

# SUBJECTIVE EVALUATION OF MUSICAL INSTRUMENT TIMBRE MODIFICATIONS

*Minna Ilmoniemi<sup>1,2</sup>, Vesa Välimäki<sup>1</sup>, and Minna Huotilainen<sup>2,3</sup>*

<sup>1</sup>Helsinki University of Technology  
Lab. of Acoustics and Audio Signal Processing  
P.O. Box 3000, FIN-02015 HUT  
Espoo, Finland  
minna.ilmoniemi@acoustics.hut.fi

<sup>2</sup>Cognitive Brain Research Unit  
Dept. of Psychology  
P.O. Box 9, FIN-00014 University of Helsinki  
Helsinki, Finland  
<sup>3</sup>Collegium of Advanced Studies  
P.O. Box 4, FIN-00014 University of Helsinki  
Helsinki, Finland

## ABSTRACT

In this paper psychoacoustic distances, or perceived differences, within sounds that differ from each other in terms of musical timbre are studied. The sounds were modified in four timbre dimensions: the ratio of even and odd harmonic components, brightness, the attack time, and the amount of noise. The psychoacoustic distances between the sounds differing only in timbre were evaluated through a subjective listening test. A set of sounds having an equal psychoacoustic distance from the reference sound was generated for a brain research experiment.

## 1. INTRODUCTION

Musical timbre is an attribute that allows a listener to discriminate sounds having the same pitch, duration, and loudness. Often timbre is used to refer to the “tone quality” or “tone color”. Timbre is a multidimensional attribute on which both the spectral content and temporal patterns have an effect. The multidimensional scaling method (MDS) [1] has been utilized in interpreting the results in timbre dimension studies. In MDS a map describing the psychoacoustic distances of sounds in a multidimensional geometric space is constructed. The dimensions of the map correspond to the different features of timbre. A short distance between two sounds in the timbre space corresponds to a high perceived similarity whereas a long distance corresponds to a high perceived dissimilarity.

In many experiments the relative amplitudes of lower and higher harmonics and the variations in amplitude envelopes have been found to contribute to the perception of timbre and similarity judgments of sounds [2][3][4]. The spectral energy distribution, which is highly correlated to the spectral centroid (i.e., the mean frequency of the spectrum) and the brightness of a sound, has a great effect on timbre. Also the nature of the attack segment is related to the timbre. The attack is important especially for the sound source identification. In musical instruments the timbre is often determined more by the onset than by the remainder (i.e., tone with the onset removed). It is known that many wind instruments cannot be identified if only the remainder is considered. However, the attributes contributing to the timbre perception are present throughout tones, not only on onsets or remainders.

In this paper psychoacoustic distances between sounds differing from each other only in timbre are studied. Timbre modifications and perceptual differences within the sounds were evaluated through a subjective listening test. The main purpose was to generate a sound set in which each timbre has an equal perceived distance from the reference. Sound sets with equal psychoacoustic distances among the sounds are needed for example in brain research tasks.

The outline of this paper is as follows. In Section 2 the generation of test sounds is explained and the description of the listening test arranged in order to evaluate timbre modifications is given. In Section 3 the listening test results are presented. Section 4 concludes the paper.

## 2. LISTENING TEST

### 2.1. Test sounds

Test signals were generated by processing a natural sound of the cello taken from the McGill University Master Samples library [5]. A cello tone was selected due to its harmonicity, attack segment properties, and amplitude envelope. The cello sound had a fundamental frequency of 524 Hz. The sampling rate used was 44.1 kHz. The sound was truncated after 200 ms from the onset and the amplitude in the final 20 ms was attenuated linearly. The musical timbre of the cello sound was altered in four dimensions that were carefully selected according to psychoacoustical research. The timbre dimensions are the ratio of even and odd harmonic components, brightness, the attack time, and the amount of noise.

Natural-sounding test sounds were reconstructed from the harmonic components that were extracted from the original cello tone [6][7]. The 30 lowest harmonics of the cello sound were extracted by using a fractional delay inverse comb filter with a resonator that picks up single harmonics. The sounds differing from each other in timbre were generated by processing and combining the extracted harmonics according to timbre parameter values.

The ratio of even and odd harmonic components was modified by amplifying even harmonics and attenuating odd harmonics, or vice versa. The brightness of the sound is related to the centroid frequency so that the higher the centroid frequency the brighter the sound. In brightness dimension the centroid frequency was altered by amplifying or attenuating the extracted harmonics exponentially depending on whether the centroid frequency was required to be higher or lower, respectively. The coefficients for modifying the centroid frequency lower and higher are illustrated in Figs. 1(a) and 1(b), respectively.

