Acta Acustica united with Acustica
Vol. 103 (2017) 288 – 298

DOI 10.3813/AAA.919057

Rearrangement of Timbre Space Due To Background Noise: Behavioural Evidence and Acoustic Correlates

Asterios Zacharakis[1)], Michael J. Terrell[2)], Andrew J. R. Simpson[3)], Konstantinos Pastiadis[1)], Joshua D. Reiss[4)]

[1)] Aristotle University of Thessaloniki, School of Music Studies, Thessaloniki, Greece. aszachar@mus.auth.gr

[2)] Independent Researcher

[3)] University of Surrey, Centre for Vision, Speech and Signal Processing, Surrey, GU2 7XH, UK.

[4)] Queen Mary University of London, Centre for Digital Music, Mile End Road, London, E1 4NS, UK.

Summary

Studies of timbre are usually conducted in a "vacuum" of perfect silence. However, in the real-world, sounds are mostly heard in the presence of competing background noise. A series of pairwise dissimilarity listening tests on musically trained participants demonstrated how different levels of background noise can cause rearrangement of timbre spaces. Furthermore, it was shown that while spectral acoustic descriptors (e.g. spectral centroid or tristimulus values) seem robust under the presence of background noise, descriptors representing deviations from purely harmonic characteristics (e.g. inharmonicity) lose their salience for the higher noise level. Such results suggest that studies of timbre may need to take background noise into account in order to enhance their validity for real world applications.

PACS no. 43.66.DFc, 43.66.Jh, 43.66.Lj, 43.75.Zz

1. Introduction

In the real world acoustic signals are rarely perceived in absolute silence. Inevitably, this has triggered studies that assess the role of the physiology of auditory periphery on masking phenomena, elucidating complex mechanisms such as excitation overlap, suppression, etc. [1, 2, 3]. Evidence concerning the effects of noise on the neural representation of various sounds emerges also from the neuroimaging literature [4, 5, 6]. Several noise-related psychophysical phenomena (auditory masking being the most prominent) are induced by complex modifications of neural discharge rate patterns as well as phase-locking alterations. Mechanisms of masking have the potential to induce distortions to auditory representations as early as the cochlea and the first stages of the auditory pathway. The representations at subsequent and more central auditory stages may also be affected. Where the acoustic features of two sources overlap in space and time they may interact to produce masking [7] and/or intermodulation [8, 7]. Intermodulation results in the introduction of temporal coherence between independent features, and hence presumably impedes the otherwise clear segregation and grouping of features which constitute timbre [8, 9]. Human perception has also been shown to recruit mechanisms of central and peripheral noise compensation. The degree of success depends on several factors, such as the type and level of noise, its spectral and temporal statistics, the nature and spectral/temporal constituents of the signals of interest, to mention only but a few. The overall cognitive attitude, may also be determined from high-level intelligent processes such as those that refer to attention, analysis of auditory scenes, etc. Consequently, the effect of noise upon the auditory representations should not be considered obvious and should be studied carefully in order to unveil details of the interactions between the various noise and signal parameters and the physiological or behavioural responses.

However, relatively few studies have dealt with issues concerning behavioural responses to musical signals in noise. Instead, most studies focus on perception of noise-degraded speech [10, 11, 12, 13]. This is expected due to the importance of accuracy and quality in speech communication. More specifically, and aside from the numerous works on behavioural responses to speech in noise (SIN) (further discussion and references can be found in [14] and [15]), during the last decade there has been increasing interest in the potentially depictive and interpretative role of more objective measures obtained from recording and imaging techniques of the central nervous system (CNS). Both western and mandarin-type (namely with pitch contour variations) syllables in noise have been investigated, mostly with a family of techniques referred as Evoked Potentials (including Auditory Brainstem Response variants

Received 27 January 2016,
accepted 13 January 2017.

such as ABR and cABR). Such studies have provided evidence of the neural transcription of auditory stimuli within early, intermediate and late CNS structures and their functional role in the perception of speech in competing environments (e.g. noise). Additionally, they have offered objective evidence (such as superiority in timing of neural code, enhanced representation of harmonics, and less degradation of the response morphology in noise [11], or enhanced neural encoding of speech F_0 [13] for observed differences in performance between various groups of listeners (such as musicians vs. non-musicians) in such demanding tasks. Such findings may be seen as new space opening for the systematic study of various remediation strategies for several types of auditory processing difficulties and disorders [16].

Additionally, D'Ausilio et al. [12] using stimulation techniques (Transcranial Magnetic Stimulation) have suggested an advanced model of the functional role of various brain regions (entangling specific motor areas) which constructively inter-operate with perception-oriented brain areas to achieve better understanding of noisy speech signals, thus enlightening previously proposed (and criticized) articulatory-motor theories and models of speech perception (for details please refer to the works of A. M. Liberman, such as [17]). To summarise, the accumulation of such studies suggests that for speech discrimination or recognition tasks, noise compensation (i.e. extraction of exploitable information from noisy signals) involves several neural mechanisms and cognitive processes.

Despite the sizeable amount of research on the physiological mechanisms of auditory perception in noise, at the best of our knowledge, relatively few studies have attempted a comprehensive behavioural exploration of timbre in noise (e.g. [18]) and the potential alterations of timbral spaces per se in relation to the properties of interfering noise (e.g. type of noise, level, spectral profile, etc.) have not yet been examined. Instead, the majority of behaviourally oriented works has merely focused on confusions between speech sounds (for an overview, see [19]). The confusion patterns imply complex alterations of underlying perceptual spaces which may generally be attributed to the degree up to which specific acoustic properties are obscured. The identification of such perceptual spaces will shed some light on interesting issues such as, for example, whether and how certain acoustic properties are perceived under noise.

The aim of the present exploratory work was to present some evidence concerning the effects of interfering noise on the perception of multi-component signals and thus, to define more specific questions for addressing in future work. More specifically, it focused on the possible role of background wideband noise level on the perceived relationships between synthetic tones containing various types of spectral components and/or modulations. This approach aimed to investigate how a number of acoustic properties was perceptually affected by background noise.

To this end, we adopted the pairwise dissimilarity rating approach which has become a norm in timbre perception research [20, 21, 22, 23, 24], whereby listeners report the perceived distances between pairs of stimuli. Our experimental design included the presentation of the synthesised sound stimuli under three different background noise level conditions designated as: *silence*, *low noise* and *high noise*. Multidimensional Scaling (MDS) analysis was then utilised to yield three timbre spaces (corresponding to the three background noise conditions) from the acquired dissimilarity ratings. A comparison of the generated perceptual spaces revealed that the spatial configurations were sensitive to the presence of high levels of background noise. Findings were interpreted in terms of acoustic characteristics that either retain or lose their relevance with perceptual dimensions under noisy conditions. Our analysis demonstrated that the predictive ability of some descriptors (e.g., inharmonicity) was eliminated for the *high noise* condition while other acoustic correlates were found to be more robust under background noise.

