IMPORTANCE OF SPECTRAL DETAIL IN MUSICAL INSTRUMENT TIMBRE

Michael D. Hall[1], *James W. Beauchamp*[2], *Andrew B. Horner*[3], *and Jennifer M. Roche*[4]

[1]Department of Psychology, James Madison University, Harrisonburg, VA
[2]School of Music and Department of Electrical & Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL
[3]Department of Computer Science, Hong Kong Univ. of Science and Technology, Hong Kong
[4]Department of Psychology, University of Memphis, Memphis, TN

ABSTRACT

Previous research has established that time-varying spectral envelope shape is critical to instrument timbre, with influences of both spectral irregularity and spectral flux (e.g., McAdams, et al., 1999). This paper describes two studies which attempt to quantify the salience of various spectrotemporal parameters. Using a triadic timbral similarity ranking task, the first study examined the relative contributions of spectrotemporal parameters for centroid- and temporally normalized static and dynamic versions of ten $E^b{}_4$ instrument tones (i.e., without and with flux). Rotations of MDS solutions indicated relevant spectrotemporal variation, but failed to converge on a particular parameter. For 2-D solutions for static tones, even/odd harmonic ratio correlated best (R = 0.78-9), while for 3-D solutions for dynamic tones, only spectral centroid variation yielded consistently high correlations (R = 0.82-3). The second study examined how timbre recognition and discrimination for six A_4 instrument tones were impacted by eliminating all but four or seven harmonics while retaining flux. In an MDS task listeners rated the certainty of whether pairs of tones were derived from the same instrument. Impoverished tones were generally recognized as the same instrument, with minimal impact on perceptual distance as long as original resonances were conveyed. Assessment of correlations between perceptual and acoustic dimensions was further aided by MDS coordinates in the absence of stress. These studies affirm the importance of spectral detail in judging timbral similarity while revealing that minimal detail may be sufficient for recognition.

1. BACKGROUND

Several critical perceptual dimensions for musical instrument timbre have been well established. Primary among these is spectral envelope shape (e.g., Krumhansl, 1989; also see Hall & Beauchamp, 2009). Several important sources of spectrotemporal variation also have been identified, including spectral flux and spectral irregularity (e.g., McAdams et al., 1999). Much of this research has relied upon multi-dimensional scaling (MDS) techniques (e.g., McAdams et al., 1995), where ratings of perceived dissimilarity between pairs of tones are used to generate a map of perceptual distance. Dimensions in the map are then correlated with acoustic measures to determine the acoustic bases for listeners' judgments. What is less clear is how much spectral detail is required for accurate recognition of musical instruments and for natural-sounding synthesis of specific timbres. This paper summarizes and re-evaluates two previous projects that address different aspects of this issue. The first experiment (Beauchamp et al., 2006) sought to identify the possible additional contributions to timbre of several spectral and spectrotemporal properties in the presence and absence of spectral flux. The second (Hall, 2009) was interested in determining whether reasonably accurate timbre recognition and source discrimination was possible given minimal distinctive information about spectral envelope shape. This was accomplished for each instrument tone by retaining only the harmonics that coincided with average spectral peaks, thereby eliminating all weaker harmonics.

These experiments collectively highlighted common, but often overlooked, concerns with reliance on traditional MDS procedures. Various alternative techniques are provided, including possible ways to maximize correlations between perceptual dimensions and acoustic measures, as well as a method for combining measures of dissimilarity and discrimination within a single task.

2. EXPERIMENT 1

Many studies have shown that average spectral centroid and temporal envelope are important for judging timbral dissimilarity (e.g., Caclin et al., 2005). In an effort to investigate the salience of higher-ordered spectrotemporal parameters, listeners were asked to judge the dissimilarity of several instrument sounds which were normalized for attack and decay times and average centroid. Data were processed by two MDS programs to show the relative positions of the instruments. Best fit straight lines were used to measure the correlation between spectrotemporal parameters (Beauchamp, 2007) measured from time-varying spectral analyses of the individual sounds.

