DERIVING A TIMBRE SPACE FOR THREE TYPES OF COMPLEX TONES VARYING IN SPECTRAL ROLL-OFF

William L. Martens[1], Mark Bassett[2] and Ella Manor[3]

Faculty of Architecture, Design and Planning
University of Sydney, Sydney NSW 2006, Australia

[1]Email: william.martens@sydney.edu.au
[2]Email: mark.bassett@sydney.edu.au
[3]Email: ella.manor@sydney.edu.au

Abstract

Three types of complex harmonic tones were created by adjusting the relative amplitudes of selected sets of even-numbered or odd-numbered harmonics, relative to the more homogenous pattern of harmonic amplitudes associated with an 'Oboe-like' timbre, which timbre comprised the first of the three types of tones employed in this study. The other two types of tones were produced either by the reduction of even-harmonic amplitudes to create a 'Clarinet-like' timbre, or by the reduction of odd-harmonic amplitudes to create an 'Organ-like' timbre. In all three cases the overall harmonic amplitude envelope could be described by a simple spectral roll-off parameter, that parameter being the rate of amplitude attenuation over increasing frequency, measured in dB/octave. For each of these three types of complex harmonic tones, further variation was introduced into the stimulus set by small adjustments in the rate of spectral roll-off in harmonic amplitude (which included four attenuation rates that ranged from 3 dB/octave to 7.5 dB/octave, in incremental steps of 1.5 dB/octave). Thus a set of 12 timbres was constructed that differed perceptually along one continuous dimension (identified with the auditory attribute termed 'sharpness') and one categorical dimension (related to perceived musical-instrument character, nominally identified as 'Oboe-like', 'Clarinet-like' and 'Organ-like'). All pairwise comparisons of these 12 timbres were presented to four listeners for evaluation in terms of overall timbral dissimilarity for each pair, without regard to particular identifiable auditory attributes.

The collected dissimilarity ratings were treated as estimates of inter-stimulus distances between the Cartesian coordinates of the stimuli configured in a two-dimensional (2D) perceptual space, which was derived using INdividual Differences SCALing (INDSCAL) analysis. INDSCAL was employed to produce two useful outputs: First it produced the abovementioned perceptual space (termed 'Stimulus Space') for the group of four listeners as a whole. Second, it produced estimates of the differences in weighting that each of the four listeners placed on the resulting dimensions (which weightings are captured by INDSCAL in terms of a 'Subject Space'). The INDSCAL analysis of data from these four listeners revealed very small differences in the perceptual salience that each dimension holds for each listener. Finally, the group Stimulus Space coordinates on the continuous perceptual dimension (identified with the 'sharpness' attribute) were modelled using a two-term regression equation that included the conventional physical measure designed to predict variation in 'sharpness' and also a term confirming significant dependence on the odd-to-even harmonic amplitude ratios of the tones.

1. Introduction

This investigation of the multidimensional perceptual differences between three families of complex harmonic tones was motivated by a need to design and generate a set of sound stimuli for a timbral ear-training program, which was initially presented in a recent paper by McKinnon-Bassett et al. [1]. While a full description of this timbral ear-training program is beyond the scope of the current paper, it is most appropriate to begin with this underlying motivation for the study presented here, so that the work is put into a proper context. In particular, it was the need for a unified scheme for parametric synthesis of a set of musical timbres that could be presented to listeners as part of an ear-training program, for which tight control of the perceptual differences between the stimuli could be achieved. Furthermore, those stimuli were designed in order to be identified readily by listeners as more or less natural sounding examples of a small assortment of musical instruments exhibiting timbral variation typical in musical performance. Ultimately, the goal was to enable parametric control over timbre synthesis along two perceptual dimensions, the first being a continuous dimension (identified with the auditory attribute termed 'sharpness') and the second being a categorical dimension (related to perceived musical-instrument character, with tones resembling Oboe, Clarinet, and Organ).

With the context for the current work established, this introduction turns to a more general treatment of the tradition of timbre research within which the present study is situated, beginning with a presentation of the scope of this research along with a technical definition for the term 'timbre' as it has come to be understood. As the timbre of musical instrument tones is generally regarded as a multidimensional perceptual phenomenon, it is difficult to give it a precise definition; nonetheless, in the most general case, a standard definition given for the term 'timbre' can be found in ANSI S1.1-1194 [2], which reads as follows:

> "That attribute of auditory sensation which enables a listener to judge that two nonidentical sounds, similarly presented and having the same loudness and pitch, are dissimilar."