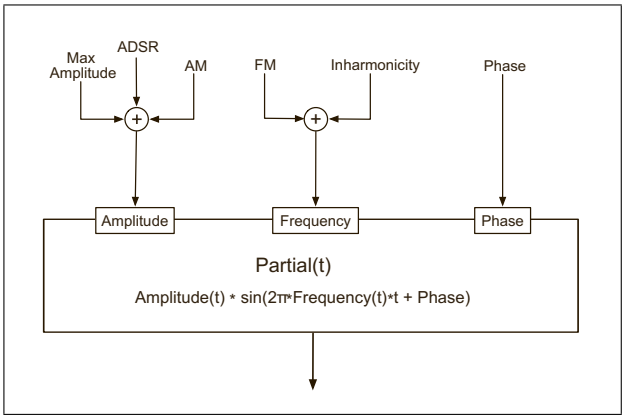


Figure 1. Partial level diagram of the additive synthesiser. The temporal development of amplitude for each partial is defined by a combination of maximum amplitude, ADSR envelope and sinusoidal amplitude modulation. The temporal development of frequency position for each partial is defined by an initial displacement from the ideal harmonic position together with a sinusoidal frequency modulation. Both AM and FM are defined by their width and frequency. Phase takes an angle from 0° to 360° as an input but this feature was not used for the preparation of this stimulus set.

2. Material and Method

2.1. Participants

Nine listeners (aged 22–41, mean age 29, 6 male and 3 female) with long term music practice (17.2 years on average, range: 10 to 25) participated in the listening test. They were all researchers from the Centre for Digital Music at Queen Mary University of London and highly aware of their hearing acuity. However, they were selected at random, without any more specific inclusion criterion (e.g. degree of technical skills/education, etc). They also had no prior training or knowledge of the test, and, consequently, they had been 'naive' [25] about it. All participants reported normal health and hearing, meeting the

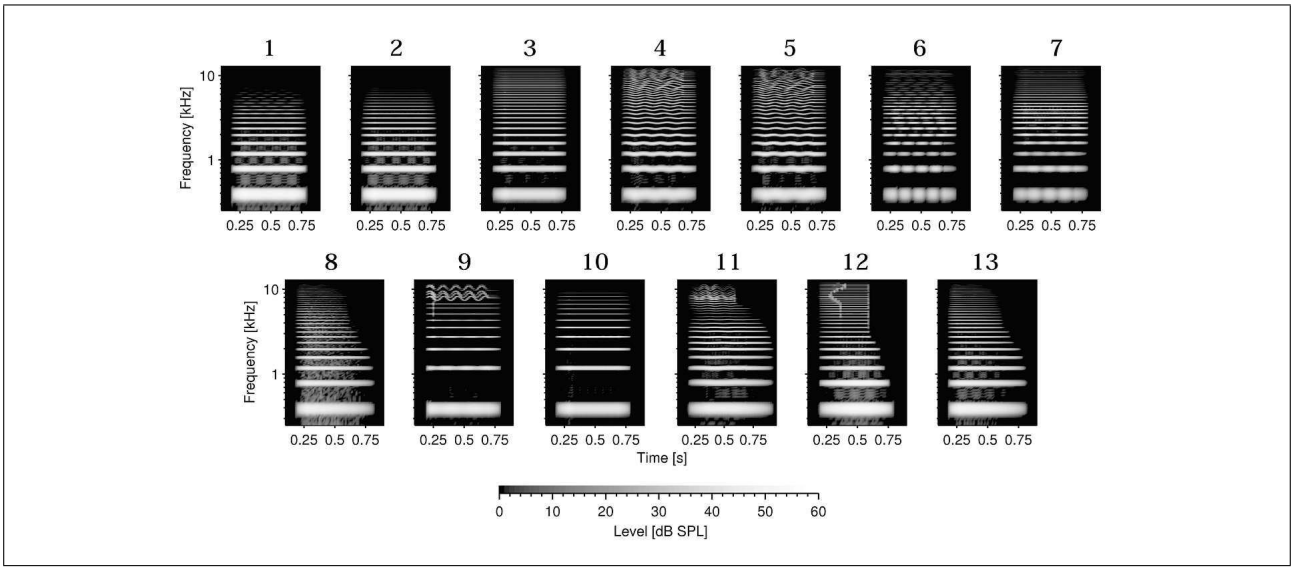


Figure 2. Stimuli spectrograms illustrating the spectro-temporal features of the stimuli. Panels **1** - **13** show the spectrograms of the thirteen respective sounds in the *silence* condition.

characteristics of otologically normal subjects as defined by ISO/FDIS 7029[1].

2.2. Stimuli and apparatus

Thirteen complex, tonal sounds were synthesised using additive synthesis. Each sound contained thirty nominal partials, which were independently controlled for; i) maximum amplitude, ii) envelope; amplitude and frequency modulation, and iii) inharmonicity. Figure 1 shows the partial module of the additive synthesiser. Each sound was 600 ms long and the inter stimuli interval was 400 ms. The fundamental frequency (f_0) was kept constant at 392 Hz (G4). Figure 2 shows the spectrograms of the 13 sound stimuli in *silence* condition that reveal the amount of variation of acoustic parameters featured in the stimulus set. Figure 3 and Table I show an example of the graphical interface settings that were applied for production of stimulus No.11 and the mean values of some extracted acoustic descriptors for each stimulus respectively. The definition of the acoustic features is given in Table IV.

Whilst the synthetic sounds we employed featured some typical characteristics of real-world musical sounds (i.e., pitch, large number of overtones, prominent harmonicity, ADSR-type temporal envelope), they did not resemble specific instruments. Hence, they were not likely to be subject to higher level and/or more abstract categorical cues to similarity, e.g. “this sound is a piano", that might enhance robustness of sound identification and therefore affect dissimilarity ratings under noisy conditions.

Prior to the listening test, the stimuli were equalised in loudness in an short listening test within the research team. The levels were adjusted separately for each condition (i.e., *silence*, *low noise*, *high noise*). One sound from the stimulus set was initially picked up as a reference and

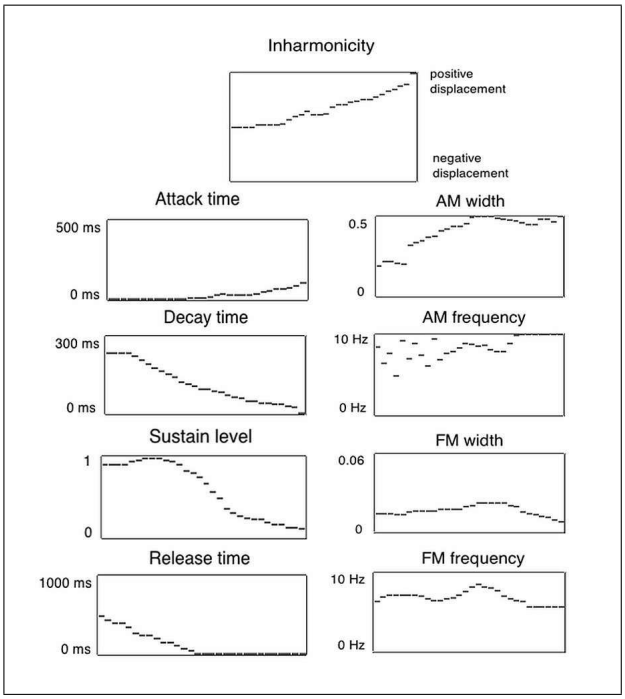


Figure 3. Settings of the additive synthesiser graphical interface for sound stimulus No.11. The upper box, labelled inharmonicity, represents harmonic displacement from the fundamental. In this case, positive displacement of the harmonic partials increases for higher partials. The left column shows the ADSR and the right column shows the frequency and width of the AM and FM for each partial. The numbering of y-axis for the AM and FM widths refers to the width variation as a proportion of the amplitude and the frequency position of the partial respectively.

