TEMPORAL THRESHOLD OF PITCH DETECTION

Chris Howe
Rithesh Kumar
Anand Mahadevan

Christopher Latina
Raja Raman

MUSI-6001
Georgia Tech Center for Music Technology

ABSTRACT

Pitch detection is one of the most vital parameters that enhance music cognition. It is a distinctive feature that allows us to separate one song from another. Pitch relationships dictate the melodic and harmonic qualities of a piece of music. This experiment focuses on characterizing the various attributes that could affect pitch detection. The experiment was validated over a setting of music enthusiasts, and the subsequent analysis and discussion are presented.

1. INTRODUCTION

Inspired by György Ligeti's 1968 piece Continuum for harpsichord, this paper investigates the threshold of pitch discernibility. "Written for harpsichord, the piece challenges the performer's ability to play extremely fast sequences of notes on two keyboards, and the listener's perception of those notes, which no longer appear as individual sounds but as part of a sonic continuum." [1] Ligeti describes the paradox of continuity in his piece, "it takes about eighteen separate sounds per second to reach the threshold where you can no longer make out individual notes". [2]

While this phenomenon is novel, the understanding of melody is inherent to musical comprehension. Using a patch created in Max/MSP, the experiment tested the distinguishability of pitch order with varying note onset timings for a set of decay times. The experiment played 60 sets of three-note pure tone sequences at multiple speeds and decay times presented in a randomized order. Participants selected the order in which they heard the pitches and submitted their answers. For each decay time, *d*, we repeated the process for a set of inter-onset frequencies, Δf, with the intention to draw out a correlation between three parameters: pitch discernibility, decay time, and inter-onset frequencies.

2. RELATED WORK

Existing work in auditory psychology investigates the effects of short-term memory on pitch perception and ability to discern auditory streams. Diana Deutsch's series of works on Short Term Memory Tones [3,4] reveal a number of thresholds regarding pitch discernibility for differences in note sequences. Further research into Albert Bregman's concept of stream segregation alluded to a psychoacoustic threshold of pitch discernibility. His classic experiments in stream segregation offer some insight into Gestalt explanations for audio perception. Bregman's experiment "Auditory Streaming and Apparent Motion: repeating cycle of six tones" presented listeners with two streams of three note sequences, one in the high frequency range and one in the low frequency range. He observed that listeners found it difficult to distinguish the note sequence pattern when the sequence was played at a rate of 10 notes per second [5]. Figure 1, below, visually describes the concept of stream segregation. On the left, the sequence appears as a singular pattern. When sped up, the high and low streams emerge. While Bregman's experiment focused on the cognition of high and low streams, this paper focuses on the threshold of cognition for a single stream. Doughty and Garner's 1948 paper, "Pitch characteristics of short tones. I. Two kinds of pitch threshold" discusses the significance of decay on pitch perception as a function of pitch frequency and intensity [6]. These results show that three thresholds exist for pitch perception in relation to decay, namely a click, a semblance of a tone, and a tone at times between 2ms and one half of a second. However, Doughty and Garner did not investigate the effect of inter-onset frequency.

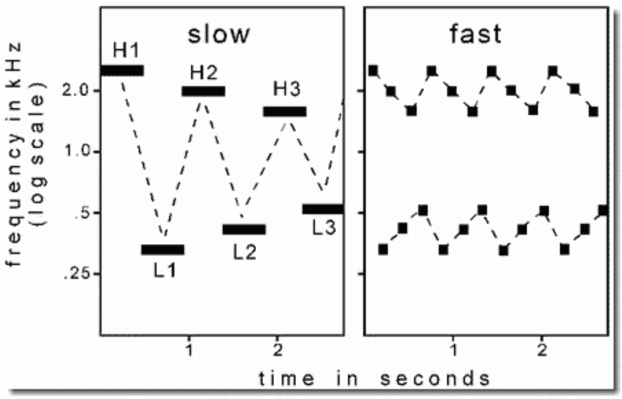

Figure 1. Stream Segregation [5]

3. HYPOTHESIS

At higher inter-onset frequencies (less time in between note onsets in the sequence), listeners will be less likely to report the correct note sequence. At inter-onset frequencies lower than 5Hz, listeners will likely always be able to distinguish the pitch sequence for pure tones.

