Discrimination of Sustained Musical Instrument Sounds Resynthesized with Randomly Altered Spectra

Andrew Horner

Department of Computer Science, Hong Kong University of Science & Technology
email: horner@cs.ust.hk

James Beauchamp

School of Music & Dept of Electrical and Computer Eng., Univ. of Illinois at Urbana-Champaign
email: j-beauch@uiuc.edu

Abstract

The perceptual salience of random spectrum alteration was investigated for musical instrument sounds. Spectral analysis of sounds from eight musical instruments (bassoon, clarinet, flute, horn, oboe, saxophone, trumpet and violin) produced time-varying harmonic amplitude data. With various amounts of random spectrum alteration applied to this data, sounds were resynthesized with errors of 1-50%. Moreover, the peak centroids of the randomly altered sounds were equalized to those of the originals. Listeners were asked to discriminate the randomly altered sounds from reference sounds resynthesized from the original data. In all eight instruments, discrimination was very good for 30 – 50% errors, moderate for 15 – 25% errors, and poor for 1-10% errors. Thus, sounds with the same harmonic amplitude-vs-time envelope shapes and peak centroid can sound different if the error is about 15% or more.

1 Introduction

It is common knowledge that musical instruments can be identified even when their spectra have been substantially altered. A trumpet is recognizable when performed in a vast cathedral or small bathroom, played through a 3D surround system or cheap PC speakers, or modified through a spectrum equalizer. These modifications, while noticeable, are easily tolerated unless perhaps large resonances occur. There seem to be some overriding aural cues that allow human listeners to classify various sounds as coming from certain common sources.

Spectral envelope distortion has been investigated in speech perception by Watkins and Makin (1994 and 1996). An earlier study on resonance modifications in music and speech by Toole and Olive (1988) concluded: “it is surprising just how much the … signal … can be modified without significantly altering perceived timbre”.

How much variance can be tolerated? If a spectrum equalizer is used to modify a sound with its levels set at random within a range of ±*x* dB, at what level of *x* would a listener begin to distinguish the modified from the original sound? At what level would it no longer be identified as a sound produced by the original instrument or type of instrument?

Two important applications potentially benefit from answering these questions. The first application is in the determination of perceptually relevant parameters in timbre research. Several previous timbre perception studies have shown spectral centroid and amplitude rise time to be highly correlated with the two principal perceptual dimensions of timbre (Grey and Gordon 1978; Iverson and Krumhansl 1993; Krimphoff, McAdams, Winsberg 1993; Lakatos 2000). However, no consensus has emerged about the best physical correlate for a third dimension of timbre. Using random spectrum alteration, musical sounds can be resynthesized while retaining rise time, spectral centroid, and the harmonic amplitude time-variant evolution of an original sound. If we produce such a group of randomly altered sounds that are perceptually dissimilar, what makes them dissimilar? The answer to this question may give us some good clues about other dimensions of timbre.

A second application of random spectrum alteration is the production of similar, yet different, timbres in music synthesis. Sampling synthesizers have often been criticized as sounding “too much the same”, since repeated notes, played at the same amplitude, typically sound exactly the same. If we know how much a sound can be changed by random spectrum alteration without destroying its identity, we can potentially produce a much more dynamic and realistic result.

In the present study, we sought to determine the extent to which different degrees of random spectrum alteration can affect the perception of synthesized sustain sounds. We measured listener discrimination with respect to several unaltered original sounds for various percentages of random alteration. For example, if we randomly rescale the partial amplitudes by ±10% (representing a 5% error on average), what will be the average discrimination? If we generate several different sounds with similar amounts of random spectrum alteration, will the discrimination be relatively constant or will it vary over a range of values? How does discrimination vary for different amounts of random spectrum alteration? Will it vary from instrument to instrument? Does listener musical experience affect discrimination of randomly altered sounds? If several sounds have similar amounts of random alteration and the same centroid and rise times, but sound dissimilar, what makes them sound different? We will attempt to address these questions.

2 Random Spectrum Alteration

Eight sustain musical instrument sounds, also used by McAdams, Beauchamp, and Meneguzzi (1999), were selected as prototype signals. They were first subjected to spectrum analysis using a computer-based phase vocoder method.

Random spectrum alteration was performed on the analysis data, after which the sounds were generated by the additive synthesis method. Random alteration was done by multiplying each harmonic amplitude by a random scalar r_k:

$$A_k^{'}(t) = r_k A_k(t), \qquad (1)$$

where harmonics in the same critical band share the same random scalar. This is tantamount to a linear stationary process. The goal of this random spectral alteration is to perturb each harmonic amplitude, without changing the spectral centroid or loudness.

