# Detection of Random Spectral Alterations of Sustained Musical Instrument Tones in Repeated Note Contexts

**Chung Lee**
The Information Systems Technology and Design Pillar,
Singapore University of Technology and Design,
20 Dover Drive, Singapore 138682
chung_lee@sutd.edu.sg

**Andrew Horner**
Department of Computer Science and Engineering,
Hong Kong University of Science and Technology,
Clear Water Bay,
Kowloon, Hong Kong
horner@cse.ust.hk

## ABSTRACT

Eight sustained musical instrument sounds were randomly altered by a time-invariant process to determine how well spectral alteration could be detected on repeated notes. Sounds were resynthesized in a series of eight 0.25-second repeated notes and spectrally altered with average spectral alterations of 8, 16, 24, 32, and 48%. Listeners were asked to discriminate each randomly altered repeated note sequence from the original unaltered sequence. The results showed that spectrally altered repeated note sequences were significantly more discriminable than single tones in comparisons of the same duration (two seconds). Non-uniform repeated note sequences were more discriminable than uniform sequences that simply repeated the same random instance.

## 1. INTRODUCTION

One of the most common criticisms of music synthesizers and soundcards is that the sound is too uniform and lacks the natural variations of acoustic instruments. In fact, if two notes are played at exactly the same amplitude and pitch on most synthesizers, the two notes would be identical. The problem is especially pronounced on repeated notes of the same pitch, such as double-tongued wind tones. Clearly there needs to be enough note-to-note variations to make each note sound different and yet in-character for the instrument.

To address this problem, Horner *et al.* proposed a linear random spectral alteration model which introduced small but noticeable timbral variations into each note [1]. The model randomly alters the spectrum of a note with controllable levels. The approach largely preserves spectral centroid and attack time, which are widely recognized as two of the most salient attributes in timbre perception [2, 3, 4, 5, 6]. However, the model has only been tested on individual, isolated tones - an important but preliminary step.



Note duration is an important factor in the detection of randomly altered spectra. Is a longer altered note more discriminable than a shorter one because listeners have more time to detect the alteration? Or, is it easier to hear alterations in shorter tones since less memory is required to make the comparison? If shorter notes are repeatedly joined together with the same duration as a single long note, which is more discriminable?

The current study seeks to test the discrimination of spectral alteration on repeated note sequences to address the above questions as well as the following issues: How much spectral alteration is needed to make the altered repeated notes distinguishable from the original? Do repeated notes make it easier or harder to hear alterations by exposing or hiding the alterations?

To answer these questions, the current study aims at enlarging knowledge about the perception of musical sounds by systematically evaluating how well listeners can discriminate spectral changes in a sequence of repeated notes. This work has wide applications in sound design for synthesizers, soundcards, and software synthesis.

### 1.1 Previous Work Done on Spectral Alteration

McAdams *et al.* investigated resynthesized tones where spectrotemporal parameters were simplified using various methods [7]. Instrument tones from different families (strings, brass and woodwinds) were tested. Listening test subjects were asked to distinguish the original instrument tones from those data-reduced using methods similar to those used by [8]. These data reductions smoothed the micro-variations in the tones.

Gunawan and Sen studied the discrimination thresholds for changes to spectral envelopes. They altered the spectral envelopes using 14 zero-phase bandpass filters with various center frequencies and bandwidths [9]. Results showed that changes to the first few lower harmonics were more audible than to higher harmonics.

Most relevant to the current study, one of the authors of the current study investigated the time-invariant alteration of musical instrument spectra, where each harmonic was multiplied by a time-invariant random scalar [1]. It was found that listeners had more difficulty discriminating alterations to instrument sounds containing more pro-

nounced spectral variations. This suggested that dynamic spectral variations increase the difficulty of detecting spectral alterations.

### 1.2 Previous Work Done on Timbre Perception in Note-to-Note Contexts

Campbell and Heller investigated the effect of melodic context on note onset perception [10]. Their stimuli were generated from performances of two-note legato phrases (F4 to A4) played on six different instruments. Listening test subjects were asked to identify the instrument of the stimuli. Based on the results, they concluded that 110-ms legato transients gave higher identification rates than either attacks, steady states, or other shorter legato transients.

