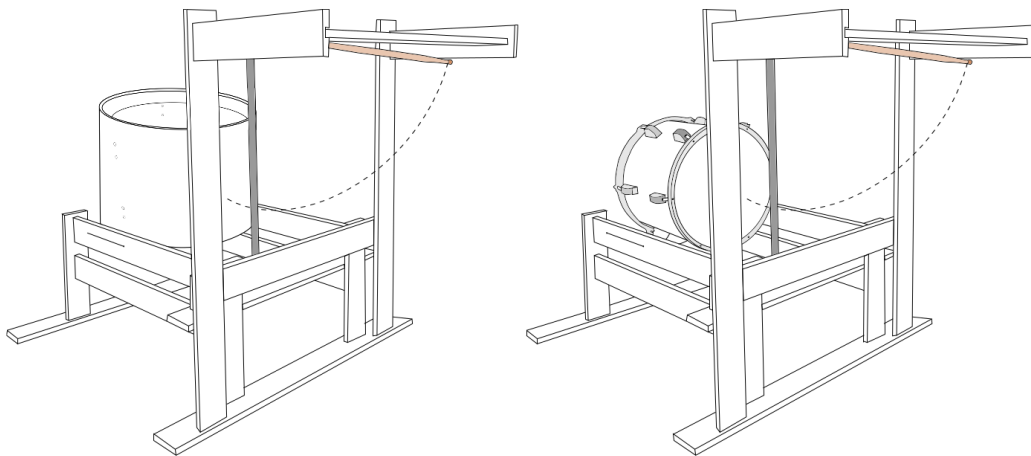


Figure 1: Pendulum-based striking apparatus used to standardize excitation during the recordings. The left panel shows the setup for shell vibration measurements, and the right panel shows the setup for assembled tom-tom recordings.

2.2 Feature Extraction

Temporal, spectral, and spectro-temporal descriptors were extracted from the recorded sounds, including spectral centroid, flatness, entropy, slope, zero-crossing rate, inharmonicity, log-attack-time, and effective duration. Features available in Librosa were computed directly, whereas the remaining descriptors followed Timbre Toolbox definitions. Most descriptors were computed using a 2048-sample Hann window with 50% overlap (Peeters et al., 2011).

3. Results and Discussion

Feature-selection results were integrated across SVM, Random Forest, ANOVA, and Mutual Information, prioritizing descriptors selected by multiple methods as robust indicators of timbral discrimination. For wood type, spectral slope and spectral entropy showed the strongest consensus, indicating that wood changes affect both spectral energy distribution and spectral complexity.

Table 1. Consensus descriptors selected across methods for each construction and assembly factor.

Construction factor	Main consensus descriptors
Wood type	Spectral Slope; Spectral Entropy; ZCR
Shell thickness	Effective Duration; Inharmonicity; Spectral Flatness
Reinforcement ring	Spectral Kurtosis; Effective Duration; RMS
Hoop	Spectral Entropy; ZCR; Effective Duration
Drumhead	Spectral Kurtosis; ZCR; Effective Duration

The micro-timbre space for wood type showed a clear separation of Copaiba from the other woods, while Teak and Araucaria appeared more closely mixed. This pattern is consistent with the physical and mechanical differences among the woods and suggests perceptually relevant differences in brightness/darkness balance and spectral complexity.

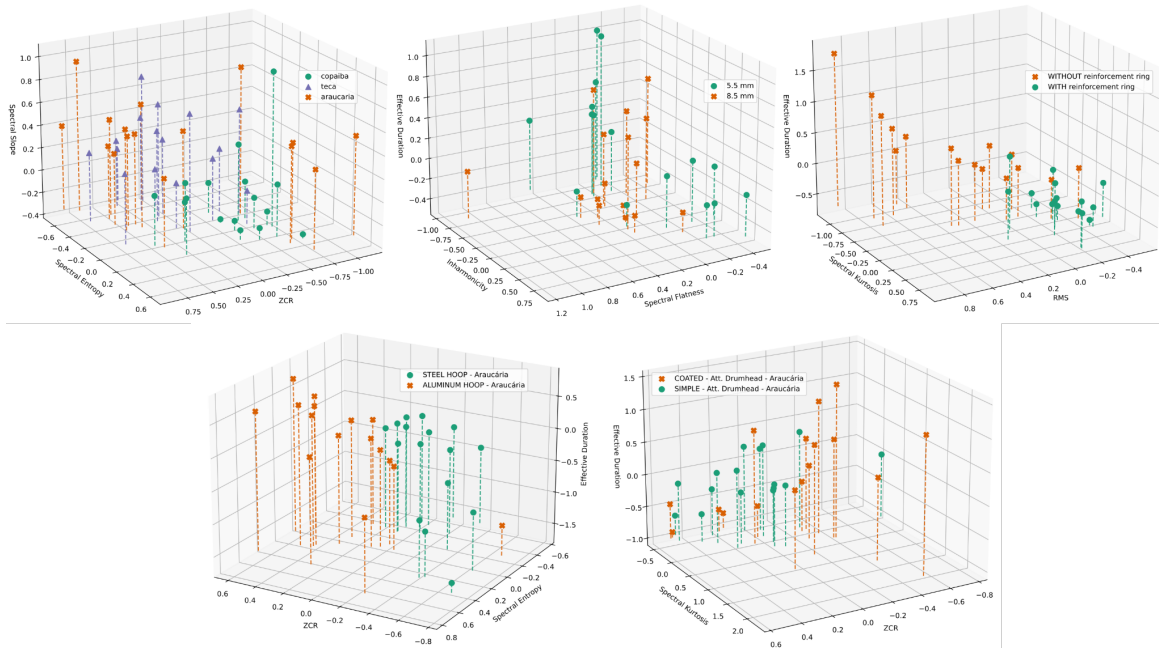


Figure 2: Micro-timbre spaces for timbral discrimination of five tom-tom construction factors: wood type (top left), shell thickness (top center), reinforcement ring (top right), hoop type (bottom left), and batter drumhead type (bottom right). Each space was constructed from the descriptors most consistently selected by the integrated feature-selection procedure.

For shell thickness, Effective Duration showed the strongest consensus across methods, indicating that thickness-related differences are expressed mainly through damping and sustain behavior.

Reinforcement-ring differences were also associated with Effective Duration, together with spectral kurtosis, suggesting a more controlled and temporally compact response when reinforcement alters shell behavior.

Hoop and drumhead comparisons further reinforced the timbral relevance of the selected descriptors. In the hoop task, the selected descriptors point to differences in harmonic control, sustain, and energy distribution. In the drumhead task, spectral kurtosis, zero-crossing rate, and effective duration suggest differences in harmonic scattering, brightness, and sustain: the clear drumhead tended to produce a more controlled, brighter, and less sustained response than the coated drumhead.

Taken together, these findings suggest that the proposed micro-timbre spaces capture stable timbral tendencies rather than merely separating categories statistically. Across the five construction factors, the selected descriptors repeatedly point to interpretable auditory dimensions—such as spectral complexity, brightness, sustain, damping, and harmonic control—indicating that subtle physical modifications in real tom-toms produce perceptually organized variation. In this sense, the contribution of the present study is not only methodological, but also timbral: it shows that controlled construction differences can be translated into compact descriptor-based representations that remain meaningfully connected to audible aspects of drum sound.

References

- Ibáñez-Arnal, M., Doménech-Ballester, L., & Sánchez-López, F. (2020). Design methodology and material evaluation criteria for membranophone shells. *Applied Acoustics*, 169, 107449. <https://doi.org/10.1016/j.apacoust.2020.107449>
- La Cerra, S. (2020, February 20). The modern drum set, part 3: Toms. Yamaha Hub. <https://hub.yamaha.com/drums/studio/part-3-toms/>
- McAdams, S., Winsberg, S., Donnadieu, S., De Soete, G., & Krimphoff, J. (1995). Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes. *Psychological Research*, 58, 177–192. <https://doi.org/10.1007/BF00419633>
- Peeters, G., Giordano, B. L., Susini, P., Misdariis, N., & McAdams, S. (2011). The Timbre Toolbox: Extracting audio descriptors from musical signals. *The Journal of the Acoustical Society of America*, 130(5), 2902–2916.
- Siedenburg, K., & McAdams, S. (2017). Four distinctions for the auditory “wastebasket” of timbre. *Frontiers in Psychology*, 8, Article 1747. <https://doi.org/10.3389/fpsyg.2017.01747>

Session 5

Negotiating Timbre between East and West:
Friction, Reflexivity, and Knowledge-Making in Cross-Cultural Timbral Studies

Negotiating Timbre Between East and West: Friction, Reflexivity, and Knowledge-Making in Cross-Cultural Timbral Studies

Yifan Huang¹, Linglan Zhu¹, and Stephen McAdams¹

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

Cross-cultural timbre research has often been framed within a comparative paradigm, assessing the extent to which perceptual, analytical, or computational frameworks—typically developed within specific Western historical and epistemological traditions—can be extended to other musical contexts. This panel proposes a shift in perspective. Rather than treating cross-cultural research primarily as a test of applicability, it argues that the cross-cultural encounter itself can function as a generative site of knowledge-making. When established frameworks meet divergent musical practices, moments of mismatch or “partial failure” need not be understood as methodological shortcomings. Instead, such frictions can render implicit assumptions visible, prompting reflexive interrogation and transformation of the frameworks themselves.

The 90-minute session is designed to foster dialogue across perception, analysis, technology, and notation. Following a brief introduction, four concise presentations (~15 minutes each) examine how conceptual, analytical, and technological frameworks developed within Western musical traditions interface with musical knowledge in Chinese music contexts. The session then moves into a moderated discussion (~30 minutes) structured around shared guiding questions, followed by an open forum inviting perspectives from the audience.

The four presentations include:

- Perceptual Semantic Study (Huang et al.) examines how a timbre semantic model originally developed for Western orchestral sounds is taken up across linguistic and cultural contexts when applied to Chinese orchestral instruments. While shared perceptual structures are observed, culturally inflected descriptive strategies invite reflection on how semantic models are interpreted and mobilized in cross-cultural research.
- Orchestration Analysis (Zhu et al.) explores the application of a perception-based orchestration framework (TOGE) to a contemporary work for piano and Chinese orchestra. By showing how hybrid instrumental textures strain existing analytical categories, this presentation highlights challenges in theorizing perceptual organization within intercultural orchestration contexts.
- Physical Modelling and Organology (Luan & Scavone) investigates the timbre of the Chinese bamboo flute (dizi) using physical modelling approaches drawn from Western musical acoustics. Engagement with material variability and nonlinear sound-producing mechanisms highlights methodological challenges that arise when modelling paradigms across instrument-making traditions.
- Digital Music Encoding (Sun et al.) presents a gesture-based MEI encoding module for jianzipu, the notation system of the Chinese guqin, in which sound production is notated through embodied actions. By foregrounding gesture, this work suggests alternative representational priorities for encoding timbre-oriented musical systems.

The Panel Discussion Will Be Structured Around Three Themes:

1. Epistemological Visibility—how encounters with other musical traditions render explicit the attributes of timbre that are typically foregrounded or marginalized within established research paradigms.
2. Generative Mismatch—what forms of musical knowledge emerge from moments of friction or “partial failure”, and how such mismatches can be understood as productive.
3. Framework Transformation—how cross-cultural timbre research actively reshapes the established framework it employs, rather than merely evaluating their applicability.

By reframing cross-cultural timbre research as an encounter that interrogates and transforms established tools, this panel moves beyond binary comparison. It argues that knowledge does not reside in a framework or a culture alone, but emerges through their interaction, contributing to a more nuanced and culturally situated understanding of timbre.

A Cross-Cultural Comparison of the Timbre-Semantic Space of Chinese Instrumental Sounds

Yifan Huang^{1†}, Lena Heng², Zijin Li³ and Stephen McAdams¹

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

² Department of Psychology, St. Francis Xavier University, Antigonish, Nova Scotia, Canada

³ Music Artificial Intelligence and Music Information Technology, Central Conservatory of Music, Beijing, China

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

Timbre is not only acoustically grounded but also culturally situated. Different musical cultures develop not only different instruments, but also different ways of using, hearing, and speaking about sound. This makes timbre a particularly revealing domain for cross-cultural music perception. When listeners encounter the sounds of an instrumental tradition with which they have different degrees of familiarity, are those sounds organized according to universal perceptual dimensions or culturally specific ones?

Direct measurements of timbre perception cannot be easily achieved, and semantic descriptions of timbre provide a way to examine how listeners conceptualize what they hear. Such descriptions can reflect relatively general auditory dimensions, such as brightness or roughness, but they may also draw on metaphors, material associations, instrument knowledge, and culturally embedded aesthetic categories. Existing models of timbre semantics have provided important tools for studying these verbal dimensions of timbre. Reymore and Huron's (2020) 20-dimensional model of timbre qualia, for example, was developed from verbal descriptions of imagined Western orchestral instrument sounds. This model offers a rich vocabulary for characterizing instrumental timbre, but its scope beyond Western orchestral instruments and English-language context remains an open question. Chinese orchestral instruments offer a valuable case through which to examine this issue. These instruments include distinctive materials, resonating bodies, playing techniques, and historically situated aesthetic associations. They therefore provide an opportunity to examine the extent to which an existing semantic model of timbre qualia can be used across instrumental traditions.

The present study investigated the timbre-semantic space of Chinese orchestral instrument sounds in two listener groups: native Mandarin-speaking listeners in Beijing and native English-speaking listeners in Montreal. We examined how shared perceptual organization and differential familiarity with an instrumental tradition jointly shape timbre semantics. This study addressed two main questions. First, can Reymore and Huron's 20-dimensional model be used meaningfully to describe Chinese orchestral timbres? Second, do Mandarin- and English-speaking listeners organize these timbres along comparable semantic dimensions?

Sixteen recorded Chinese orchestral instrument sounds were selected from the CCMusic dataset (Liu & Li, 2021). All sounds were produced at D5 in order to reduce pitch-related variation and to restrict the task to timbral differences. The stimulus set included representative sounds from four major instrument families: aerophones, bowed chordophones, plucked chordophones, and idiophones, as shown in Table 1. This selection was intended to sample a broad range of materials, excitation mechanisms, and spectrotemporal profiles within the Chinese orchestral sound world.

Two groups of participants completed the experiment. The Beijing group consisted of native Mandarin speakers ($n = 42$), and the Montreal group consisted of native English speakers ($n = 43$). Participants in both groups had a range of musical experience. On each trial, participants heard one instrumental sound and rated how well each of the 20 semantic dimensions from Reymore and Huron's model described that sound, using a seven-point continuous scale. The Montreal group completed the task in English, using the

original descriptors. The Beijing group completed the task in Mandarin Chinese, using translated descriptors. The Chinese translations were evaluated by three trained musicians who were native Mandarin speakers and fluent in English, as well as by one trained musician who was a native English speaker and fluent in Mandarin. The goal was to preserve the intended perceptual meaning of each descriptor, rather than only its literal lexical form, while ensuring that the Chinese terms sounded natural in musical description. After the rating task, participants were asked to provide any additional terms that they felt could describe the sounds. This open-ended component was included to capture descriptive strategies that might not be fully represented by the 20 predefined semantic dimensions.

Table 2: Chinese orchestral instruments used as stimuli, organized by instrument family.

Aerophones	Bowed Chordophones	Plucked Chordophones	Idiophones
梆笛Bangdi (bd)	高音板胡Gaoyin banhu (gybh)	柳琴Liuqin (lq)	编磬Bianqing (bq)
曲笛Qudi (qd)	高音胡GaoHu (gh)	琵琶Pipa (pp)	编钟Bianzhong (bz)
高音笙Gaoyin sheng (gys)	二胡Erhu (eh)	中阮Zhongruan (zr)	云锣Yunluo (yl)
高音唢呐Gaoyin suona (gysn)	中胡Zhonghu (zh)	古筝Guzheng (gz)	
		扬琴Yangqin (yq)	

Table 1: English (Reymore & Huron, 2020) and Chinese versions of the 20 semantic descriptors

English Term	Chinese Term	English Term	Chinese Term
rumbling/low	低沉的/浑厚的	ringing/long decay	带余音的/余音绕梁的
soft/singing	甜美的/歌唱性的	sparkling/brilliant	明亮的/闪亮的
watery/fluid	似水的	airy/breathy	带气息的/气声的
direct/loud	嘹亮的/响亮的	resonant/vibrant	丰满的/有共鸣的
nasal/reedy	带鼻音的	hollow	空腔的
shrill/noisy	刺耳的	woody	木质的
percussive	敲打声的/打击声的	muted/veiled	阻塞的/闭塞的
pure/clear	清澈的	sustained/even	平稳的/延续的
brassy/metallic	金属的	open	开阔的/开放的
raspy/grainy	沙哑的/粗糙的	focused/compact	集中的

In the first step of the analysis, mean ratings were examined to identify the semantic profile for each instrument and to compare these profiles across listener groups, as shown in Figure 1. The semantic-rating profiles indicated that the 20-dimensional model was broadly applicable to Chinese orchestral timbres. Participants in both groups used the descriptors systematically rather than randomly. In general, both English- and Mandarin-speaking participants showed broadly similar conceptualizations of the timbres they perceived, with comparable top- and bottom-rated descriptors for individual sounds. Differences emerged for certain terms, which may be related to translation choices and varying degrees of background knowledge about the instruments, especially for descriptors related to instrument material.

The open-ended responses revealed a complementary layer of timbre semantics. Participants in both groups provided basic perceptual descriptors, including terms corresponding to sharp, brief, calming, 尖锐, 短促, and 舒缓. However, the two groups differed in their descriptive strategies. Mandarin-speaking participants more frequently produced holistic, evaluative, and culturally anchored expressions, such as 端庄 [dignified], 优雅的 [elegant], 有韵味 [rich in aesthetic character], 民族的 [ethnic or culturally traditional], and 竹林般的 [bamboo-forest-like]. These responses suggest that familiar Chinese instrumental timbres may invite descriptions that draw not only on local acoustic attributes, but also on aesthetic categories, cultural imagery, and broader histories of musical association. English-speaking participants, by contrast, tended to provide terms referring to specific perceptual or material attributes, such as metallic, airy, or synthetic. The contrast between the rating and open-ended tasks is therefore theoretically important. Whereas the rating task revealed a partially shared semantic organization of timbre, the open-ended responses showed how linguistic and cultural familiarity can shape the elaboration and interpretation of these percepts.