Although timbre is a term that usually has been given such a negative definition (i.e., a definition stating not what timbre is, but rather what it isn't - that timbre is what differs between tones when pitch and loudness do not differ), some of its identifiable component dimensions can be listed. Listing other attributes that are not timbral attributes, such as duration, may also narrow the definition of the term further. For the discussion to follow in this paper, a distinction is made between time-variant timbral components and more global timbral components (i.e., attributes of a whole sound event, rather than its components that are discriminable over time). One such global timbral component is tone colouration; a term that may be more narrowly defined than is the term timbre. Thus, it may be useful in studying timbre to note that two musical notes played at the same pitch, loudness, and duration may also be matched in tone colouration, and yet those two musical notes may still differ in timbre. Of course, the term tone colouration may also be regarded as multidimensional. Nonetheless, at least one timbral attribute of steady sounds seems to be readily distinguished, an attribute typically identified either as 'brightness' or 'sharpness.' This point is made quite clear in the following quotation from the popular text entitled "*Psychoacoustics: Facts and Models*" which has provided a foundation for the definition of auditory attributes in this field:

> "Previously, there has been a tendency to transfer everything in steady-state sounds not related to the sensations of loudness or pitch, to a residual basket of sensations called timbre. Using this definition of timbre, it is necessary to extract from the mixture of sensations those that may be important. The sensation of "sharpness" . . . seems to be one of these." (Zwicker and Fastl [3], p. 215)

It is even more clarifying to note that tone colouration is most easily defined for steady sounds with no spectral evolution, since colouration ratings can be predicted directly from a sound's steady-state spectrum [4]. Though tone colouration is certainly not a unidimensional perceptual attribute, it certainly can be described by a lower dimensional structure than timbre can be. For example, the perceptual space associated with steady-state vowel sounds has only two highly salient dimensions,

and these are well predicted by the two prominent formant frequencies of vowel sound spectra in the region ranging from around 300 to around 3000 Hz [5]. It should be noted that these 'vowel-colouration' results were based upon the analysis of inter-stimulus distance using MultiDimensional Scaling (MDS), which is the analytical approach taken in the current research as well. Such results are not necessarily replicable in studies employing Semantic Differential (SD) analysis [6]. For example, when such vowel sounds were included in a much larger set of stimuli varying widely in spectral envelope, only the verbal descriptor 'sharpness' was found to usefully differentiate between the sounds:

> "The portion of timbre not accounted for by sharpness did not appear to be verbally describable in a psychologically usable manner. As both the tone and noise stimuli of equal pitch, loudness, and sharpness did sound quite similar, their remaining relatively small perceptual differences should best be analysed with the evidently successful MDS methods." (G. von Bismarck [6], p. 157)

As mentioned above, the current research was motivated by a need to design a set of sound stimuli for use in timbral ear-training, along with a unified scheme for parametric synthesis of this set of musical timbres. So, as a readily identifiable dimension of timbral variation, 'sharpness' seems to be a good choice as one continuum along which synthetic musical timbres were to be manipulated for the present application. The second dimension in terms of which synthetic musical timbres were manipulated here affords a categorical distinction between the timbres of different musical instruments (again, nominally identified as 'Oboe-like', 'Clarinet-like' and 'Organ-like'). In fact, it is precisely the human listener's ability to make this three-way categorical distinction between perceived musical-instrument character that was most desirable in the application of the results of this study in the design of set s of stimuli for timbral ear-training. What was uncertain, and requiring of some preliminary investigation, was the relative perceptual salience of this categorical distinction between the three synthetic musical instrument timbres, compared to the magnitude of timbral variation in 'sharpness' associated with manipulation of spectral roll-off of harmonic amplitudes. The current study was undertaken to address directly the question of the relative salience of these two perceptual dimensions through collection of global dissimilarity ratings.

Such a direct dissimilarity-based investigation of two timbral dimensions is not without precedent. A similarly simplistic study examining the relative influence of two parameters on perceptual similarity of complex tones was reported in 1969 by Plomp and Steeneken [7], which featured the now classic comparison of the relative salience of phase differences versus differences in harmonic amplitude patterns. They presented tones equal in loudness and pitch, but having harmonics amplitudes with varying attenuation rates, and having their component 10 harmonics in either sine or cosine phase. The stimuli were presented successively in triads, and the listener's task was to select from three stimuli the two that were most similar and the two that were most dissimilar. These choices made by eight listeners were tallied to create cumulative similarity indices estimating the perceptual dissimilarity for all pairwise comparisons between their eight stimuli. The results obtained using MultiDimensional Scaling (MDS) analysis on these data revealed that the maximal effect of phase on timbre was quantitatively smaller than the effect of changing the slope of the amplitude pattern by 2 dB/octave. This study by Plomp and Steeneken [7] provided some inspiration for the current study, in which multiple types of complex tones were also varied in the slope of their harmonic amplitude pattern (here, in steps of 1.5 dB/octave). But perhaps even more inspirational was Plomp's [8] audio demonstration of tone colour variation that was featured on a compact disc of demonstrations that he authored in 1998, entitled (in Dutch language) *Hoe wie horen. Over de toon die de muziek maakt* (How we hear. On the tone that makes music). Track 4 of that CD presents steady-state complex tones of the three types included in the current study (sounding as 'Oboe-like', 'Clarinet-like' and 'Organ-like'), and demonstrates the timbral variation associated with the manipulation of spectral roll-off of harmonic amplitudes (which varies in the demonstration from 0 dB/octave to 9 dB/octave, in incremental steps of 3 dB/octave).