At very short decay times, listeners will likely have difficulty reporting the correct pitch sequence because only the onset and a short attack is prominent resulting in an un-pitched click. At longer decay times, there is an overlapping or smearing effect, which interferes with clarity and order recognition.

4. EXPERIMENT DESIGN

The experiment presents a randomly selected sequential permutation of three consecutive whole tones. The tonic note is also randomly selected from A3 (220Hz), A5 (880 Hz), or A7 (3520 Hz). The pitches used are A, B and C#. The inter-onset frequencies, *Δf*, are the rate at which the notes are played. 5 Hz for example represents 1/5th of a second in between note onsets. The inter-onset frequencies tested were a range between 5 and 60 Hz incremented by steps of 5 Hz. These were ordered randomly for each class of decay times.

A set of five decay times were tested: 50ms, 100ms, 250ms, 1s and 2s. After hearing the sample, listeners order the three notes by low, middle and high and submit their answer. The Max/MSP patch documented the listener's answer, the correct answer, the tonic note, inter-onset frequencies and decay time.

5. SOFTWARE IMPLEMENTATION

5.1 Front End Interface

5.2

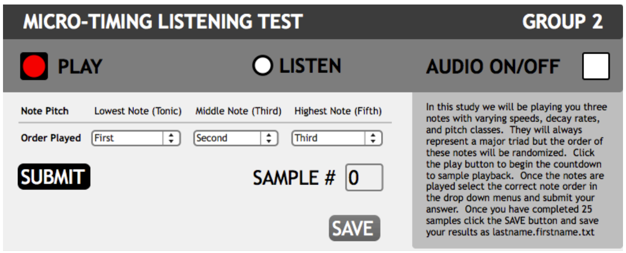

Figure 2. Screenshot of Max/MSP Listening Experiment

Subjects are asked to first turn on their audio output to ensure that the test tones are audible. Once the subject clicks the play button, a countdown provides visual cues leading up to the test tones. Then, the subject is asked to place the order of each note in relation to its pitch. Once the subject has entered their answer, they submit them to be logged into a .txt file. After all 60 test sequences have been played, the subject saves the .txt file and submits the results via email.

Figure 3 shows the initial selection of interval relationships from the *coll* MajorIntervals.txt. Once the *playBang* is received from the subject's input the intervals are randomly selected and sent to trigger the corresponding sine wave generator.

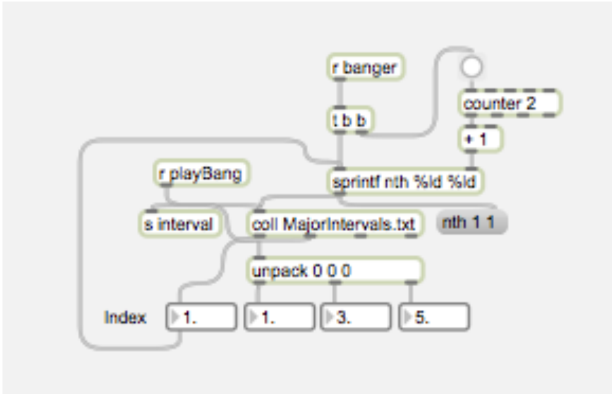

Figure 3. MajorIntervals.txt coll

Figure 4 shows the next step in the *playBang* signal path. Pitch class selection is randomly assigned via the PitchClass.txt *coll*. Assigned tonic frequencies are selected from this *coll* of three possible options; A3 (220Hz), A5 (880 Hz), or A7 (3520 Hz). The sub-patcher *p Interval* calculates the equal tempered interval frequency for each randomly selected interval sent by the MajorIntervals.txt *coll*. The output of p Interval is then routed to each corresponding sine wave generator.