[1] Otologically normal person is regarded a person in a normal state of health who is free from all signs or symptoms of ear disease and from obstructing wax in the ear canal and who has no history of undue exposure to noise, exposure to potentially substances, or familial hearing loss [26].

Table I. Mean feature values for the 13 sound stimuli in silence.

	SC_norm	SC_std	T1	T2	T3	Sp_dev	OER	Inharm.	Inharm_std	nsn	MCV
S1	3.28	0.87	0.31	0.46	0.23	0.06	2.10	0.05	$3.9 \cdot 10^{-3}$	$27.9 \cdot 10^{-3}$	$86.2 \cdot 10^{-3}$
S2	3.48	0.85	0.29	0.46	0.25	0.03	1.99	0.05	$3.9 \cdot 10^{-3}$	$28.1 \cdot 10^{-3}$	$84.7 \cdot 10^{-3}$
S3	10.46	0.35	0.10	0.26	0.64	0.11	1.17	0.17	$11.7 \cdot 10^{-3}$	$50.6 \cdot 10^{-3}$	$77.7 \cdot 10^{-3}$
S4	9.55	0.57	0.11	0.28	0.61	0.11	1.21	0.17	$50.2 \cdot 10^{-3}$	$72.3 \cdot 10^{-3}$	$78.7 \cdot 10^{-3}$
S5	8.71	0.67	0.11	0.31	0.56	0.20	1.40	0.15	$51.0 \cdot 10^{-3}$	$64.3 \cdot 10^{-3}$	$78.6 \cdot 10^{-3}$
S6	8.57	0.53	0.13	0.12	0.83	0.47	1.08	0.05	$3.8 \cdot 10^{-3}$	$50.5 \cdot 10^{-3}$	$77.5 \cdot 10^{-3}$
S7	8.93	0.51	0.05	0.10	0.86	0.14	0.99	0.14	$10.8 \cdot 10^{-3}$	$41.2 \cdot 10^{-3}$	$79.3 \cdot 10^{-3}$
S8	5.14	1.69	0.04	0.42	0.36	0.09	2.05	0.04	$4.3 \cdot 10^{-3}$	$36.3 \cdot 10^{-3}$	$79.21 \cdot 10^{-3}$
S9	7.94	1.26	0.22	0.15	0.60	0.97	3.45	0.08	$70.3 \cdot 10^{-3}$	$10.5 \cdot 10^{-3}$	$78.9 \cdot 10^{-3}$
S10	6.82	0.64	0.26	0.17	0.64	1.12	$151 \cdot 10^{6}$	0.01	$50.0 \cdot 10^{-3}$	$41.2 \cdot 10^{-3}$	$80.1 \cdot 10^{-3}$
S11	6.66	1.52	0.20	0.27	0.57	0.23	2.94	0.13	$33.5 \cdot 10^{-3}$	$56.3 \cdot 10^{-3}$	$81.8 \cdot 10^{-3}$
S12	8.30	5.17	0.16	0.29	0.43	0.10	3.94	0.05	$3.9 \cdot 10^{-3}$	$43.8 \cdot 10^{-3}$	$83.3 \cdot 10^{-3}$
S13	4.98	1.84	0.28	0.42	0.35	0.10	2.23	0.04	$4.1 \cdot 10^{-3}$	$27.3 \cdot 10^{-3}$	$81.3 \cdot 10^{-3}$

was set at a convenient[2] listening level. The rest of the stimuli were then equalised in loudness according to this reference by the first author resembling the classical up-down psychophysical procedure [28]. The equalised set was in turn evaluated by the rest of the authors. In each condition containing background noise, real-time generated white noise was presented continuously throughout the block. Figure 4 indicatively shows the effect of background noise on the spectrogram of sound stimulus No. 7.

In all three conditions, the resulting RMS playback level of the target sounds (i.e., not including the background noise) was measured to be approximately 60 dBA SPL (rms, slow response). In the *low noise* condition, the background noise level was 44 dBA SPL (rms, slow response). In the *high noise* condition, the background noise level was 68 dBA SPL (rms, slow response). Post-test, all participants reported that the level was comfortable for all stimuli and confirmed that loudness across stimuli was constant within blocks (i.e., within conditions). They also reported that the target sounds were somewhat quieter in *low noise* and considerably quieter in *high noise* conditions (though never inaudible).

The listening test was conducted under controlled conditions in an acoustically isolated listening room. Sound stimuli were presented through the use of a laptop computer, with a Tascam US122L external audio interface and a pair of Sennheiser HD600 circumaural headphones.

2.3. Procedure

Participants were asked to rate all the pairwise distances among the 13 sound stimuli within each separate condition (*silence*, *low noise* and *high noise*). Therefore, they rated the perceptual distances of 91 pairs (*same-sound* pairs included) within each of the three conditions. The rating was given using an unbounded scale (i.e., free magnitude estimation) [29] whereby they freely inserted a number of their choice to represent the overall dissimilarity of each pair, with 0 indicating a same pair[3]. The ratings were then normalised by dividing all the dissimilarities by the maximum response for each listener, thus providing a range from 0 to 1. The order of the three conditions, the order of the pairs within each condition and the order of the sounds within each pair were all presented randomly.

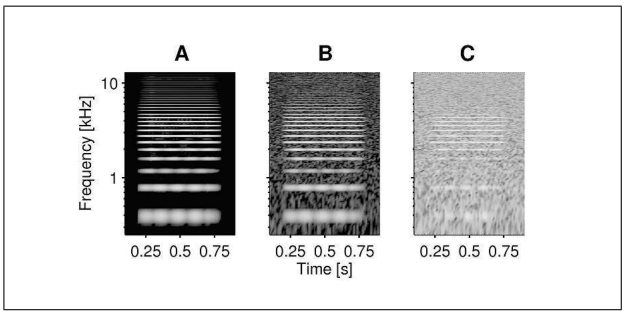


Figure 4. Background noise spectrograms showing the effect of background noise on a typical stimulus (sound index 7). **A** shows the spectrogram of the sound in the *silence* condition. **B** shows the spectrogram of the sound in the *low-noise* condition. **C** shows the spectrogram of the sound in the *high-noise* condition.