2.1 Method

Participants were 10 musically experienced undergraduates at the Hong Kong University of Science and Technology. They listened to the tones over headphones in a quiet lab. Ten sustained musical instrument tones performed at $E^b{}_4$ (311.1 Hz) served as stimulus sources: bassoon, cello, clarinet, flute, horn, oboe, recorder, alto saxophone, trumpet, and violin. Two types of tones were created via sinusoidal additive resynthesis: static (impoverished) and

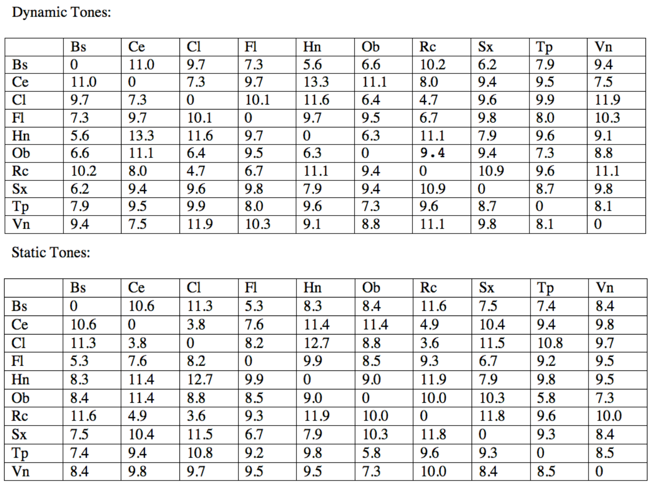
ISBN: 1 876346 62 0 © 2010 ICMPC 11

dynamic (with spectral flux). Static spectra were fixed at averages of the originals, and their temporal envelopes were trapezoids with .05 s attack and decay and 0.5 s total duration. Dynamic tones were shortened by interpolation to 2.0 s, and attack and decay segments were normalized to .05 and .15 s, respectively. Spectra were modified so that their average normalized spectral centroids were set to a common value of 3.7, and were equalized in loudness.

Triadic comparisons were used to measure dissimilarities between tones. On each trial subjects heard three different tones (ABC) and judged which pair (AB, BC, or AC) was most dissimilar. A separate test was run where the subjects judged which pair was most similar. For each instrument pair the dissimilarity score was given by number of times most dissimilar minus number of times most similar plus 9, yielding a possible range of 0 to 18.

2.2 Results and Discussion

Average dissimilarity scores for the ten subjects for the static and dynamic cases are shown in Table 1. Note that the actual scores range from 3.6 to 12.7 (static) and 4.7 to 13.3 (dynamic). For the static tones, cello, clarinet, and recorder were rated most similar (within 5.0); for the dynamic case, clarinet and recorder are close (within 5.0), but not cello. On the other hand, horn and bassoon are relatively close (within 6.0) in the dynamic case, but not in the static case. For the static tones, clarinet and horn are most dissimilar (> 12.0), whereas horn and cello are most dissimilar (> 13.0) for dynamic tones.



Dynamic Tones:

	Bs	Ce	Cl	Fl	Hn	Ob	Rc	Sx	Tp	Vn
Bs	0	11.0	9.7	7.3	5.6	6.6	10.2	6.2	7.9	9.4
Ce	11.0	0	7.3	9.7	13.3	11.1	8.0	9.4	9.5	7.5
Cl	9.7	7.3	0	10.1	11.6	6.4	4.7	9.6	9.9	11.9
Fl	7.3	9.7	10.1	0	9.7	9.5	6.7	9.8	8.0	10.3
Hn	5.6	13.3	11.6	9.7	0	6.3	11.1	7.9	9.6	9.1
Ob	6.6	11.1	6.4	9.5	6.3	0	9.4	9.4	7.3	8.8
Rc	10.2	8.0	4.7	6.7	11.1	9.4	0	10.9	9.6	11.1
Sx	6.2	9.4	9.6	9.8	7.9	9.4	10.9	0	8.7	9.8
Tp	7.9	9.5	9.9	8.0	9.6	7.3	9.6	8.7	0	8.1
Vn	9.4	7.5	11.9	10.3	9.1	8.8	11.1	9.8	8.1	0

Static Tones:

	Bs	Ce	Cl	Fl	Hn	Ob	Rc	Sx	Tp	Vn
Bs	0	10.6	11.3	5.3	8.3	8.4	11.6	7.5	7.4	8.4
Ce	10.6	0	3.8	7.6	11.4	11.4	4.9	10.4	9.4	9.8
Cl	11.3	3.8	0	8.2	12.7	8.8	3.6	11.5	10.8	9.7
Fl	5.3	7.6	8.2	0	9.9	8.5	9.3	6.7	9.2	9.5
Hn	8.3	11.4	12.7	9.9	0	9.0	11.9	7.9	9.8	9.5
Ob	8.4	11.4	8.8	8.5	9.0	0	10.0	10.3	5.8	7.3
Rc	11.6	4.9	3.6	9.3	11.9	10.0	0	11.8	9.6	10.0
Sx	7.5	10.4	11.5	6.7	7.9	10.3	11.8	0	9.3	8.4
Tp	7.4	9.4	10.8	9.2	9.8	5.8	9.6	9.3	0	8.5
Vn	8.4	9.8	9.7	9.5	9.5	7.3	10.0	8.4	8.5	0

Table 1: Dissimilarity matrices from Experiment 1, where *Bs* = bassoon, *Ce* = cello, *Cl* = clarinet, *Fl* = flute, *Hn* = horn, *Ob* = oboe, *Rc* = recorder, *Sx* = saxophone, *Tp* = trumpet, and *Vn* = violin.

SPSS and Matlab MDS programs were used to process the dissimilarity data. Data were projected on the two or three dimensions which minimized the stress.

For the static case the data were correlated with *spectral irregularity* (*SIR*) (Kendall and Carterette, 1996) and *even/odd ratio* (*E/O*), the ratio of the energies in the even and odd harmonics (Caclin et al., 2005). For the dynamic case two measures were added: *average spectral centroid variation* (*SCV*) and *spectrotemporal incoherence* (*SIN*, aka flux). For each measure straight lines were constructed which correlated best with the measure. All solutions were rotated so that the horizontal axes correlated best with the *E/O* measure. Formulas for *SIR*, *SCV*, and *SIN* are given in Beauchamp and Lakatos (2002) and Beauchamp (2007).

The two 2D solutions for the static case (stress = 0.12) are shown in Figure 1; consistent correlations were obtained for *E/O* (*R*=0.78-0.79), but differed somewhat for *SIR* (0.69-0.75). There appeared to be three major groupings of instruments: a) recorder, clarinet, cello; b) oboe, trumpet, violin; and c) bassoon, saxophone, horn, although c) was less obvious than a) and b). The groupings were reasonably consistent between the SPSS and Matlab results, but instrument positions were quite different in the two solutions.

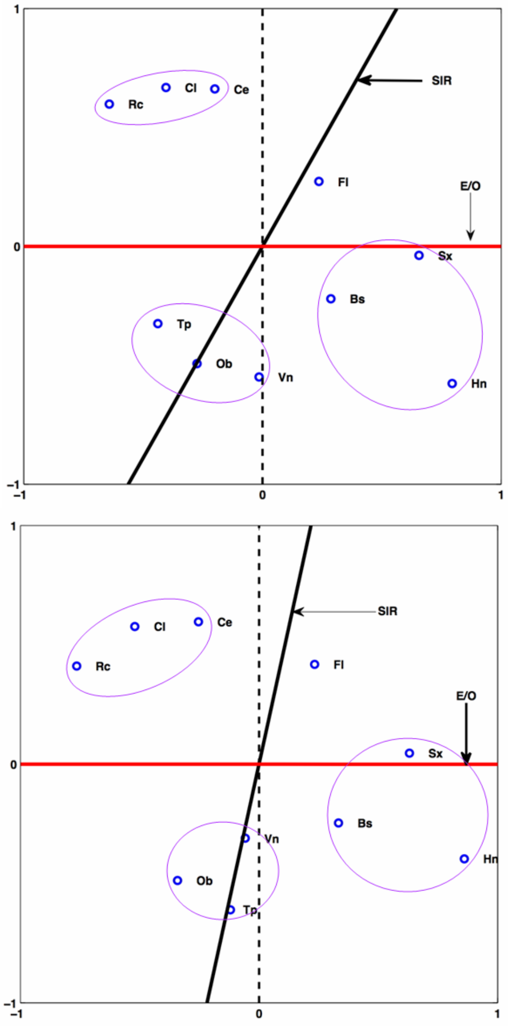


Figure 1: Two-dimensional MDS solutions using SPSS (upper) and Matlab (lower) programs for the static tones in Experiment 1.