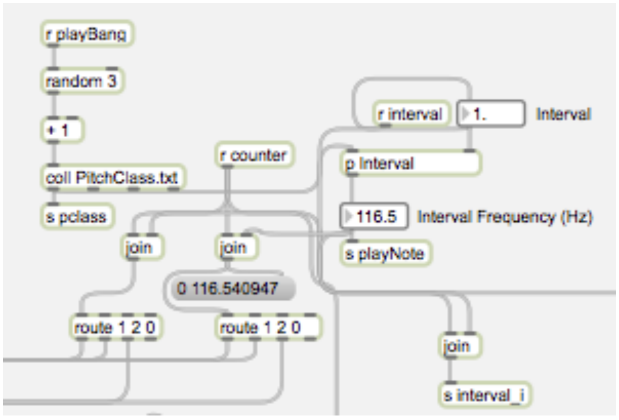

Figure 4. Random interval selection and calculation

In an attempt to compensate for variations in pitch perception loudness, a Fletcher-Munson curve system was implemented [8]. Taking into account the maximum dBFS of the system a scaling factor was applied to each sine wave output. Unfortunately, without full control over the final dB output for each subject's speaker system this compensation is fruitless. In the end, since our sine waves per set were of a similar frequency bandwidth, amplitude variations were minimal, and our results showed no favoring towards a specific frequency range. With the lack of verification over subjects' headphones, speakers or volume control, the Fletcher-Munson curve system was an attempt in vain.

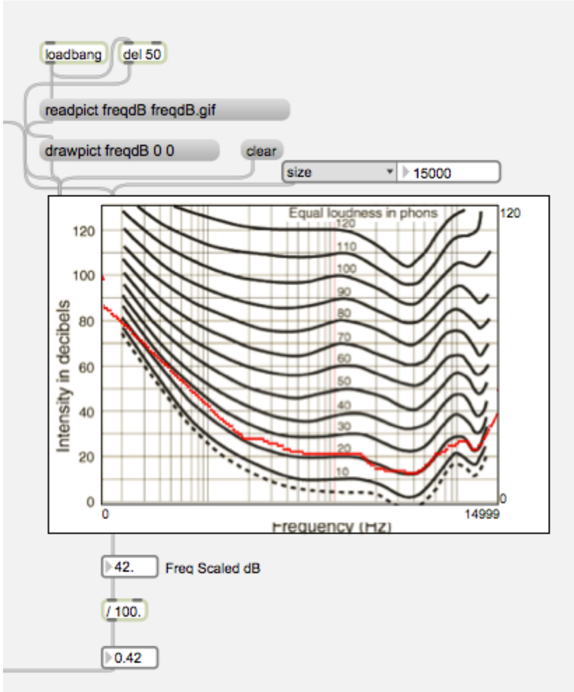

Figure 5. Fletcher-Munson curve compensation

The CNMAT external object *decaying-sinusoids~* was used for sine wave generation [9]. Input arguments for *decaying-sinusoids~* include frequency, normalized amplitude, and decay rate. In order to ensure overlapping decay times, decay rate must increase per sine wave generator. Figure 6 shows the passing of amplitude information and scaling of the decay rate per generator. Each sine wave generator is triggered in sequence from left to right with the assigned random interval and pitch class data collected earlier in the program. The output of *decaying-sinusoids~* is then sent to the audio output, or *dac~*, for the subject to hear.