Overall, our study suggests that timbre semantics involves both shared perceptual organization and culturally mediated description. The results show that Reymore and Huron's 20-dimensional timbre-qualia model provides a useful framework for characterizing Chinese orchestral timbres, and that both listener groups organized the sounds according to interpretable dimensions related to acoustic characteristics. However, the open-ended descriptions also suggest that the model may not fully capture culturally specific semantic dimensions, inviting further study. By bringing Chinese orchestral timbres into a cross-cultural semantic-rating paradigm, this study extends existing models of timbre qualia beyond the Western orchestral context and contributes to a more nuanced understanding of how listeners hear, organize, and describe culturally situated instrumental sounds.

Acknowledgements

This research was supported by the Panicao China Strategic Initiative Fund. Zijin Li was supported by the National Social Science Fund of China, Special Project for Rare and Precious Studies: Construction of a Digital Cultural Heritage Platform for Musical Instruments of Ethnic Minorities in Southwest China. We thank Bennett K. Smith for programming the bilingual experimental interface, Bryana-Alexa Liberta for volunteering in experiment running in Montreal, and Ziya Zhou for assistance with experimental setup in Beijing.

References

- 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>
- Zhaorui Liu, & Zijin Li. (2021). Music Data Sharing Platform for Computational Musicology Research (CCMUSIC DATASET) (1.1) [Data set]. *Zenodo*. <https://doi.org/10.5281/zenodo.5676893>

Notating Timbre in Jianzipu: A Gesture-Based MEI Encoding Module

Eto Sun^{1†}, Yifan Huang¹ and Ichiro Fujinaga¹

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

[†] Corresponding author: eto.sun@mail.mcgill.ca

Timbre is often implicitly specified and treated as a secondary outcome of performance in pitch-centric notation systems, such as common Western staff notation. In a broader notation context, *jianzipu* (减字谱, reduced-character notation), the notation system of the *guqin* (古琴, Chinese plucked string instrument), can be understood as adopting a timbre-oriented perspective (Gu, 2004). This study examines jianzipu as a gesture-based notation system where physical actions that produce musical sound are notated. We first present preliminary acoustic analyses and then introduce a digital encoding module that represents jianzipu as structured gesture information for computational music research.

Introduction. Jianzipu is a logographic notation system that encodes fingering and playing gestures rather than representing abstract musical pitches (Hui, 2024). It has been in continuous use since the late Tang dynasty, for approximately 1,300 years (Wang, 2006, pp. 57–58). Jianzipu is the main notation used for the *guqin*, which is one of China’s foremost solo instrumental traditions. Jianzipu differs fundamentally from Common Western Music Notation (CWMN), which treats the note as an atomic unit defined by pitch and duration rather than physical action. It also differs from Western string tablature. Although tablature uses string-fret coordinates in place of explicit pitch names, it still functions primarily as a pitch-surrogate system. Jianzipu instead prescribes sound production through a complex logographic system where each symbol encodes physical actions, including touch type, movement direction, finger identity, string position, and left-hand articulation. In this framework, pitch emerges as a derivative of physical action, whereas timbre is shaped by the notated gesture itself.

Structure of Jianzipu. The basic unit of jianzipu is the *puzi* (谱字), and a score is formed as a sequence of *puzi*. *Puzi* are generally categorized as *zhengzi* (正字), *pangzi* (旁字), and *pangzhu* (旁注) (Gu, 2004; Zha, 1958). A *zhengzi* is the principal musical event, encoding the right-hand finger and plucking gesture, the string number (*xian*), the left-hand finger, the *hui* position, the *hui* subdivision (*fen*), and the articulation gesture. *Hui* are the thirteen inlaid studs on the *guqin* soundboard that mark harmonic nodes, functioning similarly to frets on other string instruments. *Fen* refers to a decimal subdivision between adjacent *hui* specifying fine positional adjustments; for example, *qi hui liu fen* (七徽六分, 7th hui 6th fen) denotes a position six-tenths of the way from the 7th hui toward the 8th. A *pangzi* represents gestures that shape the resonance of the preceding *zhengzi*, typically performed by the left hand while the string is still vibrating. These gestures function as independent musical actions rather than simple ornaments, often evolving in multiple stages. A *pangzhu* annotates section boundaries, harmonic (*fanyin*) markers, indications of tempo or dynamics, and special techniques.

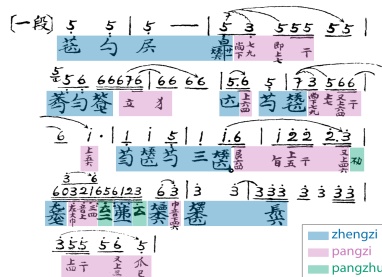


Figure 1: A jianzipu example from Liu Shui (流水) from Gu (2004, p. 367). Zhengzi, pangzi, and pangzhu are marked in different colors.

Preliminary acoustic analysis. To illustrate how timbral features can be encoded in *puzi*, we conducted a preliminary acoustic analysis of two examples: a short phrase containing five gestures (Figure 2) and an instance of the gesture *dayuan* (打圆), which consists of seven notes (Figure 3). For each example, we computed the spectrogram on an equivalent rectangular bandwidth (ERB) scale, root-mean-square (RMS) envelope, and the spectral centroid trajectory for the analysis.

In Figure 2, the five gestures in the phrase are as follows: (1) with the left thumb pressing the seventh string at the seventh *hui* position, the performer plucks the seventh string using *tiao* (挑), a right-hand index-finger technique involving nail-side contact and an outward motion away from the player; (2) the left hand then moves upward to the sixth *hui* and second *fen*; (3) the left hand advances further to the fifth *hui* and sixth *fen*; (4) the left hand retreats; and (5) the left hand returns to the previous position. Only the first gesture produces sound through the initial plucking of the string and is notated in *zhengzi*. The remaining four gestures, notated in *pangzi*, reshape the sound by moving the left-hand finger along the string while the string is still vibrating. The spectrogram shows how the spectrum changes in relation to these notated gestures. Notably, the third and fifth *puzi* produce the same pitch. However, because they involve different gestures, particularly differences in the speed of left-hand movement, they produce different spectral centroids.

The second example, shown in Figure 3, presents the sound of the gesture *dayuan* (打圆), which consists of seven notes at the same pitch. These seven notes are produced through an alternation between an inward pluck by the right-hand middle finger using the fleshy side of the fingertip and an outward pluck by the right-hand index finger using the nail on a different string. This alternating motion is repeated seven times. The spectrogram, and especially the spectral centroid, reveals acoustic variation across the successive plucks. These variations arise from several interacting factors, including the alternation between flesh-side and nail-side contact and differences between the strings being plucked. As a result, even though the seven notes share the same pitch, they exhibit rich acoustic variation.

These preliminary examples suggest that different guqin gestures can produce timbral variation and that these subtle acoustic differences are systematically recorded in *jianzipu*.

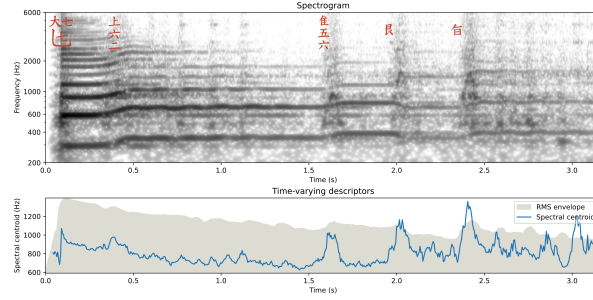


Figure 2: Spectrogram and acoustic features of a phrase from Pingsha Luoyan (平沙落雁).

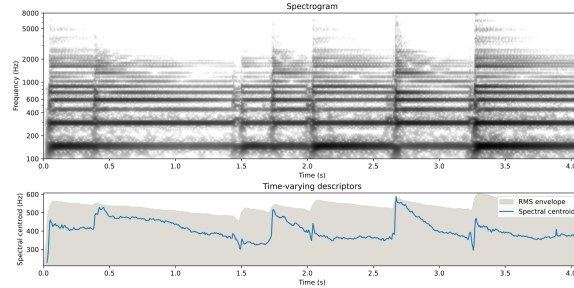


Figure 3: Spectrogram and acoustic features of the guqin gesture dayuan (打圓).

Encoding Jianzipu. Although jianzipu possesses substantial musical and historical value and remains an active notation system in contemporary practice, it lacks a formal machine-readable encoding for systematic and computational research. Current research on jianzipu digitization has mainly addressed computer-assisted typesetting, custom fonts, and pitch extraction (Hui, 2024; Liang, 2025; Ni et al., 2011). These works remain isolated without a structured encoding framework. To address this gap, we introduce a novel encoding system based on the Music Encoding Initiative (MEI) (Hankinson et al., 2011).

The proposed module, `MEI.jianzipu` (<https://github.com/One-Piece-Music/music-encoding-jianzipu>), is based on MEI version 5.1 (Music Encoding Initiative, 2025). Instead of modeling a score as a sequence of pitch-defined notes, the module represents jianzipu as a sequence of compound gesture events within the existing `<staff>` and `<layer>` framework. In the current customization, a conventional jianzipu score is modeled as a single primary staff with one layer of such events. This design treats the compound gesture encoded by each puzi as the primary unit of representation.

The module introduces three core elements to represent the distinct components of a puzi: `<zhengzi>`, `<pangzi>`, and `<pangzhu>`. For internally compound or multi-string actions, a `<zhengzi>` may contain `<zPart>` sub-elements that decompose the event into constituent parts. The schema encodes string position, tone type, and left- and right-hand techniques through attributes constrained by controlled vocabularies, datatype restrictions, and Schematron rules. By capturing these structured components, the module preserves the interactions between right-hand excitation and left-hand modulation that generate the instrument’s distinctive timbre.

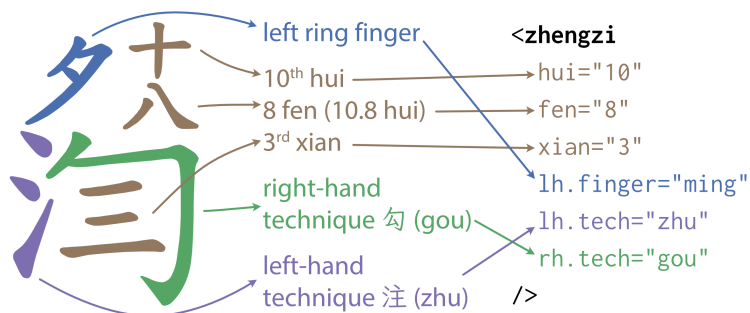


Figure 4: A zhengzi example and the corresponding MEI encoding using the proposed jianzipu module. The meaning of the zhengzi is: the left ring finger starts slightly above the 10th hui 8th fen position and slides downward to that position, while simultaneously the right hand's middle finger plucks the 3rd string toward the body.

To visualize MEI files encoded with the jianzipu module, we developed a web-based previewer (<https://github.com/One-Piece-Music/Qino>) that parses the MEI XML and renders jianzipu puzi in real time. The renderer uses the JianZiPu font created by Liang (2025). Users can edit the MEI source in one panel and inspect the rendered result in the other.

Conclusions. This study presents a preliminary acoustic analysis of guqin performance techniques and an MEI module for representing jianzipu as machine-readable gesture information. The acoustic analysis indicates that jianzipu gestures are timbre-relevant because distinct notated actions can produce sounds with different spectral and temporal properties.

The MEI module transforms notated gesture information into digitally analyzable data, from which timbre-relevant patterns can be studied in musicological and computational research. The current module has a defined scope. It prioritizes core conventions that are widely shared and remain stable in contemporary guqin practice. The present version focuses on a structural representation of puzi gestural semantics. Pitch information is therefore not integrated into the current module, because pitch can be inferred from tuning and positional information as a derivative layer. Visual details such as manuscript-specific abbreviations, contextual omissions, and historical graphic variations are not yet modeled and are left for future development.

More broadly, this work positions gesture-based encoding as a viable approach to representing timbre-oriented notation systems, offering a reference model for encoding other music traditions and informing the design of gesture-based notation for contemporary digital musical instrument (DMI) practice. For guqin studies, the encoding enables data-driven musicological research. For example, scholars have observed that *Xiaoxiang Shuiyun* (潇湘水云), a canonical guqin composition preserved in more than fifty tablatures, shows increasing complexity in left-hand techniques across its historical transmission (Liu, 2020). Encoding such sources would make it possible to examine changes in resonance-modulating gestures and pangzi density across manuscripts, offering a path toward historical analysis of timbre-related notation practice.

References

- Gu, M. (2004). 琴学备要 [Essential Studies on the Guqin] (Manuscript, Vol. 1). Shanghai Music Publishing House.
- Hankinson, A., Roland, P., & Fujinaga, I. (2011). The Music Encoding Initiative as a document-encoding framework. *Proceedings of the 12th International Society for Music Information Retrieval Conference (ISMIR 2011)*, 293–298.
- Hui, Y. (2024). A retrospective view of Chinese contributions to digital humanities in computational Guqin music research. *Computer Music Journal*, 48(2), 45–58. <https://doi.org/10.1162/COMJ.a.4>
- Liang, N. Y. (2025). *JianZiPu* (Version 3.1) [Computer software]. <https://github.com/neuralfirings/JianZiPu>
- Liu, X. (2020). 历代古琴曲谱汇考3: 潇湘水云 [A Compilation of Guqin Tablatures Across Dynasties, vol. 3: Xiaoxiang Shuiyun]. Xiling Seal Engravers' Society Publishing House.
- Music Encoding Initiative. (2025). *MEI Guidelines (Version 5.1)*. <https://music-encoding.org/guidelines/v5/content/index.html>
- Ni, E., Jiang, M., Ding, X., & Zhou, C. (2011). Handwriting input system of Chinese Guqin notation. *Journal on Computing and Cultural Heritage*, 3(3), 1–22. <https://doi.org/10.1145/1921614.1921616>
- Wang, Y. (2006). 中国传统音乐乐谱学 [The Study of Musical Notation in Traditional Chinese Music] (1st ed.). Fujian Education Press.
- Zha, F. (1958). 存见古琴指法谱字辑览 [A Comprehensive Collection of Existing Guqin Fingering Puzi] (Mimeographed edition).

Western and Chinese Musical Acoustics in Dialogue: Understanding the Timbre of the Chinese Transverse Flute (Dizi) Through Physics

Xinmeng Luan^{1†} and Gary Scavone¹

¹ Computational Acoustic Modeling Laboratory, Center for Interdisciplinary Research in Music Media and Technology, Schulich School of Music, McGill University, Montreal, Canada

[†] Corresponding author: xinmeng.luan@mail.mcgill.ca

The *dizi* is a traditional Chinese transverse flute, typically crafted from bamboo. While it shares certain structural similarities with the Western transverse flute, the *dizi* is distinguished by a unique acoustical feature: an additional tonehole covered by a thin, wrinkled membrane. This membrane is widely recognized as the primary source of the *dizi*'s characteristic bright timbre, which sets it apart from Western flutes and many other wind instruments. Despite its musical importance and long performance tradition, the physical mechanisms underlying the *dizi*'s sound production remain less thoroughly studied within the framework of modern musical acoustics.

In this presentation, we will examine the physical principles governing the *dizi* and discuss how its distinctive timbre arises from the interaction of multiple physical processes. In particular, we will focus on the roles of membrane vibration and the resonant behavior of the air column. Nonlinear membrane vibration, specifically the cubic nonlinearity, plays a key role in enriching the harmonic content of *dizi* tones and contributes significantly to its characteristic brightness. We will also discuss how the frequency content of the instrument can be examined through its input admittance, providing a snapshot of the resonator's acoustic response. This analysis naturally leads to a discussion of the cutoff frequency associated with the tonehole lattice. Moreover, we will present our recent findings on the presence of two distinct open tonehole lattices in the *dizi*: the finger-hole lattice and the end-hole lattice, and discuss their respective contributions to the instrument's acoustic behavior.

Beyond sound production, this talk will explore how a deeper physical understanding of the *dizi* can inform and support instrument making. In Western musical acoustics, physics-based modeling and acoustical analysis have increasingly contributed to improvements in intonation, timbral consistency, and playability by enabling more predictive and quantitative design. Similar approaches could benefit *dizi* craftsmanship by clarifying how membrane properties, tonehole placement, and bore geometry shape the instrument's sound, helping to bridge empirical craft traditions with physics-based insight. At the same time, *dizi* making is inherently constrained by the organic nature of bamboo. Variations in density, wall thickness, stiffness, and internal structure introduce unavoidable non-homogeneity, making precise control of intonation and timbre challenging compared to Western instruments built from more standardized materials. This material distinction raises important questions about how directly physics-based design methodologies developed in Western contexts can be transferred to instruments like the *dizi*.

This discussion will naturally lead to a broader discussion on whether methodologies developed in Western musical acoustics can be directly applied to the study and design of Chinese instruments, or whether adapted approaches are required to account for their unique characteristics. We will conclude by considering future directions of research on the physics of Chinese musical instruments and the strategies needed to advance this field.

References

- Luan, X., Wang, S., Scavone, G., & Li, Z. (2025). Acoustic characterization of the resonator in the Chinese transverse flute (dizi). *The Journal of the Acoustical Society of America*, 157(5), 3836-3847.
- Luan, X., Wang, S., Li, Z., & Scavone, G. (2022, September). Acoustical analysis of the Chinese transverse flute (dizi) using the transfer matrix method. In *Proceedings of Meetings on Acoustics* (Vol. 49, No. 1, p. 035014). Acoustical Society of America.
- Tsai, C.-G. (2003). *The Chinese membrane flute (dizi): Physics and perception of its tones* (Doctoral dissertation, Humboldt-Universität zu Berlin, Germany).
- Ng, H. H., Tan, D., Ng, T., & Chen, J. (2021). An acoustic impedance analysis of the dizi (Chinese transverse flute). In *Proceedings of the 27th International Congress on Sound and Vibration (ICSV 2021)*.
- Rossing, T. D., & Fletcher, N. H. (1991). *The physics of musical instruments*. Springer.
- Benade, A. H. (1990). *Fundamentals of musical acoustics*. Courier Corporation.
- Chaigne, A., & Kergomard, J. (2016). *Acoustics of musical instruments*. Springer.