Kendall studied the importance of different partitions in the task of instrument identification of musical phases[11]. He compared the role of attack and steady state in single-note and melodic contexts. In melodic contexts, the rate of successful identification of "steady-state only" stimuli (with attack removed) was statistically equivalent to the unaltered signals (84%). However, in single-note contexts, both the "steady-state only" (50%) and the "attack only" (51%) contexts were at the same level as the unaltered tones (54%). Kendall concluded that the perceptual importance of transients had been overstated.

### 1.3 Scope of the Current Study

In the current study, listening tests were conducted to determine the discrimination of linear random spectral alterations of repeated notes. Both uniform (i.e., using the same random instance for all repeated notes) and non-uniform (using multiple random instances in repeated notes) stimuli were tested. The single-note discrimination experiment in Horner *et al.* was re-conducted for comparison [1]. Five error levels (8%, 16%, 24%, 32%, and 48%) were included. The major objective of the current study is to compare discrimination for uniform and non-uniform random spectral alterations of repeated note sequences with single-note alterations.

Section 2 outlines the stimuli preparation for the original and altered repeated note sequences. Section 3 describes the details of the listening test. Section 4 describes the results of the test, and compares discrimination of uniform and non-uniform repeated note sequences. Finally, we discuss the implications of these results.

## 2. STIMULUS PREPARATION

### 2.1 Prototype Instrument Tones

Eight sustained musical instrument tones were selected as prototype signals for the listening test. These included tones from a bassoon, clarinet, flute, horn, oboe, saxophone, trumpet, and violin performed at approximately 311.1 Hz ($E^{b}_{4}$). They represent the wind and the bowed string families. All eight instrument tones were also used by a number of timbre studies [1, 7, 12, 13, 14]. Using these samples makes it easier to compare the results from the previous studies.

### 2.2 Preparation of Reference Tones

Frequency variations, tone duration, and loudness are potential factors in discrimination. To avoid this, they were equalized in all reference tones. The reference tones were standardized to a two-second duration by interpolating the analysis data. Next, the duration-equalized reference tones were compared, and amplitude multipliers were determined such that the tones had approximately the same loudness [15]. Finally, each harmonic's frequency was set to the exact product of its harmonic number and the fixed analysis frequency, resulting in flat equally-spaced frequency envelopes. The frequency deviations were set to zero in order to restrict listener attention to the amplitude data. More details about the preparation of reference tones were described in Horner *et al* [1].

### 2.3 Analysis Method

Instrument tones were analyzed using a phase vocoder algorithm. Harmonic amplitudes were judged (by visual inspection of the spectra) to be near-zero beyond 35 harmonics for the bassoon, oboe, and trumpet tones, so a sampling rate of 22,050 Hz was used. The other tones were sampled at 44,100 Hz (70 harmonics). More details on the analysis process are given in Beauchamp [16].

### 2.4 Preparation of Repeated Note Sequences

For comparison with the previous random alteration study [1], the duration of the repeated note sequences were set to two seconds. The duration of each note should be long enough so that the sustain can be perceived by listeners. For this reason, we decided to use eight 0.25-second notes to form the repeated note sequence. The attack and decay of all notes were equalized to 0.02 seconds so that the sustain duration was long enough for discrimination. A 0.02-second attack/decay was long enough to prevent noticeable "clicks" at the beginning and end of each repeated note. Duration, attack, and decay equalization were done using the SNDAN program [16]. Horner *et al.* gives more details about the procedure [1].

### 2.5 Random Spectral Alteration

Time-invariant random alteration was performed on the analysis data the same way as in Horner *et al.* [1] by multiplying each harmonic amplitude with a randomly selected scalar. The random instance is accepted if the relative-amplitude spectral error is within 1% of the required error level, otherwise re-picked.

Spectral centroid has been shown to be strongly correlated with one of the most prominent dimensions of timbre as derived by multidimensional scaling (MDS) experiments [3, 5, 17, 18, 19, 20]. To eliminate spectral centroid from being a factor of discrimination, random alteration instances were only accepted if the peak spectral centroid of the original and altered spectra were within 2.5% of one another.