By uniformly picking r_k in the range [1 - 2ε, 1 + 2ε], the error is expected to be approximately ε, though the actual error ε' will slightly deviate from ε. For this study, we generate 50 tones for each instrument, where the error ε ranges from 1% to 50% in increments of 1%. So, for 50% error, r_k will be picked in the range [0, 2].

Preserving the spectral centroid after random spectrum alteration has been applied provides an important group of related, yet different, timbres. To preserve spectral centroid, we iteratively tilt the altered spectra to achieve the desired centroid using Newton's method.

For loudness equalization, an amplitude multiplier was again determined such that the altered sound had a loudness of 87.4 phons. An iterative procedure adjusted the amplitude multiplier starting from a value of 1.0 until the resulting phons were within 0.1 phons of 87.4, as measured by Moore and Glasberg's loudness program (Moore, Glasberg, and Baer 1997).

The random spectrum alteration algorithm therefore consists of the following steps:

```
(1) Pick initial values for $r_k$ such that $1 - 2\varepsilon < r_k < 1 + 2\varepsilon$.
(2) Apply random alteration: $A_k'(t) = r_k A_k(t)$.
(3) Centroid equalization:
    a. Calculate the average spectra of the original and altered sounds.
    b. Calculate the spectral centroids of the average spectra from step 3a.
    c. Iteratively tilt the average altered spectrum using Newton's method by modifying $r_k$ until the centroids match.
(4) Iterative loudness equalization using Moore and Glasberg's LOUDEAS program.
(5) End
```

3 Experimental Method

The 20 subjects were undergraduate students at the Hong Kong University of Science and Technology, ranging in age from 18 to 23 years, who reported no hearing problems. They included ten "musicians" (six males, four females) and ten non-musicians (four males, six females). "Musicians" were defined as having at least five years of practice on an instrument, and "non-musicians" were defined as never having played a musical instrument. The subjects were paid for their participation.

The eight sustain instruments used belong to the air column (air reed, single reed, lip reed, double reed) and bowed string families: bassoon, clarinet, flute, horn, oboe, saxophone, trumpet, and violin. Each sound was analyzed and resynthesized using the reference analysis data with no frequency variations and no inharmonicity. With fixed harmonic frequencies, listeners were encouraged to focus their attention exclusively on the amplitude data, since they were prevented from detecting cues stemming from frequency deviations amplified by random spectrum alteration. Also, since the original sustain sounds had relatively small frequency deviations and were nearly strictly harmonic, frequency flattening had only a minor effect on the sounds' qualities.

The sounds were stored in 16-bit integer format on hard disk. All "reference sounds" (resynthesized using the analysis data and strictly fixed harmonics) were equalized for duration (2-s) and loudness (87.4 phons). The randomly altered sounds for each instrument were resynthesized by additive synthesis.

A randomly altered sound was generated for each error level from 1-50% in increments of 1%, yielding a total of 50 modified sounds for each instrument. The randomly altered sounds were also generated using strictly fixed harmonics. To compare random alterations across instruments, the same initial set of random scalars r_k were used on all eight instruments, though the scalars were slightly modified by centroid tilt-correction.

A two-alternative forced-choice (2AFC) discrimination paradigm was used. The listener heard two pairs of sounds and chose which pair was "different". Each trial structure was one of AA-AB, AB-AA, BB-BA, or BA-BB, where A represents the reference sound and B one of the 50 randomly altered sounds. This paradigm has the advantage of not being as susceptible to variations in subjects' criteria across experimental trials compared to the simpler A-B method. All four combinations were presented for each randomly altered sound. The two 2-s sounds of each pair were separated by a 500-ms silence, and the two pairs were separated by a 1-s silence. On each trial, the user was prompted with "which pair is different, 1 or 2?" and response was by the keyboard. The computer would not accept a response until at least the first pair had been played.

For each instrument, a block of 200 trials was presented to the subjects (four trial structures x 50 random alterations). Performance for each random alteration was computed on four trials for each subject. The duration of each block was about 40 minutes. Eight blocks were presented corresponding to the eight instruments. The total duration of the experiment was about six hours. Listeners took 5-10 minute breaks between blocks, and finished the test in two separate 3-hour sessions.

A program running on a PC controlled the experiment. Subjects were seated in a "quiet room" with 40 dB SPL background noise level (mostly due to the computer and air conditioning). The headphones naturally masked some of this background noise. Sound signals were converted to analog by a SoundBlaster Audigy soundcard and then presented through Sony MDR-7506 headphones at 87.4 phons. The Audigy DAC uses 24 bits with a maximum sampling rate of 96 kHz, and a 100dB S/N ratio. The sounds were actually played at 22.05 kHz or 44.1 kHz.

