Evidence for timbre space robustness to an uncontrolled online stimulus presentation

Asterios Zacharakis[1†], Ben Hayes[2], Charalampos Saitis [2] and Konstantinos Pastiadis[1]
[1] School of Music Studies, Aristotle University of Thessaloniki, Thessaloniki, Greece
[2] Centre for Digital Music, Queen Mary University of London, London, UK
[†] Corresponding author: aszachar@mus.auth.gr

Introduction

Research on timbre perception is typically conducted under controlled laboratory conditions where every effort is made to maintain stimulus presentation conditions fixed (McAdams, 2019). This conforms with the ANSI (1973) definition of timbre suggesting that in order to judge the timbre differences between a pair of sounds the rest perceptual attributes (i.e., pitch, duration and loudness) should remain unchanged. Therefore, especially in pairwise dissimilarity studies, particular care is taken to ensure that loudness is not used by participants as a criterion for judgements by equalising it across experimental stimuli. On the other hand, conducting online experiments is an increasingly favoured practice in the music perception and cognition field as targeting relevant communities can potentially provide a large number of suitable participants with relatively little time investment from the side of the experimenters (e.g., Woods et al., 2015). However, the strict requirements for stimuli preparation and presentation prevents timbre studies from conducting online experimentation. Despite the obvious difficulties in imposing equal loudness on online experiments, the different playback equipment chain (DACs, pre-amplifiers, headphones) will also almost inevitably 'colour' the sonic outcome in a different way. Despite the above limitations, in a social distancing time like this, it would be of major importance to be able to lift some of the physical requirements in order to carry on conducting behavioural research on timbre perception. Therefore, this study aims to investigate the extent to which an uncontrolled online replication of a past laboratory-conducted pairwise dissimilarity task will distort the findings.

Method

A pairwise dissimilarity study presented in Zacharakis et al. (2015) was replicated in an online experiment. Sixteen musically trained listeners with normal hearing took part in the experiment (12 male, 4 female, average age: 30.7, average years of musical practice: 16.4, std of years of musical practice: 8.1). Their task was to rate the pairwise differences among 24 musical tones (300 pairs overall) –consisting of acoustic, electric and synthetic instruments– using the free magnitude estimation method. That is, they rated the perceptual distances of 300 pairs (same pairs included) by freely typing in a number of their choice to represent dissimilarity of each pair (i.e., an unbounded scale) with 0 indicating a same pair. Prior to the main listening test, a headphone screening test similarly to Woods et al. (2017) along with a familiarisation phase and a short training phase took place to make sure participants used headphones and understood the required task adequately enough. The pairs of the main experiment were presented in random order and the presentation order within each pair was also randomised. In the beginning of the experiment the participants were asked to set a comfortable playback level and keep it constant throughout the process.

Results

The Cronbach's Alpha for this set of responses was .9, indicating a strong internal consistency of responses, albeit a little lower compared to the .94 and .96 identified for English and Greek speaking participants for the controlled experiment. Comparison of the average raw dissimilarities between the two experiments (English speaking participants) showed a very strong correlation (Pearson's $\boldsymbol{r} = \mathbf{.9}$, $p<.001$). This fact already indicates that the uncontrolled online experiment did not result in a substantial alteration of the acquired ratings overall.

A subsequent non-metric Multidimensional Scaling Analysis with dimension weighting (INDSCAL within SPSS PROXSCAL algorithm) was applied to the obtained dissimilarities. An examination of the measures-of-fit that is presented in Table 1–in comparison with the respective metrics for the controlled experiment– indicates that the data are optimally represented by a 3-dimensional model since the improvement of the measures diminishes when a fourth dimension is added.

Table 1: Measures-of-fit and their improvement for different MDS dimensionalities between the laboratory and online experiments. S-Stress is a measure of misfit. The lower the value (to a minimum of 0) the better the fit. D.A.F.: Dispersion Accounted For is a measure of fit. The higher the value (to a maximum of 1) the better the fit.

	Online				**Laboratory**			
Dimensionality	**S-Stress**	**Improv.**	**D.A.F.**	**Improv.**	**S-Stress**	**Improv.**	**D.A.F.**	**Improv.**
1D	*.32*	*-*	*.81*	*-*	*.36*	*-*	*.81*	*-*
2D	*.18*	*.14*	*.92*	*.11*	*.19*	*.17*	*.92*	*.11*
3D	*.13*	*.05*	*.95*	*.03*	*.13*	*.06*	*.95*	*.03*
4D	*.09*	*.04*	*.97*	*.02*	*.10*	*.03*	*.97*	*.02*

The comparison between the two 3-dimensional timbre spaces that resulted from the laboratory and the online replication of the experiment was based on one index for configurational similarity, namely the Tucker's congruence coefficient and a second one for assessing dimensional similarity (i.e., the direct relationships between the dimensions of the two timbre spaces), namely the modified RV coefficient. A more detailed explanation regarding the use of these metrics for timbre space comparison can be found at Zacharakis & Pastiadis (2016) and Zacharakis et al. (2017). In general, values of the Tucker's congruence coefficient greater than .92 are considered fair and values larger than .95 practically imply perfect equivalence between the compared configurations (Lorenzo-Seva & Ten Berge, 2006). The statistical significance of the congruence coefficient between the two configurations was tested using a bootstrap analysis method (Monte Carlo estimate of its expected value under chance conditions). The modified RV coefficient for overall dimensional similarity between matrices varies between 0 and 1 and should be interpreted in a similar manner to the correlation coefficient between two unidimensional variables (Abdi, 2007).