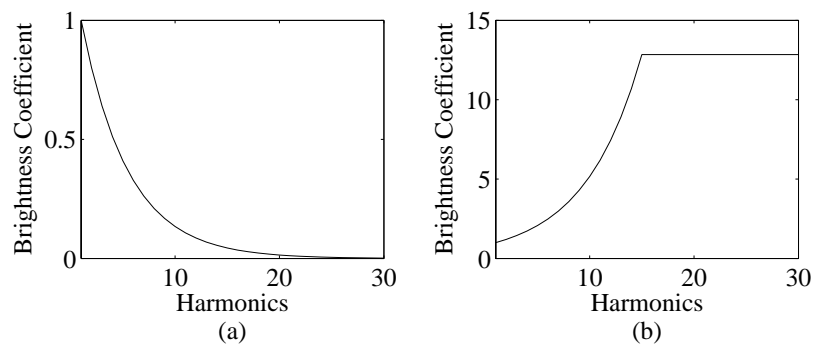


Figure 1. *The coefficients for modifying the centroid frequency (a) lower and (b) higher.*

The attack time was modified by processing the harmonic waveforms. The curves for lengthening and shortening the attack time are shown in Figs. 2(a) and 2(b), respectively. Each harmonic was operated with one of these curves. In noise dimension the background noise of the original sound was extracted by removing all harmonics so that only noise was remaining. The amount of noise was modified by adding the noise to the reference sound or by reducing it from the reference. The reference sound was reconstructed from the extracted harmonics without any processing.

Each test sound had an equal fundamental frequency (524 Hz) and duration (200 ms). The sounds were also equalized for perceived loudness so that two listeners adjusted the loudness of the sounds to be equal with interactive software. Both listeners did the equalization twice. The averages of the adjustments were used in the final equalization process.

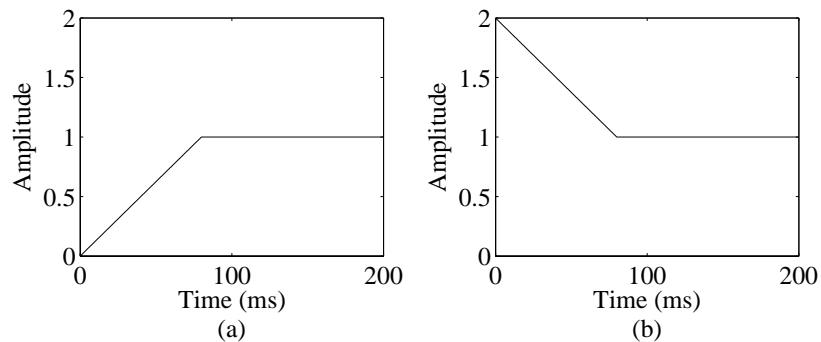


Figure 2. The envelope curves for (a) lengthening and (b) shortening the attack time.

In the listening test altogether 33 different sounds including the reference sound, i.e., eight sounds in each dimension of timbre, were used. In each dimension the parameter value controlling the timbre was smaller than in the reference sound in four sounds and correspondingly larger than in the reference in four other sounds. The parameter values were chosen in a preliminary listening. The authors defined the difference threshold for each dimension and then the other parameter values were determined so that there would be almost equal perceived differences in each dimension. The number of the test sounds was restricted so that the listening test would not take too much time but still reliable results would be obtained. In the test each sound differed from the reference with respect to one dimension only.

## 2.2. Subjects and test methods

Six subjects (two women and four men) with reportedly normal hearing participated in the listening test. The subjects were 23-30 years old with an average age of 26.3 years. Four of the subjects were the personnel of the HUT Laboratory of Acoustics and Audio Signal Processing but also two other subjects had previous experience either in acoustics or string instruments. The listening test was implemented using the experimental listening test software GuineaPig [8] and it was arranged in the listening room of the HUT Acoustics Laboratory. The sound samples were played through Sennheiser HD 580 headphones.