Prior to each condition, each listener was presented with the sounds of the entire stimulus set (within that condition) in random order, so as to become familiar with the overall timbral range. This was followed by a brief training session where listeners compared five selected pairs. The training data were discarded. Listeners were advised to maintain a consistent rating strategy throughout the experiment (i.e., to keep in mind that subsequent ratings should be scaled according to the assigned dissimilarity rating of the first pair which should be used as a reference).

[2] Combining effortless perception with safe playback level < 75 dBA [27].

[3] The free magnitude estimation method was favoured over bounded magnitude estimation as the latter introduces the following two issues during a rating procedure. Participants, not being in a position to anticipate upcoming dissimilarities, may never utilise the available range of the scale in case an even larger dissimilarity shows up later in the test. On the other hand, they may prematurely select the scale's maximum when their maximum rating should normally be appointed to an upcoming pair, thus clipping their intended response.

Listeners were permitted to listen to each pair of sounds as many times as necessary before submitting their dissimilarity rating. They were also encouraged to take regular breaks and were free to do so at any time. The overall listening test procedure, including instructions, lasted around one hour and a half for most of the participants.

3. Results

Before proceeding to the main body of the analysis we examined the internal consistency of our participant responses for each background noise condition. Cronbach's alpha was 0.87 for the *silence* condition 0.85 for the *low noise* condition and 0.94 for the *high noise* condition indicating an acceptable inter-participant reliability.

3.1. Timbre spaces for the three conditions

We subsequently used Multidimensional Scaling (MDS) [30, 31] to construct the geometric configuration of our stimuli timbre space, which allowed interpretation of dissimilarity data by Euclidean methods, e.g. the relations between the spaces and differences in their structure. Nonmetric weighted[4] MDS analysis [33, 34, 35, 36] was initially performed over a range of dimensionalities to determine the order most suitable to represent the timbre space for each presentation condition. Table II shows the evolution of two measures of fit (*Stress-I* and *DAF*) for orders of dimensionality between one and three[5]. Since the obtained Stress-I values should not exceed 0.2 which has been proposed as an acceptable maximum for such experiments [37] we adopt the 3D solution as optimal modelling of our data. Still, the Stress-I values are somewhat higher than the typical values proposed by [38] for the dimensionality and the number of points in our solution (Stress-I $\approx$ .148). However, as our data come from a sensory experiment one may tolerate such a relatively small excess due to the existence of measurement noise.

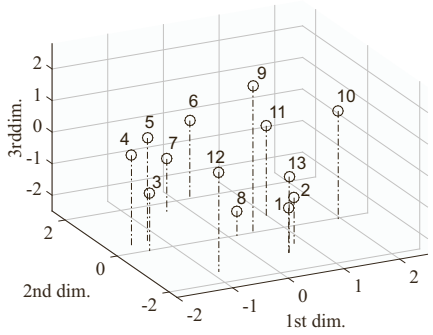
The 3-dimensional timbre spaces for each condition appear in Figure 5. Figure 6 shows the dendrograms from hierarchical clustering that elucidate the formation of stimuli relationships within the three perceptual spaces of Figure 5. The *silence* condition features four clusters of stimuli: 1-2-8-13, 3-4-5, 6-7 and 9-10-11. Stimulus 12 do not seem to group with any of those clusters while stimuli 3, 8 and 9 are loosely related with their corresponding clusters. This cluster formation is largely maintained for the *low noise* condition. The only notable differences is the breaking of cluster 1-2-8-13 into 1-2 and 8-13 and the deformation of 10-11 into 9-10 leaving stimulus 11 unclustered. Stimulus 8 differs from stimulus 13 only by an added white noise component (see Figure 2) which seems to be grouped with background noise thus making 8 and 13 essentially indistinguishable.

[4] The individual differences scaling (INDSCAL) algorithm was applied as offered by the SPSS PROXSCAL (proximity scaling) algorithm [32].

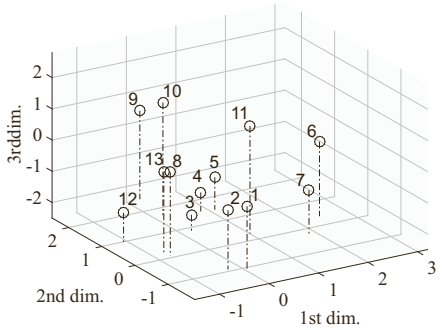
[5] *Stress-I* is a measure of misfit. The lower the value (to a minimum of 0) the better the fit. *DAF*: Dispersion Accounted for is a measure of fit. The higher the value (to a maximum of 1) the better the fit.

Table II. Measures of fit for different MDS dimensionalities for *silence*, *low noise* and *high noise* conditions.

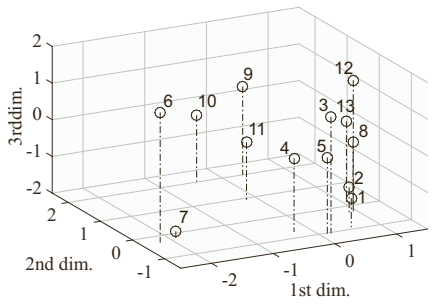
Condition	Dim.	Stress-I	Impro.	D.A.F.	Impro.
silence	1D	0.418	–	0.825	–
	2D	0.241	0.177	0.913	0.088
	3D	0.179	0.062	0.968	0.055
low noise	1D	0.406	–	0.835	–
	2D	0.253	0.153	0.935	0.100
	3D	0.177	0.076	0.968	0.033
high noise	1D	0.312	–	0.902	–
	2D	0.242	0.070	0.934	0.032
	3D	0.162	0.080	0.977	0.043



(a) Silence timbre space



(b) Low noise timbre space



(c) High noise timbre space

Figure 5. The three dimensional timbre spaces for the *silence*, *low noise* and *high noise* conditions.