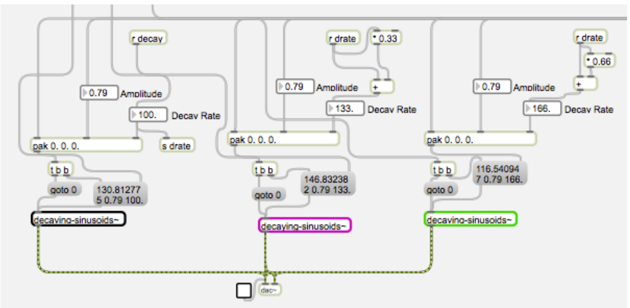

Figure 6. decaying-sinusoids~ audio output

The final step in the Micro Timing Listening Test program is logging and saving the results. Figure 7 shows the sub-patcher that logs incoming subject data per sample. For each sample set pitch class, inter onset frequency, decay rate, submitted and correct intervals are all logged. This data is indexed and collected in a *coll*, which is ultimately saved as a .txt file by the subject and submitted via email.

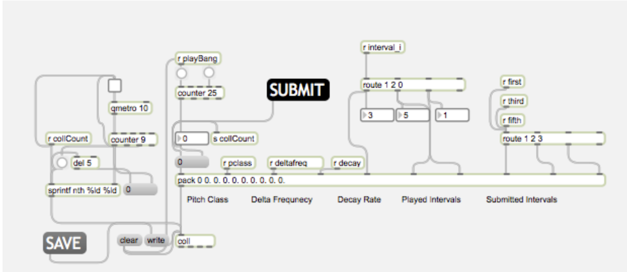

Figure 7. Data collection *coll*

6. RESULTS AND FIGURES

6.1 Results

The data set contained submissions from 26 individuals. Each had 60 trials comprised of 12 inter-onset frequencies across 5 decay times.

After calculating and plotting the mean values independently for inter-onset frequencies and decay, a general trend emerged that supported our hypothesis for inter-onset frequencies but not decay.

Further analysis with Anova confirmed the impression that the results supported our hypothesis for inter-onset frequencies but not decay time.

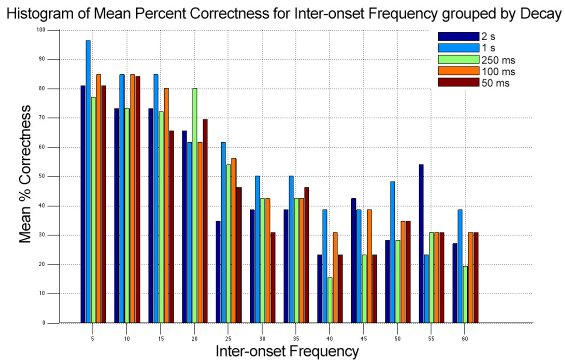

Figure 8. Grouping of decay times displayed along inter-onset frequency against percent correctness.

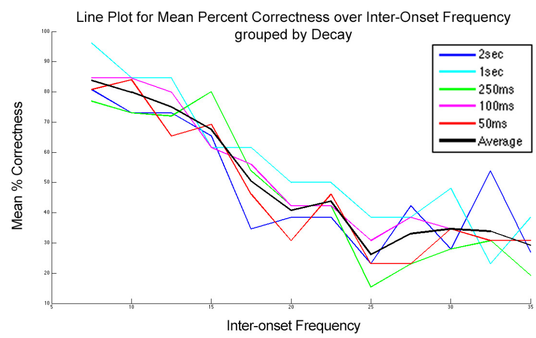

Figure 9. Percent correctness over inter-onset frequency for each decay. The black line represents the average of all decays

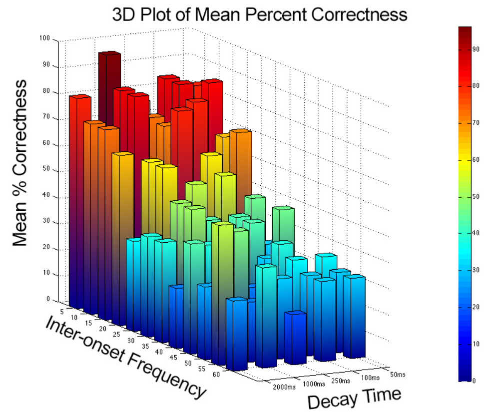

Figure 10. Map of mean percent correctness. Each decay time corresponds to the inter-onset frequency at which it was played. Each bar represents the average of all participants who submitted at the corresponding sample.

