

When NIME and ISMIR Talk Timbre

Charalampos Saitis[†], Maryam F. Torshizi, Vjosa Preniqi, Bleiz M. Del Sette, and György Fazekas
Centre for Digital Music, Queen Mary University of London, London, UK

[†] Corresponding author: c.saitis@qmul.ac.uk

Disciplines: new interfaces for musical expression, historical perspectives, ethnomusicology

Keywords: timbre, NIME, natural language processing

Introduction

Timbre is among the most evocative yet elusive attributes of music. It is through timbre that musicians can emote by manipulating the physical response of their acoustic instruments. Yet timbre is conspicuously absent from the digital luthier’s toolbox. Much synth design is still based on concepts from early analog and digital synthesis, or emulation of it using more recent techniques, while an audio engineer’s workbench is still based mainly on historical tools like oscilloscopes and signal generators. The long term goal of this project is to investigate the role of timbre in the design of sound synthesis and artificial intelligence tools that enable makers to create digital interactions with an understanding of how we listen. As a first step, we want to understand how two communities at the forefront of digital lutherie and audio signal processing currently engage with timbre in their scientific discourse. The annual International Conference on New Interfaces for Musical Expression (NIME) brings together researchers and artists working across disciplines of computer music and interaction design since 2001. Started a year earlier, the annual International Society for Music Information Retrieval (ISMIR) Conference represents the largest body of scholars and practitioners exploring new ways of intelligently analysing musical data. We conducted a linguistic analysis of papers published in NIME and ISMIR from 2001 and 2000, respectively, until 2022, which aimed to answer the following questions: How much is the term “timbre” used in each community? What is the intention for timbre within each community? What notion of timbre (e.g., acoustical, semantic, embodied) does each community operationalise?

Materials and Methods

Our methodology was based on techniques used by Jensenius (2014), Sullivan and Wanderley (2018), and Low et al. (2019). For both NIME and ISMIR, we considered their entire collection of proceedings, which are freely available for download in PDF format.[†] In total, after discarding a small number of PDF documents that were either not searchable (see below) or not convertible into plain text files, 1969 NIME papers and 2245 ISMIR papers were analysed. We first performed a keyword search through each year of proceedings to return the number of papers containing the term “timbre.” This involved a shell script (Jensenius, 2014; Sullivan & Wanderly, 2018) that returns the number and filenames of papers by year that match a given search query. It uses the `mdfind -count` terminal command which returns a spotlight search based on the OSX index of the files. Individual years were then grouped across periods reflecting the typical three-year cycle of research projects (the period covering early NIME meetings was extended to

[†] <https://www.nime.org/archives/>; <https://ismir.net/conferences/>

Table 1: Usage of terms in NIME proceedings between 2001 and 2022

Period	Papers	Timbre		Pitch		Gesture	
	#	#	%	#	%	#	%
2001–04	163	60	36.81%	100	61.35%	103	63.19%
2005–07	251	81	32.27%	136	54.18%	158	62.95%
2008–10	286	76	26.57%	154	53.85%	177	61.89%
2011–13	377	132	35.01%	208	55.17%	245	64.99%
2014–16	338	115	34.02%	203	60.06%	209	61.83%
2017–19	285	98	34.39%	166	58.25%	187	65.61%
2020–22	269	94	34.94%	170	63.20%	168	62.45%
<i>Total</i>	1969	656	33.32%	1137	57.75%	1247	63.33%
<i>Mean/year</i>	89.5	29.8	33.74%	51.7	58.13%	56.7	63.26%

Table 2: Usage of terms in ISMIR proceedings between 2000 and 2022

Period	Papers	Timbre		Pitch		Genre	
	#	#	%	#	%	#	%
2000–01	76	13	17.11%	51	67.11%	21	27.63%
2002–04	210	48	22.86%	132	62.86%	103	49.05%
2005–07	317	104	32.81%	168	53.00%	185	58.36%
2008–10	339	131	38.64%	203	59.88%	211	62.24%
2011–13	331	121	36.56%	214	64.65%	191	57.70%
2014–16	332	112	33.73%	214	64.46%	185	55.72%
2017–19	308	107	34.74%	239	77.60%	164	53.25%
2020–22	332	116	34.94%	229	68.98%	197	59.34%
<i>Total</i>	2245	752	33.50%	1450	64.59%	1257	55.99%
<i>Mean/year</i>	97.6	32.7	33.44%	63.0	64.66%	54.7	56.42%

four years; for ISMIR it was limited to two years).