Hearing Orchestration With the TOGE: A Perception-Oriented Case Study of Lei Liang's Tremors of a Memory Chord

Linglan Zhu^{1†}, Yifan Huang¹, Florian Iochem², Chen Fan², Lena Heng³, and Stephen McAdams¹

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

²ACCRA (UR3402), Université de Strasbourg, Strasbourg, France

³Department of Psychology, St. Francis Xavier University, Antigonish, Nova Scotia, Canada

[†] Corresponding author: linglan.zhu@mail.mcgill.ca

Introduction. *Tremors of a Memory Chord* (2011), composed for solo piano and Chinese orchestra by Chinese-American composer Lei Liang, offers a revealing site for studying orchestration through the lens of inter-cultural musical dialogues. The piece places the organologically distinct “timbrarium” (i.e., the global timbre content of an instrument; Cordero & Soden, 2024) of the piano with the winds, plucked strings, bowed strings, and percussion of the modern Chinese orchestra. As these instrumental worlds neither simply oppose nor smoothly assimilate to one another, *Tremors* provides a rich opportunity to examine the activity of listening to and understanding orchestration: that is, how listeners parse sound into events, streams, layers, textures, and larger segments in potentially different ways. The current study uses the Taxonomy of Orchestral Grouping Effects (TOGE; McAdams et al., 2022) not only as an analytical vocabulary, but also as an object of methodological reflection, i.e., what happens when a perception-oriented taxonomy, developed largely through examples from Western orchestral repertoire, is applied to a contemporary intercultural work.

Background and analytical protocol. The TOGE extends principles from auditory scene analysis (Bregman, 1990; McAdams & Bregman, 1979) into a taxonomy of orchestral effects. At the concurrent grouping level, sonic events may remain heterogeneous or blend through timbral augmentation or emergence. At the sequential grouping level, events may integrate into one continuous stream, segregate into multiple streams of equal importance, or stratify into layers of different prominence. At the segmental level, larger units are articulated by effects such as timbral shifts, echoes, and other discontinuities. This framework is especially useful for examining orchestration because it shifts attention from score-based instrumentation alone to the music as heard.

Method. Five trained annotators conducted a collaborative listening-based analysis in OrchView (Baril, 2023). After an introduction to the TOGE and a calibration with a short pilot passage taken from *Tremors*, group analyses were carried out on the entire piece, divided into three sections: Section 1 (mm. 1–88), Section 2 (mm. 89–173), and Section 3 (mm. 174–280). Each section was analyzed by a pair of annotators, with the first author participating in two groups. Each pair of annotators then compared annotations locally before all annotators met to discuss convergence, divergence, and problems of terminology. The procedure was designed to explore two topics: the principal orchestral grouping effects in the piece and the practical pressures placed on the TOGE for different listeners with different degrees of familiarity with traditional Chinese instruments.

Analytical results. After an extensive introduction passage featuring six sets of *muyu* (Chinese temple blocks) played tremolo, the first section establishes a noisy, flamboyant sound world in which Chinese percussion, winds, bowed and plucked strings, and piano form shifting textures. Across this section, the piano frequently occupies extreme registers where its pitch clarity weakens, while its transient or noisy components align it more with the Chinese percussion and plucked string instruments. The resulting texture often depends on perceptual stratification: different families of Chinese instruments plus the piano take turns to form foreground, middleground, and background layers whose relative perceptual saliences frequently shift over time.

The second section functions as a structural transition with a renewed timbral palette. The beginning of this section brings back motives used in the previous section, now timbrally brightened by *bangdi* and *qudi* (transverse flutes) which augment the high-register piano-percussion surface. The boisterous texture is then further intensified, with a prominent alternation of foreground and middle-ground layers among percussion, plucked strings, bowed strings, and swiftly running 16th note figures on the piano. Starting from m. 115, however, the ensemble texture abruptly stops, and the piano cadenza creates a sharp segmental rupture before the music morphs into an intimate chamber music texture associated with the “silk-and-bamboo” (丝竹乐) musical idiom. Individual lines played by a diverse set of solo instruments such as *guanzi* (cylindrical double-reed instrument), *xiao* (end-blown flute), *erhu* (bowed string instrument), *pipa* (plucked string instrument), etc., as well as the finger-muted piano attacks in the low register, are often heard as separate but mutually complementary timbral identities. Here, the piano becomes less like an antagonist and more like a collaborator, sharing attacks and damped resonances with plucked Chinese instruments while participating in a pentatonic-focused heterophonic network.

The third section returns to a musical surface that has a high density and force. Several instances of juxtaposing and combining timbres of different blocks of instrumental families can be found. For example, at m. 175, the reed instruments *suonas* (conical double-reed instrument) and *shengs* (pipe organ) form a brilliant foreground with synchronized onsets, similar registers, and shared harmonic support, resulting in a blended chimeric timbre. This moment is immediately contrasted to the ensuing high-register self-intertwined harmonic glissandi played by the *huqin* (bowed string) section, which creates a smooth, glistening surface texture. As another example of the block-wise treatment of instrumentation in this section, at m. 191, multiple percussion groups are given different local rhythms in a densely knitted texture. While they do not fuse locally, their common energy envelope and gestural direction produce a higher-level perceptual unity.

Discussion. Two recurrent musical strategies shape the piece. First, the piano sonority is repeatedly transformed so that it may either stand against or affiliate with the Chinese orchestra. In its solo cadenza passages, large registral leaps and atonal linear gestures mark formal interruptions and emphasize the piano’s concerto-like separateness. Elsewhere, extreme registers, chromatic clashes between the two hands (playing parallel melodic materials a semitone apart), rapid tremolo, string glissandi, damping, and string-plucking all weaken the piano’s canonical identity and bring it toward the noisy, complex, and highly inflected timbres of Chinese instruments. Second, the Chinese orchestra itself is treated through two broad idiomatic poles. In the tumultuous outer sections and the first half of the second section, the composer mainly utilizes block textures that evoke the features of the traditional “percussion-and-wind” music (鼓吹乐): instrumental families are overlaid or juxtaposed, producing stratified but cohesive heterogeneous unities. In the second half of the second section, the writing emulates the more intimate “silk-and-bamboo” sonority, where individual timbres retain independence within a network of melodic and timbral echoes, modal affinities, and call-and-response gestures.

Reflecting on the annotation process itself, in several passages, the most perceptually salient phenomenon was not a stable object at one TOGE level, but a moving relation among layers. For example, in some of the more complex musical textures where multiple instruments interact, local attacks on one instrument can briefly rise to the foreground, then either dissolve into audible background resonance (similar to the effect of “timbral resonance”; Touizrar & McAdams, 2019) or cut off as dry attacks. Likewise, in the heterophonic texture in the middle section of the piece, a melodic fragment on one solo instrument could be heard either as an independent stream or as part of a larger transforming surface. In OrchView, documenting every micro-shift quickly produced dense annotations that were accurate in detail but difficult to communicate. The analytical problem, therefore, is not simply one of assigning the correct taxon. It is also one of choosing the level of granularity that best captures what listeners can meaningfully

follow. *Tremors* makes this issue especially visible because Liang's writing often generates perceptual cohesion through fluctuating rather than fixed sonic identities.

For this reason, the collaborative protocol was crucial. Divergences between annotators did not merely indicate disagreement; they identified places where the music supports more than one plausible percept. These cases are valuable for future perception-based orchestration analysis because they suggest which moments may require confidence ratings, alternative annotations that are equally possible in different hearings, or notes about listeners' familiarity with the instruments and style.

Reflections on the TOGE. The analyses also revealed several limits and productive ambiguities in the TOGE's current categorical resolution. At the concurrent level, the boundary between timbral augmentation and emergence can be difficult to decide when instruments are unfamiliar, spatialized, dynamically unequal, or drawn from the same family. Some dense textures also occupy a grey zone between a blended event and a looser surface texture. This suggests that the TOGE may benefit from a more explicit account of what counts as an "event" when contemporary orchestration produces clouds or granular surfaces rather than clearly delineated discrete notes. At the sequential level, not all salient (and even relatively temporally expansive) foreground objects become "streams": short fragments, ornaments, and local highlights may organize the listening experience without sustaining a stream in the usual sense. At the segmental level, rapidly transforming textures can be heard either as stream transformation or as a chain of timbral shifts, depending on attention and familiarity. These multi-stable percepts are not analytical "failures". Rather, they expose how listener training, cultural familiarity, repeated listening, and attentional strategy affect the grouping effects that the TOGE seeks to describe.

Conclusion. *Tremors of a Memory Chord* shows how contemporary intercultural orchestration can organize form through a continual negotiation among various orchestration strategies such as blend, heterogeneity, stratification, and segmental contrast. The study confirms the TOGE's value as a listening-based vocabulary for orchestration, while also showing that its application is not neutral: the framework must be situated with respect to the instruments, idioms, listeners, and levels of listening effort involved. For future TOGE-based work, especially corpus building and multi-listener annotation, these findings point to the need to record not only which orchestral effects are heard, but also how strongly they are heard, how stable they are across listeners, and how they change with repeated listening. In this sense, the case study proposes a modest methodological shift: TOGE annotations should be read not only as labels for orchestral effects, but also as evidence of how listeners negotiate the boundaries between perceptual categories in real musical contexts.

References

- Baril, F. F. (2023). OrchView (Version 1.5.6) [MacOS]. <https://www.orchview.com/>
- Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. MIT Press.
- Cordero, V., & Soden, K. (2024). Timbrarium. *Timbre and Orchestration Resource (TOR)*. <https://timbreandorchestration.org/blog/timbrarium>
- McAdams, S., & Bregman, A. (1979). Hearing musical streams. *Computer Music Journal*, 3(4), 26–60.
- McAdams, S., Goodchild, M., & Soden, K. (2022). A taxonomy of orchestral grouping effects derived from principles of auditory perception. *Music Theory Online*, 28(3).
- Touizar, M., & McAdams, S. (2019). Aspects perceptifs dans l'orchestration de *The Angel of Death* de Roger Reynolds: timbre et groupement auditif. In P. Lalitte (Ed.), *Musique et cognition: Perspectives pour l'analyse et la performance musicales* (pp. 55–78). Editions universitaires de Dijon.

Session 6

Panel in Timbre and the Politics of Artificial Creativity

Cat in Loop: A Case Study on Timbre, Imagination and AI

Weilu Ge^{1†} and Nithya Shikarpur²

¹ Department of Music, Harvard University, Cambridge, Massachusetts, United States

² EECS, Massachusetts Institute of Technology, Cambridge, Massachusetts, United States

[†] Corresponding author: weiluge@harvard.edu

Timbre is increasingly understood not as a fixed acoustic attribute, but as a dynamic and relational field shaped through perception, interaction, and imagination (Wessel, 1979; Dolan and Rehding 2018, Morrison 2024). Building on this expanded view, this report presents an artist–engineering collaboration that explores timbre as an emergent, speculative space through real-time human–AI music co-creation.

We introduce Cat in Loop, a practice-based research project developed through an artist-engineer-performer collaboration with the Catinblack Ensemble, a low-string trio (viola, cello and double bass) during a one-week ensemble residency at Harvard University. The project complicates the conventional articulation of *techné*, the practical knowledge generated through making, and *epistémè*, the theoretical knowledge generated through reflection, as each mode of knowing is actively performed and circulates across the participants in what Hayles (2017) describes as a cognitive assemblage. Using an interactive generative AI music system, the project explores the collective construction of an imaginative CAT-like timbral and gestural vocabulary. Rather than beginning with predefined sonic targets, it foregrounds shared imagination and embodied listening as primary compositional tools and interfaces.

Within this framework, CAT functions as a compositional practice that proposes a cognitive model of what timbral perception might be and what kind of musical imagination might emerge from it (Rosenboom, 1987). The artist works as a composer-director that brings the conceptual vision and the propositional model that orientates the work; the engineer operationalizes this vision through technical implementation and the tuning of the AI model; the performers contribute their embodied knowledge of their instruments and their sense of being together as an ensemble. CAT is not understood as a fixed representational space that structures timbral qualities but a reorientation that is inhabited and reshaped by the assemblage. The neural network system also embodies this refusal of representational logic. Where current generative audio models treat timbre as a latent attribute to be captured, optimized, and reproduced (Morreale et al., 2025), the network here operates as an interface that adapts to the situated listening, making timbral relations available for imagination rather than representation.

The AI system runs in MaxMSP and integrates VampNet (Garcia et al., 2023), a masked non-autoregressive transformer model optimized for low-latency musical interaction. This design enables a continuous generative response rather than sequential or pre-rendered outputs, supporting an improvisatory setting in which human musicians and the model co-construct and steer the timbral space.

Our approach contrasts with current generative AI music applications such as Lyria (Lyria Team et al., 2025), Stable Audio (Evans et al., 2024), Suno, and Udio, which typically treat timbre as a latent attribute or open-ended text prompt optimized for stylistic coherence. While such systems demonstrate impressive generative capacity, they often constrain timbral exploration by prioritizing reproducibility and perceptual plausibility. In contrast, our project frames timbre as a shared playground for embodied human–AI interaction, where musicians and generative models co-imagine and renegotiate musical possibilities. “CAT” functions as a collectively inhabited abstraction, centering an imaginative figure across composers, performers, engineers, and the generative model.



Fig 1. photo from the HGNM residency concert performance. From left to right: Sofia von Atzingen (Viola), Javad Javadzade (Double Bass), and Ema Grčman (Cello). John Knowles Paine Concert Hall, Harvard University, April 2025

The *Cat in loop* project begins with a layered process of shared vocabulary generation. Through directed acting, listening, playing, and imagining, the ensemble and the AI model collaboratively constructs an embodied CAT figure that is fictional yet deeply connected to lived experience, instrumental identity, and affect. This imaginative CAT is informed by multiple layers of timbre-oriented compositions:

1. An initial curated dataset combining field recordings of real cat vocalizations, cat-inspired synthetic sounds, and composed musical fragments.
2. Iterative Finetuning of and Interaction With the AI Model, Vampnet.
3. A series of guided rehearsals and improvisations in which musicians develop personalized CAT vocabularies grounded in their embodied relationships with their instruments, the ensemble and their personal experience with a CAT. Using this vocabulary, musicians improvise with the AI system, probing its boundaries and responsiveness.
4. The model is subsequently retrained on a hybrid dataset combining musician-generated and composer-curated materials, reinforcing timbre as a negotiated and co-evolving space. A score is developed to guide and facilitate dialogue between all parties.
5. Notation is introduced in the later rehearsal stage to document logistics for the final performance and to help the performers retain ideas that emerged during the composition-improvisation process. As shown in fig 2 below, it takes a visual and metaphorical form, documenting relational and contextual information rather than symbolic content.

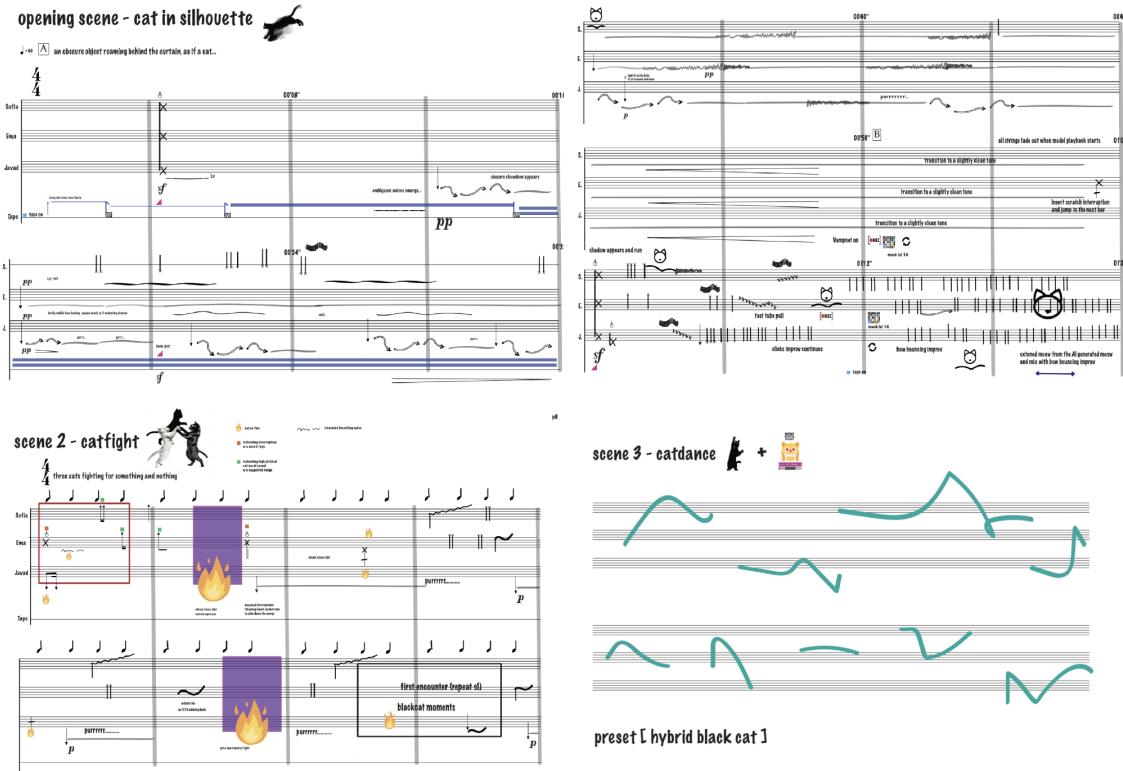


Fig 2. notation examples

The collaboration process unsettles the conventional models of labor division and hierarchical authorship in concert production. What emerges instead is a collaborative ecology in which the CAT, the performers, the neural model, the engineer, and the artist operate as parallel layers with distributed agency. The artist-engineer-performer role division here is not strict but fluid. Although the participants bring specialized knowledge from different perspectives, each also holds competencies that cross into the others' designated domains. The artist holds extensive technical expertise in computer music and system design; the engineer is a practicing vocalist and improviser with substantial experience in performance and music-making; the performers bring deep improvisational practice and a working familiarity with the technical system. The boundary permeability described above is therefore not a definitive condition of the collaboration but a shifting dynamic. As the collaboration evolves, each participant plays an indispensable role in shaping the outcome through what might be called performative listening from a distinct perceptual position.

Cat in Loop proposes an experimental, practice-based approach to AI music research that treats timbre as an open, relational space shaped by collective subjectivity, memory, and play. Within this project, timbre operates as a medium of negotiation between embodied listening and algorithmic perception, mediated through interfaces that shape imagination as much as control. The CAT emerges only through collaboration, as musicians, engineers, and the generative model, mediated by an artist-director, collectively construct a shared CAT figure that develops as a subject-in-progress through a continuous flux of perception, action, and memory.