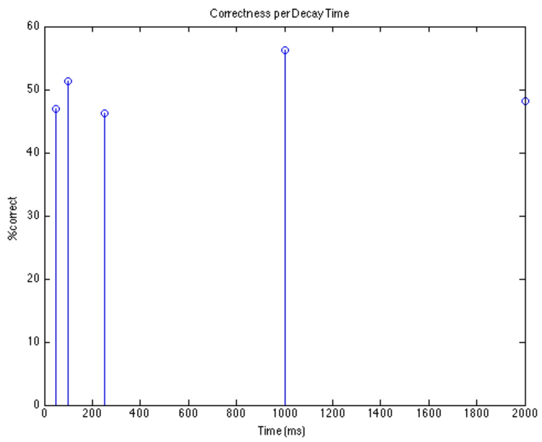

Figure 11. Percent correctness for 5 decay times

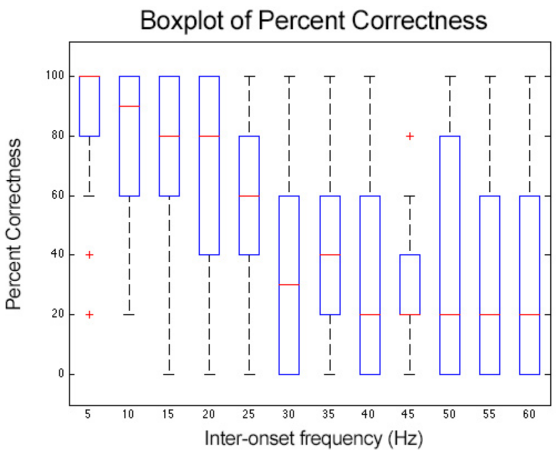

Figure 12. Boxplot displaying median percent correctness for inter-onset frequency.

Further analysis assuming a normal distribution for percent correctness of inter-onset frequencies supported this trend. A boxplot seen in Figure 12 using the standard coverage of 99.3% displays the median of each inter-onset frequencies, 25th and 75th percentiles, coverage and outliers. Outliers are indicated with the red + symbol. The whiskers display the range of values. At 5 Hz, the boxplot properly displayed outliers who were unable to perceive the pitch order at a very slow rate. At this rate it is nontrivial to perceive the pitch order. It properly indicates that the median value is 100% correctness, which aligns with the hypothesis for inter-onset frequency.

7. ANALYSIS

7.1 Anova Analysis

Analysis of variance, or Anova, is an effective statistical tool to analyze the effect of a particular variable on a given normal distribution. The parameter under scrutiny is categorized into different groups of variation. The procedure reveals to what extent the parameter affects the distribution and whether or not the different categorized groups have the same mean value. In our experiment, we performed anova analysis on two different parameters - The decay time and the inter-onset frequency.

7.2 Null Hypothesis

A generic way of beginning anova analysis is to establish a null hypothesis and verify if it can be statistically disproven. Our null hypothesis claims that all the groups belong to the same distribution, or in other words, possess the same mean value. The critical F value (statistical threshold) was determined from the computed sum of square ratios and degrees of freedom across the groups with a confidence value of 95%. (Alpha = 0.05).

7.3 Results for Decay Time

The results for decay time were not in inline with our expectation. From the results, we cannot disprove the null hypothesis since F < Fcrit. The calculated p value is extremely high and it suggests that the chance of a random sample resulting in a different mean value is very low. Typically, a p Value < 0.01 is required to confidently disprove a null hypothesis and establish concrete conclusion.