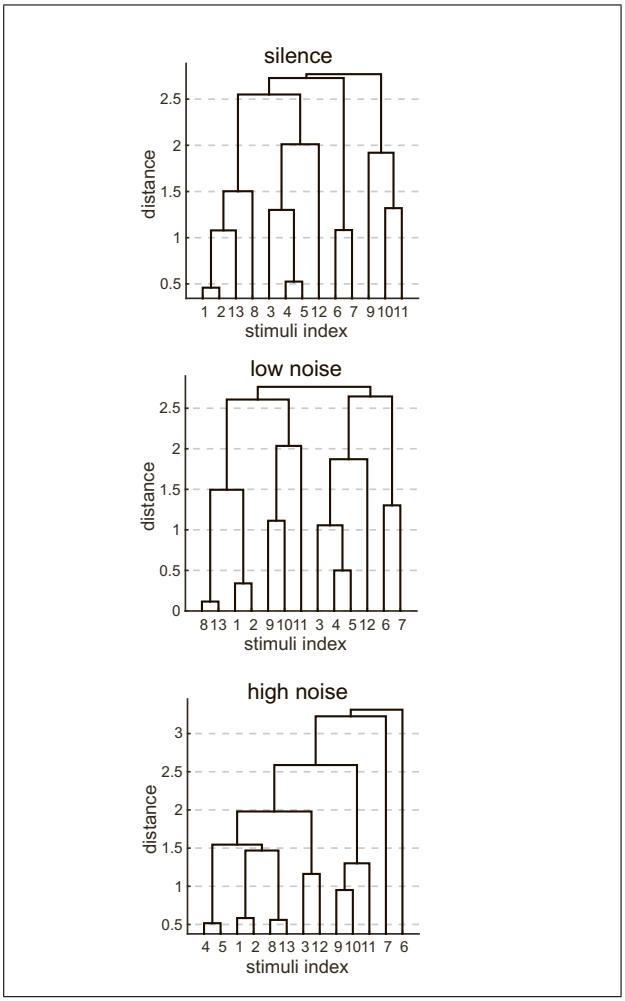


Figure 6. Dendrograms from hierarchical cluster analysis of *silence*, *low noise* and *high noise* conditions. The index numbers on the abscissa represent the thirteen stimuli used for the experiment.

The configuration of the *high noise* condition is mainly characterised by the formation of one major cluster of stimuli suggesting that differences between sounds were less perceivable and by the increased distances of stimuli 6 and 7 which implies that the presence of strong AM in the low frequencies was robust under noise.

The next two subsections will examine the extent to which the above qualitatively described variations are translated into statistically significant configurational and dimensional differences between the timbre spaces and the potential acoustic interpretation of these differences.

3.2. Configurational and dimensional similarity between perceptual spaces

In this section the relationships between the perceptual spaces which were obtained through non-metric MDS (NMDS) of the listeners' responses for each one of the three background noise conditions (e.g. *silence*, *low noise* and *high noise*) are investigated in terms of their configurational similarity, wherein the examined sounds represented the objects of the configurations. The configurational similarity reflects the similarity of the solid shapes defined by the swarms of objects within the spaces. Any global form of similarity between spaces should also take into account the orientation of the swarms relative to the axes and the scales of the spaces, and henceforth will be called dimensional similarity, thus ascribing the notion of the direct relationships between the dimensions of the extracted NMDS spaces.

Similar to the approach in [29], the configurational similarity between spaces was judged by two indices computed from the distances between the objects; the Tucker's Congruence coefficient [39, 38] for ranked distances and the m^2 statistic for Procrustes analysis [40, 41, 42]. The rank-based congruence coefficient offers a narration of gross similarity of ?shapes? which are formed by the swarms of objects in the compared MDS spaces, whereas the m^2 provides a rigorous quantification. The explicit or implicit use of ranks in the assessment of goodness-of-fit has been extensively reported in the literature, covering aspects of the MDS problem from as early as the estimation of stress (e.g. rank-images method), [38, 43] to the estimation of similarity between distance matrices (e.g. rank-based Mantel tests, CADM, etc. see [44, 45, 46]). The m^2 resembles a measure of alienation 1-r^2 (where r is the correlation coefficient between the sequences of within-space distances of the two examined spaces) [47, 48]. The exploitation of both indices was chosen for reasons of completeness and by the fact that, generally, no single measure of configurational similarity may be considered as globally adequate to depict the relationship between two examined spaces [49, 38].

As a guideline, for the congruence coefficient, values larger than 0.92 are considered good/fair (more loosely, values 0.85-.94 may also be considered good/fair according to [50]), and values larger than 0.95 practically imply perfect equivalence between configurations [51]. 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) [52, 53]. Regarding the m^2 statistic, values < 0.75 (based on recommendations for $r^2 > 0.25$ as described in [54]) signify a large effect size.

For the m^2 statistic, an approach which follows randomization testing and is suited to Procrustes analysis under the name of PROcrustean randomization TEST (PROTEST) has been proposed [55, 56, 47]. According to this, significance is tested using a large number of random permutations of the original data. The statistical significance of r^2 (derived from m^2) has only been investigated in few studies [57, 58, 49] which showed that critical values for r^2 varied with dimensionality of configurations and number of objects.

Table III summarizes the values of the congruence coefficient and m^2 for the relationships of configurations between all examined spaces. The configurations of *silence* and *low noise* spaces show a higher degree of similarity (congruence coefficient = 0.95, well above the statistical significance of $p = .05$, and $m^2 = 0.16$ highly significant).

Table III. Congruence coefficients, m^2, r^2 and RV-mod for the mutual relationships across timbre spaces as described in the schema of Figure 5. ($^{**}p < .01$). m^2 significance testing with PROTEST method. CC: Congruence coefficient (expected value, SD), Expected chance value, estimated by bootstrap with 10000 runs.

Relationship	CC	$\mathbf{m^2}$	$\mathbf{r^2}$	RV-mod
Silence-LN	0.95 (.85, 0.02)	0.16^{**}	0.84	0.79^{**}
Silence-HN	0.89 (.80, 0.02)	0.39^{**}	0.60	0.53^{**}
LN-HN	0.90 (.81, 0.02)	0.36^{**}	0.64	0.58^{**}

Between *silence* and *high noise* both the congruence coefficient and m^2 are worse than between *silence* and *low noise* (congruence coefficient = 0.89, $m^2 = .39$, both statistically significant). Similar results are observed between the *low noise* and *high noise* conditions. Despite the fact that the configurational similarity seems to be generally retained, it is clear that a considerable deterioration has taken place for high levels of noise.

Next, we proceeded with an assessment of the dimensional similarity between spaces, namely the relationships between the dimensions of the extracted NMDS spaces. Instead of one-by-one comparisons between the respective dimensions, we chose to follow an ensemble relationship approach. That is, we relied on techniques that attempt to assess the association between two data tables, where rows represent the individual objects (sounds) within each space, and columns (as variables) represent the dimensions of the space (i.e. table of sound coordinates on each respective dimension).

Out of the several approaches that have been proposed we selected the modified RV coefficient [59, 60, 61, 62] as a measure of overall (dimensional) similarity between the spaces. The RV coefficient for matrices plays a role analogous to the correlation coefficient between two variables.

The last column of Table III presents the modified RV coefficient between the MDS spaces. The RV coefficient is 0.79 between the *silence* and *low noise* spaces. However, and in a similar manner to the configurational similarity results, it is evident that it drops considerably when the comparisons refer to the *high noise* condition (*low noise* vs. *high noise*: 0.58, *silence* vs. *high noise*: 0.53). This drop of the dimensional similarity for the *high noise* condition is in accordance with the previously reported drop of configurational similarity.

3.3. Acoustic correlates

Trying to explain the differences in the spatial configurations exhibited in the *high noise* condition we examined the correlations between some acoustic features (measuring spectral content, spectral fine structure, spectrotemporal characteristics and inharmonicity) and the timbre space dimensions for all conditions. The acoustic features were extracted for the *silence* condition using the spectral modelling synthesis (SMS) MATLAB platform [66]. The window length applied was 2,048 samples ($f_s = 44.1$ kHz) with an overlapping factor of 93.75%, the zero padding factor was 2, and 30 harmonic partials were extracted for all sounds. Apart from the mean value, the standard deviation of each acoustic descriptor was also computed in an effort to capture elements of the time-variant behaviour of the sounds. Table IV presents the abbreviations and the definitions of all the features that exhibited significant correlations with the perceptual dimensions.

