UNIVERSITY OF HARTFORD ACOUSTICS

A Division of
The Engineering Applications Center
College of Engineering, Technology, & Architecture

Final Report: Just Noticeable Difference of Clarity Index (C80), Phase 2

Clothilde Giacomoni & Christopher Jasinski
Study 2a

Scott Edwards & Daniel Ignatiuk
Study 2b

Caitlin Ormsbee & Adam Wells
Study 2c

with
Michelle C. Vigeant, Ph.D.
Robert D. Celmer, Ph.D., P.E.

This report was prepared for the Paul S. Veneklasen Research Foundation, whose support made this student project possible.

Report # EAC-2011-26 June 30, 2011

LEGAL NOTICE

Overview of Work Completed Under PSVRF Grant 10.01

The primary goal of PSVRF Grant 10.01 was to carry out further investigations on the just noticeable difference (JND) of clarity index (C80) building upon the work in the first phase, Study 1. In order to achieve this goal, the proposal consisted of three separate studies:

- Study 2a: Investigate the effect of test method on C80 JND results
- Study 2b: Investigate auditory memory in terms of motif length for use in subjective tests
- Study 2c: Investigate C80 JND building on the results from Studies 1 and 2

Given the relatively extensive scope of these studies, the work was carried out by three teams of two acoustics students over two years. Studies 2a and 2b were carried out in Fall 2009, and Study 2c was carried out in Fall 2010 into Spring 2011. The undergraduate students who assisted with the work are: Clothilde Giacomoni ’09 and Chris Jasinski ’10 (Study 2a), Scott Edwards ‘09 and Daniel Ignatiuk ’09 (Study 2b), and Caitlin Orsmbee ’12 and Adam Wells ’11 (Study 2c). All students carried out the work for course credit in ES 493, Engineering Research. Jasinski was a junior when he worked on Study 2a and was able to provide assistance with the set-up and knowledge transfer as a senior to the students who worked on Study 2c. He will also be presenting a summary of the work at the ASA San Diego meeting in Fall 2011.

The results of these studies were shared with the research community as follows:

- <u>Study 2a</u>: Giacomoni, C.B., Jasinski, C.M., Celmer, R.D. and Vigeant, M.C. (2010). “Comparison of two just-noticeable-difference test methods for clarity index (C80) (A).” *J. Acoust. Soc. Am.*, Baltimore, MD, **127**: 2027
- <u>Study 2b</u>: Edwards, S.S., Ignatiuk, D.A., Celmer, R.D., and Vigeant, M.C. (2010). “The effect of motif length in reverberation-time listening tests using the ABX test method (A).” *J. Acoust. Soc. Am.*, Baltimore, MD, **127**: 2027
- <u>Summary of Studies 1 & 2a</u>: Vigeant, M.C. and Celmer, R.D. (2010.) “Effect of experimental design on the results of clarity-index just-noticeable-difference listening tests.” *Int. Cong. Acoust.* (Sydney, Australia), 7 pages
- <u>Study 2c</u>: Wells, A.P., Ormsbee, C.I., Celmer, R.D., and Vigeant, M.C. (2011). “Just-noticeable-difference of clarity index (C80) using real-time switching between test signals (A).” *J. Acoust. Soc. Am.*, Seattle, WA, **129**: 2534

The details of each study are presented in separate sections of this report as co-written by the students involved in each study. An overall conclusion of all three studies is included at the end of the report.

ABSTRACT

The goal of this project was to investigate the just noticeable difference (JND) for clarity index (C80) since only a handful of studies have been carried out on this topic, which have some limitations. In order to conduct a study with a well-grounded experimental design, preliminary work was completed. This work consisted of examining the effect of the testing method on the measured C80 JND (Study 2a) and determining a suitable length for a musical motif, or short musical clip, for use in these tests (Study 2b). The subjective listening tests for all three studies were carried out in the University of Hartford's anechoic chamber. The test signals were generated using a Yamaha Digital Mixing Engine (DME) digital signal processing unit and short motifs were convolved in real-time with the signals. Each signal was presented via eight spatially arranged loudspeakers in the chamber with the listener position centered among the speakers.

The purpose of the first preliminary study (Study 2a) was to determine how the C80 JND varies as a function of the test procedure. The work carried out under the previous grant in Spring 2009 employed the test method of requiring the subjects to listen to all of signal A and then all of signal B before deciding if the signals were the same or different in terms of clarity (Method 1). In Study 2a, Method 1 was compared to Method 2, which consisted of allowing subjects to compare the signals by switching between them in real-time before giving their response. The results of the Spring 2009 study (Study 1, PSVRF Grant 08.06) were somewhat inconclusive as it was determined that the C80 differences between the test signals were too small, making the test extremely difficult. Rather than using differences ranging between 0.0 to 3.0 dB, in Study 2a the differences were changed to 0.0, 3.0, 5.0 and 7.0 dB. Another variable evaluated in this study was to determine if there was a significant difference between test participants with musical training only and subjects with critical listening training as recording engineering students who also have musical training. A total of 11 subjects participated in this pilot study, with seven having critical listening experience. No significant difference was found between the two participant types. However, a significant effect of test method order was found. In particular, the subjects who started with Method 1 for the first half of the test and then used Method 2 for the remainder of the test gave the most accurate results, in terms of the lowest percentage reporting a difference when no difference in signals was presented and the largest percentage reporting a difference at the maximum difference of 9 dB. The results of this pilot study suggested an increased C80 JND of 3.8 dB, which is much higher than the preliminary result found in the Spring 2009 study of 1.6 dB and 1.0 dB from the other studies reported in the literature.