### 2.6 Choosing Random Instances for the Listening Test

Due to the random nature of random spectral alteration, Horner *et al.* [1] included ten random instances for each error level and instrument in their listening test. This averages out outlier discrimination scores. However, to limit the excessive length of this listening test, multiple random instances were not feasible. For each error level and instrument, we chose the random instance from the previous study by Horner *et al.* [1] which had a discrimination score closest to the average. The chosen instance was shortened to 0.25 seconds, with its attack and decay equalized, and repeated eight times for uniform sequences. For non-uniform sequences, we discarded the two most extreme outliers of the ten random instances in Horner *et al.* [1] and used the other eight instruments. The eight chosen random instances were shortened to 0.25 seconds, with their attacks and decays equalized, and randomly placed in a non-uniform sequence. [1]

## 3. EXPERIMENTAL METHOD

### 3.1 Subjects

Thirty listeners participated in our experiment. They were undergraduate students at the Hong Kong University of Science and Technology, ranging in age from 17 to 23 years, who reported no hearing problems. They had 5 to 15 years experience playing a musical instrument, with a mean of 8.8 years. The listeners were paid to compensate for their time spent in the experiment.

### 3.2 Stimuli

The eight musical instrument sounds were stored in 16-bit integer format on a hard disk. All "reference" sounds (resynthesized using the analysis data with strictly fixed harmonic frequencies) were equalized for duration and loudness. Five error levels (8, 16, 24, 32, and 48%) for the three sets of tones (single-note, uniform, and non-uniform repeated note sequences) gave a total of 15 modified sounds for each instrument. Using the Moore-Glasberg loudness program [15], it was confirmed that the loudness of the altered sounds matched that of the reference sounds within 2 phons.

### 3.3 Test Procedure

Following a number of related previous studies [1, 9, 14], a two-alternative forced-choice (2AFC) discrimination paradigm was used. Each listener heard two pairs of sounds and chose which pair was different. Each trial structure was one of AA-AB, AB-AA, AA-BA, or BA-AA, where A represents the reference sound and B represents one of the altered sounds. This paradigm has the advantage of not being as susceptible to variations in the listeners criteria across experimental trials as compared to the simpler A-B method. All four combinations were presented for each altered sound. The sounds of each pair were separated by a 500–ms silence, and the two pairs were separated by a 1-s silence. For each trial the user was prompted with "which pair is different, 1 or 2?," and the response was given by the keyboard. The computer would not accept a response until all four sounds in a trial had been played. For the complete listening test 480 trials were presented to each listener (four trial structures $\times$ five error levels $\times$ three set of tones $\times$ eight instruments). The order of presentation of these 480 trials was randomized.

For each altered sound the discrimination performance was averaged using the results of the four trials. Because these four trials were presented in random order within the 480 trials, the effects of possible learning were averaged out. The same trials were presented to each listener, although in a different random order. The duration of the test was less than 120 minutes, including two 5-minute compulsory rests after finishing 160 and 320 trials of the listening test. A custom program written in Java ran on an Intel PC to control the experiment.

Listeners were seated in a "quiet" room with less than 40 dB SPL background noise level (mostly due to computers and air conditioning). The covering of the ears by the headphones also provided an additional reduction of the noise level. Sound signals were converted to analog by a SoundBlaster X-Fi Xtreme Audio soundcard and then presented through Sony MDR-7506 headphones at a level of approximately 75 dB SPL as measured with a sound-level meter. The X-Fi Xtreme Audio DAC utilized 24 bits with a maximum sampling rate of 96,000 Hz and a 108-dB S/N ratio. The sounds were actually played at 22,050 or 44,100 Hz. At the beginning of the experiment each listener read the instructions and asked any necessary questions of the experimenter. Five test trials (chosen at random) were presented prior to the data trials for each instrument.

## 4. RESULTS

### 4.1 Postscreening of Subjects

To ensure the quality of the statistical data, postscreening of the subjects was necessary. Eight single-note sounds with a 48% error level, which were easily discriminable in our previous study [1], were used for post screening. Thirty-two trials were used for this purpose (four trial structures $\times$ one error level $\times$ eight instruments). These altered sounds were perfectly discriminable in the previous study [1], and subjects were expected to discriminate at least 26 out of 32 of them. Twenty-six out of thirty subjects were selected for statistical analysis.