The listening test was an A/B scale (hidden reference) test. In the test each modified sound and the reference were played in pairs but the subjects did not know which one was the reference. Also a pair with two reference sounds was included. The subjects were allowed to listen to the sounds as many times as they wanted and they could also have breaks whenever they wanted. The subjects were told that the sounds differ from each other only in timbre. They were asked to evaluate how different sample B is compared to sample A. The subjects evaluated the psychoacoustic distance of each pair on a continuous scale from 0 (sounds are the same) to 10 (sounds are totally different). The evaluation was done by moving a slider on a computer screen. The user interface of the listening test is shown in Fig. 3.

In the listening test, 33 different sound pairs were included. Each pair was presented four times so that sample A was the reference twice and sample B was the reference also twice. Thus, altogether 132 sound pairs were presented in the test. Two different lists of sound pairs were generated from which three subjects had one and three subjects the other. The sound pairs were presented in random order.

In the beginning of the test the subjects had the possibility to hear some of the sounds and note differences between them. A sound pair with two reference sounds and eight pairs including the reference and the sounds having the most different parameter values in each dimension (i.e., two sounds in each dimension) were presented to the subjects. It was told if the sounds were the same or totally different. After this instruction there was a training session of six sound pairs where the subjects could practice the evaluation before the actual test. The subjects were asked if the volume level was optimal for them or if they wanted to adjust it. The listening test with instruction and training sessions took about 40 minutes per subject. After the test the subjects were asked for comments on the test and sounds.

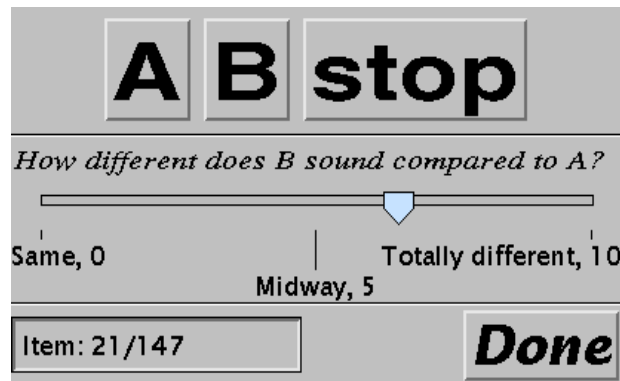


Figure 3. A graphical user interface of the listening test.

### 3. RESULTS

In this section the results of the listening test are presented. The results tell how the timbre modifications were perceived and how the psychoacoustic distances between the sounds were evaluated.

The averages of the psychoacoustic distances evaluated by comparing the sounds to the reference sound are given in Table 1. The results are presented separately for each dimension so that the timbre parameter value increases with increasing index number. The index numbers 1-4 correspond to the sounds having a smaller parameter value than the reference, index 5 corresponds to the reference sound paired with itself, and indices 6-9 correspond to the sounds having a greater parameter value than the reference. The reference sound compared to the reference sound itself gave the psychoacoustic distance of 0.7. This reveals the noise level of the test and the accuracy of the listeners. In the scale of 0-10 the psychoacoustic distance of 0.7 is very small, which means that the subjects have evaluated the perceptual differences quite accurately.

Table 1. *Psychoacoustic distances of the sounds compared to the reference sound.*

Index	Harmonics	Brightness	Attack Time	Noise	Comment
1	8.7	9.8	5.2	7.1	Smallest parameter value
2	7.2	8.5	4.8	7.2	
3	4.9	6.9	3.8	5.6	
4	2.1	4.3	2.6	1.5	
5	0.7	0.7	0.7	0.7	Reference
6	1.3	5.0	2.0	4.6	
7	4.0	6.8	2.5	6.1	
8	5.7	8.3	4.0	6.8	
9	8.6	9.3	2.6	8.4	Largest parameter value

The psychoacoustic distances with a confidence interval of 95% are shown in Fig. 4. The results are presented separately for each dimension so that the parameter controlling the timbre is increasing for each point, being smaller than in the reference sound for the first four points (see Table 1, indices 1-4) and greater than in the reference for the last four points (see Table 1, indices 6-9). The middle point of each dimension is the reference sound paired with itself (see Table 1, index 5). For more detailed results, see [6].

The subjects were asked to give comments on the test. The general impression and comment was that it was hard to compare the sounds because they were different in terms of so different attributes.