2. Methods

2.1 Listeners

Four normal-hearing listeners participated in the dissimilarity-rating task. Three of the listeners were the authors of this paper, and one additional listener was included who was naive with regard to the goals of the current experiment.

2.2 Stimuli

Each of the 12 steady-state complex tones explored in this experiment can be described as a periodic fluctuation of sound pressure *p*, over time *t*, and can be represented by the following equation:

$$p(t) = \sum_{n=1}^{N} a_n \sin(2\pi nft + \varphi_n) \quad (1)$$

The tone colour of such steady-state complex tones primarily depends upon the pattern of harmonic amplitudes a_1, a_2, a_3, …, a_n, and the phase pattern, φ_1, φ_2, φ_3, …, φ_n, as described in more depth in the companion paper that reports the preliminary results upon which the current work builds [9]. As in those previous experiments, the 'Oboe-like' waveform was produced by progressively summing up 32 component harmonics to generate a single cycle of a complex waveform (i.e., containing 32 harmonics of varying amplitude), but here those waveforms were generated at only four spectral roll-off values that ranged from 3 dB/octave to 7.5 dB/octave, in incremental steps of 1.5 dB/octave). As in the previous study [9], the fundamental frequency of all stimuli was set to 311 Hz (i.e., a musical pitch of D#4). Three types of synthetic tones were produced from sets of 4 complex waveforms using a conventional Attack-Decay-Sustain-Release (ADSR) envelope, which attempts to shape the tone's temporal characteristics to match those observed in musical instruments. The set of 'Clarinet-like' waveforms contained energy only at the odd harmonic frequencies, and the set of 'Organ-like' waveforms contain energy at the first and third harmonics, but otherwise contained energy only at the even harmonics (i.e., at harmonic numbers 1, 2, 3, 4, 6, 8, 10, …, 32). The left-side panel of Figure 1 shows how the three waveforms appear at a spectral roll-off value of 6 dB/octave.

2.3 Procedure

Listeners were required to provide global dissimilarity ratings for all pairwise comparisons of the above-described complex tones. The 12 tones were presented via Sennheiser HD600 headphones at a comfortable listening level (nominally 75 dB SL). Each listener completed three blocks of 132 trials, which is the number of comparisons resulting from the exclusion of the diagonal entries of the 12 x 12 matrix of dissimilarities (i.e., excluding all comparisons between identical stimuli). For each pair of tones, listeners recorded their dissimilarity ratings using the onscreen Graphical User Interface (GUI) that is pictured in the right side panel of Figure 1. The sound stimuli were presented serially, separated by a 500-ms delay. On-screen instructions prompted listeners to indicate how similar they thought the stimuli sounded, with the leftmost response indicating that the stimuli sounded maximally dissimilar, and the rightmost response indicating that the stimuli sounded most similar. All listeners had to develop their own criterion for the anchoring point of maximal dissimilarity during an initial practice run in which all 132 pairwise comparison trials were completed. After the initial practice run of 132 trials, each listener completed an additional two runs of 132 trials. The dissimilarity data matrices produced by each listener in these final two runs were averaged to produce a single dissimilarity data matrix for each listener, which provided the input for subsequent analysis.