### 4.2 Effects and Interactions of Error Level and Instrument

Discrimination scores for single-note, uniform, and non-uniform stimuli were computed for each error level for each instrument across the four trial structures for each listener. Because the presentation order of the four trials was randomized, any potential effects of learning were averaged out. Figure 1 shows the scores averaged over all instruments plotted against error level, with 95% confidence intervals indicated by the vertical bars. (A 95% confidence

[1] Listening test samples can be downloaded at `http://imleechung.wordpress.com/2014/06/26/repeated-notes-discrimination/`

interval means that if the listening test was re-run, the average discrimination score would have a 95% chance to lie in the interval)

The discrimination scores of the three types of stimuli (single-note, uniform, and non-uniform) were significantly different for error levels up to 24%. The non-uniform discrimination scores were consistently highest while the single-note discrimination scores were consistently lowest. All three converged to near-perfect discrimination (i.e., 90% discrimination scores) when the error level increased to 48%.

Figure 2 shows average discrimination plotted against error level for each individual instrument for single-note stimuli. The discrimination scores were similar though lower than the previous study done by Horner *et al* [1]. Moreover, the discrimination scores of the violin and trumpet did not yet reach to near-perfect at 48%.

For uniform repeated note sequences (Figure 3), the discrimination scores were generally higher than for single-notes, and mostly above 0.7. The deviation among the instruments at the 8% error level was the greatest, and the discrimination scores converged to near-perfect discrimination at 48%.

For non-uniform discrimination (Figure 4), scores quickly converged to near-perfect discrimination when the error level increased above 8%. The discrimination score of the violin was an outlier at 8%, with most of the others above 0.8. The bowed string sound and its high spectral incoherence probably explains the outlier.

ANOVA analysis of the results used instrument, error level, and type of stimuli as repeated measures to test the main effects of instrument (8 instruments), error level (5 error levels: 8%, 16%, 24%, 32%, and 48%), type of stimuli (single-note, uniform and non-uniform) and their two-way interactions (see Table 1). The main effects of instrument, error level, type of stimuli, and their two-way effects were confirmed by both parametric and non-parametric ANOVA analysis.

## 5. DISCUSSION

Our results showed that random spectral alteration was more discriminable in eight 0.25-second repeated notes than in a single two-second note (Figure 1), especially for low error levels. Although the sustain part of repeated notes were relatively short, alterations on repeated notes were more discriminable. Perhaps listeners found it easier to find differences in the repeated attacks and decays.

Alterations in non-uniform sequences were more discriminable than in uniform sequences (Figure 1). Perhaps because listeners had more opportunities to hear the differences in the repeated notes.

For direct comparison, we re-ran the single-note listening test in Horner *et al* [1]. The discrimination scores were lower than in the original experiment, especially for violin and trumpet on the 48% error level (Figure 2). This was probably due to the much smaller number of instrument instances presented to subjects (ten instances in the previous study and any one instance in the current study).

The violin had a dramatically lower discrimination scores compared to the other instruments at the 8% error level for non-uniform sequences (Figure 4). Other than being the only string instrument, the violin also had the highest spectral incoherence which may also be the reason for the significantly lower discrimination score. The relatively strong spectral variations of the violin effectively hid the small 8% spectral alterations that were more apparent in other instruments.

The current study has extended our understanding of timbre discrimination from single-note to repeated note contexts. Future studies can carry this further in the investigation of more complicated note-to-note contexts, and eventually to full melodic contexts.

**Acknowledgments**

This work was supported by the Hong Kong Research Grants Council grants 613112 and SUTD-MIT International Design Center Grant (IDG31200107 / IDD11200105 / IDD61200103).

## 6. REFERENCES

[1] A. B. Horner, J. W. Beauchamp, and R. H. Y. So, "Detection of random alterations to time-varying musical instrument spectra," *J. Acoust. Soc. Am.*, vol. 116, pp. 1800–1810, 2004.

[2] J. M. Grey and J. A. Moorer, "Perceptual evaluations of synthesized musical instrument tones," *J. Acoust. Soc. Am.*, vol. 62, no. 2, pp. 454–462, 1977.

[3] P. Iverson and C. L. Krumhansl, "Isolating the dynamic attributes of musical timbre," *J. Acoust. Soc. Am.*, vol. 94, no. 5, pp. 2595–2603, 1993.

