Perceptual and semantic scaling of FM synthesis timbres: Common dimensions and the role of expertise

Ben Hayes, Charalampos Saitis, György Fazekas

Centre for Digital Music, Queen Mary University of London, United Kingdom
{b.j.hayes, c.saitis, g.fazekas}@qmul.ac.uk

Background

Electronic sound has a rich history, yet timbre research has typically focused on the sounds of physical instruments, while synthesised sound is often relegated to functional roles like recreating acoustic timbres. Recent work (Hayes & Saitis, 2020; Wallmark et al., 2019) has studied the semantic associations of synthesised sounds through prompted synthesis paradigms. Results suggest that descriptive prompts exert a relatively consistent influence over the sound design process, but further work directly examining the perceptual structure of these sounds is warranted.

Aims

We applied two classical paradigms for studying timbre perception with the aim of identifying the perceptually salient acoustic attributes of sounds produced by frequency modulation synthesis. We also aimed to test the luminance-texture-mass timbre semantic model (Zacharakis et al., 2014) in this domain. Finally, we aimed to identify effects of prior music or synthesis experience on these results.

Method

In the first experiment, twenty-four participants provided pairwise dissimilarity ratings across 12 sounds. Stimuli were created in a previous study (Hayes & Saitis, 2020) by experienced sound designers using an FM synthesiser.

In the second, a native English-speaking subset of fifteen participants from the first experiment rated the same stimuli along 27 unipolar adjective scales. The adjectives were mined from a text corpus scraped from the popular synthesis forum *MuffWiggler*. Forum posts were filtered to a frequency sorted list of tokens co-occurring in bigrams with the words *sound*, *sounding*, *tone*, and *timbre*, from which adjectives were retained. Two raters independently pruned these according to a set of exclusion criteria, yielding the final set of adjectives.

All participants completed a preliminary headphone screening task based on the phase cancellation of sinusoids (Woods et al., 2017).

Results

Multidimensional scaling (MDS) was applied to all dissimilarity ratings, and to subgroups created by bisecting participants according to their musicality (GoldMSI) and synthesis experience (self-reported). Individual-differences scaling suggested inconsistency in the all-participants space, but strong consistency within each subgroup. Goodness of fit measures supported 3-dimensional solutions in all cases.

Exploratory factor analysis (EFA) on semantic ratings revealed two factors: one showing strong loadings for mass and darkness, and the other with texture. These are congruent with the luminance-texture-mass model, but their order suggests that mass better characterises stimulus variance.

Tucker's congruence coefficient (TCC), the modified RV coefficient, and Procrustes residual m^2 (with PROTEST permutation analysis) computed between perceptual MDS spaces suggested differences between experts and non-experts in the perceptual organisation of stimuli. The same metrics computed between semantic EFA spaces suggested that application of semantic descriptors was consistent across groups (TCC ≥ 0.97, modified RV ≥ 0.95, $m^2 \leq 0.06$). These results are presented in Table 1.

MDS and EFA dimensions each correlated strongly with multiple distinct acoustic features, spanning spectral, spectrotemporal, and purely temporal properties, suggesting that no individual feature is sufficient to describe any one dimension. The semantic mass factor showed strong correlations with all MDS spaces' first dimensions, and these correlated strongly with the second principal component of stimuli modulation power spectra.

Conclusions

Results suggest that discrimination of abstract electronic timbres may rely on attributes distinct from those used with acoustic timbres, and that these vary with expertise. However, the use of semantic descriptors is similar to that of acoustic instruments, and is invariant to expertise. This suggests that certain perceivable attributes of our stimuli may not be adequately captured by the adjective set.

Table 1: Measures of configurational similarity between subgroup MDS and EFA dimensions

| | Metric | | | | | |
| | Perceptual (MDS) | | | Semantic (EFA) | | |
Pairing	TCC	RV (mod)	m^2	TCC	RV (mod)	m^2
All × Musicians	0.49	0.07	0.76	0.99	0.98	0.02 ***
All × Non-musicians	0.93	0.81	0.13 ***	0.99	0.98	0.02 ***
All × Synthesists	0.69	0.43	0.52 **	0.99	1.00	0.01 ***
All × Non-synthesists	0.94	0.82	0.12 ***	0.98	0.98	0.04 ***
Musicians × Non-musicians	0.61	0.25	0.63 *	0.97	0.95	0.06 ***
Synthesists × Non-synthesists	0.69	0.46	0.53 **	0.98	0.97	0.05 ***

PROTEST (m^2) significance: $p < 0.05$ (*), 0.01 (**), 0.001 (***)

Implications

Studying the perception of synthesised sound can broaden our conception of timbre, allowing for insights beyond the constraints of physical instruments. Our results justify further such research into synthesised timbres, and the role of expertise in timbre perception. Such studies can also improve musical synthesis tools by providing a basis for bridging perception and synthesiser control.

References

Hayes, B., & Saitis, C. (2020). There's more to timbre than musical instruments: Semantic dimensions of FM sounds. *Proceedings of the 2nd International Conference on Timbre*. Timbre, Thessaloniki, Greece (Online).

Wallmark, Z., Frank, R. J., & Nghiem, L. (2019). Creating novel tones from adjectives: An exploratory study using FM synthesis. *Psychomusicology: Music, Mind, and Brain*. https://doi.org/10.1037/pmu0000240

Woods, K. J. P., Siegel, M. H., Traer, J., & McDermott, J. H. (2017). Headphone screening to facilitate web-based auditory experiments. *Attention, Perception, & Psychophysics*, *79*(7), 2064–2072. https://doi.org/10.3758/s13414-017-1361-2

Zacharakis, A., Pastiadis, K., & Reiss, J. D. (2014). An Interlanguage Study of Musical Timbre Semantic Dimensions and Their Acoustic Correlates. *Music Perception: An Interdisciplinary Journal*, *31*(4), 339–358. https://doi.org/10.1525/mp.2014.31.4.339

Keywords: timbre, semantics, factor analysis, multidimensional scaling, synthesis