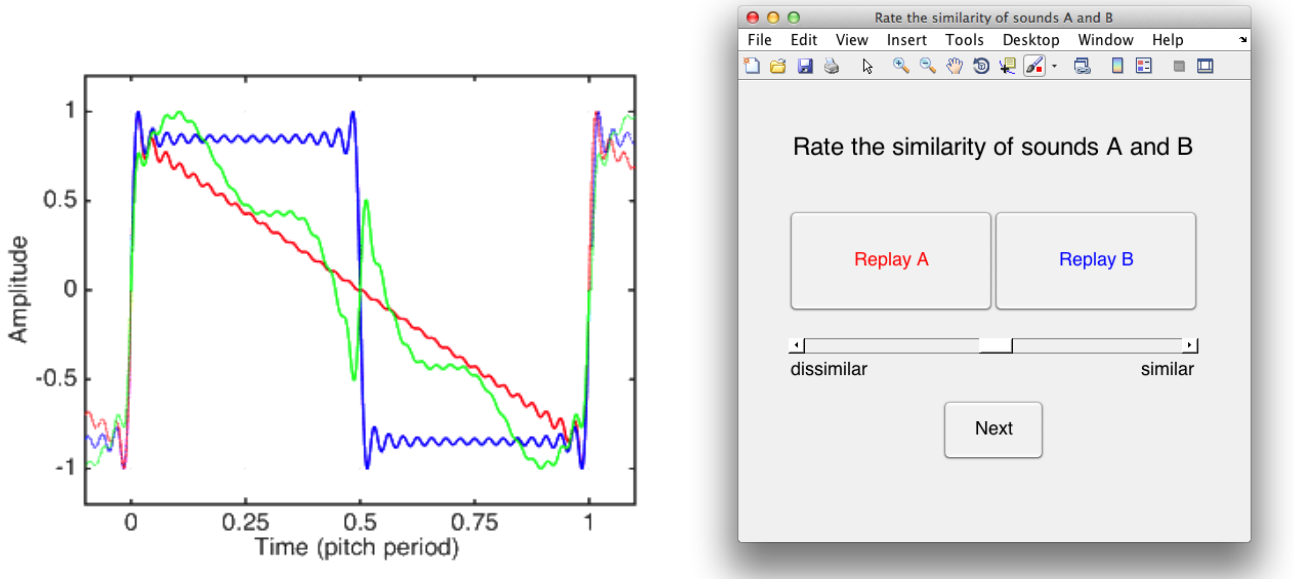

Figure 1. Left: Three waveforms exemplifying the timbral types presented in the current study, generated with spectral roll-off of 6 dB/octave (red, blue and green curves for the 'Oboe-like', 'Clarinet-like' and 'Organ-like' musical tones, respectively). Right: The Graphical User Interface (GUI) used in this study to allow listeners to indicate the inter-stimulus dissimilarity for each pair of stimuli.

3. Results

To begin with, the analysis of dissimilarity data produced by one of the four listeners will be examined separately (i.e., for this one individual only). Subsequently, an INDSCAL analysis will be performed upon the combined dissimilarity data matrices collected from all four listeners. The first step in such an analysis is typically to examine the quality of the assumed relationship between the data and the derived perceptual space. Each stimulus is assigned coordinates in the Stimulus Space, and the distances between the points denoted by these coordinates should show a high correlation with the scaled disparities calculated from the input inter-stimulus dissimilarities. The conventional means for examining whether such a correlation exists between the scaled disparities and the inter-stimulus dissimilarities is the Shepard plot, two examples of which are shown in Figure 2.

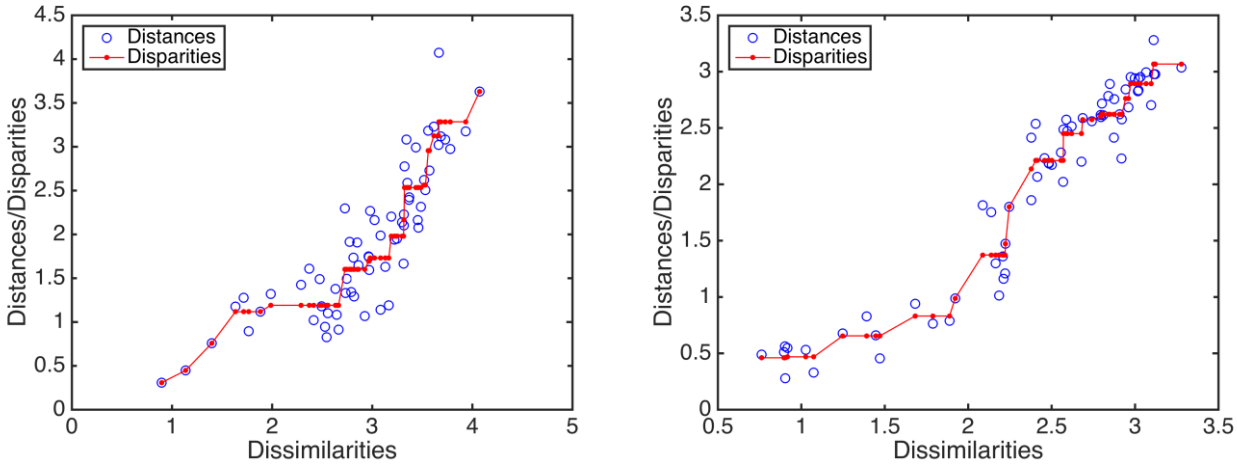

Figure 2. The left panel shows a Shepard plot for one listener, illustrating the general agreement between the scaled disparities and the inter-stimulus distance as a function of the inter-stimulus dissimilarities that are the sole source of data from which the MultiDimensional Scaling (MDS) analysis derives its output configuration termed 'Stimulus Space.' The right panel shows a Shepard plot produced by MDS analysis of the combined dissimilarity data collected from all four listeners.