To understand how “timbre” is used in each conference corpus from one three-year period to the other, we continued with concordance and collocation analyses (Gablasova et al., 2017) of those papers that mentioned “timbre” within each period, which yielded a ranked list of closely associated terms. PDF files were converted to plain text format using the free AntFileConverter application (Anthony, 2022b). For text concordance and collocation, we used the free AntConc application (Anthony, 2022a). We looked for co-occurrences within five words to the left of “timbre” and five to the right, thus identifying looser word associations than the n-gram approach. Collocates were initially determined by maximum likelihood with Bonferroni adjustment. Results were then filtered manually to further remove generic or irrelevant terms.

Findings

Keyword Occurrence

Tables 1 and 2 report the number and proportion of NIME and ISMIR papers, respectively, that mention “timbre.” An initial observation is that the two distributions are very similar, on average 33.5% of all papers

in both corpora, with only small fluctuations from period to period but also from year to year (not reported). For comparison purposes, we also report the usage of “pitch” (both corpora), “gesture” (NIME only) and “genre” (ISMIR only). Gesture dominates the NIME discourse (Jensenius, 2014), and genre has been a topic of constant interest to ISMIR (Low et al., 2019, Fig. 5). Pitch has long been favoured over timbre in music scholarship (though this is changing, see Dolan & Rehding, 2021), which often informs research practices in digital instrument design, music informatics, and machine listening. Therefore, we expected “pitch” to occur more frequently than “timbre” in the two corpora, and this is indeed what we observed: +24.39% or 1.8 times more in NIME and +31.21% or 2 times more in ISMIR (averaged across years). Our findings corroborate the centrality of the notion of gesture in NIME research (on average 5% more papers mentioned gesture over pitch). They further reveal a similarly salient role for pitch in the ISMIR literature (about 8.2% more papers mentioned pitch over genre; the difference was even higher for other “hot” topics such as similarity, rhythm, or recognition).

Concordance and Collocation

Tables 3 and 4 present ranked (by the number of papers that mention “timbre”) lists of words closely associated with timbre within each analysis period. In NIME, notions of *controlling* timbre *changes* or *variation* or *spaces* (Wessel, 1979) are prevalent across all periods—the three most frequent words in each period are always a combination of these four terms. The concept of timbre *space* also appears regularly in ISMIR, which is concerned primarily with *modelling* timbre *features* (of audio signals) or *similarity* (between audio signals). Here a shift in how the ISMIR community engages with timbre can be observed around 2017, from a more supporting role in tasks such as *instrument recognition* and music *similarity* until then to a more prominent place in deep learning based audio *synthesis* tasks such as timbre *transfer* (Hayes et al., 2021) and *controllable* sound design (Diaz et al., 2023) since about 2017. Terms related to semantic aspects of timbre (e.g., *words*, *verbal*, *descriptors*) appear in most NIME periods, but in ISMIR they are more frequent in the second decade of the conference. However, in contrast to empirical evidence of the role of timbre in musical emotion and mood (Hailstone et al., 2009; McAdams et al., 2017), we found only one related term (*affect*) and only in the NIME corpus (2011–13). Similarly, Tables 3 and 4 each contain only one word related to *gesture*. While the latter may not be a regular topic in MIR research, for NIME this finding suggests that its central scientific discourse of gesture may not consider issues of sensory experience and embodiment (cf. Jensenius, 2014, Table 3)—or if it does, timbre appears to be excluded from such considerations.