The purpose of the second preliminary study (Study 2b) was to investigate motif length in terms of auditory memory for psychoacoustics tests. No previous studies in the literature were found that provided any guidelines in terms of an appropriate motif length for use in these types of tests. In particular, how auditory memory affects the accuracy of subjects' responses. Reverberation time rather than clarity index was varied between signals, since the concept of reverberance is easier to explain to test subjects than is clarity. Research in the field of auditory memory suggests there is both short auditory memory of approximately 200 ms and long auditory memory lasting several seconds (Cowan 1984 Psych. Bul. 96:341-370); thus signals used in architectural acoustics psychoacoustics test fall into the latter category. A study was carried out using two motifs, an orchestral motif and a percussion motif, with three different lengths: 5s, 7s and 10s, where the successively longer motifs contained the same passages as the shorter ones. The signals were created with reverberation times ranging between 1.0 to 1.5 s. The subjects were presented three signals, A, B and X, where the difference in reverberation time ($\Delta T30$) ranged between 0.3-0.5s. The 25 subjects were required to identify which signal A or B matched the test signal X. The results of the test did not provide a clear conclusion since interaction effects were found between motif length and $\Delta T30$, and motif type and $\Delta T30$. However, for the smallest difference of 0.3 s, subjects did provide significantly more correct responses for the shorter motifs of 5 and 7 s. For the tests using the percussion motif and the largest difference of 0.5 s, the longer motif lengths of 7 and 10 s resulted in significantly more correct responses.

Using the results of Studies 2a and 2b, a comprehensive study was carried out to again investigate the C80 JND, Study 2c. The test procedure for this experiment consisted of an extended training period using Method 1 and then the subjects completed the actual test using Method 2. A maximum motif length of 10 s was set based on Study 2b and also previous work carried out by the PI. As with the Spring and Fall 2009 C80 JND studies (Studies 1 and 2a), two C80 base cases were used: For Base Case 1, C80 at 1 kHz was set to -3 dB with T30 of 1.9 s and Base Case 2 had a C80 of +1 dB with T30 of 1.5 s. As with all previous studies, highly trained musicians with a minimum of 5 years of musical experience and maximum hearing thresholds of 15 dB HL were used as test subjects. A total of 28 subjects participated in the study. The number of differences between signals was increased to a total of six, with the maximum difference extended to 9.0 dB. The results of the study yielded a C80 JND of 4.0 dB, similar to the results of Study 2a.

Master Table of Contents

Study 2a: Effect of Test Method on

Clarity Index (C80)

Just Noticeable Difference (JND)

by

Clothilde Giacomoni

Christopher Jasinski

with

Michelle Vigeant, Ph.D.

Robert D. Celmer, Ph.D., P.E.

Director, Acoustics Laboratory

University of Hartford

Study 2a Table of Contents

STUDY 2A LIST OF FIGURES

STUDY 2A LIST OF TABLES

1.0 INTRODUCTION

In the field of architectural acoustics, many measures have been developed and tested to quantify the perception of the room's acoustics. The most common parameter is reverberation time (T30), the time it takes for sound to decay 60 dB. A parameter that is in general inversely related to T30 is clarity index (C80), the ratio of the early energy in the first 80 ms to the late energy after 80 ms, where the latter is assumed to be the reverberant energy (Eqn. 1). The smallest detectable change or just noticeable difference (JND) is defined when the difference is detected 50% of the time. The C80 JND is of interest since it will help researchers and designers know they've achieved a significant change in C80 due to making design changes. The purpose of this investigation is to determine the effect of two different test methods on the resulting C80 JND.

$$C80 = 10\log_{10}\left[\frac{\int_{0}^{80ms} p^2(t)dt}{\int_{80ms}^{\infty} p^2(t)dt}\right] \quad \text{[dB]} \qquad (1)$$