Table 2: Metrics for configurational and dimensional similarity between the two spatial configurations. The expected value and the standard deviation of the congruence coefficient were estimated through bootstrapping with 10000 runs (**: significance at the .01 level).

	Congruence coefficient (expected value, standard deviation)	**RV-mod**
Laboratory timbre space vs. online timbre space	.97 (.90, .008)	$.71^{**}$

A comparative assessment of the two metrics that are presented in Table 2 revealed a strong similarity both at configurational and at dimensional level. This can be also confirmed by inspection of the timbre spaces themselves shown in Figure 1. Although some specific differences may exist between the two configurations, the main characteristics such as the distinct major clusters between the impulsive and the continuous sounds, notable outliers (e.g., the double bass, the bowedpad, the marimba or the saxophone) and certain smaller clusters (e.g., Acid-Moog-Saxophone, Violin-Cello) are preserved.

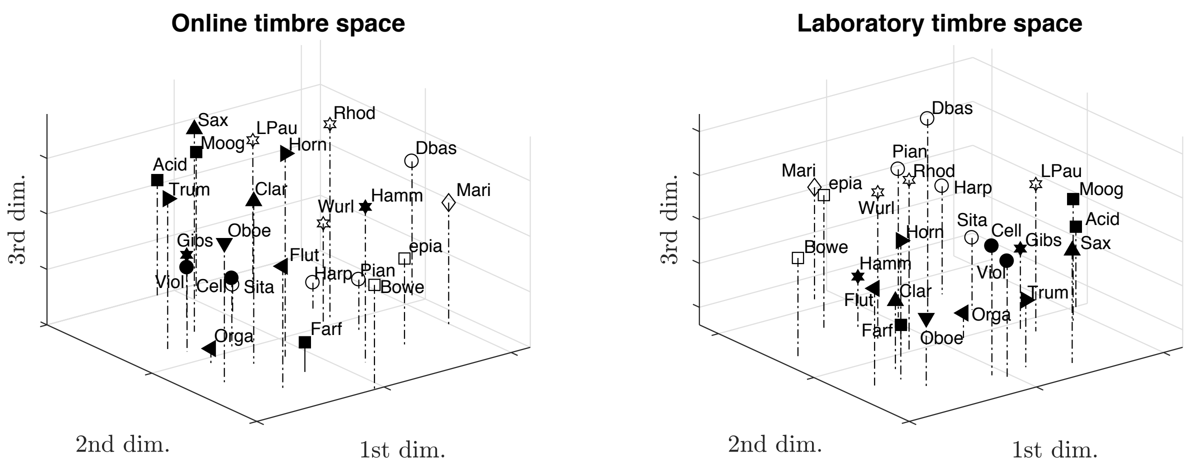

Figure 1: The two 3-dimensional timbre spaces for each experiment. The uncontrolled timbre space that came from the online experiment is shown on the left and the controlled that came from the laboratory experiment on the right. Black symbols: continuant, white symbols: impulsive, ▲: Single reed, ▼: Double reed, ◄: Aerophone, ►: Lip reed, ●: Chordophone, ♦: Idiophone, ⋆: Electrophone , ■: Synthesizer. Abbreviations of instrument names, Acid: Acid, Bowe: Bowedpad, Clar: clarinet, DBas: double bass pizzicato, epia: electric piano (rhodes), Farf: Farfisa, Flut: flute, Gibs: Gibson guitar, Hamm: Hammond, Horn: French horn, Harp: Harpsichord, LPau: Les Paul Gibson guitar, Mari: marimba, Moog: Moog, Oboe: oboe, Orga: pipe organ, Pian: piano, Rhod: Rhodes piano, Sax: saxophone, Sita: sitar, Trum: trumpet, Cell: cello, Viol: violin, Wurl: Wurlitzer.