Perspectives

We have reported an analysis of published literature spanning more than twenty years of scientific research in digital music instruments (DMIs) and music information retrieval. The present findings support the initial claim that there is still limited timbre-based practice in the digital music making creative workflow. The fact that there are almost twice as many papers mentioning “pitch” than there are discussing “timbre” in both corpora implies that current design practices in DMIs and MIR systems encode assumptions about musical space that are primarily based on analytical concepts from music theory, such as the idea that music is made by discrete onset and release events. This contrasts with converging recent musicological theories which posit that timbre in contemporary music making and listening practices is as or even more important than pitch-based concepts like melody and harmony (Fink et al., 2018; Solomos, 2019). In other words,

Table 3: Terms collocated with “timbre” in analysed NIME papers
(brackets report the number of papers that contain the collocate)

2001–04	2005–07	2011–13	2014–16	2017–19	2020–22
control (17)	control (18)	control (27)	control (23)	control (23)	control (14)
space (11)	space (10)	changes (19)	changes (22)	variation (8)	changes (9)
variation (3)	variation (8)	variations (6)	space (15)	space (8)	space (8)
features (2)	articulation (7)	voice (6)	features (7)	variety (5)	articulation (4)
descriptors (1)	changes (6)	similarity (4)	navigation (6)	manipulation (5)	transpositions (2)
	modulation (5)	affect (4)	recognition (5)	feature (4)	descriptors (2)
2008–10	gestural (4)	remapping (3)	description (3)	articulation (4)	explorer (1)
control (12)	recognition (3)		morphology (3)	learned (3)	bowing (1)
changes (8)	words (2)		morphing (2)	preserving (2)	
space (4)	classification (2)		attributes (2)	performative (1)	
nuances (3)	vocalic (1)		matching (2)	phrasing (1)	
voice (2)			semantic (1)		
remapping (1)			spatialization (1)		
deviations (1)			visualization (1)		
plucking (1)			unintended (1)		

timbre matters in digital music culture. Musical interaction—whether designing sounds or querying sound databases—and timbre perception are intimately linked processes (Hayes et al., 2022; Lam & Saitis, 2021), yet past and present scientific discourse in NIME and ISMIR does not adequately account for this relation.

The present analysis allowed us to quickly gain a high-level understanding of trends from a large collection of documents. In future work we aim to revisit the compiled documents to extract more nuanced information on digital music communities and timbre-based practice, including fine-tuning our analysis pipeline (e.g., search for syntactical variations of the same root word, expand window span of collocates, use dependency parsing techniques, use knowledge graphs) and expanding it on other relevant literature (e.g., International Computer Music Conference, Computer Music Journal, Sound and Music Computing Conference).

References

- Anthony, L. (2022). AntConc (Version 4.2.0) [Computer Software]. Tokyo, Japan: Waseda University. Available from <https://www.laurenceanthony.net/software>
- Anthony, L. (2022). AntFileConverter (Version 2.0.2) [Computer Software]. Tokyo, Japan: Waseda University. Available from <https://www.laurenceanthony.net/software>
- Diaz, R., Saitis, C., & Sandler, M. (2023). Interactive neural resonators. *In Proc. Int. Conf. on New Interfaces for Musical Expression (NIME)*, 1–5, Mexico City, Mexico.
- Dolan, E. I., & Rehding, A. (Eds.). (2021). *The Oxford Handbook of Timbre*. Oxford University Press.
- Fink, R., Latour, M., & Wallmark, Z. (Eds.). (2018). *The relentless pursuit of tone: Timbre in popular music*. Oxford University Press.
- Gablasova, D., Brezina, V., & McEnery, T. (2017). Collocations in corpus-based language learning research: Identifying, comparing, and interpreting the evidence. *Language Learning*, 67(S1), 155-179.
- Grachten, M., Schedl, M., Pohle, T., & Widmer, G. (2009). The ISMIR Cloud: A Decade of ISMIR Conferences at Your Fingertips. *In Proc. 10th Int. Society for Music Information Retrieval Conf. (ISMIR)*, 63–68, Kobe, Japan.