[4] J. M. Grey, "Multidimensional perceptual scaling of musical timbres," *J. Acoust. Soc. Am.*, vol. 61, pp. 1270–1277, 1977.

[5] J. M. Grey and J. W. Gordon, "Perceptual effects of spectral modification on musical timbres," *J. Acoust. Soc. Am.*, vol. 63, pp. 1493–1500, 1978.

[6] S. McAdams, S. Winsberg, S. Donnadieu, G. De Soete, and J. Krimphoff, "Perceptual scaling of synthesized musical timbres : Common dimensions, specificities, and latent subject classes," *Psychological Research*, vol. 58, pp. 177–192, 1995.

[7] S. Mcadams, J. W. Beauchamp, and S. Meneguzzi, "Discrimination of musical instrument sounds resynthesized with simplified spectrotemporal parameters," *J. Acoust. Soc. Am.*, vol. 105, no. 2, pp. 882–897, 1999.

[8] G. R. Charbonneau, "Timbre and the perceptual effects of three types of data reduction," *Computer Music J.*, vol. 5, pp. 10–19, 1981.

[9] D. Gunawan and D. Sen, "Spectral envelope sensitivity of musical instrument sounds," *J. Acoust. Soc. Am.*, vol. 123, pp. 500–506, 2007.

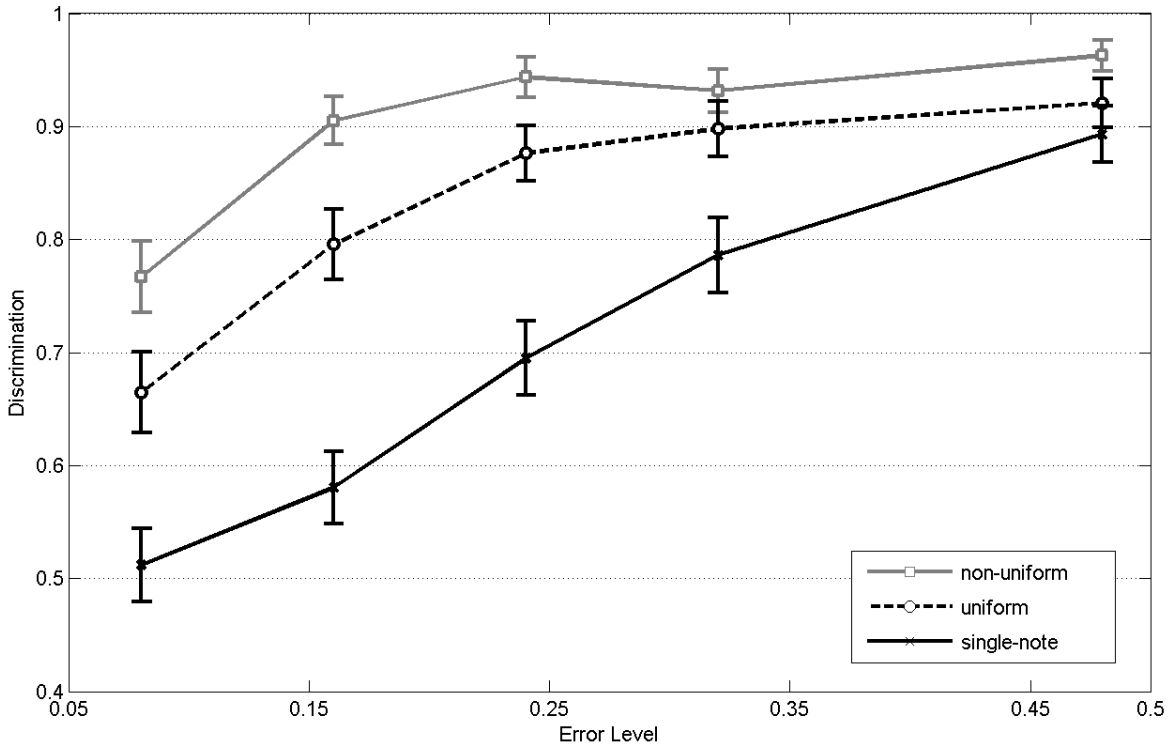


**Figure 1**. Mean discrimination scores for the eight instruments versus error level for single-note, uniform, and non-uniform stimuli. The vertical bars indicate 95% confidence intervals.