At the beginning of the experiment, the subject read instructions and asked any necessary questions of the experimenter. Five test trials (chosen at random from the altered tones) were presented before the data trials for each instrument. The order of presentation of the 200 trials was random within each block, and the order of presentation of the instruments was randomized for each subject.

4 Results

Discrimination scores were computed for each random alteration across the four trial structures for each subject. Scores averaged over the 20 subjects for the 50 random alterations on all eight instruments are shown in Figure 1. For errors up to 10%, almost all scores are in the range 40-60%. The scores are around the "indistinguishability level" of 50% that corresponds to random guessing. The range is wider and more variable for intermediate errors between 15-25%, where the scores cover nearly the full range from 50-100%. Intermediate errors correspond to "somewhat distinguishable" cases. For errors more than 30%, most scores are above 90%, and are "very distinguishable". An underlying S-curve is clearly visible in the trend. Most points are close to the trend. Since the same initial random scalars were used in all eight instruments, a few points seem to be outliers. The most obvious examples are the outlying 27% and 36% errors.

Figure 2 shows the individual instrument trends are in close agreement, varying by no more than 10%. The flute is the most different, with a 5-10% lower trend over the range 5-25%. Excluding the flute, the biggest difference is how quickly the trends converge for errors between 15-40%.

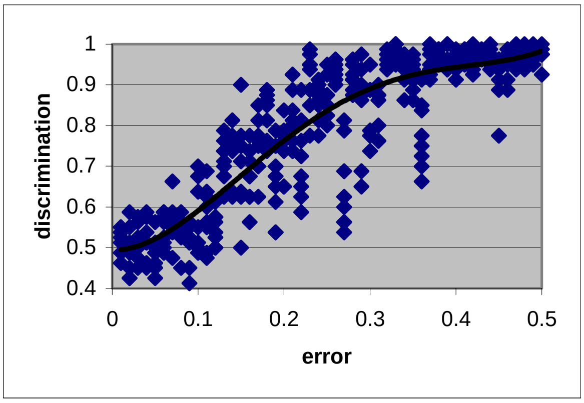

Figure 1. Mean subject discrimination scores for randomly altered sounds vs. stimuli error level for all eight instruments (the solid line shows the 4th order polynomial trend).

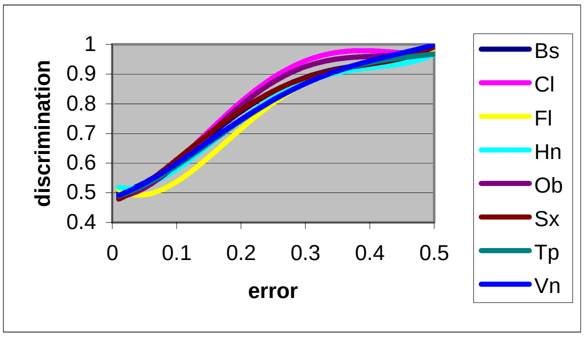

Figure 2. Discrimination trends vs. stimuli error for the eight instruments.

Figure 3 shows the trend plus the standard deviations and maximum deviations. The solid center line represents the trend, the inner lines around it represent the trend plus and minus the standard deviation, and the outer lines represent the smoothed range of maximum deviations. The standard deviations range from 5-10%, with the larger deviations occurring at errors between 10-30%. The maximum deviations cover the full range from 50-100 for errors between 20-25%. Theoretically, it is possible for even large errors to have discrimination scores as low as 50%, since there is a very small chance all the random scalars will be chosen with zero values. However, the probability is very small and decreases as the error is increased.

Figure 4 separates average discrimination scores for musicians and non-musicians. The trends shows that the musicians discriminated randomly altered sounds from reference sounds slightly better overall than the non-musicians. The largest discrimination difference is 4% for errors between 20-30%. The difference is negligible for errors less than 12% and more than 38%.

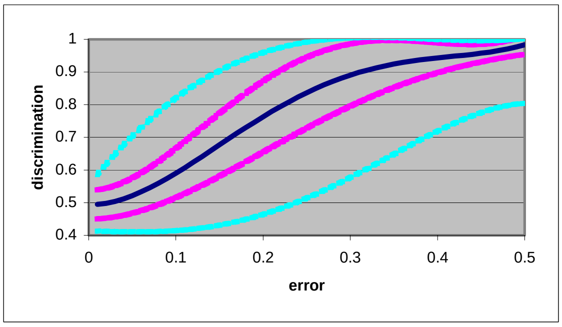

Figure 3. Discrimination trends. The solid center line represents the trend, the inner lines around it represent the trend plus and minus the standard deviation, and the outer lines represent the smoothed range of maximum deviations.