Groups	Count	Sum	Average	Variance
Decay = 50ms	26	1253.03	48.19347	259.5995
Decay = 100ms	26	1462.879	56.26457	207.2341
Decay = 250ms	26	1203.788	46.29953	159.7946
Decay = 1000ms	26	1330.303	51.1655	221.1309
Decay = 2000ms	26	1218.939	46.88228	263.33

Table 1. Variance table for decay time

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	1741.868	4	435.467	1.95964	0.104738	2.444174
Within Groups	27777.23	125	222.2178			
Total	29519.09	129				

Table 2. One-way Anova for decay time

7.4 Results for Inter-onset frequency

The results for inter-onset frequency as our analysis variable, however, strongly support our hypothesis. The F value in this case, is much greater than F crit by a fairly large margin. This effectively disproves our null hypothesis, thereby reinforcing the fact that the columnar data of percentage correctness do NOT belong to the same population. The p Value is of the order of -16 and is a very strong sign pointing towards the dependency of the distribution on variation in delta frequency.

Groups	Count	Sum	Average	Variance
Δf = 5Hz	26	2180	83.84615385	672.6153846
Δf = 10Hz	26	2080	80	608
Δf = 15Hz	26	1960	75.38461538	937.8461538
Δf = 20Hz	26	1740	66.92307692	1310.153846
Δf = 25Hz	26	1315	50.57692308	1372.653846
Δf = 30Hz	26	1060	40.76923077	1391.384615
Δf = 35Hz	26	1140	43.84615385	1024.615385
Δf = 40Hz	26	680	26.15384615	1016.615385
Δf = 45Hz	26	860	33.07692308	830.1538462
Δf = 50Hz	26	935	35.96153846	1392.038462
Δf = 55Hz	26	880	33.84615385	1144.615385
Δf = 60Hz	26	760	29.23076923	1191.384615

Table 3. Variance table for inter-onset frequency

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	125947.7	11	11449.7	10.65	1.49E-16	1.8206
Within Groups	322301.9	300	1074.3			
Total	448249.6	311				

Table 4. One-way Anova for inter-onset frequency

7.5 Two Way Anova

The analysis so far, takes only a single variable into account. Anova is also capable of determining how a distribution is affected by multiple factors. To study the effect of inter-onset frequency and decay time together on our distribution, the grouping of our data was refactored. Each row now represents the percentage correctness amongst our subjects for a particular value of inter-onset frequency and each column represents the percentage correctness for a given decay time.

Groups	Count	Sum	Average	Variance
Δf = 35Hz	5	219.2307692	43.84615385	19.23076923
Δf = 40Hz	5	130.7692308	26.15384615	76.92307692
Δf = 45Hz	5	165.3846154	33.07692308	85.79881657
Δf = 50Hz	5	173.2307692	34.64615385	66.66745562
Δf = 55Hz	5	169.2307692	33.84615385	136.0946746
Δf = 60Hz	5	146.1538462	29.23076923	48.81656805
Decay = 2000ms	12	578	48.16666667	411.5006276
Decay = 1000ms	12	674.9230769	56.24358974	495.7609826
Decay = 250ms	12	556.9230769	46.41025641	576.6856733
Decay = 100ms	12	616.7692308	51.3974359	458.1622736
Decay = 50ms	12	564.7692308	47.06410256	493.7513

Table 5. Row (Δf) and column (decay) variance table

Source of Variation	**SS**	**df**	**MS**	**F**	**P-value**	**F crit**
Rows – inter-onset frequency	2438 8.26	11	2217.1	40. 5	2.22E- 19	2.01 4
Columns - Decay Time	788.3 882	4	197.09	3.6 0	0.01259 9	2.58 3
Error	2406. 206	44	54.686			
Total	2758 2.86	59				

Table 6. Two-way Anova for *Δf* and decay interaction with percentage correctness

These results reinforce the values obtained from the single factor analysis. As expected the p Value for inter-onset frequency strongly suggests that it is positively influencing the results as per our hypothesis. Given that inter-onset frequency is changing with each row, the results for decay time (varying across each column) are slightly more encouraging and theoretically, the null hypothesis can be disproven (F > F crit). However, the p Value is hovering around 0.01 and hence, the statistical evidence is insufficient to draw out a solid conclusion on decay time affecting percentage correctness. The obtained value could very well be due to outliers in the data set, which Anova, by itself, is incapable of unearthing.

