Perception of Musical Sources with Impoverished Spectral Envelopes

Michael D. Hall
Dept. of Psychology, James Madison University, MSC 7704, Harrisonburg, Virginia, USA, 22807 hallmd@jmu.edu

1. INTRODUCTION

Musical instrument timbre is multi-dimensional, with spectral, temporal, and spectro-temporal components (e.g., McAdams, Winsberg, Donnadieu, De Soete, & Krimphoff, 1995). Envelope shape has been argued to be a critical spectral dimension (e.g., Krumhansl, 1989). Recent timbre research has shown that listeners rely upon envelope shape instead of brightness (Hall & Beauchamp, 2009).

How much spectral envelope detail is required to permit timbre recognition? One way to evaluate this question is to gradually eliminate spectral peaks and assess when timbre shifts occur. This has been done for timbre discrimination (e.g., see Beauchamp, Horner, & Ayers, 2006), revealing a broad range of performance that depends upon instrument. An alternative method of manipulating spectral detail was pursued in the current investigation. A set of resynthesized musical tones was generated at a common pitch such that all spectral information was eliminated except at harmonics that occurred at average spectral peaks. The impact of these manipulations on timbre identification and discrimination was then evaluated as a function of musical instrument.

2. METHOD

Participants were 9 undergraduates who had a mean of 7.5 years of performance training on a musical instrument (1-11 years). Eighteen tones (44.1 kHz, 16-bit) 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 using Camel Audio's *Alchemy*, an additive VST plug-in based on spectral modeling synthesis (Serra & Smith, 1990). Phase differences and variation from mean F_0 were eliminated. Tones were equated for loudness and compressed to 1755 *ms* while retaining amplitude envelopes. Three tones were synthesized for each instrument. One included all harmonics from the original tone. A second version had 7 harmonics—F_0 plus 6 harmonics with higher mean dB than adjacent harmonics. If a tone had fewer spectral peaks, then the remaining harmonics had the highest mean amplitude. A final version had 4 harmonics—F_0 plus three harmonics selected as described for the 7-harmonic series.

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

3. RESULTS

3.1 Instrument Identification

Mean timbre identification accuracy was determined for each stimulus and listener. Grand means and corresponding standard errors are displayed in Figure 1. Reduction to 7 harmonics had minimal impact on identification, and 4-harmonic stimuli were typically identified well above chance. Accuracy decreased for 4-harmonic stimuli (.61 v. .76-.77; Bonferroni $p < .05$), which contributed to a main effect of number of harmonics, $F(2,16) = 11.98$, $p < .001$.

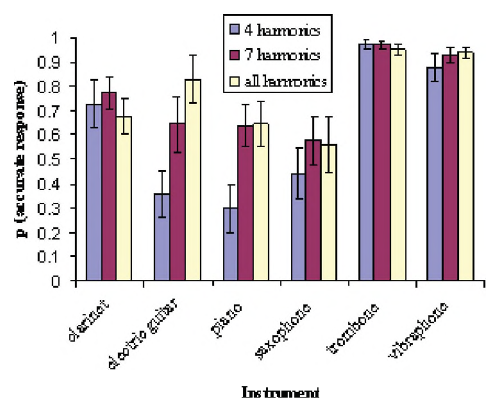

Figure 1. Mean accuracy of timbre identification.

Accuracy differed with instrument [$F(5,40) = 13.08$, $p < .0001$], and the impact of harmonic reduction depended on instrument, $F(10,80) = 3.89$, $p < .001$. 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. While saxophone tones with few harmonics were increasingly confused with clarinet, similar confusion occurred for all-harmonic tones. Significant reductions in accuracy for 4-harmonic stimuli ($p < .05$) were observed for piano and guitar. 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. There were frequent piano responses, and a significant increase in vibraphone responses for the 4-harmonic tone.

3.2 Timbre Discrimination Ratings

Mean ratings of perceptual distance were submitted to multi-dimensional scaling (MDS) using ALSCAL. The resulting 2-dimensional solution is displayed in Figure 2 ($r^2 > .94$, stress < .12). Ratings of intact clarinet, trombone, and vibraphone tones with corresponding 4-harmonic tones were less than for other stimuli ($p < .05$). For the trombone and vibraphone, these ratings were significantly below 4.5 (X^2 $p < .05$ and .01), the boundary between timbres. For the vibraphone these ratings did not significantly differ from ratings of identical tones (1.07 v. 1.04). Ratings from same-instrument comparisons also were lowest for the saxophone and guitar. For the all-harmonic piano standard, ratings for the 4-harmonic piano were significantly less than for all tones except guitar tones or the intact vibraphone. The only instrument that may have shifted out of category following harmonic reduction was the saxophone ($M = 4.61$), but its all-harmonic tone also was confused with clarinet (< 4.5).