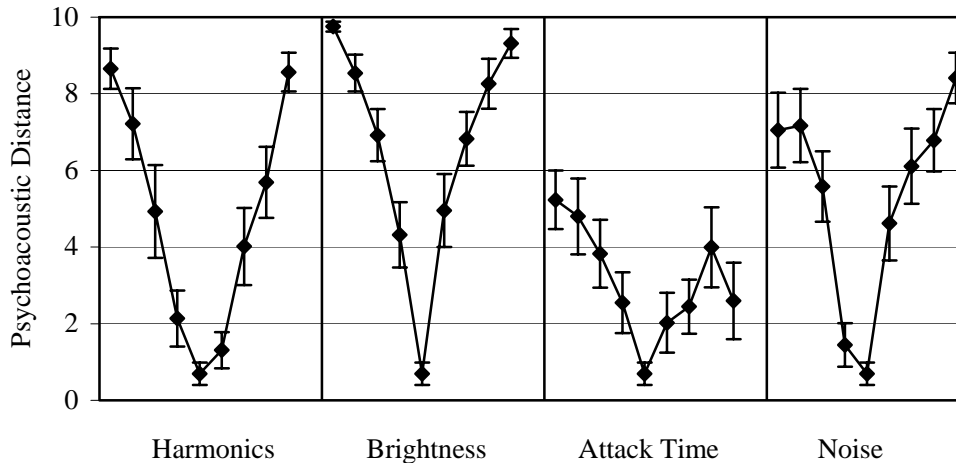


Figure 4. Psychoacoustic distances of the sounds compared to the reference with 95% confidence intervals.

From Fig. 4 it is evident that the required set of sounds can be generated by choosing sounds and timbre parameter values from the same level of psychoacoustic distance. In harmonic, brightness, and noise dimensions the timbre modifications and differences in sounds were observable and large enough. Also in attack time dimension observable differences were obtained but we did not succeed in having large enough differences in this dimension. We generated the required sound set by choosing two sounds for harmonic, brightness, and noise dimensions: one with the parameter value that is smaller than in the reference sound and the other that is greater than in the reference. We chose the sounds with a psychoacoustic distance of 7. For harmonic dimension the sound with index number of 2 (see Table 1) was chosen, the other parameter value was interpolated between the sounds having indices 8 and 9. For brightness dimension the sounds with indices 3 and 7 were chosen, and for noise dimension the indices were 1 and 8. The required sound set includes also the reference sound.

#### 4. CONCLUSIONS

In this paper, the perceived differences between sounds differing only in terms of timbre were evaluated. The sounds differed from each other in four timbre dimensions. A subjective listening test was arranged in order to evaluate timbre modifications and the results of the test were regarded as psychoacoustic distances between the sounds. A set of sounds having an equal psychoacoustic distance from the reference sound was generated. Sounds from only the three best dimensions were chosen to the final sound set. The sound set was generated for a brain research experiment where different dimensions are also combined so that the sounds differ from the reference in one to three timbre dimensions.

Within this study it was observed that it is possible to generate and evaluate sounds that are modified in different timbre dimensions. It is also possible to generate a sound set consisting of the sounds that are perceived as equally different compared to the reference.

#### 5. REFERENCES

- [1] Kruskal, J. B. "Nonmetric multidimensional scaling: A numerical method," *Psychometrika*, 29:115-129, 1964.
- [2] Grey, J. M. "Multidimensional perceptual scaling of musical timbres," *J. Acoust. Soc. Am.*, 61(5):1270-1277, May 1977.

- [3] Wessel, D. L. "Timbre space as a musical control structure," *Computer Music J.*, 3(2):45-52, June 1979.
- [4] Iverson, P. and Krumhansl, C. L. "Isolating the dynamic attributes of musical timbre," *J. Acoust. Soc. Am.*, 94(5):2595-2603, November 1993.
- [5] Opolko, F. and Wapnick, J. *McGill University Master Samples*, McGill University, Montreal, Quebec, Canada, 1987.
- [6] Ilmoniemi, M. *Modification and Brain Recordings of Musical Instrument Tones*, Master's Thesis, Helsinki University of Technology, Laboratory of Acoustics and Audio Signal Processing, Espoo, Finland, May 2004. Available at <http://www.acoustics.hut.fi/publications/>.
- [7] Välimäki, V., Ilmoniemi, M., and Huotilainen, M. "Decomposition and modification of musical instrument sounds using a fractional delay allpass filter," *Accepted for Publication in the 6<sup>th</sup> Nordic Signal Processing Symposium*, Espoo, Finland, June 9-11, 2004.
- [8] Hynninen, J. and Zacharov, N. "GuineaPig – A generic subjective test system for multichannel audio," in *Proc. 106<sup>th</sup> Audio Engineering Society Convention*, Munich, Germany, May 1999.