8. DISCUSSION

Important caveats of the experiment were the tonic frequency, decay compensation and the loudness curve. The experiment randomly selected a tonic frequency from A3, A5, or A7. While existing research shows that it is more difficult to distinguish pitches in higher registers [10], this research focuses on the relationship between inter-onset time and decay. The variation in tonic frequency is to ensure that the results are not specific to one range of the spectrum.

We compensated for decay length as shown in Figure 6 to prevent the ability to distinguish offsets. The patch lengthens the second note by 1/3 of its decay time and the initial note by 2/3 of its decay time. This may or may not guarantee an overlap of notes based on the inter-onset frequency. This surely altered the results, and in hindsight leaving each pitch's decay time equal would have been sufficient.

The notion of a threshold of distinguishability was apparent when beginning the experiment. While the mean percent correctness charts in Figures 8 and 9 do not allude to this notion, the Boxplot in Figure 12 comes closer. The boxplot is quantized to display the percentage of correct answers out of the 5 samples at each *Δf*, per user. When *Δf*, exceeds 35 Hz, the median value consistently stays at 20%. This value is close to the chance of guessing the correct permutation, which is a 1 in 6 chance.

Lastly, the disadvantages of Anova include its rigidity in respect to the data. Anova does not explicitly point out any outliers. It has to be inferred or assumed by inspection. Anova also may not be accurate if the distributions are not normal. Anova, however, is useful to statistically justify our hypothesis and design decisions at a broader level.

9. CONCLUSION

In future iterations of the experiment, a larger group of more varied participants would give a more accurate representation of a real population. Furthermore, a larger range of inter-onset frequencies and decay times may offer more insight into thresholds of cognition. The current duration limit of approximately 20 minutes also limited the number of listening samples for each participant. Compensation is a possible tactic for collecting a wider range of results for each participant and ensuring accuracy.

An analysis of variance showed that our hypothesis predicting an inverse correlation between inter-onset frequency and distinguishability of pure tone sequences was correct. For decay, the Anova results could not invalidate the null hypothesis. The statistical evidence does not conclude that decay time affects distinguishability. With further expansion of the experiment, a more distinct threshold of distinguishability may emerge.

10. REFERENCES

[1] A. Bauer : "Tone-Color, Movement, Changing Harmonic Planes: Cognition, Constraints and Conceptual Blends", *The Pleasure of Modernist Music: Listening, Meaning, Intention, Ideology*. Boydell & Brewer, 2004.

[2] R. J. de Cisneros. "Continuum, expanded." 2011 - 2013.

[3] D. Deutsch: *The Psychology of Music*, Academic Press, New York, 1982.

[4] D. Deutsch, Short Term Memory. Available at: http://deutsch.ucsd.edu/psychology/pages.php?i=209

[5] A. S. Bregman. *Auditory Scene Analysis*. MIT Press, Cambridge, MA, 1990.

[6] J. M. Doughty and W. R. Garner. "Pitch characteristics of short tones. II. Pitch as a function of tonal duration." *Journal Of Experimental Psychology* 38, no. 4, pp. 478-494. PsycINFO, EBSCOhost (accessed December 9, 2014).

[7] B. C. J. Moore: *An Introduction to the Psychology of Hearing*, Academic Press, London, 2003 p 255 - Perception of sequences of sounds

[8] H. Fletcher and W.A Munson: "Loudness, its definition, measurement and calculation", *Journal of the Acoustic Society of America* 5, 82-108 (1933).

[9] decaying-sinusoids~ | CNMAT. (n.d.). Retrieved December 7, 2014. Available at: http://cnmat.berkeley.edu/patch/4072

[10] B. C. J. Moore: “Frequency difference limens for short ‐ duration tones,” The Journal of the Acoustical Society of America, 54, 610-619 (1973), DOI:http://dx.doi.org/10.1121/1.1913640