Acknowledgements

We would like to extend our gratitude towards Harvard Group of New Music (HGNM), Harvard Music Department for facilitating this residency. We thank Sofia von Atzingen, Ema Grčman, Javad Javadzade from the Catinblack Ensemble for embracing this project and helping to see it through. Additionally we would like to thank Hugo Flores Garcia, the creator of Vampnet for his help with technical set up and debugging sessions. We are also grateful to Prof. Anna Huang for her support and encouragement.

References

- Dolan, E. I., & Rehding, A. (2018). *Timbre: Alternative Histories and Possible Futures for the Study of Music*.
- Evans, Z., Carr, C. J., Taylor, J., Hawley, S. H., & Pons, J. (2024, February). Fast timing-conditioned latent audio diffusion. In *the International Conference on Machine Learning*.
- Garcia, H. F., Seetharaman, P., Kumar, R., & Pardo, B. (2023). Vampnet: Music generation via masked acoustic token modeling. International Society For Music Information Retrieval, Milan, Italy.
- Hayles, N. K. (2017). *Unthought: The Power of the Cognitive Nonconscious*. University of Chicago Press.
- Morreale, F., Martinez-Ramirez, M. A., Masu, R., Liao, W., & Mitsufuji, Y. (2025). Reductive, exclusionary, normalising: the limits of generative AI music. *Transactions of the International Society for Music Information Retrieval*, 8(1).
- Morrison, L. (2024). Timbre space: On the flat history of a multidimensional metaphor. *Music & Science*, 7, 20592043241268720.
- Rehding, A. (2021). Timbre/Techne. *The Oxford Handbook of Timbre*, 381.
- Rosenboom, D. (1987). Cognitive Modelling and Musical Composition in the Twentieth Century: A Prolegomenon. *Perspectives of New Music*, 439-446.
- Lyria Team, L., Caillon, A., McWilliams, B., Tarakajian, C., Simon, I., Manco, I., ... & Roberts, A. (2025). Live music models. *arXiv preprint arXiv:2508.04651*.
- Wessel, D. L. (1979). Timbre space as a musical control structure. *Computer music journal*, 45-52.

Timbre at the Interface: Negotiating Audio Representations in the Age of Artificial Creativity,

panel led by Charalampos Saitis with Kai Siedenburg, Landon Morrison, Dominic Thibault, Hee Seng Kye, Weilu Ge, Nithya Shikarpur, Steven Lewis

Session 7

Timbre and Sound Spatialization

Composing Timbre in Space: Spatialisation as a Timbral Parameter in Higher-Order Ambisonics

James Bagshaw^{1†}

¹Leeds Arts University, Leeds, United Kingdom

[†] Corresponding author: james.bagshaw@leeds-art.ac.uk

1. Introduction

This paper presents spatialisation in Higher-Order Ambisonics (HOA) as a timbral operation rather than a process applied to pre-formed timbral materials. In modern spatial audio practices, timbre and space are treated as related but separable domains: spatial technical processing typically distributes, positions and moves sounds whose timbral identity has already been determined through the processes of composition. When composition begins natively in HOA, however, that distinction becomes blurred, as spatial thinking becomes integral to compositional thought rather than separable from it (Barrett, 2002). Spatial decisions about distance, scale, motion and the construction of sound spaces shape spectral perception, textural identity, gesture, and form from the earliest stages of compositional thought (Smalley, 1997, 2007).

This paper considers timbre from the context of electroacoustic discourse, where space and timbre are already intrinsically linked, and framed as concurrent yet unequal parameters of composition (Nyström, 2011). The paper then describes three spatial-timbral strategies drawn from a portfolio of HOA compositions, analysed through the author's Ambisonic Spatial Canvas framework (Bagshaw, 2026), considering the perceptual interrelation of spatial behaviours and timbral materiality. It finally positions the Canvas as a broader apparatus for articulating timbral behaviour, reflecting on perceptual evaluation and pedagogical implications.

2. Spatialisation as a Timbral Operation

Smalley's (1997, 2007) seminal works on spectromorphology and space-form treat spatial behaviour as integral to the perception of sound as gesture and texture, and the related notion of source bonding implies that timbral identity is interlinked with the spatial cues that locate sound in real and imagined environments. Nyström (2011) extends this thinking in arguing that the composer's relationship to space is itself an integral component of compositional thought, rather than a distinct layer of compositional or technical process. Barrett (2019) extends this into native HOA practice, replacing the dominant 'point in space' paradigm with an 'image in space' approach, grounded in Schaeffer's concept of allure – the timbral contour of a sound object. Normandeau's (2009) technique of timbre spatialisation offers the most direct precedent, in the distribution of sound's spectral components across space becomes an extension of composition. This paper proposes something more specific, and complementary: rather than mapping a fixed timbre onto space, it treats spatial operations themselves (i.e. distance, motion, diffusion) as timbral ones. In context, conventional timbral-design operations have direct spatial counterparts that produce perceptually equivalent transformations in modern spatial music.

Three technical and compositional pairings make this link. Filtering, in spectral terms, removes or attenuates frequency content; spatially, distance and the encoding of distance cues attenuate higher-frequency content, altering the perceived spectral envelope and shifting the apparent materiality of a sound without any spectral edit being made (Moylan, 2012). Spectral density, in sound design, is shaped through layering and processing; in HOA it is shaped through the distribution and decorrelation of sources across the field and their subsequent spatial relationships, producing composite spatial textures whose perceptual identity exceeds that of any individual sound source (Nyström, 2011). Lastly, dynamic

contrast has a spatial counterpart in the contraction and expansion of the sound field itself: the same materials can be heard as present, egocentric, and weighty when focused – and conversely airy and dissipated when diffused – with no core change to their spectral content (Smalley, 2007).

HOA is a particularly suitable technical environment for this approach, permitting sound field authoring rather than channel-based or speaker-specific content placement. This loudspeaker-agnostic framework, together with scalable spatial resolution through increasing Ambisonic order and the full-sphere representation of the sound field, permits space to be treated as a continuous and malleable musical parameter, rather than a set of discrete placement considerations (Zotter & Frank, 2019). Timbral identity can therefore be shaped through how sound occupies and moves through the sound field.

3. The Ambisonic Spatial Canvas: Spatial-Timbral Strategies in Practice

The strategies that follow are presented through the Ambisonic Spatial Canvas (Bagshaw, 2025), a framework developed from a wider portfolio of HOA compositions. The Canvas defines Ambisonic space in terminology derived from electroacoustic language – Immediate Space, External Space and Nested Spatial Environments – and provides a vocabulary for spatial density, motion trajectories and degrees of localisation within a three-dimensional sound field. Adapted from acousmatic graphical scoring traditions (Barrett, 2002; Blackburn, 2011), it functions both as a sketching tool during composition and as a reflective analytical tool afterwards. Its purpose here is narrower than its full methodological role: to discuss timbral behaviour spatially, so that decisions about distance, scale, density and motion can be discussed using a shared vocabulary. The strategies illustrated are drawn from this portfolio, including *Snow* (2017), *Helisynths* (2018), *Machine* (2021), and *Revolutions* (2021), with the visual representations adapted from analyses of those works.

Strategy A: Focused versus diffuse spaces. Diffusion and decorrelation transform timbral identity in ways that filtering or spectral processing alone cannot reproduce. A focused source (shown left) presents a concentrated spectral image with clear localisation; the same source distributed and decorrelated across the field (right) becomes spectrally re-textured, fundamentally altering its timbral makeup and envelopment characteristics. See *audio example A* taken from *Snow* (2017), whereby a sound field focuses by folding centrally, then diffuses across external Ambisonic space.

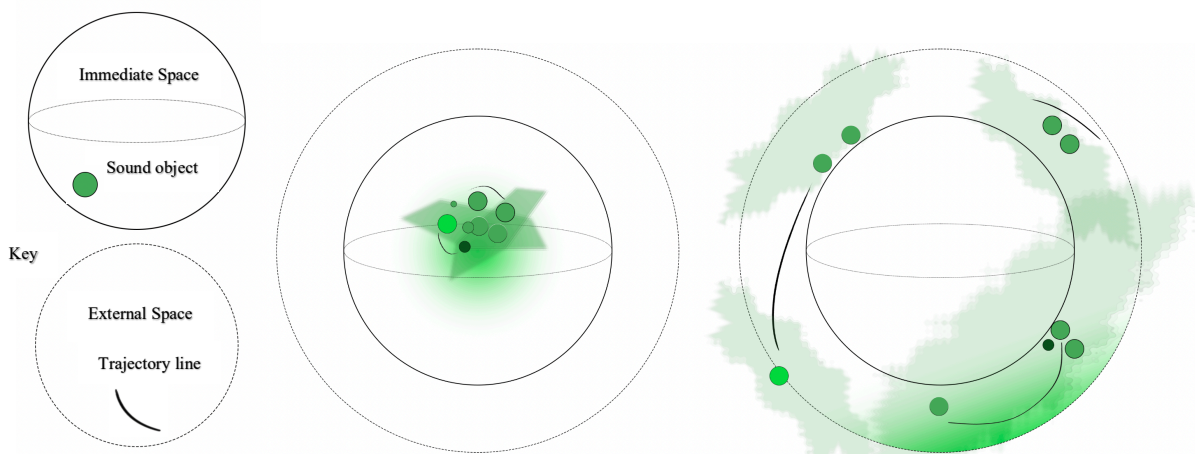


Figure 1: Visual representation of focused space (left) versus diffuse space (right).

Strategy B: Pointillistic versus textural spatialisation. The choice between localised, granular spatial events and broad-brush textural distribution is itself a timbral decision. Pointillistic spatialisation

constructs a series of discrete spatial identities through individual sound events and the spectral detail of each; textural spatialisation produces a composite timbre whose identity forms from the combined behaviour of multiple distributed sources (Hagan, 2008). The transition between these modes is a timbral transformation, comparable compositionally to a shift in spectral content, producing spatial contrast. See *audio examples B and C*, taken from *Machine* (2021), showing both pointillistic and textural composition approaches employed within the same piece.

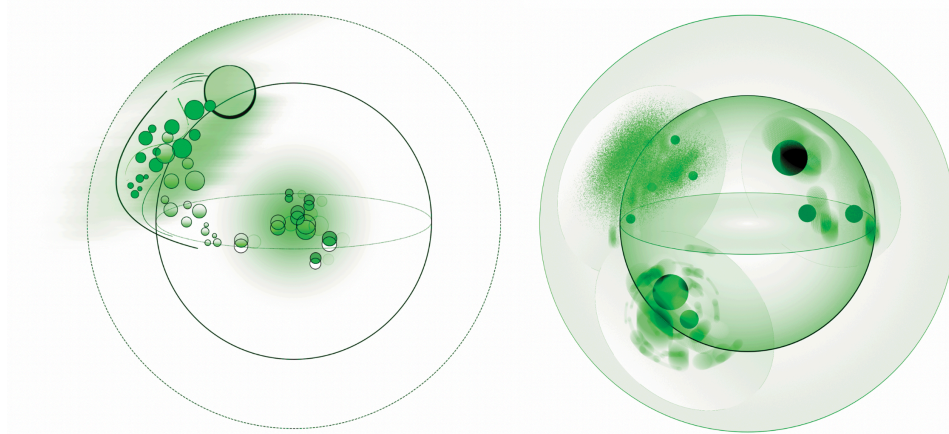


Figure 2: Visual representation of pointillistic (left) versus textural (right) spatialisation.

Strategy C: Spatial motion trajectories as timbral gestures. Motion within the sound field can constitute timbral identity rather than modulate it. For example, a spiralling, accelerating or decelerating trajectory contributes spectral and textural shape that is inseparable from the sound's identity at any given moment; the gesture is the timbre. Smalley's (1997) account of motion in spectromorphology grounds this, but composing natively in HOA brings such gestures into the compositional foreground, rather than treating them as spatial transformations applied to fixed materials. See *audio examples D and E*, taken from *Helisynths* (2018) and *Revolutions* (2021), as examples of localisable and textural trajectories.



Figure 3: Visual representation of localisable trajectory (left) versus textural trajectory (right).

4. Perceptual Grounding, Implications and Future Directions

The strategies above rest on a perceptual claim: spatial behaviour changes the apparent materiality of sound. Perceived distance and envelopment transform sonic qualities such as weight, presence and

granularity in ways listeners understand as timbral, even when no spectral edit has been made. Moylan (2012) describes distance perception as compositionally significant for this reason; Chagas and Fontaine (2024) extend this into HOA territory, arguing that timbral diffusion across 3D space produces sound structures whose perceptual identity is shaped by spatial behaviour rather than solely spectral content.

The argument is currently framed by analogy, drawing upon the composer's reflective listening and established perceptual literature rather than on listener-based testing of the relationships proposed. The framework's purpose is to treat timbre and space as collaborative parameters in composition, exploring technical processing and aesthetic intent simultaneously. The communication of spatial ideas is often divided between technical instruction (encoding, decoding, rendering) and compositional development (form, gesture, material). A spatial-timbral framing bridges this division: Ambisonic fluency is not a pre-requisite, and the ideas explored within HOA act as a domain of timbral thought rather than technical reproduction methodology. The Canvas vocabulary is portable in this sense, as it visually describes what is heard and composed, not what is computed. Treating timbre as a dynamic spatial construct rather than a fixed instrumental or sound design decision opens compositional territory that conventional studio practice tends to leave unexplored. In HOA, what space does to a sound is what the sound becomes; the spatial and timbral are not separable domains, but a single compositional surface.

Supplementary Materials

Audio examples can be accessed at: <https://doi.org/10.6084/m9.figshare.32301519>

References

- Bagshaw, J. (2025). The Ambisonic Canvas: A Portfolio of Spatial Music Compositions in the Ambisonic Domain. *Doctoral thesis*, University of Hull.
- Bagshaw, J. (2026). The Ambisonic Spatial Canvas: A practice-as-research framework for spatial music in Higher-Order Ambisonics. (*Forthcoming*) *Proceedings of the Spatial Audio Gathering 2026*, Leeds.
- Barrett, N. (2002). Spatio-musical composition strategies. *Organised Sound*, 7(3), 313–323.
- Barrett, N. (2019). Composing Images in Space: Schaeffer's Allure Projected in Higher-Order Ambisonics. *Proceedings of the International Computer Music Conference (ICMC)*, Dublin.
- Blackburn, M. (2011). The Visual Sound-Shapes of Spectromorphology: An illustrative guide to composition. *Organised Sound*, 16(1), 5–13.
- Chagas, P.C. & Fontaine, K. (2024). Integrating Higher Order Ambisonics and Cognitive Assemblages in Immersive Sound Composition. *Proceedings of the International Computer Music Conference (ICMC)*, Seoul, South Korea.
- Hagan, K. L. (2008). Textural composition and its space. In *Proceedings of the 5th Sound and Music Computing Conference (SMC)*, Berlin.
- Moylan, W. (2012). Considering space in recorded music. In *The Art of Record Production: An Introductory Reader for a New Academic Field* (1st ed., pp. 136–162). Routledge.
- Normandeau, R. (2009). Timbre Spatialisation: The medium is the space. *Organised Sound*, 14(3), 277–285.
- Nyström, E. (2011). Textons and the Propagation of Space in Acousmatic Music. *Organised Sound*, 16(1), 14–26.
- Smalley, D. (1997). Spectromorphology: Explaining sound-shapes. *Organised Sound*, 2(2), 107–126.
- Smalley, D. (2007). Space-form and the acousmatic image. *Organised Sound*, 12(1), 35–58.
- Zotter, F., & Frank, M. (2019). *Ambisonics: A practical 3D audio theory for recording, studio production, sound reinforcement, and virtual reality*. Springer.

Resonance and Depth: Composing Timbral Identity and Spatial Form in an Artist-in-the-Loop Co-Creative Multi-Agent Percussive Performance

Keon Ju Maverick Lee

School of Interactive Arts and Technology, Simon Fraser University, Vancouver, British Columbia, Canada

Corresponding author: keon_maverick@sfu.ca

This research-creation [1] artistic report presents *Resonance and Depth*, a 9-minute immersive electroacoustic/acousmatic [2] performance that treats timbre and space as primary compositional materials. Built from live electronic percussion and a personal corpus recorded, curated, and composed by the author, including metals, skins, frictional noise, resonant objects, and chromatic percussive materials, the work investigates how timbre-as-structure can emerge through real-time recombination rather than conventional harmonic development. Rather than positioning rhythm, pitch, or harmony as the main carriers of form, *Resonance and Depth* foregrounds timbral morphology, spatial movement, and performer–system interaction.

Situated within a research-creation methodology [1], the work develops artistic practice, technical experimentation, and theoretical reflection together. The performance is not simply a demonstration of a technological system, but an inquiry into how performer identity can be retained, transformed, and redistributed through co-creative machine processes. The personal corpus functions as both musical memory and compositional material: it carries traces of the performer’s instrumental vocabulary while becoming available for algorithmic recombination, fragmentation, and spatial reconfiguration.

The piece is realised as a musical multi-agent society/system [3,4] designed for artist-in-the-loop music improvisation in a co-creative space [5]. At the macro scale, an evolving generative ambient layer shapes atmosphere, density, and long-form spectral drift. At the micro scale, the performer’s rhythmic gestures—short transients, rolls, frictional attacks, and resonance tails—drive a call-and-response relationship with the agents, producing tension–release passages across the performance. The agents reshape the performer’s material into evolving textures, while the performer responds to the changing timbral field, timing, density, and spatial behaviour.

The performance workflow is coordinated by a conductor agent that supports structured improvisation across the piece. It monitors scene-level changes such as density, energy, texture, musical tension, and agent states, then automates behavioural shifts across the system. This enables musical acts to emerge dynamically while preserving an overall formal arc. The workflow can be summarised as follows:

- **Human performer / Drumbeam–multipad percussion interface:** Provides embodied rhythmic input through electronic percussion while extending MIDI control from discrete note triggering to continuous gestural control of timbre, morphology, and instrumental identity.
- **MACAT:** Recombines the author’s personal corpus as a timbral memory space.
- **MACataRT:** Generates audio-mosaicing responses from the live stream and corpus.

- **MIDI-based agent:** Translates performance activity into symbolic output using audio-to-MIDI transcription for machine listening and Karplus–Strong sound synthesis.
- **Spatialisation layer:** Distributes sound objects through trajectory, diffusion, depth, width, and apparent-size controls.
- **Conductor agent:** Coordinates scene changes and automates agent behaviours for the structured music improvisation, enabling transitions between sparse ambience, dense call-and-response, close mirroring, and offers more autonomy for the human performer during the performance.