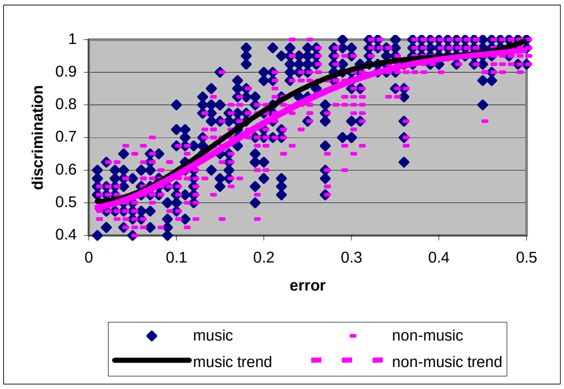

Figure 4. Musician and non-musician average discrimination scores for all eight instruments (labeled as “music” and “non-music” respectively) vs. stimuli error. The lines show the 4th order polynomial trends.

5 Conclusions

The average discrimination data show a clear monotonic increasing dependence on the amount of random spectrum alteration. The discrimination data is relatively unaffected by musical training of the listeners and is only slightly affected by instrument sound. However, musical training seems to help when the randomly altered sounds are moderately different from the original, as is the case with 20-30% errors.

It was surprising to the authors how forgiving the ear is to spectral changes made by random spectrum alteration. A previous spectral matching study (Horner, Beauchamp, and Haken 1993) indicated that good perceptual matches to wind instrument sounds could be achieved with errors of less than 4%, whereas an 8% error was usually highly discriminable. For this study, we initially planned to test error values in increments of 2% (2%, 4%, 6%, 8%, and 10%), fully expecting near-100% distinguishability with 8% error. However, in informal listening tests, we found that we could not distinguish randomly altered sounds with 10% errors from the corresponding resynthesized original sounds. Eventually we found that 20% errors were audible for some instruments and that 30% resulted in clearly and consistently distinguishable sounds.

The formal discrimination results illustrated in Figures 1 - 4 support our revised expectations: 10% error sounds are nearly identical to the originals, 30% error sounds are nearly always distinguishable from the originals, and intermediate errors bridge these extremes. Random alteration with 20% error seemed to generate the best collection of “similar, yet different” sounds compared to the originals.

Random spectral alteration provides an efficient method for generating sets of musical sounds with the same rise time and spectral centroid. Such sounds are audibly distinct if the error is about 15% or more.

6 Acknowledgments

This work was supported in part by the Hong Kong Research Grant Council's Projects HKUST6089/01E and HKUST6194/02E. We would like to thank Simon Cheuk-Wai Wun for his excellent listening test program, and Wu Yuk Ying and Francis Hing-Cheung Wong for managing the listening test.

References

Grey, J. M., and Gordon, J. W. 1978. "Perceptual effects of spectral modification on musical timbres," *J. Acoust. Soc. Am.* 63:1493-1500.

Horner, A., Beauchamp, J., and Haken, L. 1993. "Methods for Multiple Wavetable Synthesis of Musical Instrument Tones," *J. Audio Eng. Soc.* 41:336-356.

Iverson, P., and Krumhansl, C. L. 1993. "Isolating the dynamic attributes of musical timbre," *J. Acoust. Soc. Am.* 94:2595-2603.

Krimphoff, J., McAdams, S., and Winsberg, S. 1994. "Caractérisation du timbre des sons complexes. II. Analyses acoustiques et quantification psychophysique," *J. de Physique* 4(C5):625-628.

Lakatos, S. 2000. "A common perceptual space for harmonic and percussive timbres," *Perception & Psychophysics* 62:1426-1439.

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

Moore, B. C. J., Glasberg, B. R., Baer, T. 1997. "A model for the prediction of thresholds, loudness and partial loudness", *J. Audio Eng. Soc.* 45:224-240.

Toole, F. E., and Olive, S. E. 1988. "The modification of timbre by resonances: perception and measurement," *J. Audio Eng. Soc.* 36:122-142.

Watkins, A. J., and Makin, S. J. 1994. "Perceptual compensation for speaker differences and for spectral-envelope distortion," *J. Acoust. Soc. Am.* 96:1263-1282.

Watkins, A. J., and Makin, S. J. 1996. "Effects of spectral contrast on perceptual compensation for spectral-envelope distortion," *J. Acoust. Soc. Am.* 99:3749-3757.