Table V presents all the acoustic descriptors that featured significant correlations with the dimensions of the three timbre spaces. Interestingly, while the energy distribution of harmonic partials (the normalised spectral centroid and the tristimulus values), spectral fine structure (spectral deviation) and some spectrotemporal information (standard deviation of the SC) retain their association with dimensions of all spaces, inharmonicity and its standard deviation, do not seem to have predictive ability for any of the *high noise* space dimensions. This presents some evidence that the configurational and dimensional differences observed between the *high noise* space and the other two could be attributed to a degradation of perception regarding inharmonic and noisy components of the target sounds.

4. Discussion

The main goal of this study was to demonstrate that background noise can cause rearrangement of timbre spaces of complex tones. Through MDS analysis of dissimilarity data, we displayed and compared the organisation of the perceptual spaces between three listening conditions regarding the level of background noise, namely *silence*, *low noise* and *high noise*. Quantitative measures show that both dimensional and configurational congruence decreases significantly between *silence* and *high noise* conditions. Although this may appear as an expected finding, it had not been previously shown in a timbre space level. As noted in the introduction, background noise may affect perception at various processing stages. However, we do not offer an in-depth psychophysical analysis or modelling of the possible mechanisms that may be involved in our results. Our aim was to provide with evidence and build upon a perspective of treatment for important perceptual phenomena that characterise the relationships between auditory objects and signal and/or noise features.

Subsequently we attempted to highlight the way that the identified psychological dimensions are altered at higher noise levels by considering the extent to which acoustic correlates of the *silence* space are preserved in noisy timbre spaces. Whereas all three dimensions for the *silence* and *low noise* conditions exhibited several acoustic correlates, *high noise* dimensions were less explained. More specifically, the 3rd *high noise* dimension showed no correlation with any of the examined acoustic descriptors and the 2nd dimension correlated merely with odd even ratio. The standard deviation of the spectral centroid along with odd even ratio exhibited an increasing influence at higher levels of background noise. Although a designation

Table IV. Abbreviations and definitions of the significant audio features.

Feature	Abbreviation	Explanation
Normalised Spectral Centroid	SC_norm	normalised barycenter of the harmonic spectrum [63]
Tristimulus 1, 2, and 3	T1, T2, T3	Relative amplitudes of the 1st, the 2nd to 4th and the 5th to the rest harmonics [64]
Mean Coefficient of Variation	MCV	Variation of the first 9 harmonics over time [65]
SC standard deviation	SC_std	Standard deviation of SC over time
Spectral deviation	Sp_dev	Metric of the harmonic spectrum fine structure [63]
Odd Even Ratio	OER	Ratio of the energy contained in odd versus even harmonics [63]
Inharmonicity	Inharm.	Metric of the frequency displacement of partials relatively to a purely harmonic sequence [63]
Inharm. standard deviation	Inharm_std	Standard deviation of inharmonicity over time

Table V. Spearman's correlation coefficients between acoustic descriptors and perceptual dimensions of the three spaces ($^{*}p < .05$, $^{**}p < .01$). Only significant correlations are depicted.

	Sil_1	Sil_2	Sil_3	LN_1	LN_2	LN_3	HN_1	HN_2	HN_3
SC_norm	-0.78**	0.55*		0.56*		0.59*	0.56*		
SC_std		0.59*		0.69**			0.76**		
T1		0.81**		0.76**			0.74**		
T2		0.68*		0.66*			0.75**		
T3		0.71**		0.71**			0.82**		
Sp_dev		0.60*	0.76**	0.64*	0.70**		0.71**		
OER	0.58*	0.56*				0.61*		0.69**	
Inharm.	0.70**	0.62*		0.63*		0.64*			
Inharm_std			0.56*		0.65*				
MCV	0.60*	0.65*		0.60*					

of acoustically derived labels for the perceptual dimensions does not seem straightforward, it is clearly demonstrated that acoustic features like inharmonicity and its standard deviation, along wi th the mean coefficient of variation lose their predictive ability for the *high noise* condition. A possible explanation could be that such effects can be attributed to inharmonic, amplitude modulating and noise-like characteristics of the target sound being incorporated into the background noise [67, 9, 68, 69]. This finding could be further examined utilising the questioning of previous works regarding the sensitivity for detection of changes (differences) between timbral entities [70] and the way the contrasts of timbral features are altered under the presence of noise.

For the moment, our approach provides a behaviour-based treatment of the way and the degree to which acoustic information is retained and contributes to the formation of perceptual relations between sounds (e.g. perceptual spaces) in noisy conditions. As future work, we propose the exploitation of computational auditory modelling for the analysis of auditory representations as a means for the study of acoustic information processing throughout the auditory pathway (e.g. cochlear models, STRFs, etc. [71, 72, 73, 74]. Such an approach will be of interest for several applications (e.g. signal processing and telecommunications, psychoacoustics, special education and communication, etc.) As examples, we could consider the improvement of methods for the assessment of Auditory Processing Disorders (APD) [10] or the incorporation of musical education/experience as a beneficial factor for speech perception in noise [11, 75]. Further, the investigation of relationships between auditory modelling based representations and the perceptual spaces will also facilitate interpretations on the share of specific physiological, psychophysical or, possibly, even cognitive phenomena and mechanisms (such as masking, grouping, etc.) in the formation of perception. A necessary following step is to use a combination of more realistic background noises (traffic, car engine, electrical appliances noises etc.) and stimuli (natural sound sources such as musical instruments or speech).

Finally, our previous work has demonstrated that there exists a clear relation between timbre perception and its semantic dimensions [76, 29]. Thus, a potential research direction would be to explore whether the effects of background noise are limited to timbre perception or are also extended to the domain of semantics.

5. Conclusion

In this article, we investigated the robustness of a timbre space for different levels of background noise. We acquired pairwise dissimilarity ratings on a group of synthesised musical tones under three different background noise conditions: *silence*, *low noise* and *high noise*. A comparison of the generated perceptual spaces for the three different conditions revealed that both psychological dimensions and configurations are sensitive to the presence of

higher levels of background noise. Although the questioning of whether the observed changes take place in the periphery, as a result of auditory masking, or reflect higher level processes is intriguing, our adopted approach remains at the level of listeners' responses. Additionally, we sought to explain the above findings in terms of acoustic correlates of perceptual dimensions under noisy conditions. Features that capture various aspects of our stimuli (spectral, spectrotemporal, inharmonicity, etc.) were extracted for the *silence* condition and were subsequently correlated with the perceptual dimensions for all three conditions. This analysis demonstrated that the predictive ability of features representing deviation from pure harmonicity was eliminated for the *high noise* condition while others (mainly describing static spectra) were proven more robust.