Agent 1, MACAT [6], acts as the primary timbral recombination engine. It traverses a personal corpus as a timbre space, sequencing fragments into emergent phrases that remain connected to the performer’s instrumental identity while being algorithmically recontextualised. A synthesis layer, including resampling, reversal, pitch transformations, and related sound-design processes, further reshapes retrieved materials into malleable sound objects, sustaining a continuity between instrumental causality and acousmatic ambiguity.

Agent 2, MACataRT [6], introduces a musique-concrète counter-voice through audio mosaicing [7] for creating reactive music accompaniment. It listens to the live stream and reassembles corpus fragments by feature similarity, producing doubles, shadows, echoes, and swarms around the performer and another agent’s audio inputs. This agent can move from close mirroring to more autonomous textural divergence, creating an intermediate space between accompaniment and autonomous agency.

Agent 3, a MIDI-based agent presented in the musical agents tutorial [8], listens to the audio agents through audio-to-MIDI transcription and/or direct MIDI input. By translating audio-agent activity and live percussive input into symbolic musical information, this agent introduces a MIDI-domain layer of mediation between sound, gesture, and system response. Its MIDI output is rendered through MIDI-based percussion and a Karplus–Strong string synthesiser [9], producing a distinctive timbral palette that contrasts with, while also resonating against, the metallic, skin-based, and frictional materials of the original corpus. This symbolic-to-synthetic layer explores how percussive and acousmatic materials can be reinterpreted through symbolic control, physically inspired synthesis, and MIDI-driven timbral articulation, a topic that remains relatively underexplored compared with audio-domain approaches.

This MIDI-based layer also reframes spatial agency in relation to recent audio-domain work on spatial awareness and influence in co-creative agents. Gatti and Fiorini [10] propose an automated corpus-spatialisation system for Somax2 in which audio descriptors, transient behaviour, and agent roles are mapped to spatial trajectories and inter-agent influence. In *Resonance and Depth*, a related idea is translated into the MIDI domain: notes, velocity, timing, CC/MPE data, and percussive control streams operate not only as triggers, but also as spatial–timbral control signals. This extends spatial influence from audio-feature-driven spatialisation toward symbolic, performative control, where MIDI becomes a medium for composing timbral identity and spatial form.

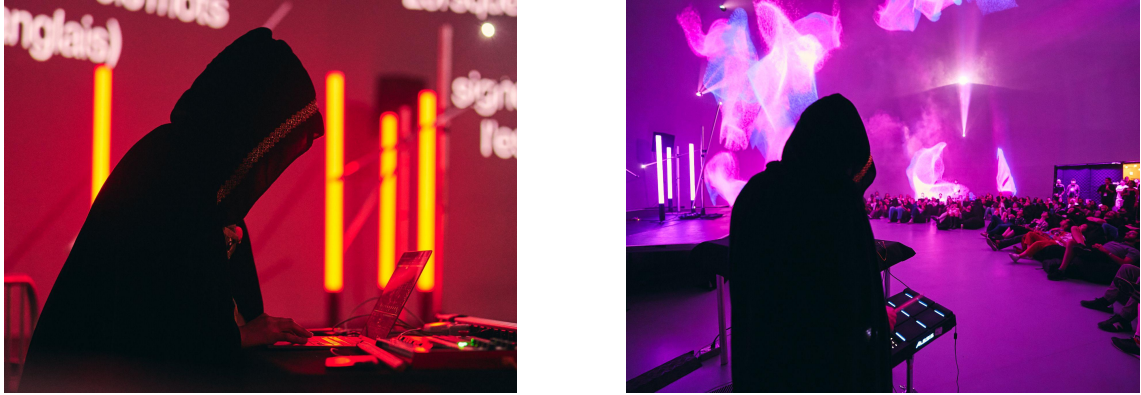


Figure 1: Documentation of 93.5 mix immersive dome performance at the société des arts technologiques during the MUTEK Montréal Festival, 2025. Photo credit: Bruno Aiello-Destombes.

Spatialisation is treated as composition rather than post-mix, following Normandeau's argument that space can function as a primary musical parameter in acousmatic music and that timbre itself can be distributed across loudspeakers as "timbre spatialisation" [11]. This approach extends the author's earlier sound spatialisation practice developed during an artist residency at the Société des arts technologiques (SAT), where immersive diffusion and spatial motion were explored as compositional materials within dome-based immersive auditory environments. In *Resonance and Depth*, this framework is expanded through musical agents and an object-based 3D approach [12], where each sound source is authored with evolving position, trajectory, depth, width, apparent size, and diffusion. The original corpus is designed for the 93.5-loudspeaker dome, where stochastic motion and diffusion trajectories act as timbral gestures: distance, orbit, elevation, and spatial density become parameters of phrasing and form.

Here, spatialisation is part of a dynamic co-creative ecology in which human performance, generative systems, concatenative processes, and spatial behaviours influence one another in real-time. The incorporation of musical agents transforms the immersive environment into an adaptive performance ecosystem where timbral recombination, mosaicing, and spatial trajectories become active participants in the musical discourse. For portability, the same object scene can be re-imagined across formats; this performance version is provided as stereo/binaural documentation, with an optional 5.1 theatre art mix or spatial dome mix if an audio system is available.

The Drumbeam instrument integrated with a percussive multipad supports the exploration of timbral identity in the MIDI domain by extending percussive gesture beyond discrete note triggering into continuous, multidimensional control. Parameters such as strike intensity, spatial position, and gestural variation can be mapped to MIDI CC or MPE data, allowing the performer to shape timbre as an expressive continuum. Within the broader system, rhythmic performance becomes a practice of timbral articulation, morphological variation, and instrumental redefinition within symbolic control systems.

For the Timbre conference, the work contributes three main perspectives. First, it proposes a practice-based method for crafting authored corpora that retain performer identity within concatenative sound systems [13]. Second, it presents a model of co-creative timbral agency in which recombination, mosaicing, and symbolic mediation act as improvisational partners rather than as purely automatic generation processes. Third, it offers an immersive case study of timbre spatialisation, showing how

psychoacoustic attributes can be composed alongside instrumental gesture to produce musical form. Through these contributions, *Resonance and Depth* frames timbre as a generative, spatial, and relational force through which performer, machine, and listening environment co-produce the work.

References

- [1] Stévanec, S., & Lacasse, S. (2017). *Research-creation in music and the arts: Towards a collaborative interdiscipline*. Routledge.
- [2] Waters, S. (2016). Beyond the acousmatic: Hybrid tendencies in electroacoustic music. In *Music, electronic media and culture* (pp. 56–84). Routledge.
- [3] Eigenfeldt, A., & Pasquier, P. (2012). Creative agents, curatorial agents, and human-agent interaction in *Coming Together*. In *Proceedings of Sound and Music Computing* (pp. 181–186).
- [4] Tatar, K., & Pasquier, P. (2019). Musical agents: A typology and state of the art towards musical metacreation. *Journal of New Music Research*, 48(1), 56–105.
- [5] Thelle, N. J. W., & Wærstad, B. I. G. (2023, June). Co-creative spaces: The machine as a collaborator. In *Proceedings of the International Conference on New Interfaces for Musical Expression*. Universidad Autónoma Metropolitana.
- [6] Lee, K. J. M., & Pasquier, P. (2024, December). Musical agent systems: MACAT and MACataRT. In *Proceedings of the Creativity and Generative AI Workshop at NeurIPS 2024*.
- [7] Lazier, A., & Cook, P. (2003, September). MOSIEVIUS: Feature-driven interactive audio mosaicing. In *Proceedings of the Digital Audio Effects (DAFx) Conference*.
- [8] Lee, K. J. M., Pasquier, P., Nika, J., Bujard, B., Eigenfeldt, A., Thelle, N. J., ... & Bown, O. (2025, September). Musical agents tutorial. In *Proceedings of the AI Music Creativity Conference*.
- [9] Karplus, K., & Strong, A. (1983). Digital synthesis of plucked-string and drum timbres. *Computer Music Journal*, 7(2), 43–55.
- [10] Gatti, A. M., & Fiorini, M. (2025, July). Spatial awareness and influence for co-creative agents: Introducing automated corpus spatialization in Somax2. In *Proceedings of the Sound and Music Computing Conference (SMC 2025)*.
- [11] Normandeau, R. (2009). Timbre spatialisation: The medium is the space. *Organised Sound*, 14(3), 277–285. <https://doi.org/10.1017/S1355771809990094>
- [12] Normandeau, R., Bélanger, O., Lengelé, C., & Ledoux, D. (2018, August). SPATGRIS/SERVERGRIS, creative tools for 2D and 3D sound spatialization. In *Proceedings of the International Computer Music Conference*.
- [13] Schwarz, D. (2006). Concatenative sound synthesis: The early years. *Journal of New Music Research*, 35(1), 3–22.

Timbre Spatialisation: The Medium Is Space

Robert Normandeau^{1†}

¹ Université de Montréal, Québec, Canada

[†] Corresponding author: robert.normandeau@umontreal.ca

Space is presented here as a musical parameter that can now be considered just as important in the composition of acousmatic music as the other musical parameters found in instrumental music. Certain aspects of sound spatialisation can be considered unique to the acousmatic language. The sound spatialisation presented here is an immersive spatialisation, in which listeners are placed in a setting where loudspeakers surround them. I present my cycle of work and my research, in which I have developed a new way of composing space that I call *Timbre Spatialization*.

Timbre Spatialization

With instrumental music of the 1960s, composers explored spatialisation, creating works that assigned performers to different locations within the concert hall (Stockhausen's *Gruppen* or *Carré*, Xenakis's *Persephassa*, among others). However, these works are limited to the timbre of the instruments: a violin on the left side will always sound like a violin on the left side. Sound and the source of projection are linked. The distinctive feature of the acousmatic medium is its virtuality: the sound and the source of projection are not linked.

A loudspeaker can reproduce any type of timbre. Furthermore, today, with the right software, all these sounds can be localized at any point within any group of loudspeakers. What is unique about electroacoustic music is the ability to fragment sound spectra across a network of loudspeakers. When playing a violin, the instrument's entire spectrum is heard, whereas with multichannel electroacoustic music, the timbre can be distributed across all available virtual points within the defined space.

This is what I call timbre spatialisation: the entire spectrum of a sound is recombined only virtually within the concert hall space. Each point represents only a part of the whole. This is not a conception of space that is added at the end of the compositional process—a common approach, especially today with multitrack software—but rather a true composed spatialisation. It is a musical parameter unique to acousmatic music.

The Source: The Cinema Language

In 1992, I submitted the first doctoral dissertation on acousmatic music at the Université de Montreal. It was titled *A Cinema for the Ear*. It wasn't so much the iconic and referential aspects of cinema that interested me at the time, but rather the techniques of composition—that is, cinematic grammar. One of the prevailing ideas at the time was the one that drew an analogy between cinema and electroacoustics. Although this idea had been mentioned many times, no one had developed it systematically.

Theoretical developments surrounding the concept of cinematic language led us to believe that it was at this level that the most significant elements of film studies lay. And while not all of these elements are directly applicable to electroacoustics, the approach that led to their development is nonetheless highly instructive, and it is undoubtedly possible to draw inspiration from it to develop a personal approach of a similar nature.

The Grammar of Cinema

Analyzing the various elements that make up the grammar of cinema is, in fact, an easily accessible and achievable first step in establishing a connection between acousmatic and cinema. The following discussion draws parallels between these two artistic disciplines. Some of them are immediately applicable—and indeed, numerous examples can be found in the chapter devoted to the analysis of works—while others are more a matter of personal interpretation. Our aim was to draw on the many texts devoted to the analysis of cinema and film in order to validate equivalences between the two artistic genres. We identified twelve potential cinematic writing techniques that could be transposed to acousmatic: the shot, the zoom, the tracking shot and the pan, the high-angle and low-angle shots, editing, colour, lighting, and framing.

Clair de terre (1999), the Work Central to My Dissertation

Although my dissertation was submitted in 1992, the work that truly illustrated my argument is *Clair de terre*, created in 1999. Structured around twelve movements that draw on elements of cinematic grammar, I will focus here on just two: Primary and Secondary Colours and Ternary Colours. The main analogy between sound and light in terms of colour, without delving into the details of perceptual analysis, is the spectral relationship. Low frequencies, in both domains, correspond to red and warm tones, whereas, conversely, violet is found in the high frequencies and is associated with cool tones.

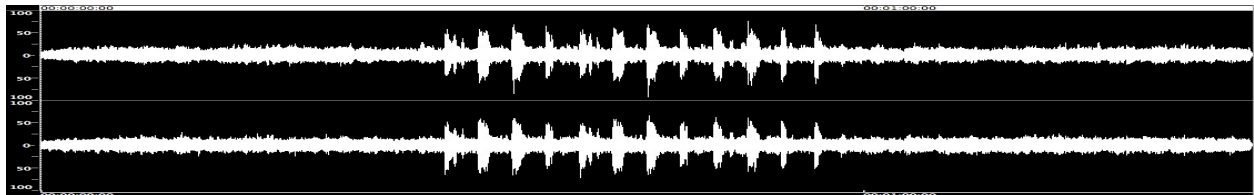
To realize these two movements, I had the idea of using bandpass filters that divided the spectrum into as many bands as there were different “colours” of sound. These filters were recorded on separate tracks and played throughout the space via different speakers. These two movements turned out to be the seeds of what would become a complete cycle of work.

StrinGDberg V2 (2025) 18'

A work commissioned by the GRM in Paris in 2001. Two years after *Clair de terre*, I decided to take this idea of deconstructing complete spectra into smaller units and spatializing them independently a step further. The only sound sources in *StrinGDberg* are a hurdy-gurdy and a cello. The piece is constructed using two superimposed layers. The first layer consists of a single recording of a hurdy-gurdy improvisation lasting about one minute. A second layer, made up of cello sounds, adds rhythm to the work, as well as a strong dramatic dimension at the end.

Sound Processing

While playing the hurdy-gurdy, the musician improvised a three-part sequence: improvisation 1, melody, improvisation 2. I primarily used the middle section to compose the piece. From the central section of this improvisation, I selected twelve “consonants”—the note attacks—and twelve “vowels”—the sustained passages between the attacks.



I had previously composed a cycle of works based on onomatopoeias. And I had explored the concepts of consonants and vowels in depth. Here, I retained only their main characteristics: vowels determine the timbre of the voice, and consonants, its articulation.

The consonants and vowels were then “frozen”—insofar as sound can be frozen—or at least extended, in order to provide sound material that, in the case of consonants, rarely exceeds 100 ms. All twenty-four files were filtered through four dynamic bandpass filters, whose parameters evolved over the course of the sections and over time. The bandwidth of each filter widened over the course of the piece, and the centre frequency constantly varied. This means that the sound is heavily filtered overall at the beginning of the piece and ends at a point where the entire spectrum is unfiltered.

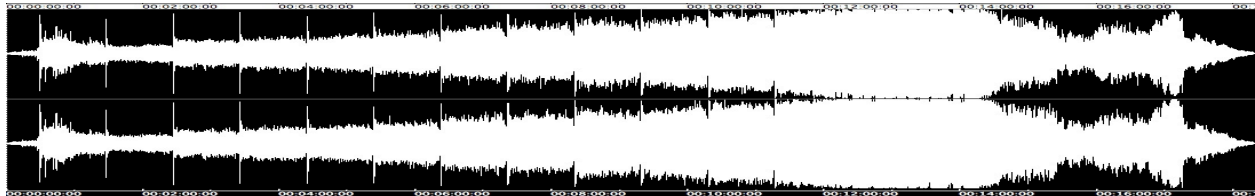


Figure 2: The general structure of *StrinGDberg*

To the eye, it appears to be a long crescendo. But to the ear, this is not quite the case, even though the sound swells throughout the piece. It is, rather, a timbral progression. Starting out sparse and filtered, the timbre becomes more complex and richer as time goes on.

In *StrinGDberg*, the form, duration, rhythms, and proportions are derived from the original improvised melody. All the sound files in the work were created and organized with multichannel playback in mind. This is another characteristic of what I call timbre spatialisation. The various filtered tracks are assigned to different speakers—sixteen in the original version, thirty-eight in the most recent version—and are constantly in motion. The final mix then takes shape within the acoustics of the concert hall and differently for each listener. There is no front-back or left-right orientation.

This resolves the balance issues caused by the listener’s proximity to a specific speaker, as the sounds constantly change and evolve in each speaker. Ideally, the piece is not presented in a conventional concert hall, but in a space where people can move around during the performance.

The work originally had two versions: the premiere at the GRM in Paris in 2001 and the revised version from 2003 premiered in Aarhus (Denmark).

***Bourdons* (2021–25) and *StrinGDberg* V2 (2025)**

When Jean-François Rivest asked me to write a piece for the Université de Montréal Orchestra (OUM), I wanted to adapt one of my pieces from the cycle for this large ensemble. I chose *StrinGDberg*. In *Bourdons*, I transposed this idea of filtering to the orchestra. The musicians are invited to play with as little timbre as possible at the beginning of the piece until they make their instruments scream, squeak, and screech at the climax. For the audience, it is an opportunity to (re)discover the timbre of these instruments, which are otherwise so familiar. The electroacoustic component envelops the orchestra through a dome-shaped device that completely surrounds it. The octobass of the Montreal Symphony Orchestra supports the musicians by playing the low drones.

The process of adapting the work for *Bourdons* led to a revision of *StrinGDberg*. Indeed, this piece was composed in 2001 using tools from that era. By 2025, technology had evolved considerably, and the

resulting sound quality is significantly superior today. The filters from 2001 have been replaced by newer versions that preserve the acoustic qualities of the original instruments. It is this new version that will be presented at the concert.

Perspective

We still have a great deal to explore about space perception, and I think that this research should be made in parallel with music composition. This research could include experiments on how we perceive sound movement in 3D, how we perceive the same musical gesture presented at the front of the audience compared to the back, how we perceive low frequency content from different locations, what the perceptual threshold regarding location angle is and thus, how many speakers we need to feel completely immersed by the sound. In other words, what is relevant in terms of musical perception in an immersive environment?