For the 2D solutions of the dynamic tone case (see Figure 2), where the *SCV* and *SIN* correlates were introduced, the stresses jumped to 0.15-0.17. *E/O* and *SCV* correlations are at *R* = 0.69-0.71 and 0.68, respectively, whereas the *SIN* and *SIR* correlations are much lower at *R* = 0.53-0.56 and 0.39-0.40, respectively. Despite the increased stress levels, there seems to be a basic agreement between the positions of the instruments between the SPSS and Matlab solutions. For example, flute, trumpet, and violin fall along the *SIN* line in about the same positions, and the recorder-clarinet-cello and sax-bassoon-horn constellations, as well as the oboe's position, are very similar.

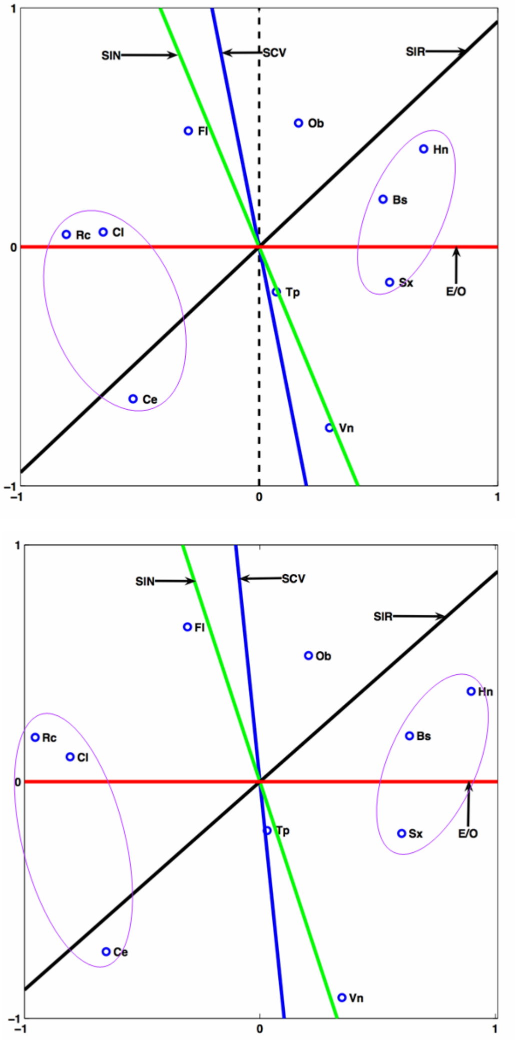


Figure 2: Two-dimensional MDS solutions from SPSS (upper) and Matlab (lower) for Dynamic tones in Experiment 1.

Stresses were much lower (.095) for the 3D solutions of the dynamic tone case (see Figure 3), although the worthiness of instrument groupings are harder to assess without being able to change the viewing angle. Interestingly, the SPSS and Matlab solutions were very different in terms of the positions of the instruments as well as the best-fit lines relative to the *E/O* correlation line. Except for *SCV*, correlations disagreed: for *E/O*, *R* = 0.82 (SPSS), 0.68 (Matlab); for *SCV*, *R* = 0.83 (SPSS), 0.82 (Matlab); for *SIN*, *R* = 0.53 (SPSS), 0.83 (Matlab); for SIR, *R* = 0.82 (SPSS), 0.71 (Matlab). While all of the parameters correlated well in at least one solution, this also demonstrates that radically different solutions can yield the same stress, hindering decisions about which parameters best describe musical sounds. There is also a disagreement between the 2D and 3D solutions in that the "winners" for 2D seem to be *E/O* and *SIN*, whereas for 3D the "winners" (on average) are *SCV* and *SIR*.