Discussion

This study investigated whether or not controlling for playback level and frequency colouration due to differences in playback equipment significantly affects the perceptual timbre space of a certain set of stimuli. It was motivated by the difficulty in having physical access to participants in order to conduct controlled laboratory experiments on timbre perception due to the outbreak of the COVID-19 global pandemic. Our analysis showed that the dissimilarity ratings obtained through the online test featured high internal consistency and, most importantly, were robust to the uncontrolled experimental conditions, at least for familiar instrumental tones and musically trained participants with no hearing impairments. This naturally led to a timbre space that exhibited strong similarity with the one resulted from the laboratory conditions. To put this in perspective, the configurational similarity quantified by the Tucker's congruence coefficient between the online and laboratory spaces was .97 when the respective value between the same laboratory condition and one other identical laboratory condition featuring participants that simply spoke a different native language was found to be .98 (Zacharakis et al. 2015).

Since this experiment concerns a more or less familiar set of sounds, it could be argued that the higher level cognitive mechanism for timbre recognition that has been shown to play a role in dissimilarity judgments (Siedenburg et al., 2016) and which seems to be also evident in the current timbre spaces, could account for the robustness of comparative timbre judgements to presentation conditions. Thus, if the categorical information conveyed by familiar stimuli was able to counterbalance the lack of controlled presentation conditions the next step would be to examine whether diminishing such information through the modification of known stimuli or the presentation of synthetic unfamiliar stimuli altogether would result in timbre space variability for an uncontrolled experimental condition. Indeed, there is already some evidence that presentation conditions (i.e., the existence of background noise) rearranges the timbre space of unfamiliar synthetic tones. In addition, there is also some evidence that the degree of exploitation of categorical information may be mediated by musical expertise (Siedenburg & McAdams, 2017). Therefore, despite this first positive step, future work needs to examine variables such as the balance between acoustic and categorical information present in the stimulus set as well as the level of musical expertise in order to come up with a comprehensive online testing protocol for the typically highly controlled task of pairwise dissimilarity rating. Although in this case the mean raw dissimilarities were already strongly correlated, an additional point of interest would be to test the suggested remedial properties of the non-metric MDS algorithm on noisy data (Shepard, 1966; Young, 1970).

Acknowledgments

This study was supported by a post-doctoral scholarship issued by the Greek State Scholarships Foundation (grant title: "Subsidy for post-doctoral researchers", contract number: 2016-050-050-3-8116), which was co-funded by the European Social Fund and the Greek State. The authors wish to thank the participants of the online experiment.

References

Abdi, H. (2007). The RV coefficient and congruence coefficient. In N. Salkind (ed.), *Encyclopedia of measurement and statistics*, (pp. *849*-853). Thousand Oaks: CA, Sage.

ANSI. (1973). *Psychoacoustical Terminology*, S3.20-1973, New York: American National Standards Institute.

Lorenzo-Seva, U., & Ten Berge, J. M. (2006). Tucker's congruence coefficient as a meaningful index of factor similarity. *Methodology*, *2*(2), 57-64.

McAdams S. (2019). The Perceptual Representation of Timbre. In: Siedenburg K., Saitis C., McAdams S., Popper A., Fay R. (eds), *Timbre: Acoustics, Perception, and Cognition* (pp. 23-57). Springer Handbook of Auditory Research, vol 69. Springer, Cham.

Shepard, R. N. (1966). Metric structures in ordinal data. *Journal of Mathematical Psychology*, *3*(2), 287-315.

Siedenburg, K., Jones-Mollerup, K., & McAdams, S. (2016). Acoustic and categorical dissimilarity of musical timbre: Evidence from asymmetries between acoustic and chimeric sounds. *Frontiers in Psychology*, *6*, 1977.

Siedenburg, K., & McAdams, S. (2017). The role of long-term familiarity and attentional maintenance in short-term memory for timbre. *Memory*, *25*(4), 550-564.

Woods, K. J., Siegel, M. H., Traer, J., & McDermott, J. H. (2017). Headphone screening to facilitate web-based auditory experiments. *Attention, Perception, & Psychophysics*, *79*(7), 2064-2072.

Woods, A. T., Velasco, C., Levitan, C. A., Wan, X., & Spence, C. (2015). Conducting perception research over the internet: a tutorial review. *PeerJ*, *3*, e1058.

Young, F. W. (1970). Nonmetric multidimensional scaling: Recovery of metric information. *Psychometrika*, *35*(4), 455-473.

Zacharakis, A., Pastiadis, K., & Reiss, J. D. (2015). An interlanguage unification of musical timbre: Bridging semantic, perceptual, and acoustic dimensions. *Music Perception*, *32*(4), 394-412.

Zacharakis, A., & Pastiadis, K. (2016). Revisiting the luminance-texture-mass model for musical timbre semantics: A confirmatory approach and perspectives of extension. *Journal of the Audio Engineering Society*, *64*(9), 636-645.

Zacharakis, A., Terrell, M. J., Simpson, A. J., Pastiadis, K., & Reiss, J. D. (2017). Rearrangement of Timbre Space Due To Background Noise: Behavioural Evidence and Acoustic Correlates. *Acta Acustica united with Acustica*, *103*(2), 288-298.