Tools

In 2009, I founded the Groupe de Recherche en Immersion Spatiale, the GRIS (Research Group in Spatial Immersion). We developed many tools to help composers explore immersive and multichannel works. Among these, the most recent is SpatGRIS, which has reached its fourth version since 2017. We have surpassed 22,000 downloads. Given that this is a highly specialized tool, this can be considered a real success in our field. It is easy to use, free, and open source. We are continuing a series of explorations of sound space, such as Varèse's Philips Pavilion, Stockhausen's Osaka Pavilion, Kupper's Sound Cupolas, the Klangdome in Karlsruhe, and the SARC in Belfast. It is estimated that there are at least twenty domes around the world today dedicated to immersive music.

A Cycle of Works

Following this premiere, I felt compelled to continue in this direction, as the idea of timbral spatialisation seemed so promising to me. I therefore embarked on a cycle in which each new work explores a different aspect of this concept.

- *Éden* (2003). The key difference is that, in *Éden*, many different timbres are layered on top of one another, and there is no progression over time—the entire spectrum is always present. Only the “inner” nature of the sounds, the micro-variations change over time.
- *Palindrome* (2006–09). This work was designed so that the listening experience is the same in both directions, including the channels of the stereo files, the spatialisation structure of the 24 tracks, the levels of the various mixing curves, and the musical phrases.
- *Kuppel* (2006). It is composed of three sound materials recorded in Karlsruhe (Germany) during a stay in the fall of 2005. These include outdoor recordings of the city's various trams, recordings of the Markplatz carillon, and indoor recordings of a creaking door at the ZKM. It's an homage to the city where the work was produced.
- *Anadliad* (2010). The breath is that of the two instruments used in the piece. First, the bagpipes. Then the pibgorn, a Gaelic wind instrument, was recorded during a residency in Bangor in 2008. And finally, the breath of the gods—those who cause the Welsh coast to be swept by intense winds and storms.

- *La part des anges* (The Angels' Share) (2012). The “angels’ share” is that volatile portion of alcohol that ages and slowly evaporates through the walls of the barrel. Here, the material that evaporates consists of a cappella songs from different cultures and historical periods.
- *L'engloutissement* (The Engulfment) (2013–19). The only sound treatment used here was extreme slow motion, and the primary material was water. It is a tribute to the American artist Bill Viola (1951–2024), whose retrospective at the Grand Palais in Paris in 2014 deeply moved me. *L'engloutissement* is an immersive work, intended to be presented in a dome of loudspeakers.
- *Le ravissement* (The Rapture) (2019–21). What is distorted here is time. Time and spectres are stretched here beyond the thresholds of perception. And figuratively speaking, what is reduced, what is enchanted, is the identity of the musical sources used, which come from a Jorane (singer and cellist) performance in which I participated in 2019.
- *Rail* (2019-21). It is not so much a representation of a passing train—so often depicted in electronic music—as it is the auditory perception of a passenger inside the train. Hence the sensation of muffled, dull, rumbling sound material. Recordings of trains from Canada, the United States, and Mexico come together to seal a North American treaty.

References

- Normandeau, R. (2021). *Le ravissement, Tunnel azur, Melancholia, L'engloutissement, Rail, Mélancolie*, empreintes DIGITALes, Canada (IMED-21174).
- Normandeau, R. (2014). *La part des anges, Kuppel, Pluies noires, Hamlet-Machine without Actors Baobabs, Dômes*, empreintes DIGITALes, Canada (IMED-14128).
- Normandeau, R. (2012). *Palimpseste, Murmures, Jeu de langues, Anadliad, Palindrome, Palimpsestes*, empreintes DIGITALes, Canada (IMED-12116).
- Normandeau, R. (2005). *Chorus, Éden, Hamlet-Machine with Actors, Puzzle, StrinGDberg, Puzzles*, empreintes DIGITALes, Canada (IMED-0575).
- Normandeau, R. (2001). *Clair de terre, Malina, Erinyes, Clair de terre*, empreintes DIGITALes, Canada (IMED-0157).
- Normandeau, R. (2009). *Timbre Spatialisation: The medium is the space*. In *Organised Sound*, Vol 14.3 (pp. 277-285).

Session 8

Percussion and Rhythms

Timbre and Rhythmic Parsing: Spectral Centroid and Temporal Envelope Effects on Prominence and Grouping

Joshua Rosner^{1,2†}, Stephen McAdams^{1,2}, and Michael Wagner³

¹ Music Research, McGill University, Montréal, Québec, Canada

² Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT),
Montréal, Québec, Canada

³ Linguistics, McGill University, Montréal, Québec, Canada

[†] Corresponding author: joshua.rosner@mail.mcgill.ca

While empirical investigation of rhythm's interaction with pitch height, intensity, and duration dates back to the turn of the 20th century (Bolton, 1894; Woodrow, 1911), the role of timbre in rhythm remains underexplored. Previous research has demonstrated that rhythm perception is affected by changes in timbre (Wessel, 1979) and identified timbral changes as causes of accentuation (Fraisse, 1982). Singh's (1997) findings suggest that changes in spectral complexity can significantly influence the perception of rhythm and meter, emphasizing the multifaceted role of timbre in shaping meter. While London (2012) conceptualizes meter as an attentional process, the ways in which attention to timbre informs grouping and prominence decisions invite further investigation. Following recent work in linguistics (Moghiseh et al., 2023), we posit that listeners parse both speech and music along two perceptual dimensions: prominence (accent/metrical strength) and grouping (boundary detection) (Lerdahl & Jackendoff, 1983; McAdams, 2019). Here, we test whether two acoustic correlates of timbre—spectral centroid and temporal-envelope shape (attack/decay)—systematically bias prominence and grouping perception in isochronous binary tone sequences, controlling tone duration, pitch, and loudness.

Experiment 1 asked whether spectral centroid, a primary acoustic correlate of perceived brightness (Schubert & Wolfe, 2006), biases rhythmic parsing. We tested six conditions with different spectral centroids of complex harmonic tones (fixed within a sequence) compared to a baseline centroid. Baseline tones were MIDI 60, with an acoustic duration of 125 ms and a 150-ms offset-to-onset inter-stimulus interval (ISI). Brightness was manipulated through synthesis using a patch from Caclin et al. (2005) by varying harmonic rank from a baseline of 1.50 to six higher levels: 1.75, 2.04, 2.38, 2.77, 3.23, and 3.77, corresponding to spectral centroids from 466.98 to 993.02 Hz. Because increasing spectral centroid can affect perceived loudness, brighter tones were loudness-matched using an algorithm from Wood (2015). For each condition, we created two alternating sequences comprising nine baseline tones and nine manipulated tones at a different spectral centroid from the baseline; the two sequences differed only in concatenation order, concluding either with a baseline tone or with a manipulated tone. 126 online participants (96 passed headphone screening) heard 12 sequences in total (6 conditions $\times$ 2 orders). Participants were recruited via Prolific and restricted to native English speakers born and currently residing in Canada or the United States, with approval ratings of 98% or higher. Adapting Wagner's (2022) paradigm, participants answered two forced-choice questions about the final sound of each sequence: (1) was the final tone first or last in a pair (grouping/finality), and (2) was the final tone more or less prominent. Question order was counterbalanced across participants. The broader experiment also included duration, loudness, and pitch-height manipulations as benchmark cues, but the present report focuses solely on brightness.

Experiment 2 probed the role of temporal-envelope cues on rhythmic parsing. We synthesized nine unique tones by crossing three attack times and three decay times: 10, 45, 80 ms. Effective duration (i.e., the amount of time the energy envelope is above -20 dBFS, see Kazazis et al., 2022) was held constant at 175 ms, with a perceptual ISI of 200 ms; acoustic durations varied to maintain equivalent effective duration. Participants heard 81 sequences formed from all binary combinations of the nine tones. Both

experiments controlled for overall effective duration and perceived loudness. 100 online participants (65 passed headphone screening) completed the same grouping and prominence judgments as in Experiment 1, with question order counterbalanced across participants. Participants were recruited using the same Prolific restrictions as in Experiment 1. Primary analyses used the full sample to preserve the balanced question-order design. Prominence and grouping/finality responses were analyzed using binomial mixed-effects logistic regression models with directional relative attack, directional relative decay, question order, and their interactions as fixed effects, and participant as a random intercept. Relative attack and relative decay were coded directionally: -1 when the final tone had a shorter attack/decay than the preceding tone, 0 when the two tones were equal, and 1 when the final tone had a longer attack/decay.

In Experiment 1, final tones carrying the brightness manipulation were more likely to be judged “more prominent” than baseline final tones ($\beta = 0.78$; $SE = 0.13$; $p < .001$) and were also more likely to be judged “last” in a pair ($\beta = 0.79$; $SE = 0.13$; $p < .001$) (see Figures 1–2).

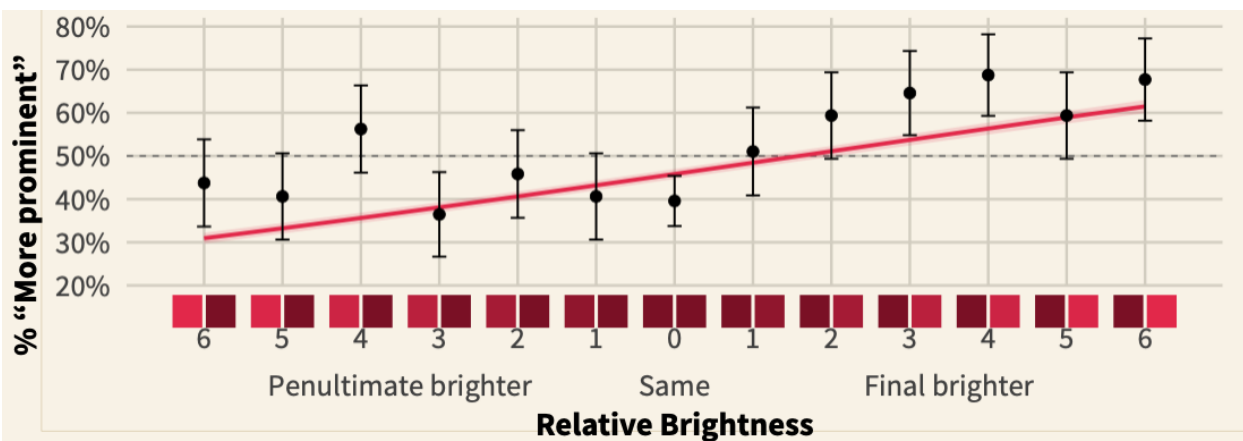


Figure 1: Proportion of Experiment 1 trials in which participants judged the final tone as “more prominent,” plotted by relative brightness. Negative values indicate that the penultimate tone was brighter than the final tone, 0 indicates equal brightness, and positive values indicate that the final tone was brighter. Error bars show 95% binomial confidence intervals; the line shows model-predicted values.

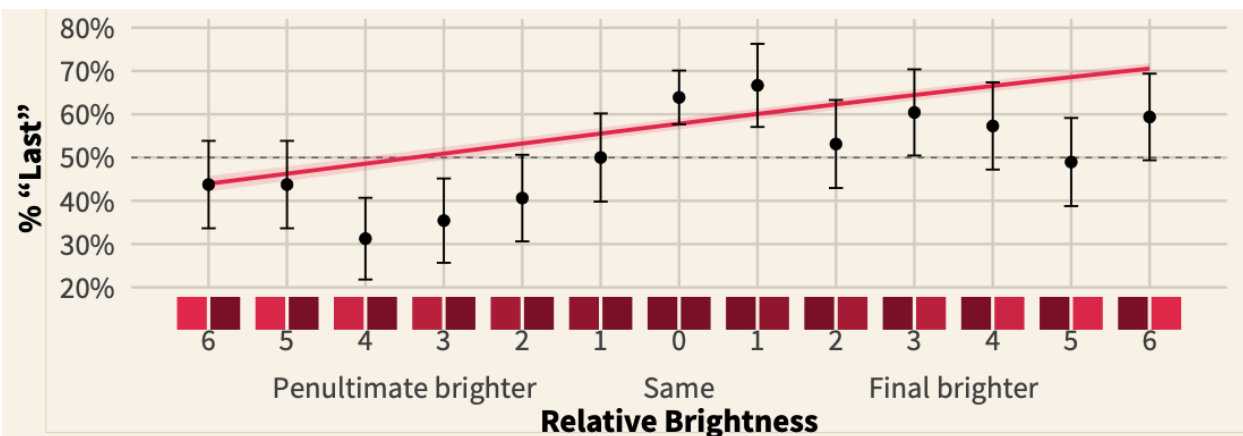


Figure 2: Proportion of Experiment 1 trials in which participants judged the final tone as “last,” plotted by relative brightness. Negative values indicate that the penultimate tone was brighter than the final tone, 0 indicates equal brightness, and positive values indicate that the final tone was brighter. Error bars show 95% binomial confidence intervals; the line shows model predicted values.

In Experiment 2, shorter relative attack increased the probability that the final tone was judged “more prominent” ($\beta = -0.40$; $SE = 0.12$; $p < .001$) (Figure 3). Relative decay did not independently predict prominence. For grouping, longer relative decay conditionally cued finality: it increased the probability that the final tone was judged “last” when the final tone did not also have a longer attack than the preceding tone ($\beta = 0.15$; $SE = 0.07$; $p = .035$), but not when the final tone had a longer attack.

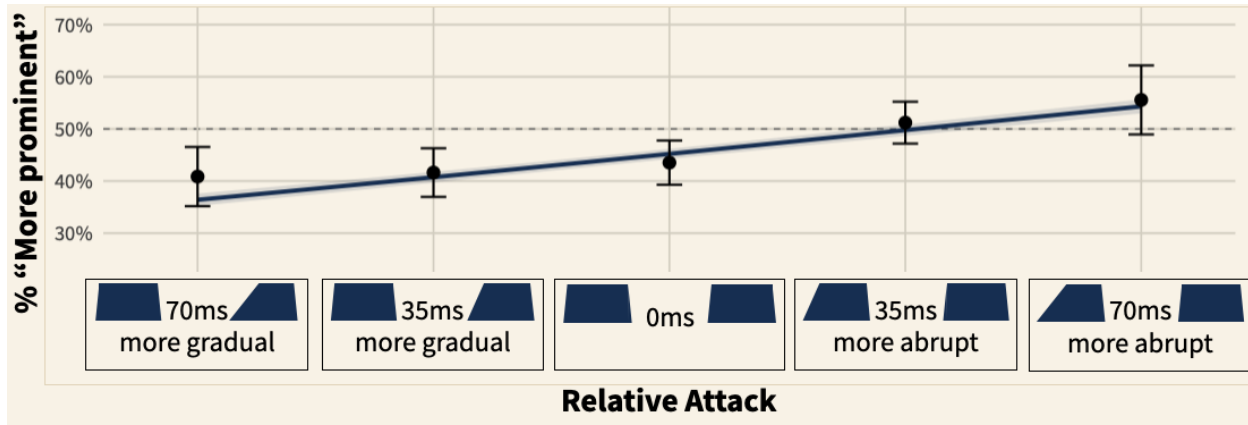


Figure 3: Proportion of Experiment 2 trials in which participants judged the final tone as “more prominent,” plotted by relative attack. Relative attack was coded directionally: More gradual indicates that the final tone had a longer attack than the preceding tone, 0 ms indicates equal attack times, and more abrupt indicates that the final tone had a shorter attack. Error bars show 95% binomial confidence intervals; the line shows model-predicted values.

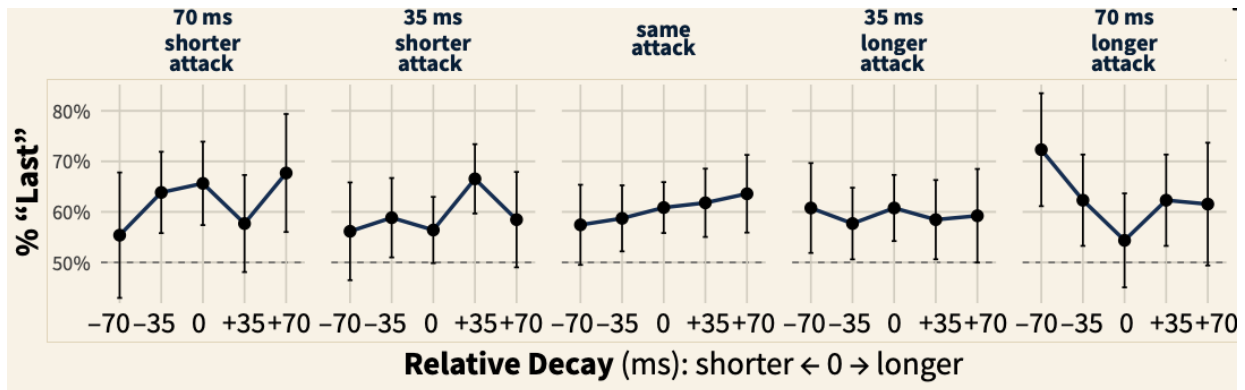


Figure 4: Proportion of Experiment 2 trials in which participants judged the final tone as “last,” plotted by relative decay and separated by relative attack. Relative decay was coded directionally: Negative values indicate that the final tone had a shorter decay than the preceding tone, 0 indicates equal decay times, and positive values indicate that the final tone had a longer decay. Error bars show 95% binomial confidence intervals.

Diagnostic analyses showed that grouping and prominence decisions were not explained by small timing discrepancies used to achieve perceptual isochrony. Contrary to predictions based on the literature, acoustically longer final tones were less likely to be judged “more prominent” ($\beta = -0.36$; $SE = 0.10$; $p < .001$), suggesting that effective duration, and perhaps perceptual duration more broadly, may be more meaningful than acoustic duration for rhythmic decisions.

Our results show that the acoustic cues of attack, decay, and spectral centroid are recruited for distinct rhythmic percepts: Higher spectral centroid served as a cue to perceived prominence and perceived

finality, and faster attacks served as a cue to prominence, consistent with evidence that popular music lyrics place brighter vowels on beats 2 and 4 (Ammirante and Rovetti, 2021), suggesting a timbral mechanism that can reinforce backbeat accents (beats 2 and 4 in 4/4), typically articulated by the timbrally sharper snare drum (Biamonte, 2014). By contrast, longer decays provide a conditional cue to finality: participants used decay as a grouping cue only when the final tone did not have a longer attack, consistent with evidence that listeners rely more on onsets than on decay transients in timbre-based judgments such as instrument identification (Clark et al., 1963). Our results support the view that rhythm perception reflects the recovery of at least two partially separable perceptual dimensions—grouping and prominence—rather than a single “foot type” percept (Wagner, 2022). While supplementing previous findings, our research also invites the consideration of sounds as multi-dimensional percepts, rather than reducing cueing to unitary dimensions like “duration” or “loudness” in the study of speech and music segmentation.