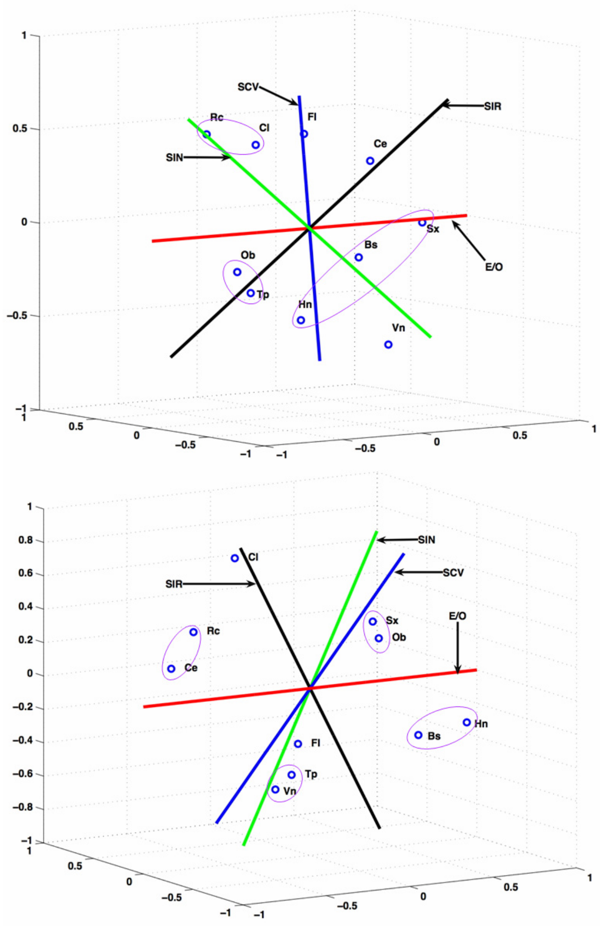


Figure 3: SPSS (upper) and Matlab (lower) solutions for the Dynamic tones.

3. EXPERIMENT 2

An alternative method of manipulating spectral detail was pursued in Experiment 2: exaggeration of spectral peaks. Toward this end, a set of resynthesized musical tones was generated at a common pitch that included tones where all spectral information was eliminated except for harmonics occurring at average spectral peaks. The impact of these manipulations on timbre identification, as well as on discrimination in conjunction with MDS, was then evaluated as a function of musical instrument.

3.1 Method

Participants were 9 undergraduates who had a mean of 7.5 years of musical training (1-11 years). Eighteen tones were derived from A_4 samples from the MUMS database (Opolko & Wapnick, 1987) for piano, vibraphone, electric guitar, tenor trombone, saxophone, and E^b clarinet. Tones were resynthesized according to the additive component of spectral modeling synthesis (Serra & Smith, 1990) in Camel Audio's *Alchemy*. Phase differences and variation from mean F_0 were eliminated. Loudness and duration were equated while retaining amplitude envelopes. Three tones were synthesized for each instrument: one with all harmonics, a 7-harmonic version (F_0 plus 6 harmonics with higher mean dB than adjacent harmonics), and a 4-harmonic version (F_0 plus 3 harmonics selected in the same manner). If a tone had insufficient spectral peaks, then remaining harmonics had the highest mean amplitude.

Listeners completed two tasks. In timbre identification, listeners indicated which instrument each tone was derived from (10 random repetitions/tone). An instrument discrimination task was restricted to all- and 4-harmonic tones. Listeners rated whether tone pairs were from the same instrument (6 repetitions/pair). Ratings of *1* to *4* indicated "same", and *5-8*, "different". Higher ratings indicated greater differences [*1* ("identical")-*8* ("very different")].

3.2 Results and Discussion

Instrument Identification. Mean timbre identification accuracy was determined for each stimulus and listener. Grand means and corresponding standard errors are displayed in Figure 4. Reduction to 7 harmonics had minimal impact on identification, and 4-harmonic tones were typically identified well above chance.

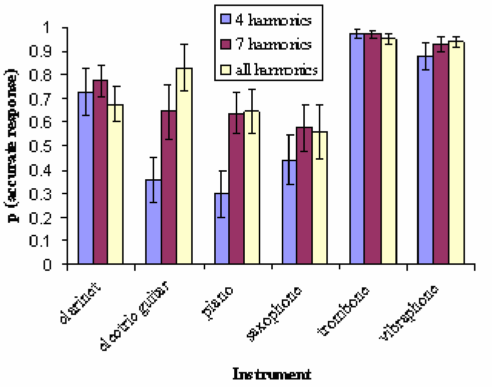


Figure 4: Mean accuracy (and corresponding standard error bars) for timbre identification in Experiment 2.