Acknowledgement

This work was partially supported by the EPSRC grant EP/E045235, Platform Grant for the Centre for Digital Music.

References

[1] B. Delgutte: Physiological mechanisms of psychophysical masking: Observations from auditory-nerve fibers. J. Acoust. Soc. Am. **87** (1990) 791–809.

[2] A. Recio-Spinoso, N. P. Cooper: Masking of sounds by a background noise–cochlear mechanical correlates. The Journal of Physiology **591** (2013) 2705–2721.

[3] W. S. Rhode, C. D. Geisler, D. T. Kennedy: Auditory nerve fiber response to wide-band noise and tone combinations. Journal of Neurophysiology **41** (1978) 692–704.

[4] F. Liang, L. Bai, H. W. Tao, L. I. Zhang, Z. Xiao: Thresholding of auditory cortical representation by background noise. Frontiers in Neural Circuits **8**.

[5] J. A. Costalupes: Representation of tones in noise in the responses of auditory nerve fibers in cats. I. Comparison with detection thresholds. The Journal of Neuroscience **5** (1985) 3261–3269.

[6] M. B. Sachs, H. F. Voigt, E. D. Young: Auditory nerve representation of vowels in background noise. Journal of Neurophysiology **50** (1983) 27–45.

[7] J. H. McDermott, D. Wrobleski, A. J. Oxenham: Recovering sound sources from embedded repetition. Proceedings of the National Academy of Sciences **108** (2011) 1188–1193.

[8] M. A. Stone, B. C. J. Moore, C. Füllgrabe, A. C. Hinton: Multichannel fast-acting dynamic range compression hinders performance by young, normal-hearing listeners in a two-talker separation task. J. Audio Eng. Soc. **57** (2009) 532–546.

[9] D. Deutsch: The Psychology of Music, Academic Press, San Diego, chapter Grouping mechanisms in music. 2nd edition, 1999, 299–348.

[10] E. Skoe, N. Kraus: Auditory brain stem response to complex sounds: a tutorial. Ear and Hearing **31** (2010) 302–324.

[11] A. Parbery-Clark, E. Skoe, N. Kraus: Musical experience limits the degradative effects of background noise on the neural processing of sound. The Journal of Neuroscience **29** (2009) 14 100–14 107.

[12] A. D'Ausilio, I. Bufalari, P. Salmas, L. Fadiga: The role of the motor system in discriminating normal and degraded speech sounds. Cortex **48** (2012) 882–887.

[13] J. H. Song, E. Skoe, K. Banai, N. Kraus:. Perception of speech in noise: Neural correlates. Journal of Cognitive Neuroscience **23** (2011) 2268–2279.

[14] R. H. Wilson, R. A. McArdle, S. L. Smith: An evaluation of the bkb-sin, hint, quicksin, win materials on listeners with normal hearing and listeners with hearing loss. Journal of Speech, Language, Hearing Research **50** (2007) 844–856.

[15] J. L. Sperry, T. L. Wiley, M. R. Chial: Word recognition performance in various background competitors. Journal-American Academy of Audiology **8** (1997) 71–80.

[16] J. Slater, E. Skoe, D. L. Strait, S. O'Connell, E. Thompson, N. Kraus: Music training improves speech-in-noise perception: Longitudinal evidence from a community-based music program. Behavioural brain research **291** (2015) 244–252.

[17] A. M. Liberman, I. G. Mattingly: The motor theory of speech perception revised. Cognition **21** (1985) 1–36.

[18] B. C. Moore, A. Sek: Discrimination of modulation type (amplitude modulation or frequency modulation) with and without background noise. J. Acoust. Soc. Am. **96** (1994) 726–732.

[19] S. A. Phatak, J. B. Allen: Consonant and vowel confusions in speech-weighted noise. J. Acoust. Soc. Am. **121** (2007) 2312–2326.

[20] J. M. Grey: Multidimensional perceptual scaling of musical timbres. J. Acoust. Soc. Am. **61** (1977) 1270–1277.

[21] R. A. Kendall, E. C. Carterette: Perceptual scaling of simultaneous wind instrument timbres. Music Perc. **8** (1991) 369–404.

[22] P. Iverson, C. L. Krumhansl: Isolating the dynamic attributes of musical timbre. J. Acoust. Soc. Am. **94** (1993) 2595–2603.

[23] S. McAdams, S. Winsberg, S. Donnadieu, G. D. Soete, J. Krimphoff: Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, latent subject classes. Psychological Research **58** (1995) 177–192.

[24] A. Caclin, S. McAdams, B. K. Smith, S. Winsberg: Acoustic correlates of timbre space dimensions: A confirmatory study using synthetic tones. J. Acoust. Soc. Am. **118** (2005) 471–482.

[25] N. Zacharov, V.-V. Mattila: GLS-A generalised listener selection procedure. In Audio Engineering Society Convention 110. Audio Engineering Society, 2001.

[26] International organization for standardization. acoustics - statistical distribution of hearing thresholds related to age and gender. 2000.

[27] W. Melnick: Human temporary threshold shift (tts) and damage risk. J. Acoust. Soc. Am. **90** (1991) 147–154.

[28] S. A. Gelfand: Hearing: An introduction to psychological and physiological acoustics, CRC Press, 2016.

[29] A. Zacharakis, K. Pastiadis, J. D. Reiss: An interlanguage unification of musical timbre: bridging semantic, perceptual and acoustic dimensions. Music Perc. **32** (2015) 394–412.

[30] R. N. Shepard: The analysis of proximities: Multidimensional scaling with an unknown distance function: I. Psychometrika **27** (1962) 125–140.

[31] R. N. Shepard: The analysis of proximities: Multidimensional scaling with an unknown distance function: II. Psychometrika **27** (1962) 219–246.

[32] J. J. Meulman, W. J. Heiser: PASW Categories 18, Chapter 3. SPSS Inc., Chicago, 2008.

[33] J. B. Kruskal: Multidimensional scaling by optimizing goodness-of-fit to a nonmetric hypothesis. Psychometrika **29** (1964) 1–28.

[34] J. B. Kruskal: Nonmetric multidimensional scaling: a numerical method. Psychometrika **29** (1964) 115–130.

[35] R. N. Shepard: Metric structures in ordinal data. Journal of Mathematical Psychology **3** (1966) 287–315.

[36] F. W. Young: Nonmetric multidimensional scaling: Recovery of metric information. Psychometrica **35** (1970) 455–473.

[37] S. Chollet, D. Valentin, H. Abdi: Novel Techniques in Sensory Characterization and Consumer Profiling, CRC Press, chapter Free Sorting Task, 2014, 207–228.

[38] I. Borg, P. J. F. Groenen: Modern Multidimensional Scaling: Theory and Applications, Springer, New York. 2nd edition, 2005, 1–614.

[39] L. R. Tucker: A method for synthesis of factor analysis studies. Technical report, DTIC Document, Washington, DC: Department of the Army. 1951. [Personnel Research Section Report No. 984].