References

- Ammirante, P., & Rovetti, J. (2021). Bright vowels are favoured on weak beats in popular music lyrics. *Journal of New Music Research*, 50(3), 259–265. <https://doi.org/10.1080/09298215.2021.1936076>
- Biamonte, N. (2014). Formal functions of metric dissonance in rock music. *Music Theory Online*, 20(2). <https://doi.org/10.30535/mto.20.2.1>
- Bolton, T. L. (1894). Rhythm. *The American Journal of Psychology*, 6(2), 145–238. <https://doi.org/10.2307/1410948>
- Caclin, A., McAdams, S., Smith, B. K., & Winsberg, S. (2005). Acoustic correlates of timbre space dimensions: A confirmatory study using synthetic tones. *The Journal of the Acoustical Society of America*, 118(1), 471–482. <https://doi.org/10.1121/1.1929229>
- Clark, M., Jr., Luce, D., Abrams, R., Schlossberg, H., & Rome, J. (1963). Preliminary experiments on the aural significance of parts of tones of orchestral instruments and on choral tones. *Journal of the Audio Engineering Society*, 11(1), 45–54.
- Fraisse, P. (1982). Rhythm and tempo. In D. Deutsch (Ed.), *The psychology of music* (pp. 149–180). Academic Press.
- Kazazis, S., Depalle, P., & McAdams, S. (2022). “The Timbre Toolbox User’s Manual”, <https://github.com/MPCL-McGill/TimbreToolbox-R2021a> (Last viewed 2026/05/12).
- Lerdahl, F., & Jackendoff, R. S. (1983). *A generative theory of tonal music*. MIT Press.
- London, J. (2012). *Hearing in time: Psychological aspects of musical meter* (2nd ed.). 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. Fay (Eds.), *Timbre: Acoustics, perception, and cognition* (pp. 211–243). Springer. https://doi.org/10.1007/978-3-030-14832-4_8
- Moghiseh, E., Sonderegger, M., & Wagner, M. (2023). The iambic-trochaic law without iambs or trochees: Parsing speech for grouping and prominence. *The Journal of the Acoustical Society of America*, 153(2), 1108–1129. <https://doi.org/10.1121/10.0017170>
- Schubert, E., & Wolfe, J. (2006). Does timbral brightness scale with frequency and spectral centroid? *Acta Acustica united with Acustica*, 92(5), 820–825.
- Singh, P. (1997). The role of timbre, pitch, and loudness changes in determining perceived metrical structure [Abstract]. *The Journal of the Acoustical Society of America*, 101(5), 3167. <https://doi.org/10.1121/1.419119>
- Wagner, M. (2022). Two-dimensional parsing of the acoustic stream explains the iambic–trochaic law. *Psychological Review*, 129(2), 268–288. <https://doi.org/10.1037/rev0000302>

- Wessel, D. L. (1979). Timbre space as a musical control structure. *Computer Music Journal*, 3(2), 45–52.
<https://doi.org/10.2307/3680283>
- Wood, C. (2015). *Perception of timbre intervals* [Master's thesis, McGill University].
https://www.mcgill.ca/mpcl/files/mpcl/wood_2015_mathesis.pdf
- Woodrow, H. (1911). The role of pitch in rhythm. *Psychological Review*, 18(1), 54–77.
<https://doi.org/10.1037/h0075201>

A Multidimensional Scaling Timbre Space for Percussive Musical Instruments Generated by Physical Modeling Synthesis

Marcelo Caetano^{1†} and Josh Mitchell¹

¹ Department of Music, University of California San Diego, San Diego, CA, USA

[†] Corresponding author: mcaetano@ucsd.edu

Improvements in physical modeling synthesis have made it possible to generate realistic sounds from several musical instruments in real time (Bilbao 2012; Jossic et al. 2017; Paté et al. 2017; Trévisan et al. 2017; Bank & Chabassier 2019; Giordano & Chatziioannou 2021; Fontana et al. 2024; Miliković et al. 2025). *Pianoteq*⁴ is a real-time interactive synthesizer based on physical modeling capable of generating percussive sounds from several musical instruments, such as the piano, the guitar, and the harp. Additionally, *Pianoteq* allows morphing between different instruments because these are generated by the same underlying physical model. Therefore, morphing with physical modeling synthesis in *Pianoteq* simply requires matching corresponding parameters and interpolating parameter values. However, linear interpolations of physical modeling parameters seldom result in perceptually intermediate morphs. This work is the first part of a study that will investigate whether the morphs produced by *Pianoteq* are perceptually intermediate, in the sense that they should correspond to hybrid instruments with intermediate timbral characteristics.

Previous studies have investigated the timbre of sound combinations such as hybrids, spectral interpolations, and morphs. Krumhansl (1989) used FM synthesis to create sounds that were “hybrids”, with the “head” of one instrument and the “tail” of another one. The positions of the hybrid sounds in the three-dimensional multidimensional scaling (MDS) timbre space obtained tended to be between the instruments used to create them. Horner et al. (2009) used spectral interpolations between pairs of musical instruments to investigate discrimination as a function of spectral difference. They found that changes to some instruments were more easily discriminable than others, and that some instruments had a stronger effect on discrimination. Carral (2011) used spectral interpolation to produce intermediate steps between sounds of a trombone played with two different mouthpieces and then used these interpolations to study the just noticeable difference (JND) between pairs of these sounds, which she found to be 0.86dB for the 50% JND. Hikichi and Osaka (2001) investigated sound interpolation of a simplified physical model to morph between piano and guitar sounds. They proposed interpolation strategies to obtain intermediate timbres, and they built an MDS timbre space to visualize where the morphs are perceived to be. They found that using the spectral centroid to guide the interpolations produces more perceptually intermediate morphs.

This study comprises two parts, the first part will build an MDS timbre space with musical instrument sounds synthesized with *Pianoteq*, and the second part will add morphs to that timbre space to investigate where these morphs are perceived to be relative to the musical instruments used to make them. We expect perceptually intermediate morphs to appear between the morphed instruments in their underlying timbre space. We hypothesize that deviations from the ideal relative position of the morphs will give important information not only about the interpolation strategy, but also about the perceptual dimensions of timbre connected to the physical parameters of the model. This work reports preliminary results from the first part.

Pianoteq is a virtual piano synthesizer based on physical modeling that is also capable of emulating other percussive musical instruments based on similar physical principles as the piano, such as plucked strings

⁴ <https://www.modartt.com/>

or idiophones. Several types of pianos, from acoustic grand pianos and upright pianos to electric pianos, can be obtained by adjusting the parameters of the model to emulate specific models, such as the Steinway Model D. Additionally, plucked strings such as the harp and acoustic guitar as well as percussion instruments such as vibraphones and xylophones can be obtained. *Pianoteq* allows great flexibility and expressivity, allowing the performer to manually adjust not only musical parameters of each instrument, such as tuning and temperament, but also physical parameters, such as mallet hardness for percussion and soundboard impedance for acoustic pianos. We selected ten musical instruments that represent the percussive gamut that *Pianoteq* emulates, including two pianos (Steinway acoustic and Wurlitzer electric), four chromatic percussions (church bells, kalimba, marimba, and vibraphone), two plucked strings (acoustic guitar and harp), a harpsichord, and a steel drum. The note selected was the C4, approximately 261 Hz, with durations below 400ms, typically 100ms except for instruments that have a characteristic long resonance tail, such as the harp and the kalimba. All sounds were normalized to -16dB RMS with Audacity⁵ prior to the experiment.

The experiment collected pairwise dissimilarity data for all 55 possible pairs, including the 10 identical pairs (unknown to the participants). Participants first heard the 10 sounds in random order to be exposed to the degree of variation, followed by a couple of training trials, and finally by the 55 experimental trials. The order of presentation of the pairs was randomized, and so was the order of presentation of the sounds in a pair. On each trial, the pair was presented with a 500ms silent interval between them. The participants were asked to move a horizontal slider along the scale from “identical” to “very different” to the position that best reflected their judgment of the dissimilarity between the sounds. The slider was initially positioned in the middle, and it had to be moved at least once to allow proceeding to the next trial. The listener could replay the sound pair and adjust the slider at will until satisfied with the judgment. Listeners were instructed to use the entire scale and to use headphones during the experiment.

In total, we collected data from 29 participants, 26 of whom declared previous musical experience (including 4 professional musicians), and 3 declared no musical experience. Their ages ranged from 18 to 45 years old, and none reported hearing problems. All participants volunteered to take the test online on Qualtrics⁶, which was considered exempt by the institutional review board of UC San Diego. We used the ratings for the identical pairs as data quality control, and we excluded 3 participants who rated more than 20% of the identical pairs above 10% of the minimum of the scale. Additionally, we calculated the pairwise correlation between the data matrices for all 29 participants and did a hierarchical cluster analysis of the correlation matrix to further identify participants whose ratings are distant from the others. Seven participants clustered apart from the others, including those who did not rate the identical pairs consistently close to 0. Finally, we decided to remove these 7 participants and only analyze the data from the 22 remaining participants.

We used three-way SMACOF (de Leeuw & Mair 2009) to build an MDS timbre space from the dissimilarities by the 22 participants with both INDSCAL and identity as constraint. Identity is used for model selection, where the goal is to decide the dimensionality of the MDS timbre space, whereas INDSCAL is used for the final configuration. Thus, we ran SMACOF with the number of dimensions varying from 1 to 7 and then we calculated the coefficient of determination R^2 and the Aikake information criterion (AIC) for each model (Elliott et al. 2013). R^2 is a measure of goodness of fit and AIC is a distance between the selected model and the “true” unknown model. Additionally, we plotted

⁵ <https://www.audacityteam.org/>

⁶ https://ucsd.co1.qualtrics.com/jfe/form/SV_eaikIgbSXC63hxc

Shepard diagrams, which indicate badness of fit, for each model to inspect the residual error. All criteria indicated that a model with 3 dimensions is the best compromise between model fit and interpretability.

References

- Bilbao, S. (2012). Time domain simulation and sound synthesis for the snare drum. *J. Acoust. Soc. Am.* 131(1), 914–925. <https://doi.org/10.1121/1.3651240>
- Chabassier, J., Chaigne, A., & Joly, P. (2013) Modeling and simulation of a grand piano. *J. Acoust. Soc. Am.* 134(1), 648–665. <https://doi.org/10.1121/1.4809649>
- Jossic, M., Mamou-Mani, A., Chomette, B., Roze, D., Ollivier, F., & Josserand, C. (2017) Modal active control of Chinese gongs. *J. Acoust. Soc. Am.* 141(6), 4567–4578. <https://doi.org/10.1121/1.4985108>
- Paté, A., Le Carrou, J. L., Givois, A., & Roy, A. (2017) Influence of plectrum shape and jack velocity on the sound of the harpsichord: An experimental study. *J. Acoust. Soc. Am.* 141(3), 1523–1534. <https://doi.org/10.1121/1.4976955>
- Trévisan, B., Ege, K., & Laulagnet, B. (2017) A modal approach to piano soundboard vibroacoustic behavior. *J. Acoust. Soc. Am.* 141(2), 690–709. <https://doi.org/10.1121/1.4974860>
- Bank, B., & Chabassier, J. (2019) Model-Based Digital Pianos: From Physics to Sound Synthesis. *IEEE Signal Processing Magazine*, 36(1), 103–114, <https://doi.org/10.1109/MSP.2018.2872349>
- Giordano, N., & Chatziioannou, V. (2021) Status and future of modeling of musical instruments: Introduction to the JASA special issue. *J. Acoust. Soc. Am.* 150(3), 2294–2301. <https://doi.org/10.1121/10.0006439>
- Fontana, F., Järveläinen, H., Papetti, S., & De Pra, Y. (2024) Acoustic cues of keyboard mechanics enable auditory localization of upright piano tones. *J. Acoust. Soc. Am.* 156(1), 164–175. <https://doi.org/10.1121/10.0026484>
- Miljković, T., Bjelić, M., Čertić, J., & Pavlović, D. S. (2025) Estimation of harp string inharmonicity influenced by phantom partials. *J. Acoust. Soc. Am.* 158(4), 3187–3202. <https://doi.org/10.1121/10.0039660>
- Krumhansl, C. L. (1989). Why is musical timbre so hard to understand? In S. Nielzen, O. Olsson (Eds), *Structure and perception of electroacoustic sound and music*. (pp. 43-53), New York, USA: Excerpta Medica.
- Horner, A. B., Beauchamp, J. W., & So, R. (2009) Detection of time-varying harmonic amplitude alterations due to spectral interpolations between musical instrument tones. *J. Acoust. Soc. Am.* 125(1), 492–502. <https://doi.org/10.1121/1.3025916>
- Carral, S. (2011). Determining the just noticeable difference in timbre through spectral morphing: a trombone example. *Acta Acustica united with Acustica*. 97(3), 466–76.
- Hikichi, T. & Osaka, N. (2001). Sound timbre interpolation based on physical modeling, *Acoustical Science and Technology*, 22(2), 101–111. <https://doi.org/10.1250/ast.22.101>
- Elliott, T. M., Hamilton, S. L., & Theunissen, F. E. (2013) Acoustic structure of the five perceptual dimensions of timbre in orchestral instrument tones. *J. Acoust. Soc. Am.* 133(1), 389–404. <https://doi.org/10.1121/1.4770244>
- de Leeuw, J., & Mair, P. (2009). Multidimensional Scaling Using Majorization: SMACOF in R. *Journal of Statistical Software*, 31(3), 1–30. <https://doi.org/10.18637/jss.v031.i03>

Mapping the Concert Snare Drum: A Sound Catalogue Through Improvisation

Tarcísio Braga^{1,2 †}

¹ Department of Music Performance, McGill University, Montréal, Québec, Canada

² Curso de Música, Amazon State University (UEA), Manaus, Amazonas, Brazil

[†] Corresponding author: tarcisio.braga@mail.mcgill.ca

Mapping the Concert Snare Drum is an artistic-research project that involves transcribing and analyzing my own snare drum solo improvised performances in order to develop a sound catalogue that demonstrates a range of sound production possibilities on the snare drum. It seeks to gain analytical tools capable of articulating nuances that are often described through vague or metaphorical language. The catalogue integrates material from orchestral, rudimental, and contemporary snare drum contexts, which are well represented in existing literature (Blades 1992; Gauthreaux 1989; Stone 1935), with Brazilian practices which remain largely underexplored in scholarly work outside Brazil. My main research question is situated at the intersection of improvisation, composition, and performance, problematizing the various dialogic relationships between technique and spontaneity.

Mapping the Concert Snare Drum consists of a report of the first phase of my doctoral research, which involved performing an original solo snare drum piece called *Tambor*. This piece aims to integrate the environmental and acoustic properties of specific locations into an immersive experience. *Tambor*, meaning *drum* in Portuguese, originated from freely improvised snare drum warm-ups. Colleagues often remarked that they resembled composed pieces, which inspired me to take them more seriously. In 2024-2025, *Tambor* was performed seven times⁷, in both live and studio contexts. These performances generated a substantial body of high-quality audio and multi-angle video material that form the basis of an analytical work, comprising a close examination of subtle and refined elements of my personal technique. The analysis also includes identifying large-scale formal structures and sectional organization of the piece, including the timbral progression from bare hands to brushes to sticks.

Timbre, or tone color, is “the perceived sound resulting from the prominence of certain overtones (partials) over others, [...] giving each instrument its unique color” (Solomon 2002, 55). I expand this by using composer Stuart S. Smith’s “doing more with less” concept which emphasizes exploring the “enormous color potential” of a single instrument (Baker 2004, 13). Solomon categories of timbre alterations will also be used in the descriptions: dynamics, beater choice, placement of attack, and muting (Solomon 2002).

In *Tambor* I explore Afro-Brazilian and West-African hand drum influences, by transposing them to the snare drum, treating different kinds of slaps as rimshots, and experimenting with their tonal variations through different stick placements. I also incorporate Brazilian snare drumming traditions from *maracatu*, *samba*, and *frevo*, in which the snare drum (also called *caixa* or *tarol*) is an essential instrument. The connection of Brazilian snare drumming with improvisation happens at the pedagogical and performance practice levels. Most Brazilian musical traditions are transmitted primarily through informal learning, which according to Almeida, includes processes like observation, imitation, repetition, and improvisation based on what is seen and heard (2009, 12). *Tambor* positions itself as a bridge between diverse

⁷ Including performances at: McGill Percussion Studio Concert, McGill Performance Simulator, Recording at IMLAV (Immersive Media Lab) Studio, 7th COBS Student Colloquium, Canadian Percussion Network (CPN) Conference in Winnipeg, CIRMMT’s MMR (Multimedia Room), Tarcísio’s D.Mus. Recital 1.

drumming traditions, connecting Brazilian snare drum practices (deeply rooted in popular, folkloric, and traditional styles) with classical and contemporary percussion.

The performance of *Tambor* also features an interplay with a quadraphonic “tape” composed of processed samples from my personal collection of field recordings made in the Brazilian Amazon rainforest. For this paper I will be analysing materials specifically from my Doctorate Recital 1 Performance. Since this is the latest performance of *Tambor*, it is an example of how the process of repeated improvisation has brought maturity, not only to the solo snare section of the piece but also to the quadraphonic tape counterpart (as seen in *Figure 1*).

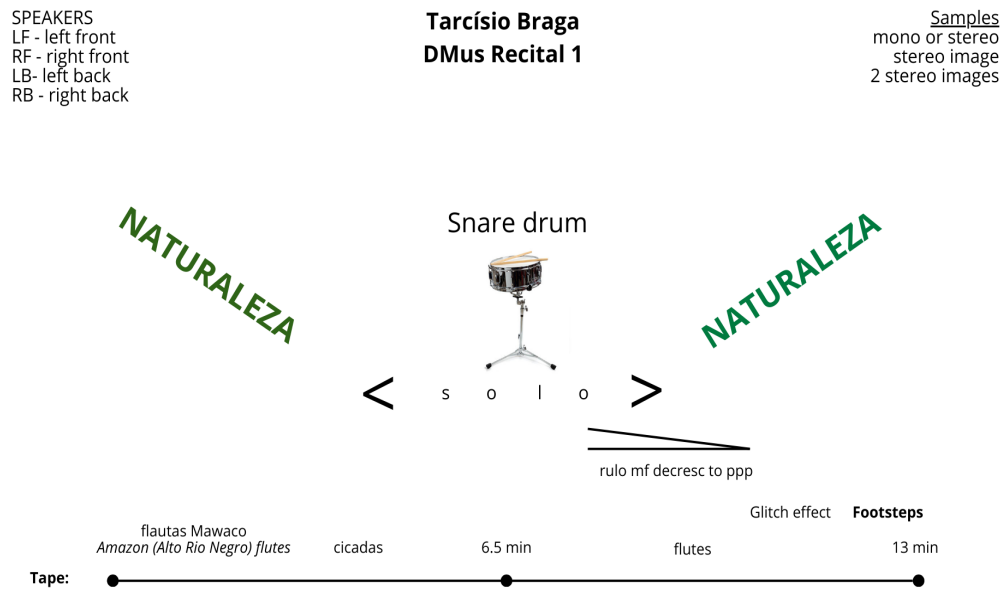


Figure 1: Temporal timeline of quadraphonic tape with large-scale formal structures and sections from *Tambor* performance at the Doctoral Recital 1 Performance

For the same performance, the following timbral progression was identified (as seen in Table 1). The abbreviation “R” stands for right hand and “L” stands for left hand.