[10] W. C. Campbell and J. J. Heller, "Convergence procedures for investigating music listening tasks," *Bulletin of the Council for Research in Music Education*, no. 59, pp. pp. 18–23, 1979.

[11] R. A. Kendall, "The role of acoustic signal partitions in listener categorization of musical phrases," *Music Perception: An Interdisciplinary Journal*, vol. 4, no. 2, pp. 185–213, 1986.

[12] J. W. Beauchamp, A. B. Horner, H. Koehn, and M. Bay, "Multidimensional scaling analysis of centroid- and attack/decay-normalized musical instrument sounds," *J. Acoust. Soc. Am.*, vol. 120, no. 5, p. 3276, 2006.

[13] A. B. Horner, J. W. Beauchamp, and R. H. Y. So, "Evaluation of mel-band and mfcc-based error metrics for correspondence to discrimination of spectrally altered musical instrument sounds," *J. Audio Eng. Soc*, vol. 59, no. 5, pp. 290–303, 2011.

[14] C. Lee and A. B. Horner, "Discrimination of mp3-compressed musical instrument tones," *J. Audio Eng. Soc*, vol. 58, no. 6, pp. 487–497, 2010.

[15] B. C. J. Moore, B. R. Glasberg, and T. Baer, "A model for the prediction of thresholds, loudness, and partial loudness," *J. Audio Eng. Soc*, vol. 45, no. 4, pp. 224–240, 1997.

[16] J. W. Beauchamp, *Analysis, Synthesis, and Perception of Musical Sounds*. Springer New York, 2007, ch. Analysis and Synthesis of Musical Instrument Sounds, pp. 1–89.

[17] D. L. Wessel, "Timbre space as a musical control structure," *Computer Music J.*, vol. 3, no. 2, pp. 45–52, 1979.

[18] C. L. Krumhansl, *Structure and Perception of Electroacoustic Sounds and Music*. Excerpta Medica, Amsterdam, 1989, ch. Why is musical timbre so hard to understand ?, pp. 43–53.

[19] R. A. Kendall and E. C. Carterette, "Difference thresholds for timbre related to spectral centroid," in *Proc. 4th International Conference on Music, Perception and Cognition, Montreal, Faculty of Music, McGill University*, 1996, pp. 91–95.

[20] S. Lakatos, "A common perceptual space for harmonic and percussive timbres," *Perception & Psychophysics*, vol. 62, no. 7, pp. 1426–1439, 2000.