[40] J. C. Gower: A general coefficient of similarity and some of it's properties. Biometrics **27** (1971) 857–871.

[41] J. C. Gower: Generalized procrustes analysis. Psychometrika **40** (1975) 33–51.

[42] J. C. Gower, G. B. Dijksterhuis: Procrustes Problems. Oxford University Press, Oxford, New York, 2004.

[43] M. C. Hout, S. D. Goldinger, K. J. Brady: MM-MDS: A multidimensional scaling database with similarity ratings for 240 object categories from the massive memory picture database. PLoS ONE **9** (2014) e112 644.

[44] P. Legendre, F.-J. Lapointe: Assessing congruence among distance matrices: Single-malt scotch whiskies revisited. Australian & New Zealand Journal of Statistics **46** (2004) 615 – 629.

[45] J. W. Schneider, P. Borlund: Matrix comparison, Part 2: Measuring the resemblance between proximity measures or ordination results by use of the mantel and procrustes statistics. Journal of the American Society for Information Science and Technology **58** (2007) 1596–1609.

[46] V. Campbell, P. Legendre, F.-J. Lapointe: The performance of the congruence among distance matrices (CADM) test in phylogenetic analysis. BMC Evolutionary Biology **11**.

[47] P. R. Peres-Neto, D. A. Jackson: How well do multivariate data sets match? The advantages of a Procrustean superimposition approach over the Mantel test. Oecologia **129** (2001) 169–178.

[48] J. Oksanen: Multivariate analysis of ecological communities in R: vegan tutorial. University Oulu, Finland, 2013.

[49] I. Borg, D. Leutner: Measuring the similarity of MDS configurations. Multivariate Behavioral Research **20** (1985) 325–334.

[50] J. M. F. t. Berge: Rotation to perfect congruence and the cross validation of component weights across populations. Multivariate Behavioral Research **21** (1986) 41–64.

[51] U. Lorenzo-Seva, J. M. F. t. Berge: Tucker's congruence coefficient as a meaningful index of factor similarity. Methodology: European Journal of Research Methods for The Behavioral and Social Sciences **2** (2006) 57–64.

[52] B. Efron, R. J. Tibshirani: An Introduction to the Bootstrap. Chapman and Hall/CRC, New York, softcover reprint of the original 1st ed. 1993 edition, 1994.

[53] F. Cutzu, S. Edelman: Faithful representation of similarities among three-dimensional shapes in human vision. Proceedings of the National Academy of Sciences of the United States of America **93** (1996) 12 046–12 050.

[54] P. D. Ellis: The Essential Guide to Effect Sizes: Statistical Power, Meta-Analysis, the Interpretation of Research Results, Cambridge University Press, Cambridge: New York. 1st edition, 2010, 31–44.

[55] D. Jackson: PROTEST: a PROcrustean randomization TEST of community environment concordance. Ecoscience. Sainte-Foy **2** (1995) 297–303.

[56] P. Legendre, L. F. J. Legendre: Numerical Ecology, Elsevier, New York, 1998, 546.

[57] F. M. Andrews, R. F. Inglehart: The structure of subjective well-being in nine western societies. Social Indicators Research **6** (1979) 73–90.

[58] R. Langeheine: Statistical evaluation of measures of fit in the Lingoes-Borg procrustean individual differences scaling. Psychometrika **47** (1982) 427–442.

[59] H. Abdi: The RV coefficient and the congruence coefficient. In N. Salkind, ed., Encyclopedia of Measurement and Statistics, SAGE Publications, Inc., 2455 Teller Road, Thousand Oaks, California, 91320, United States, 2007.

[60] A. K. Smilde, H. a. L. Kiers, S. Bijlsma, C. M. Rubingh, M. J. van Erk: Matrix correlations for high-dimensional data: the modified RV-coefficient. Bioinformatics (Oxford, England) **25** (2009) 401–405.

[61] C.-D. Mayer, J. Lorent, G. W. Horgan: Exploratory analysis of multiple omics datasets using the adjusted RV coefficient. Statistical Applications in Genetics and Molecular Biology **10** (2011) 1–27.

[62] J. Josse, S. Holmes: Measures of dependence between random vectors and tests of independence. Literature review. arXiv preprint arXiv:1307.7383 0.

[63] G. Peeters, B. L. Giordano, P. Susini, N. Misdariis, S. McAdams: The Timbre Toolbox: Extracting acoustic descriptors from musical signals. J. Acoust. Soc. Am. **130** (2011) 2902–2916.

[64] H. Pollard, E. Jansson: A tristimulus method for the specification of musical timbre. Acustica **51** (1982) 162–171.

[65] R. A. Kendall, E. C. Carterette: Verbal attributes of simultaneous wind instrument timbres: II. Adjectives induced from Piston's Orchestration. Music Perc. **10** (1993b) 469–502.

[66] X. Amatriain, J. Bonada, A. Loscos, X. Serra: Spectral Processing, John Wiley and Sons Ltd, Chichester, England, 2002, 373–438.

[67] A. S. Bregman: Auditory scene analysis: The perceptual organization of sound, Cambridge, MA: MIT Press, chapter Integration of Simultaneous Auditory Components, 1994, 213–394.

[68] S. A. Shamma, M. Elhilali, C. Micheyl: Temporal coherence and attention in auditory scene analysis. Trends in neurosciences **34** (2011) 114–123.

[69] S. Teki, M. Chait, S. Kumar, S. Shamma, T. D. Griffiths: Segregation of complex acoustic scenes based on temporal coherence. eLife **2** (2013) e00 699.

[70] R. A. Kendall:. Difference thresholds for timbre related to amplitude spectra of complex sounds. Ph.D. thesis, University of Kansas, 1975.

[71] R. Meddis, E. Lopez-Poveda, R. R. Fay, A. Popper: Computational Models of the Auditory System. Springer, New York, London, 2010.

[72] B. N. Pasley, S. V. David, N. Mesgarani, A. Flinker, S. A. Shamma, N. E. Crone, R. T. Knight, E. F. Chang: Reconstructing speech from human auditory cortex. PLoS Biol **10** (2012) e1001 251.

[73] D. J. Klein, P. König, K. P. Körding: Sparse spectrotemporal coding of sounds. EURASIP Journal on Advances in Signal Processing (2003) 1–9.

[74] D. A. Depireux, J. Z. Simon, D. J. Klein, S. A. Shamma: Spectro-temporal response field characterization with dynamic ripples in ferret primary auditory cortex. Journal of Neurophysiology **85** (2001) 1220–1234.

[75] F. Liu, A. R. Maggu, J. C. Y. Lau, P. C. M. Wong: Brainstem encoding of speech and musical stimuli in congenital amusia: evidence from Cantonese speakers. Frontiers in Human Neuroscience **8** (2015) 1029.

[76] A. Zacharakis, K. Pastiadis, J. D. Reiss: An interlanguage study of musical timbre semantic dimensions and their acoustic correlates. Music Perc. **31** (2014) 339–358.