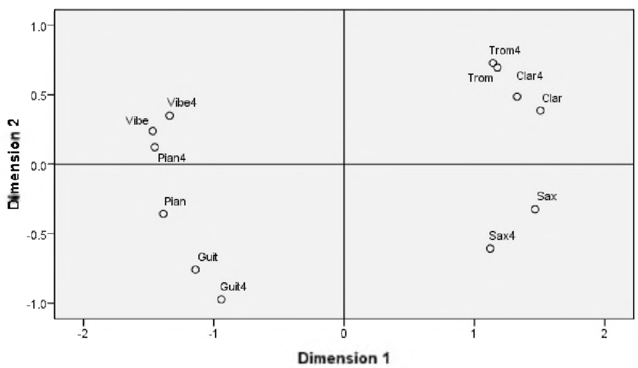

Figure 2. MDS solution based upon timbre discrimination.

MDS coordinates were evaluated for correlations with several acoustic measures, including mean mel-frequency cepstral coefficient (MFCC), mean spectral centroid, and spectral irregularity (see www.vamp-plugins.org), as well as (log) rise time in *ms*. Dimension 1 was best predicted by spectral irregularity ($r = .94$, $p < .0001$), and MFCC ($r = .94$, $p < .0001$; rise time $r = .77$, $p < .01$). Dimension 2 was moderately correlated with rise time ($r = .53$, $p < .05$), but likely reflects responses to spectral differences, including reductions in spectral complexity with the stimulus manipulation. For example, tones with fewer spectral peaks (i.e., with gradual spectral changes) gathered toward the top of the axis. Furthermore, 4-harmonic tones generally are located higher on the axis than their intact counterparts. The exceptions in Figure 2 (guitar and saxophone) reveal stress in the solution; ordering of these 4- and all-harmonic tones was reversed on the corresponding axis in the 3-D solution, which accounted for less than 3 percent more variance.

4. DISCUSSION

Several findings suggest retention of important source information despite a minimal number of harmonics. Stimuli with few harmonics were typically still primarily perceived as the same instrument across tasks, and observed perceptual shifts were limited to instruments (piano and guitar) that had greater spectral complexity than could be approximated by the limited number of frequencies. Furthermore, these timbres were often confused with instruments that shared aspects of production (e.g., brief attacks followed by gradually decaying spectral information). Thus, it appears that timbre information can be maintained as long as harmonics reflect natural resonances of the instrument that produces them.

These findings are not likely solely due to temporal properties (e.g., rise time). After all, one dimension in the MDS solution from discrimination was best accounted for by spectral or spectro-temporal measures. Although a strong acoustic correlate was not obtained from the included measures for the remaining dimension, it also appears to be best accounted for by spectral information (e.g., diffuseness of spectral peaks). Additionally, reductions in instrument recognition with decreasing harmonics could only reflect responses to changes in spectral envelope shape.

The maintenance of timbre with limited spectral information complements preceding work. For example, it has been demonstrated that timbres become more discriminable from their original counterparts with sufficient smoothing of the spectral envelope, and smoothed envelopes are difficult to distinguish, presumably due to an absence of well-defined spectral peaks (Beauchamp, Horner, & Ayers, 2006). The general findings reported here also can be regarded as a nonspeech analogue to similar speech demonstrations. For example, in sinewave speech phonemes are accurately reported given only three frequencies that correspond to formant center frequencies (e.g., see Remez, Rubin, Pisoni, & Carrell, 1981). In fact, sentence transcription in sinewave speech is more accurate when harmonically related frequencies are used (e.g., Toth & Kocsor, 2003) like those in the current investigation. As a result, the findings reported here could represent further evidence that listeners focus on intense spectral information in vibratory sources.

REFERENCES

Beauchamp, J. W., Horner, A. B., & Ayers, L. (2006). Effect of critical band data reduction on musical instrument sounds. *Proceedings of the 9th International Conference on Music Perception and Cognition (ICMPC 07)*, Bologna, Italy, 51-52.

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., Winsberg, S., Donnadieu, S., De Soete, G. & 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.

Remez, R. E., Rubin, P. E., Pisoni, D. B., & Carrell, T. D. (1981). Speech perception without traditional speech cues. *Science*, *212*, 947-950.

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.

Toth, L. & Kocsor, A. (2003). Alternatives to sine-wave speech. *EUROSPEECH-2003*, 2073-2076.