Accuracy decreased for 4-harmonic tones (.61 v. .76-.77; Bonferroni $p < .05$) and differed with instrument [$F(5,40) = 13.08$, $p < .0001$], as did the impact of harmonic reduction, $F(10,80) = 3.89$, $p < .001$. Accuracy was not significantly reduced with fewer harmonics for trombone, vibraphone, clarinet, or saxophone. Trombone and vibraphone were not confused with other instruments, and clarinet tones were reliably identified. Saxophone tones were frequently confused with clarinet. Reductions in accuracy were observed for 4-harmonic piano and guitar tones ($p < .05$). For piano there was a corresponding increase in vibraphone responses ($p < .05$). Reducing guitar harmonics produced responses for sources with brief attacks and less spectral irregularity (piano and vibraphone).

Timbre Discrimination Ratings. One benefit of the rating method used in Experiment 2 is that it permitted simultaneous assessment of perceptual distance through MDS and pair-wise discrimination. For the latter measures, ratings 1-4 were treated as "same" responses and ratings of 5-8 were treated as "different" responses. For each participant, discrimination sensitivity (d') was calculated for each instrument compared with its corresponding 4-harmonic tone according to a differencing model. Thus, higher d' scores reflected greater sensitivity to harmonic reduction.

Mean d' scores (and standard error bars) for comparisons of intact and corresponding 4-harmonic stimuli are displayed in Figure 5. As can be seen in the figure, sensitivity changed with instrument, $F(5,40) = 9.17$, $p < .0001$, revealing instrument-specific impacts of fewer harmonics. Participants discriminated reduced-harmonic versions of the guitar, piano, and saxophone tones from their intact counterparts more often than for the clarinet, trombone, or vibraphone tones ($p < .05$). In fact, participants were not sensitive to differences between 4-harmonic and intact versions of the trombone or the vibraphone ($d' = 0$).

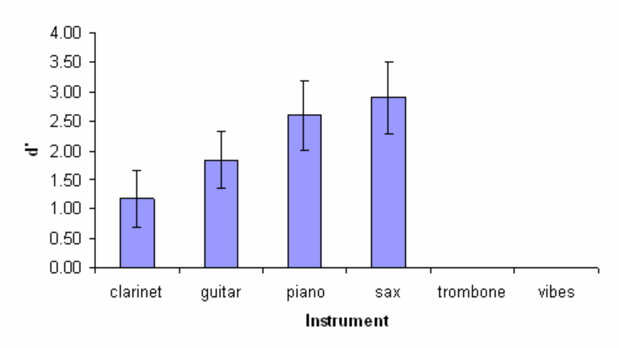


Figure 5: Mean sensitivity (and standard errors) to harmonic reduction as a function of instrument in Experiment 2.

Mean ratings were initially submitted to MDS in SPSS (*ALSCAL*). The resulting 2D solution is displayed at the top of Figure 6 ($R^2 > .94$, stress < .12); each instrument is indicated by its first letter, followed by "4" for any 4-harmonic tone. Consistent with the sensitivity data, ratings of intact clarinet, trombone, and vibraphone tones with corresponding 4-harmonic tones were lower than for other stimuli ($p < .05$). For the trombone and vibraphone these ratings were below the timbre boundary of 4.5 (X^2 $p < .05$ and .01), and for the vibraphone did not significantly differ from ratings of identical tones (1.07 v. 1.04). Only the saxophone may have shifted out of category with harmonic reduction ($M = 4.61$), but its intact tone also was confused with clarinet (< 4.5).

MDS coordinates were evaluated for correlations with mean mel-frequency cepstral coefficient (*MFCC*), mean spectral centroid, and spectral irregularity, as well as (log) rise time in *ms*. As in Experiment 1, we note basic shortcomings of the MDS method. Dimension 2 appears to be related to reductions in spectral complexity, with tones containing fewer spectral peaks generally located higher on the axis, as well as 4-harmonic tones relative to their intact counterparts. Yet, only a moderate correlation with rise time was found (R = .53, p < .05). Unfortunately, it became apparent that stress in the MDS solution differentially impacted ratings for particular pairs of stimuli. This was revealed by the reversed position of intact and 4-harmonic versions of guitar and saxophone along dimension 2, which was absent from even the 3-D solution that accounted for only 3 percent more variance.