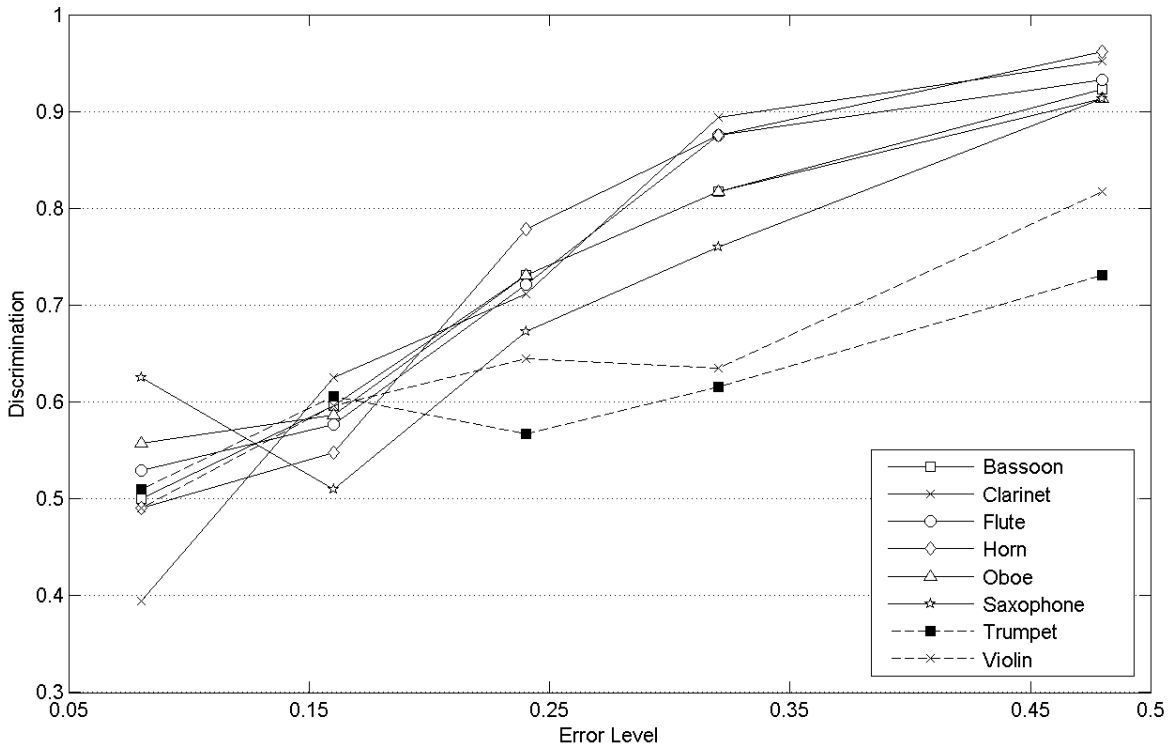


**Figure 2**. Average discrimination scores versus error level for the eight instruments for single-note stimuli.

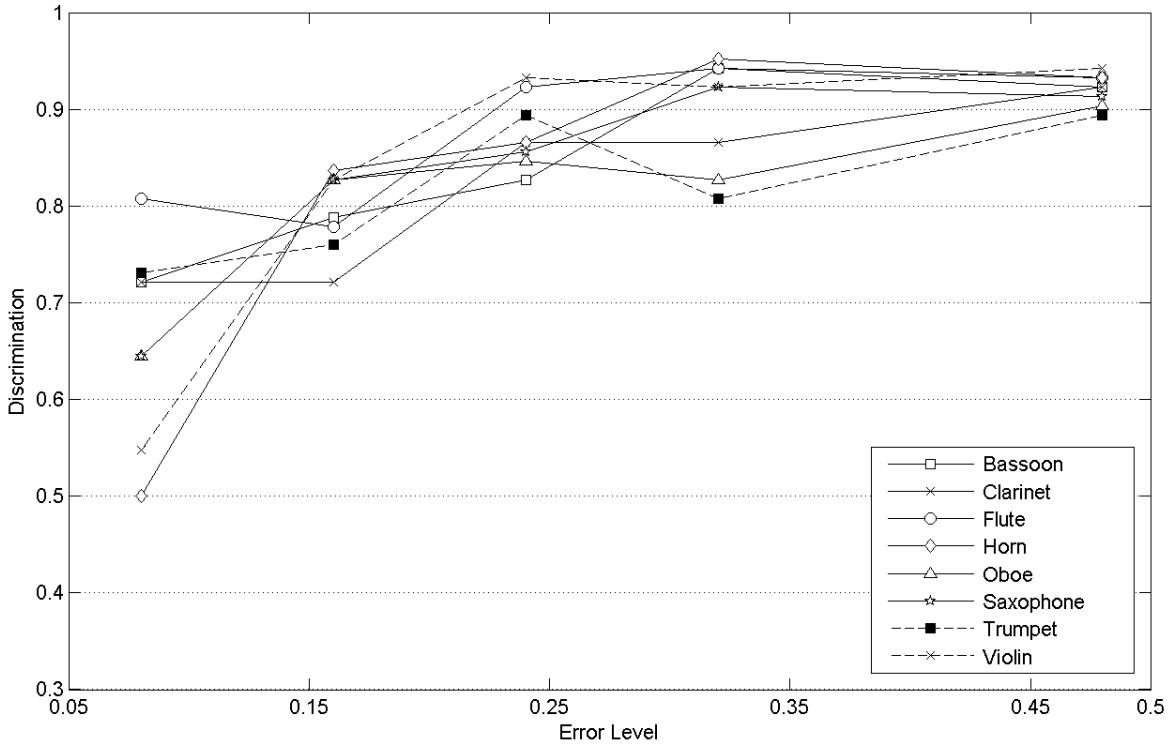


**Figure 3**. Average discrimination scores versus error level for the eight instruments for uniform stimuli.

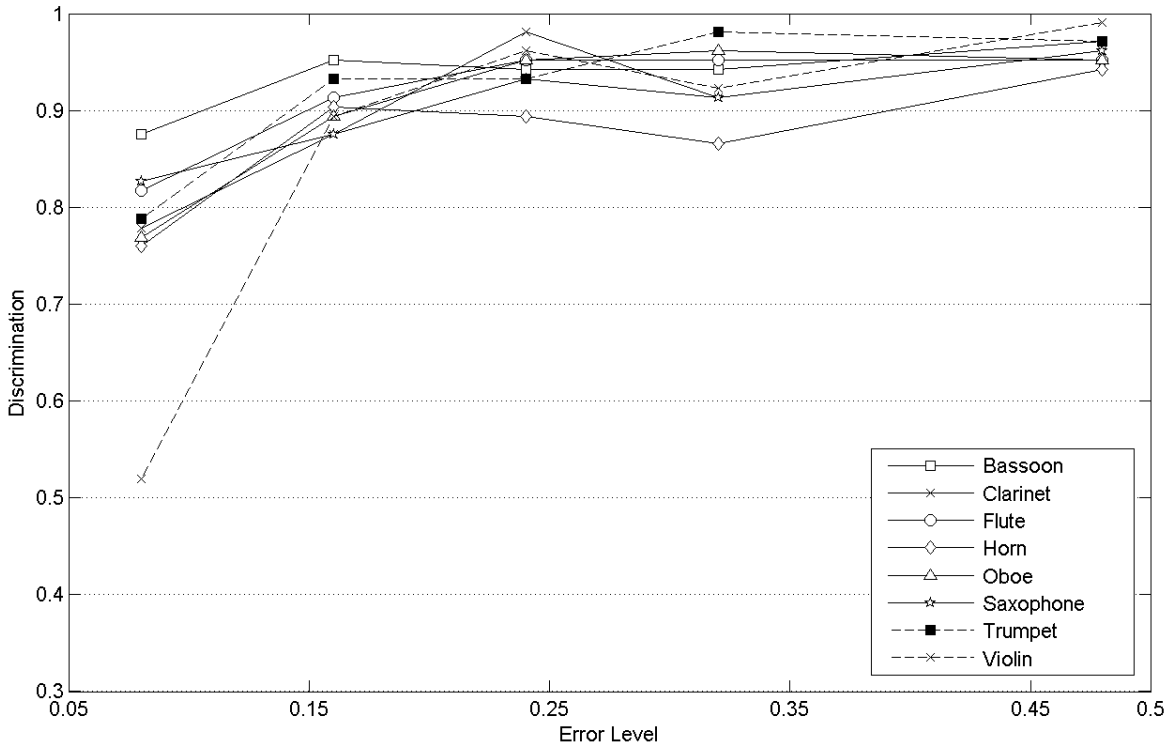


**Figure 4**. Average discrimination scores versus error level for the eight instruments for non-uniform stimuli.

| Source | DF | Sum of Square | Mean Square | F Value | Prob>F |
|---|---|---|---|---|---|
| Instrument | 7 | 1.282 | 0.183 | 4.766 | <0.0001 |
| Error Level | 4 | 29.137 | 7.284 | 189.503 | <0.0001 |
| Stimuli Type | 2 | 23.413 | 11.707 | 304.550 | <0.0001 |
| Instrument & Error Level | 28 | 2.849 | 0.102 | 2.647 | <0.0001 |
| Instrument & Stimuli Type | 14 | 1.693 | 0.121 | 3.146 | 0.0001 |
| Error Level & Stimuli Type | 8 | 4.582 | 0.573 | 14.899 | <0.0001 |
| Measurement Error | 3056 | 117.469 | 0.038 | | |
| Corrected Total | 3119 | 180.424 | | | |

**Table 1**. ANOVA table illustrating the main effects and two-way interactions of instrument (eight instruments), error level (five levels: 8, 16, 24, 32, and 48%), and stimuli type (three types: single-note, uniform, and non-uniform) on data collected from 30 listeners participating in the discrimination experiment. Data are the percentage of correct discrimination scores (100%, 75%, 50%, 25%, and 0%) over each group of four trials. All main effects are confirmed with non-parametric Friedman ANOVA by ranks.