The left panel of Figure 3 shows the configuration of the 2D Stimulus Space that was derived by an MDS analysis of the averaged dissimilarity ratings produced during two sessions, in each of which one listener gave responses for all pairwise comparisons of 12 stimuli. The inter-stimulus distances effectively are based upon four dissimilarity ratings produced by this listener, since the order in which pairs of stimuli were presented was counterbalanced in order to avoid order effects (each pair was session presented in each session in both orders). A non-metric MDS analysis was performed using the Matlab function *'mdscale'* with the goodness-of-fit criterion to minimize *'stress'* (which is the default for this routine, and is normalised by the sum of squares of the inter-point distances). Results for the one listener shown in the left panel of Figure 3 indicate a *stress* value of 0.113, which is consistent with a reasonably good fit between inter-stimulus dissimilarities and distances. Also, a visual inspection of the resulting 2D Stimulus Space for this one listener reveals an easily interpretable configuration of points. The MDS-derived coordinates of the 12 stimuli on the first, most salient dimension of the Space clearly correspond to the variation in perceived sharpness expected given the experimental manipulation of the rate of spectral roll-off in harmonic amplitude (which included four attenuation rates that ranged from 3 dB/octave to 7.5 dB/octave. For each of the three types of timbres presented, the Dimension 1 coordinates along this presumably continuous perceptual dimension correspond to monotonically increasing magnitude for the auditory attribute identified as 'sharpness.' The diamond-shaped plotting symbols labelled with numerals 1 through 4 show this monotonic increase for the 'Oboe-like' timbre, as do the square and circular plotting symbols for 'Clarinet-like' and 'Organ-like' timbres, respectively. The Dimension 2 coordinates, on the other hand, seem to correspond to a categorical distinction between the three timbral types presented, as was expected between the groups of four stimuli sharing common even or odd harmonic amplitude patterns, but varying in sharpness within each group.

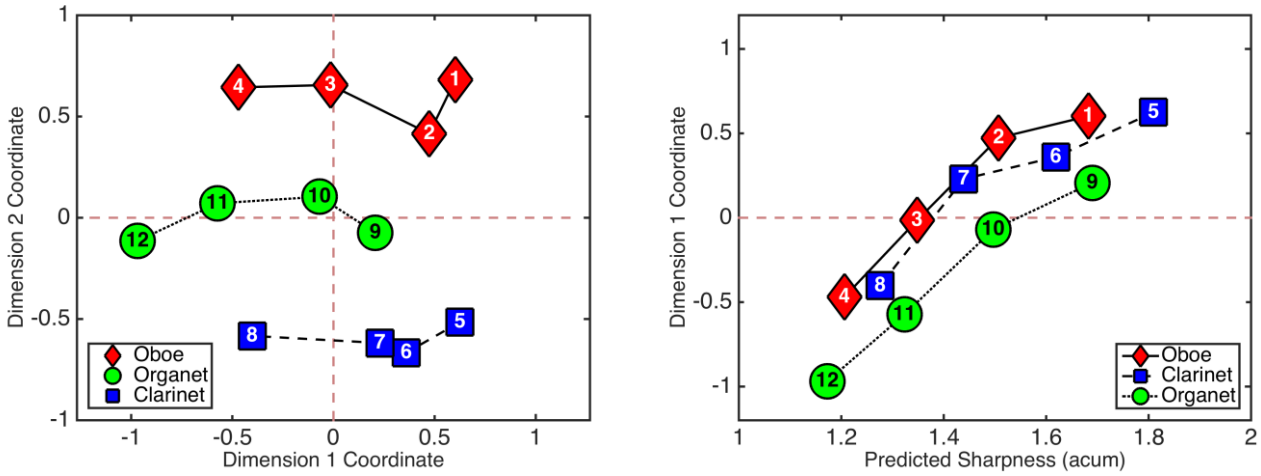

Figure 3. The left panel shows the 2D Stimulus Space derived by an MDS analysis of the dissimilarity ratings produced by one listener for all pairwise comparisons of 12 stimuli. The right panel plots as a function of the predicted sharpness of those 12 stimuli, the coordinates of the stimuli on the first, most salient dimension of the MDS-derived Stimulus Space. In both of these graphs, distinct plotting symbols are used to differentiate between the three types of complex harmonic tones, and these are connected via distinct types of lines according to the increasing spectral roll-off values.

The next question that begs to be asked is whether the MDS-derived coordinates in this perceptual space can be predicted from an analysis of the physical stimuli. It was expected that such a prediction would be successful for the Dimension 1 coordinates using is a conventional measurement that Zwicker and Fastl [3] describe in Chapter 9 of their book "*Psychoacoustics: Facts and Models.*" Their predicted sharpness values are anchored to the sharpness of a reference stimulus, and measured using a unit termed the *acum* (which means "sharp" in Latin). The right panel of Figure 3 plots the MDS-derived coordinates on Stimulus Space Dimension 1 as a function of the predicted sharpness values of the 12 stimuli.