To minimize the impact of the algorithm on unique impacts of stress on estimated perceptual distance, a method was developed to allow correlations with acoustic measures in the absence of stress—i.e., based upon the original mean ratings. A 2D solution from Matlab was compared against a version that allowed the maximum number of dimensions (11). The latter represents the mean ratings, but with several dimensions depicting error variance, and therefore not being meaningful contributors to those ratings.

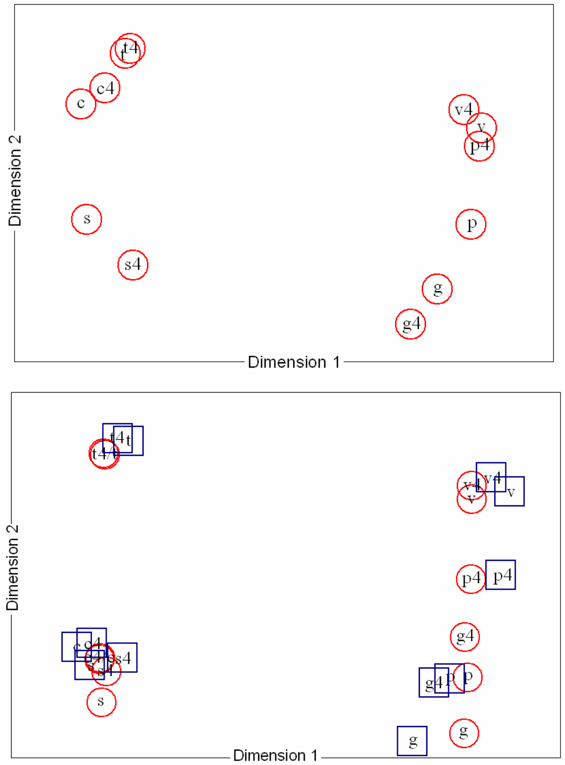


Figure 6: For Experiment 2, 2D MDS solution (circles) from SPSS (upper) and Matlab (lower), along with corresponding coordinates from the unstressed (11D) solution (squares). In the figure *t* = trombone, *c* = clarinet, *s* = saxophone, *v* = vibraphone, *p* = piano, and *g* = guitar, and *4* indicates a 4-harmonic version of the given instrument tone.

Both 2D and 11D solutions are displayed at the bottom of Figure 6. Two dimensions of the 11D solution strongly correlated with the 2D solution's dimensions (R = .99 and .98, respectively). Dimension 1 was best predicted by spectral irregularity (R = -.93, p < .0001), and *MFCC* (R = -.94, p < .0001), and also correlated fairly well with rise time (R = -.75, p < .01). Consistent with suppositions about spectral complexity, dimension 2 was correlated with mean spectral centroid (R = -.64, p < .05), although similarly strong correlations with rise time were still obtained (R = .62, p < .05). Correlations were comparable to the 2D solution, as indicated by the close alignment of the corresponding stimuli in Figure 6, as well as to the SPSS solution (e.g., for dimension 1, R = .94, .94, and .77 for irregularity, *MFCC*, and rise time, respectively).

4. GENERAL DISCUSSION

Several conclusions can be reached from these experiments about spectral contributions to timbre. Experiment 1 revealed a few fundamental dimensions beyond those that were previously identified. Primary among these is the ratio of energy across even and odd harmonics, which correlated with performance as well as, or better than, other dimensions (e.g., irregularity). Furthermore, for dynamic tones centroid variation correlated better with judgments than spectral incoherence (flux).

Experiment 2 further indicated that minimal spectral detail is often adequate for accurate timbre recognition. Stimuli with a low number of harmonics were primarily perceived as the intended instrument, and timbre shifts were limited to instruments with more spectral complexity than could be effectively captured by four harmonics. Thus, timbre can be maintained despite tremendous signal reduction as long as the remaining energy preserves natural resonances.