Table 4: Terms collocated with “timbre” in analysed ISMIR papers
(brackets report the number of papers that contain the collocate)

2000–01	2002–04	2011–13	2014–16	2017–19	2020–22
modelling (2)	similarity (8)	features (43)	features (45)	features (25)	features (28)
instrument (2)	features (8)	modelling (12)	similarity (18)	instrument (15)	transfer (16)
	space (4)	descriptors (9)	invariant (9)	changes (9)	instrument (15)
	instrument (4)	changes (9)	descriptors (8)	learned (8)	synthesis (14)
		instrument (8)	electronic (5)	space (8)	disentangled (12)
		invariant (6)	variations (5)	synthesis (7)	space (10)
similarity (30)	features (44)	verbal; descr. (6)	MFCC (5)	representations (7)	changes (8)
features (27)	similarity (28)	instrumentation (6)	vocal (3)	invariant (6)	control (8)
instrument (14)	changes (11)	articulation (5)	variability (3)	regularized (4)	similarity (8)
modelling (14)	instrument (10)	space (5)	shape (3)	variational (4)	MFCC (6)
recognition (8)	modelling (7)	robustness (5)	stability (2)	disentangled (2)	variational (6)
polyphonic (7)	space (7)	vocal (4)	syllables (2)	transfer (2)	variations (6)
gestures (5)	game (2)	distance (2)	shifts (2)	discrimination (1)	descriptors (5)
attributes (5)	streaming (1)		attributes (1)		regularized (5)
articulation (4)	metrics (1)				articulation (4)
singing (2)	varieties (1)				semantic (4)
inference (1)					translation (3)
classifiers (1)					overtones (3)
					encoder (2)

- Hailstone, J. C., Omar, R., Henley, S. M., Frost, C., Kenward, M. G., & Warren, J. D. (2009). It's not what you play, it's how you play it: Timbre affects perception of emotion in music. *Quarterly Journal of Experimental Psychology*, 62(11), 2141–2155.
- Hayes, B., Saitis, C., & Fazekas, G. (2021). Neural waveshaping synthesis. *In Proc. 22nd Int. Society for Music Information Retrieval Conf. (ISMIR)*, 254–261, online.
- Hayes, B., Saitis, C., & Fazekas, G. (2022). Disembodied Timbres: A Study on Semantically Prompted FM Synthesis. *Journal of the Audio Engineering Society*, 70(5), 373–391.
- Jensenius, A. R. (2014). To gesture or not? An analysis of terminology in NIME proceedings 2001–2013. *In Proc. Int. Conf. on New Interfaces for Musical Expression (NIME)*, 217–220, London, UK.
- Lam, J. R., & Saitis, C. (2021). Intuitive Timbre Explorer for education purposes and perceptual studies. *In Proc. Int. Conf. on New Interfaces for Musical Expression (NIME)*, 1–9, online.
- Low, T., Hentschel, C., Polley, S., Das, A., Sack, H., Nürnberger, A., & Stober, S. (2019). The ISMIR Explorer: A Visual Interface for Exploring 20 Years of ISMIR Publications. *In Proc. 20th Int. Society for Music Information Retrieval Conf. (ISMIR)*, 754–760, Delft, The Netherlands.
- McAdams, S., Douglas, C., & Vempala, N. N. (2017). Perception and modeling of affective qualities of musical instrument sounds across pitch registers. *Frontiers in psychology*, 8, 153.
- Solomos, M. (2019). *From music to sound: The emergence of sound in 20th-and 21st-century music*. Routledge.
- Sullivan, J., & Wanderley, M. (2018). Stability, Reliability, Compatibility: Reviewing 40 Years of NIME Design. *In Proc. 15th Sound and Music Computing Conference (SMC)*, 308–315, Limassol, Cyprus.
- Wessel, D. L. (1979). Timbre space as a musical control structure. *Computer Music Journal* 3(2), 45–52.