The *acum* scaling is referenced to the perceived sharpness of a narrow-band noise one critical-band wide with a centre frequency of 1 kHz having a sound pressure level of 60 dB. The perceived sharpness of this narrow-band stimulus establishes the 1 *acum* point on a perceptual scale for the sharpness of other stimuli, which should be heard as less sharp than any of the 12 stimuli presented in this study. Most interesting to note in the current results is that the sharpness predictor values, although highly correlated with Dimension 1 coordinates for this one listener, do not show an offset in sharpness values corresponding to the Stimulus Space offset for 'Organ-like' timbres (plotted using circular symbols in Figure 3). This mismatch between predicted and obtained coordinates along the 'sharpness' continuum suggests that some adjustment for timbral type might be appropriate in this case. It will be of interest, then, to determine whether a similar mismatch is observed in the combined results of more listeners. Furthermore, it may be that whether some adjustment to the prediction could be successful using a physical measure of the differences between timbral types, such as that proposed in the Manor et al. [9] companion paper.

For the analysis of dissimilarity data produced by all four of the listeners who participated in the current study, an INDSCAL analysis was performed upon the juxtaposed dissimilarity data matrices. Whereas the MDS results for the one listener indicated a *stress* value of 0.113, the INDSCAL results for four listeners indicated a *stress* value of .062, which is a considerably better fit between inter-stimulus dissimilarities and distances than was found for a single individual. That being said, the INDSCAL-derived Stimulus Space shown in the left panel of Figure 4, which takes into account data from all four listeners, was not as readily interpretable as was the configuration of points observed in the MDS-derived coordinates that were based upon data from a single listener. Nonetheless, the coordinates of the stimuli on the first, most salient dimension of the INDSCAL-derived Stimulus Space are strictly monotonically related to predicted sharpness, as is clearly seen in the right panel of Figure 4. More striking in Figure 4 than the results that were shown in Figure 3, for a single listener, is the even wider separation in the combined results from four listeners between the Dimension 1 coordinates of the 'Organ-like' timbres (plotted using circular symbols) and the Dimension 1 coordinates of the other two types of timbres ('Clarinet-like' and 'Oboe-like'). For the current work, it is an important goal to determine whether this mismatch in perceived 'sharpness' (as quantified in terms of Dimension 1 coordinates) can be incorporated into a prediction model for the set of 12 stimuli presented here.

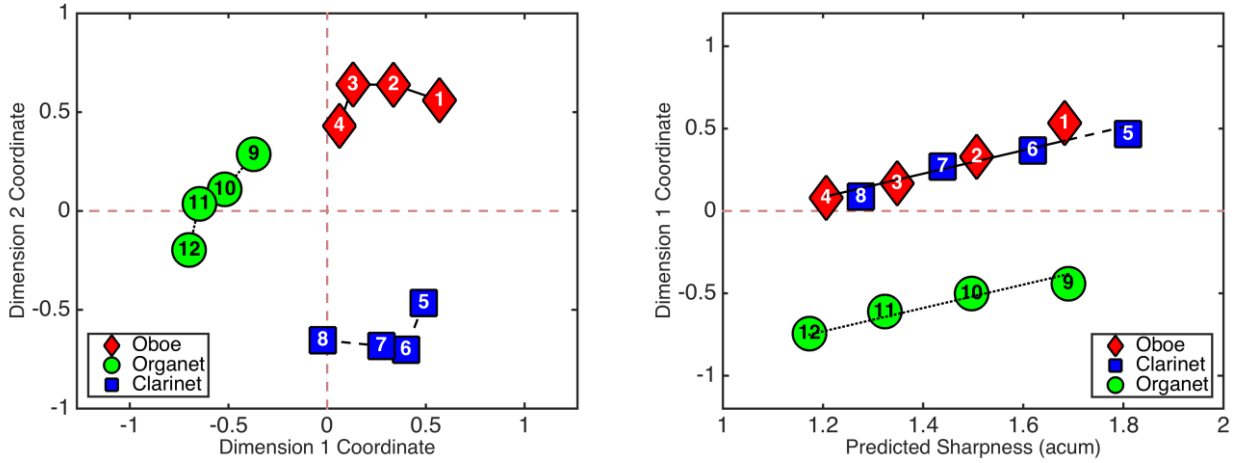

Figure 4. The left panel shows the 2D Stimulus Space derived by an INDSCAL analysis of the dissimilarity ratings produced by all four listeners for pairwise comparisons of 12 stimuli, using the same plotting symbols as in Figure 3. The right panel plots as a function of the predicted sharpness of those 12 stimuli, the coordinates of the stimuli on the first, most salient dimension of the INDSCAL-derived Stimulus Space. Note that the lines connecting the plotting symbols in the right panel result from the below-described multiple regression analysis.

Table 1. The different weights that each listener put on the two dimensions of the group Stimulus Space that was derived by an INDSCAL analysis of the dissimilarity data collected from four listeners.