Conclusions from both experiments were negatively impacted by reliance on MDS procedures. For example, in Experiment 1 SPSS and MatLab solutions often produced contrasting results, as exemplified by the dynamic tones's 3D solutions (see Figure 3). This was also demonstrated in Experiment 2 (circles across top and bottom panels of Figure 6). Additionally, some dimensions within the obtained MDS solutions did not point to a particular acoustic parameter as the basis of timbre judgments. In Experiment 1 no one spectrotemporal parameter (beyond average spectral centroid and attack/decay times, which were normalized across the stimuli) stood out as the best correlate. Likewise, in Experiment 2, while one perceptual dimension was found to correlate very well with both mean *MFCC* and spectral irregularity, the remaining dimension produced similar correlations across very different parameters, mean spectral centroid and rise time.

Finally, both experiments reveal potentially useful alternatives to MDS researchers. Experiment 1 demonstrated that correlations with an acoustic measure can be maximized by permitting axes to deviate from those displayed in the MDS solution. Some apparent limitations also might be overcome by minimizing differential effects of stress. Whereas different MDS algorithms were shown to yield different solutions, we found that permitting the maximum number of dimensions (as in Experiment 2) provided an untransformed depiction of perceptual distances consistent across software platforms. Comparisons of such solutions with those involving fewer dimensions should permit reliable determination of critical perceptual dimensions in a way that does not alter estimated perceptual distance for specific stimulus pairs. Future research will ultimately determine whether this alternative procedure can be effectively applied across multiple studies and research contexts.

5. ACKOWLEDGEMENTS

We would like to give our sincere thanks to Mert Bay, who did the correlations and prepared the MDS diagrams for Figures 1 - 3. We would also like to acknowledge the Hong Kong Research Council for their support of the Experiment 1 project under grant 613508.

6. REFERENCES

Beauchamp, J. (2007). Analysis and synthesis of musical sounds. In J. W. Beauchamp (Ed.), *Analysis, Synthesis, and Perception of Musical Sounds* (pp. 1-89). New York: Springer.

Beauchamp, J., Horner, A., Koehn, H.-K., & Bay, M. (2006). Multidimensional scaling analysis of centroid- and attack/decay-normalized musical instrument sounds. (abstract) *J. Acoust. Soc. Am.*, *120*(5), pt. 2, 3276.

Beauchamp, J. W. & Lakatos, S. (2002). New spectro-temporal measures of musical instrument sounds used for a study of timbral similarity of rise-time- and centroid-normalized musical sounds. *Proc. 4th Int. Conf. Music Perception and Cognition*. (Univ. of New South Wales, Sydney, Australia), pp. 592-595.

Caclin, A., McAdams, S., Smith, B. K. & Winsberg, S. (2005). Acoustic correlates of timbre space dimensions: A confirmatory study using synthetic tones. *J. Acoustic Soc. Am.*, *118*(1), 471–482.

Hall, M. D. (2009). Perception of musical sources with impoverished spectral envelopes. *Proceedings of Acoustics Week in Canada 2009*. Canadian Acoustics Association.

Hall, M. D. & Beauchamp, J. W. (2009). Clarifying spectral and temporal dimensions of musical instrument timbre. *Canadian Acoustics*, *37*(1), 3 - 22.

Kendall, R. A. & Carterette, E. C. (1996). Difference thresholds for timbre related to spectral centroid. *Proc. 4th Int. Conf. Music Perception and Cognition*. (Faculty of Music, McGill Univ., Montreal), pp. 91-95.

Krumhansl, C. L. (1989). Why is musical timbre so hard to understand? In S. Nielsen and O. Olsson (Eds.), *Structure and Perception of Electroacoustic Sound and Music* (pp. 43-53). Amsterdam: Elsevier.

McAdams, S., Beauchamp, J. W. and Meneguzzi, S. (1999). Discrimination of musical instrument sounds resynthesized with simplified spectrotemporal parameters. *J. Acoust. Soc. Am.* ***105***(2), pt. 1, 882-897.

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

Opolko, F., & Wapnick, J. (1987). McGill University Master Samples [CD-ROM]. Montreal, Quebec, Canada: jwapnick@music.mcgill.ca.

Serra, X. & Smith, J. (1990). Spectral modeling synthesis: A sound analysis/synthesis based on a deterministic plus stochastic decomposition. *Computer Music Journal*, *14*(4), 12-24.