Table 1: Timbral progression in *Tambor* (Doctoral Recital 1 Performance).

Stages of timbral progression	Description
Stage 1: Bare-hands technique (R / L)	Using bare hands on the snare drum brings me back to Afro-Brazilian <i>atabaques</i> and West-African <i>djembe</i> traditions of hand drumming, that differ from stick-based percussion practice. It invokes and becomes a way of honoring a common ancestry of drumming, reminding us that the snare drum remains part of a broader drumming lineage.
Stage 2: (Hybrid) Brush (R) + Bare Hand (L)	Using brush and hand acted as a transitional section of the solo.

Stage 3: (Dual) regular brush technique (R / L)	Here brushes were used, not so much connected to the jazz tradition, but rather to evoke fluid, constant, wide, feelings related to nature in the Amazon (big rivers, widened sense of time, endless and dense tree canopy of the Amazon).
Stage 4: Hybrid - Stick (R) + Bare hand (L)	Using stick and hand combination brings me back to the Afro-Brazilian <i>Ketu</i> candomblé tradition, where <i>aguidavis</i> and hand are used, and also to the <i>repinique</i> drum tradition of samba schools. A compilation video featuring Stage 4 excerpts from the 2024–2025 Tambor performances can be viewed here: https://www.youtube.com/watch?v=u1ybDEBydm8
Stage 5: (Dual) regular stick technique (R / L)	This was the longer section of the solo section, using a regular dual stick technique is a reference to orchestral, rudimental, contemporary, and Brazilian traditions such as <i>maracatu</i> , <i>frevo</i> , <i>samba</i> , <i>dobrado</i> , <i>banda de pifanos</i> , and <i>marchinha</i> .

The creation of a sound catalogue involves processes such as mapping sounds, refining their descriptions, and employing diverse notational approaches. Sound exploration on solo improvisation often leads to generating extended techniques, and a sound catalogue provides a structured way to organize and remember them. According to McGill Jazz Drums Professor John Hollenbeck (J. Hollenbeck, personal communication, Fall of 2025), a sound catalogue should include three main categories, as is shown on the table below:

Table 2: Excerpt of sound catalogue from Doctoral Recital 1 Performance of Tambor
<https://youtu.be/L8Y3zngy3RQ?si=PcgcAfdXv4YjU1jE>

Minutes	Name	How the sound is made	Image, emotional reaction, and/or memory
1' 32"	<i>Igarapé</i> (stream)	Touching 5 fingernails on the batter drum head surface edge, scrape nails gently in circular motion around the head. Scrape freely at different dynamics / speeds: constant, slowing down, speeding up.	Movement of water on small ponds
1' 52"	<i>Gota</i> (drop)	With the distal joints touching the rim, tap with index finger pad (not tip) on edge of head, immediately lifting finger after stroke, to get a full ringing sound of the head.	Rain starting to fall
1' 56"	<i>Chuva</i> (rain)	Strike the head quickly using four finger pads, in this order: little (pinky) finger, ring finger, middle finger, index finger of left hand, and index finger of right hand (optional: alternate between index and middle finger of right hand as the ending finger). The goal is a continuous rolling sound.	Rainfall in the Amazon forest
2' 03"	<i>Madeira</i> (wood)	Touching 5 fingernails of the right hand on the batter drum head surface edge, scratch laterally side to side	Saw cutting through wood
2' 24"	<i>Água</i> (water)	Using a plastic brush (such as “Vater Percussion Little Monster”), perform constant circular motions around batter drum head surface edge	Movement of water

In a cyclical process, the knowledge derived from this analytical phase is reinjected into my artistic practice, in a movement going from practice to analysis and back to practice, constituting the core of the artistic research methodology that guides the project. Timbre-based pieces combining composition and improvisation can be built from the catalogue, forming narratives from the imagery of sounds. These catalogue-based pieces can be fully notated or flexible sketches, combining composition and improvisation. I plan to include unconventional, extended, graphic, and alternative notational systems (Blades 1992; Solomon 2002), with the possibility of developing new notation.

This project will provide new pedagogical and creative tools for students, percussionists, drummers, and composers, enabling them to achieve a wider range of expression, and to navigate more effectively the spectrum of tone color and sound production on the snare drum. This project also expands creative and pedagogical horizons by integrating improvisation and traditional Brazilian drumming techniques into solo concert snare drum.

References

- Almeida, J. L. S. (2009). Ensino/Aprendizagem Dos Alabês: Uma experiência nos terreiros Ilê Axé Oxumarê e Zoogodô Bogum Malê Rundó. [Doctoral dissertation in Music Education, Universidade Federal da Bahia]. Repositório Institucional da UFBA. <https://repositorio.ufba.br/handle/ri/9093>
- Baker, J. C. (2004). The Snare Drum as a Solo Concert Instrument. [Doctoral dissertation, University of North Texas]. University of North Texas Libraries. https://digital.library.unt.edu/ark:/67531/metadc4721/m2/1/high_res_d/dissertation.pdf
- Blades, J. (1992). Percussion Instruments and Their History. Bold Strummer.
- Gauthreaux, G. II (1989). Orchestral Snare Drum Performance: An Historical Study. [Doctoral dissertation, Louisiana State University]. LSU Scholarly Repository. https://repository.lsu.edu/gradschool_disstheses/4715/
- Solomon, S. Z. (2002). How to Write for Percussion: A Comprehensive Guide to Percussion Composition. Oxford University Press.
- Stone, George Lawrence. (1935). Stick Control. Alfred Publishing.

Percussion Patterns, Rhyming Poetry, and the Expectation of Recurring Timbres: Case Studies From Sri Lanka

Eshantha Peiris^{1†}

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

[†] Corresponding author: eshantha.j.peiris@gmail.com

Introduction

Scholars have long sought connections between music and language, comparing the use of formal devices such as pitch, rhythm, and syntax (Jackendoff 2011). Another phenomenon that links music and language is that they both create a variety of expectations in those who produce and perceive them. This paper highlights a specific type of expectation in music and language perception that has been largely overlooked; namely that of timbre-based patterning. Examining percussion music and rhyming poetry through the same analytic lens, I hypothesize that—at a lower level of cognition parallel to the processing of meaning and affect—they are all perceived through a listener’s short-term memory simply as patterns of timbres.

Expanding on theorizations of metric projection (Hasty 1997) and timing expectations in global cyclic musics (Roeder 2025), this paper proposes a cognitive model for how attentive listeners (with varying degrees of enculturation) attend to unfolding timbral patterns in both music and language. This model is depicted in Figure 1, where the boxes are heard sequentially from left to right, and each color represents a different timbre. According to this model, the second appearance (*) of the red timbre activates an echoic memory of its first appearance and raises the expectation that there will another recurrence of the red timbre following the same number of intervening timbres (even with varying proportions of duration). When this expectation is fulfilled (when the arrow completes at **), it raises an expectation of a further recurrence (shown with the dotted line).⁸ Testing this empirically might involve asking listeners to predict an expected sound in the third repetition of a short series of contrasting timbres. Here, I present this argument by analyzing two examples of traditional music from Sri Lanka, identifying timbral parallelisms in the music and words and speculating about how they might raise and satisfy cognitive expectations in listeners who have different levels of attentiveness and enculturation.

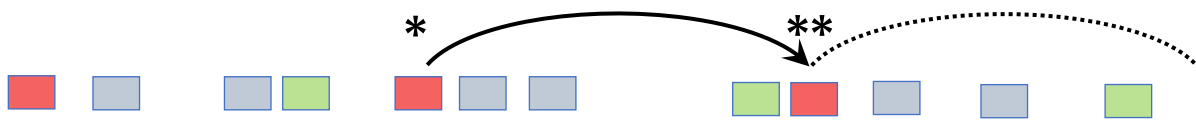


Figure 1: Modeling a listener’s cognitive projection of sequentially recurring timbres.

Timbral Patterning in Music

Unlike with pitched instruments, which typically maintain a consistent timbre when creating melodies with contrasting pitches, patterns for drums often feature moment-to-moment contrasts in timbre. The Sri Lankan ceremonial drumming piece *Maṅgul Bera*—which is played in loose unison with elastic timing on *gāṭa beraya* drums—features phrases built from such contrasting drum-stroke timbres. Figure 2 is a time-proportional transcription of an [excerpt from a performance of *Maṅgul Bera*](#),⁹ in which the phrase

⁸ While the arrows and dotted lines are inspired by Hasty’s (1997) arrows of durational projection, they are repurposed here as arrows of “sequential projection.”

⁹ This performance at a ritual is led by drummer Muddanave Sunil.

represented by the drum mnemonic syllables *tarikiṭa kuṇḍa kuṇḍa jīm jīm ta jīm kuṇḍa kuṇḍa ta jīm* is played three times in a row. The three brackets in Figure 2 segment the three phrases. Since the drum strokes in the phrase do not map neatly onto a grid of isochronous pulses, and the timing proportions between strokes are not identical across repetitions, I suggest it likely that some listeners will lean more on sequential expectations than on timing expectations when orienting to the unfolding pattern.¹⁰ In other words, listeners who recognize that the sequence of drum strokes is repeating are afforded the ability to predict *which* drum stroke (or motif) is likely to be played next, but not necessarily predict *when*. I further hypothesize that particularly salient drum strokes can function as markers of cyclicity that clarify the unfolding musical structure for a listener. As highlighted in Figure 2, I posit that the four quicker drum strokes *ta-ri-ki-ṭa* that begin the three respective phrases (in blue) draw attention to the undamped, low drum strokes that follow them (in green) and the retrospectively add emphasis to the undamped, high drum strokes that precede them (in purple).

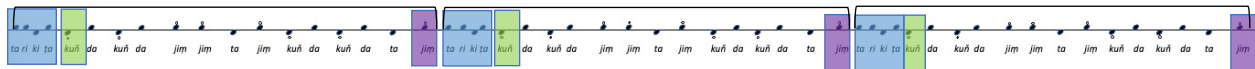


Figure 2: Salient drum strokes that mark cyclicity, in an [excerpt from Maṅgul Bera](#).

In addition to such “bottom-up” cognitive processing of the drum pattern as it unfolds in real time, enculturated listeners will also bring “top-down” schematic expectations to the listening experience (Nattiez 2007; Clayton 2020), which afford enhanced predictability about how the drum pattern is likely to play out. These genre-specific schematic expectations could stem from passively acquired knowledge of stock drum phrases and their associated timing conventions, explicit awareness of dance steps linked to these phrases, and/or embodied procedural knowhow of playing such phrases on the drum.

Timbral Patterning in Language

Given that phonemes in spoken language are differentiated by contrasting formant-based vocal timbres, phonological parallelisms such as rhyme can also be characterized as recurring phonemic timbres in sung/spoken performance that can raise and satisfy cognitive expectations in listeners, similar to recurring timbres in percussion music. Rhymes that appear at the end of parallel poetic lines can be particularly salient. An empirical experiment to test the perception of such patterning might involve comparing attentional peaks in EEG data of listeners attending to different versions of a rhyming poem in an unfamiliar language, with EEG data from listening to cyclic drum patterns.

Even if they may not mark cyclicity as strongly as end rhymes, phonological parallelisms such as assonance, alliteration, and consonance can also raise and satisfy cognitive expectations of recurrence. The sung poem [Sāwulā Vannama](#) (transliterated in Figure 3) begins with an introductory verse of non-lexical syllables; this verse serves as a template for the placement of prominent phonemes within the poetic meter in subsequent verses with Sinhala text, which are sung to the same melody.¹¹ Treating nasal consonants as a single phonemic category, I have highlighted the alliterative nasal consonants (“m” and “n”) that appear in parallel positions within the poetic meter (i.e., at the end of poetic clauses) and its associated melody. I argue that attentive listeners—especially those who have schematic expectations for some sort of phonemic parallels in this genre of poetry—will develop short-term dynamic expectations for the recurrence of nasal consonants at the end of poetic clauses.

¹⁰ John Roeder (2025) argues for such sequential expectations in pieces with much slower tempi.

¹¹ In performance, these verses are typically separated by a drummed interlude.

Introductory Verse With Non-Lexical Vocables:

 
tam tam dana tā nam / dana tam dana tā nam / tam tam dana tā nat tei / tam da nā tam da nā nā

Verse 1 with Sinhala lyrics:

suriñdun pera sitan balayut pāvatiñdu / saha gos kala yuden / asu – ran jaya no-la-di

Verse 2 with Sinhala lyrics:

bara at gena kagi / ena kontavi vali / asuran kala yude / jaya gena kañda kumaru

Figure 3: Nasal consonants at parallel positions in the poetic meter, in [Sāwulā Vannama](#).

In performances of *Sāwulā Vannama*, the music features a hierarchical musical meter and a polyphonic texture, such that the perception of recurring phonemic timbres will be just one of many factors contributing to the listener experience of repetition, alongside the accompanying drum pattern and repeating sequences of sung pitches and durations. In turn, these sonic patterns will be perceived alongside the higher-level processing of semantic meaning from the words (for those who understand the language). While my analytic method resembles Lerdahl's hierarchical model of syllabic sound color (2026) in its focus on phonological relationships such as rhyme, assonance, and alliteration, it differs in its emphasis on cognitive expectations of recurrence and proposed similarity to percussion timbres.

Conclusion

The above analyses have focused primarily on dynamic expectations created and reevaluated in real time. Such processes of predictive listening would be most applicable to a first-time listener who is attentive to sonic patterning but is not familiar with the musical style. However, assuming that different kinds of expectations can occur in parallel (Huron 2006), I suggest it likely that such dynamic processes of predictive listening remain active even if an enculturated listener possesses schematic expectations from prior familiarity with the musical (or poetic) genre or veridical expectations from knowing exactly how the composition (or a particular recording of it) will go. This acknowledgement of enculturated knowledge serves as a reminder that the ethnographic interpretations associated with ethnomusicological research methods are essential to a more complete understanding of cross-cultural processes of music and language cognition, complementing theories from music theory, phonology, and music psychology.

The questions raised by this paper have implications for a variety of scholarly debates, including whether timbre is cognized pre-attentively (Fales 2002), whether timbre can carry syntax (Meyer 1989), whether contrasting timbres necessarily provoke segmentation in listener perception or can be integrated into a single auditory stream (McAdams 2019), whether the shared reliance on timbral patterns in music and speech suggests shared neurological processing resources and/or shared evolutionary ancestry (Brown 2017; Lerdahl 2026), the role of timbral patterning in infant cognitive development (Lau et al. 2021), the relationship of rhyming poetry to ordinary speech (Tsur 1996; Obermeier et al. 2013), and whether timbre can be coded in relative terms (e.g., as timbre intervals or timbre contours) (McAdams and Cunibile 1992).

References

- Brown, S. (2017). A joint prosodic origin in language and music. *Frontiers in Psychology*, 8, Article 1894. <https://doi.org/10.3389/fpsyg.2017.01894>
- Clayton, M. (2020). Theory and practice of long-form non-isochronous meters: the case of the North Indian *Rūpak Tāl*. *Music Theory Online* 26(1). <https://doi.org/10.30535/mt.26.1.2>
- Fales, C. (2002). The paradox of timbre, *Ethnomusicology* 46(1), 56–95. <https://doi.org/10.2307/852808>
- Hasty, C. (1997). *Meter as rhythm*. Oxford University Press.
- Huron, D. (2006). *Sweet anticipation: music and the psychology of expectation*. MIT Press.
- Jackendoff, R. (2011). Music and language. In T. Gracyk, & A. Kania (Eds), *The Routledge companion to philosophy and music*. (pp. 101-112). Routledge.
- Lau, B. K., Oxenham, A. J., & Werner, L. A. (2021). Infant pitch and timbre discrimination in the presence of variation in the other dimension. *Journal of the Association for Research in Otolaryngology: JARO*, 22(6), 693–702. <https://doi.org/10.1007/s10162-021-00807-1>
- Lerdahl, F. (2026). *The sounds of poetry viewed as music*. Columbia University Press.
- McAdams, S. (2019). Timbre as a structuring force in music. In ed. K. Siedenburg, C. Saitis, S. McAdams, A. N. Popper and R. R. Fay (Eds), *Timbre: acoustics, perception, and cognition*. (pp. 211–243). Springer. https://doi.org/10.1007/978-3-030-14832-4_8
- McAdams, S., & Cunible, J. C. (1992). Perception of timbral analogies. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 336(1278), 383–389. <https://doi.org/10.1098/rstb.1992.0072>
- Meyer, L. B. (1989). *Style and music: theory, history, and ideology*. University of Chicago Press.
- Nattiez J.-J. (2007) Le timbre est-il un paramètre secondaire? [Is timbre a secondary parameter?]. *Les Cahiers de la Société Québécoise de Recherche en Musique* 9(1–2):13–24
- Obermeier, C., Menninghaus, W., von Koppenfels, M., Raettig, T., Schmidt-Kassow, M., Otterbein, S., and Kotz, S. A. (2013). Aesthetic and emotional effects of meter and rhyme in poetry. *Frontiers in Psychology*, 4, Article 10. <https://doi.org/10.3389/fpsyg.2013.00010>
- Roeder, J. (2025). Slow 5-cycles and their Special Properties heard in World Music. *Analytical Approaches to World Musics* 13(1). <https://doi.org/10.5281/zenodo.15762649>
- Tsur, R. (1996). Rhyme and Cognitive Poetics. *Poetics Today*, 17(1), 55–87. <https://doi.org/10.2307/1773252>