	Dimension	
Listener Number (*i*)	***s*=1**	***s*=2**
Listener 1	0.481	0.444
Listener 2	0.478	0.440
Listener 3	0.474	0.459
Listener 4	0.457	0.473

Prior to the attempt to account for the observed variation in the 'sharpness' attribute, the question is addressed whether four listeners differed significantly on the weight that each put on the two different dimensions of the group Stimulus Space that was derived by the INDSCAL analysis. The two weights that were generated by INDSCAL for each listener are shown in Table 1, which were calculated as follows: For each of individual listener *i* of *O* cases, dissimilarity judgments between stimulus *j* and stimulus *k* are collected for all pairwise comparisons of a given set of *M* stimuli. The first INDSCAL computation, given this *M*-by-*M* matrix of input dissimilarity judgments, is to create a matrix ***D*** of distance estimates (with elements d_{ijk}) between stimulus *j* and stimulus *k* for each individual case *i*. Two output data matrices, ***W*** and ***Y***, are created from these distance estimates according to the following model (which for the current analysis assumed a Euclidean distance metric):

$$d_{ijk} = \sum_{s=1}^{p} \sqrt{w_{is}(y_{js} - y_{ks})^2} \tag{2}$$

The output data matrix ***Y*** contains the points y_{js} for each stimulus *j* on each dimension *s* of the group Stimulus Space of dimensionality *p*. The output data matrix ***W*** contains the weights w_{is} for each individual case *i* on each dimension *s* of that derived Stimulus Space. These weights define a unique point for each individual listener *i* on each dimension *s* of the Subject Space with the same dimensionality *p* as that of the group Stimulus Space. Note that the values show in Table 1 are quite similar across listeners, and so there is little evidence for a difference in the spatial structure underlying the dissimilarity judgments produced by each individual listener. Therefore, the group Stimulus Space configuration was assumed to provide a good estimate of the configuration of points that might be found were data from many listeners to be collected.

Finally, assuming that the observed variation along the first dimension of the group Stimulus Space corresponds primarily to variation in the 'sharpness' attribute, a multiple regression model was fit to the coordinates of the stimuli along the Dimension 1 coordinates shown in Figure 4, as a function of two predictor variables: The first was the conventional weighted spectral centroid measure provided by Zwicker and Fastl [3] that has been termed here 'predicted sharpness,' and the second was a categorical dummy variable corresponding to the presence of odd harmonics in the stimulus, such as the 'Oboe-like' and 'Clarinet-like' musical tones versus the 'Organ-like' musical tones, with an absence of odd harmonics. It should be noted that when only predicted sharpness is included in the regression equation, the coefficient of determination was quite low, at R^2=0.21. Setting the dummy variable to zero for 'Organ-like' tones, and to one otherwise, produced a much better fitting prediction equation, with a coefficient of determination of R^2=0.99 (The regression lines fit to the coordinates for each timbral type appear in the right panel of Figure 4). Such a close fit might seem surprising; however, it is an obvious consequence of how different the Dimension 1 coordinates were for complex tones with similar predicted sharpness but very different odd versus even harmonic amplitudes. What remains to be discussed here is what this result might mean for subsequent listening tests that employ these stimuli in the context of timbral ear training, which is taken up in the final section of this paper.

4. Discussion and Conclusion

It is clear from the results of both the MDS analysis of dissimilarity data collected for a single listener, and from the INDSCAL analysis of data from four listeners, that a simple 2D Stimulus Space is able to capture a similarity structure that corresponds well with the two manipulated synthesis parameters. Also, the INDSCAL analysis of data from the four listeners revealed that differences in the perceptual salience of the two dimensions were negligible, as illustrated by the Subject Space weights listed in Table 1. Therefore, the group Stimulus Space coordinates on the continuous perceptual dimension (identified with the 'sharpness' attribute) were modelled successfully using a two-term regression equation that accounted for 99% of the variance on this dimension. The predictability of the INDSCAL-derived coordinates here suggests that listeners are able to 'hear out' the variation in the 'sharpness' attribute as a separate feature of the timbre of members of a set of complex tones that also vary in character, that character being identified as 'Oboe-like,' 'Clarinet-like' and 'Organ-like' for the set of 12 stimuli presented in the current study. That being said, it is also clear that the conventional predictor of perceived sharpness provided by Zwicker and Fastl [3] must be adjusted for complex tones that are identified as 'Organ-like' relative to the other two timbral types, and so a simple Cartesian composition of the two independent dimensions is not indicated here. This implies that a timbral ear training application requiring a factorial combination of the timbral factors investigated here will also not be designed in such a straightforward manner. Nonetheless, the current results do inform designers how to proceed to ensure stimuli intended for such timbral ear training are perceptually distinct from one another along these two dimensions. The complexity of this result also raises a final question that should be discussed here – a question which has to do with how to associate physical predictors with a Stimulus Space for timbre, which is a topic that has been addressed by many studies, but still resists simple solution (see Krumhansl [10] for a good overview).

A closely paper by Terasawa et al. [10] also described a derived perceptual space for the timbre of steady-state complex tones, and detailed an objective metric that took into account perceptual orthogonality of timbral dimensions, and also measured the quality of timbre interpolation. Although their study included a detailed investigation of spectral fine-structure, there was no selective manipulation of odd versus even harmonic amplitudes, and therefore their measurements do not bear upon the Dimension 2 coordinates in the current study. They determined that a spectral measure based on Mel-Frequency Cepstral Coefficients (MFCC) provided a good foundation for a timbre space prediction model. The criteria they applied in evaluating the prediction model included linearity and orthogonality, which they regarded as most important in establishing a basic perceptual structure onto which additional features might be added. One obvious additional feature to add would be the distinctions that were introduced in the current study via manipulation of odd versus even harmonic amplitudes, which is a clearly identifiable timbral feature that is often missing from automated feature extraction for musical sound (see for example [12]). Of course, feature extraction based upon the absence of even harmonic energy is likely to fail given that such energy is not typically so attenuated in actual clarinet tones [13]. Note also that this feature need not be regarded as strictly categorical. Indeed, a gradual interpolation between 'Oboe-like' and 'Clarinet-like' timbres has been observed in the recent study by Manor et al. [9] (the current paper's companion paper).

Although a smooth interpolation could be heard as the 'Oboe-like' tone's even-harmonics were gradually attenuated to around 9 dB, when the attenuation reached around 12 dB the timbral character became most clearly 'Clarinet-like' for most listeners. Although there was good agreement between listeners on the shift in identification from one timbral type to another, it was certainly not the case in that study that the 'Oboe-like' and 'Clarinet-like' timbres were categorically perceived. In fact, it is the absence of a strictly categorical perception of 'Oboe-like,' 'Clarinet-like' and 'Organ-like,' timbres that lead to the proposal of a related timbre identification task that could play a role in evaluating the results of a timbral ear-training program. Unlike a timbral ear-training program in which the task on which listeners are trained is the task on which listeners are subsequently tested, as found in the program developed by Quesnel [14], an alternative task requiring identification of 'Oboe-like' versus 'Clarinet-like' versus 'Organ-like' timbres can present a problem featuring differences subtle enough that perfect performance is never expected even for well-trained listeners, and therefore might provide a sensitive measure for detecting behavioural changes resulting from timbral ear-training.

References

[1] McKinnon-Bassett, M., Martens, W. and Cabrera, D. "Experimental comparison of two versions of a technical ear training program: Transfer of training on tone color identification to a dissimilarity-rating task", *Proceedings of the Audio Engineering Society Conference 50th International Conference: Audio Education*, Murfreesboro, USA, 25-27 July 2013.

[2] American National Standard: Acoustical Terminology, ANSI S1.1-1194 (ASA 111-1994), American National Standards Insitute, New York, 1994.

[3] Zwicker, E. and Fastl, H. *Psychoacoustics: Facts and Models*, second edition, Springer-Verlag, 1990.

[4] von Bismarck, G. "Sharpness as an attribute of the timbre of steady sounds", *Acta Acustica united with Acustica*, **25**(3), 159-172, (1974).

[5] Klein, W., Plomp, R. and Pols. L.C.W. "Vowel spectra, vowel spaces, and vowel identification", *Journal of the Acoustical Society of America*, **48**(4), 999-1009, (1970).

[6] von Bismarck, G. "Timbre of steady sounds: A factorial investigation of its verbal attributes", *Acta Acustica united with Acustica*, **30**(3), 146-159, (1974).

[7] Plomp, R. and Steeneken, H. J. M. "Effect of phase on the timbre of complex tones", *Journal of the Acoustical Society of America*, **46**, 409-421, (1969).

[8] Plomp, R. *Hoe wij horen: Over de toon die de muziek maakt*, (in Dutch), *How we hear. On the tone that makes music*, Breukelen, 1998.

[9] Manor, E., Martens, W.L. and Bassett, M. "Determining the even harmonic attenuation at which the 'clarinet-like' timbre of complex tones becomes dominant", *Proceedings of Acoustics 2015*, Hunter Valley, Australia, 15-18 November 2015.

[10] Krumhansl, C.L. "Why is musical timbre so hard to understand?" in *Structure and Perception of Electroacoustic Sound and Music: Proceedings of the Marcus Wallenberg Symposium*, Lund, Sweden, 21-28 August 1988, pp. 43-53.

[11] Terasawa, H., Slaney, M. and Berger, J. "Perceptual distance in timbre space", *Proceedings of ICAD 05: Eleventh Meeting of the International Conference on Auditory Display*, Limerick, Ireland, 6-9 July 2005.

[12] Lartillot, O. and Toiviainen. P. "A Matlab toolbox for musical feature extraction from audio", *Proceedings of the International Conference on Digital Audio Effects*, Montreal, QC, 2007, pp. 237-244.

[13] Dickens, P., France, R., Smith, J. and Wolfe. J. "Clarinet acoustics: Introducing a compendium of impedance and sound spectra", *Acoustics Australia* **35**, 1-17, (2007).

[14] Quesnel, R. "Timbral ear trainer: Adaptive, interactive training of listening skills for the evaluation of timbre differences", *Proceedings of the 100th Convention of the Audio Engineering Society*, Copenhagen, Denmark, May 1996.