2.0 BACKGROUND INFORMATION

The earliest results reported from a subjective study to determine the JND for clarity index was published in 1993 by Cox, Davies, and Lam (1). The signal generation equipment available at the time was quite limited, which resulted in a simulated impulse response with very few individual early reflections. Due to the low density of reflections, the calculation of C80 of each signal wasn't very reliable since it could vary significantly depending on the arrival time of a distinct reflection shortly before or after the 80 ms cut-off. To avoid this potential problem, center time (T_c) was measured instead and then C80 was calculated using this result. Two motifs were selected from Handel's Water Music and Mendelssohn's Symphony no. 3 in A minor and were convolved with the simulated sound field. The combined signals were played back using eight loudspeakers in an anechoic chamber. Seven to ten subjects who were musically trained or regularly attended concerts were used. While changes in C80 were being observed, reverberation time (T30) was kept constant at 2 s, early decay time (EDT) at 1.8 s, early lateral energy fraction (LF) at 0.27, and the overall level at 79 dBA. The test subjects confirmed that they only perceived changes in clarity, while the rest of the parameters did not sound like they changed. There was a large disparity between the C80 JND for the two motifs; the Handel motif was found to have a JND of 0.44 dB, while the Mendelssohn motif was found to have a JND of 0.92 dB. The final reported C80 JND was 0.67 dB, an average of these two results.

Bradley, Reich and Norcross (2) conducted a subjective study to determine the just noticeable difference (JND) of C50, the clarity index for speech where the integration limit is changed from 80 ms to 50 ms. A similar signal generation and playback system to the previous study was used, however with upgraded equipment that allowed for a significant increase in individual early reflections. The signals were convolved with recorded speech. A common phrase and 50 different test words were used in the different sound fields. Three C50 base cases (-3 dB, 1 dB, 5 dB) were used to represent low, mid and high levels of clarity. Each base case was compared to C50 differences of 0.0, 0.5, 1.0, 1.5, 2.5, 4.0 dB, and 0.0 dB again at the highest value in that base case.

For each question, two sound fields with different clarity indices were presented to individual test subjects and their task was to determine if the two signals sounded the same or different. The experimental set-up allowed for subjects to switch in real-time between the signals within each pair presented. Ten subjects were used, which represented a broad range of listeners, based on their average age, hearing and listening experience. Each subject had a short practice session, where they were told that the highest and lowest differences were included but the specific signals were not identified. As the difference in C50 between signals increased, the percentage of subjects who perceived the difference increased. The three sets of results, based on the three base cases, were averaged and an estimate of the C50 JND was found to be 1.1 dB, based on the fact that there was no statistically significant difference among the three C50 cases. Due to the previously established linear relationship between C50 and C80, the C80 JND was extrapolated from these results and was found to be 0.9 dB.

A C80 JND study was carried out at the University of Hartford in Spring 2009, funded by the previous PSVRF Grant 08.06 (3). The goal of that study was to carry out a C80 JND study using modern-day equipment and employing a much larger sample size, 51 test subjects in total, than had been used in either of the previous studies. The experimental design paralleled that of Bradley *et al*'s design [2] and the differences between signals were set to close to 1.0 dB based on the previously reported results. Two base cases, rather than three, were used in the study. For Base Case 1, C80 at 1 kHz was set to -3 dB with T30 of 2.1 s and Base Case 2 had a C80 of +1 dB with T30 of 1.6 s. Each base case was presented along with a signal that had a difference of 0.0, 0.5, 0.8, 1.0, 1.2, 1.5, or 3.0 dB. The simulated sound fields were generated using a Yamaha Digital Mixing Engine, Type DME64n, a digital signal processing unit, and convolved with short musical anechoic motifs, Bizet's *L'Arlesienne Suite No. 2*, Weber's *Theme* for solo cello, and Handel's *Water Music*. The combined signals were presented to subjects using an array of eight Genelec loudspeakers, Type 8030A, in University of Hartford's anechoic chamber. Subjects were presented two signals and were asked if they sounded the same or different in terms of clarity, which was defined as how clear each individual note sounds, and also how clear the note sounds relative to the subsequent note. The experimental set-up required subjects to listen each signal in its entirety, one after the other, before answering.

The results were not as expected, in that about 40 to 50 % of the subjects reported hearing a difference when there wasn't one and only about 60% reported hearing a difference for the pair of signals with the largest difference of 3.0 dB. These results lead the investigators to conclude that the test was very difficult and that subjects were likely guessing for many of their responses. The data set was filtered to only include the results from the subjects who 65% of the time reported not hearing/hearing a difference for the extreme cases of 0.0 and 3.0 dB, respectively. This data filtration reduced the data set to 17 test subjects and yielded an overall C80 JND of 1.6 dB, averaged over both base cases and the three motifs. Given the difficulty of the test and also the likelihood that the C80 JND is above 1.0 dB, further work needed to be conducted.

3.0 EXPERIMENTAL PROCEDURE

3.1 Overview of Experimental Method

The primary goal of this study was to evaluate the effect of test method on the C80 JND. The secondary goal of this study was to evaluate whether sound recording students (majoring in Music Production & Technology) who are learning the skill of critical listening are more adept at detecting differences than students with musical training only. In order to provide test subjects with signals with more readily detectable differences in C80 than the previous study, the range of differences was increased from 3.0 to 7.0 dB. The same two base cases were used, where Base Case 1 had a C80 value of -3.0 dB at 1 kHz and T30 of 2.1 s and Base Case 2 had a C80 value of +1.0 dB at 1 kHz and T30 of 1.6 s. However, the number of differences was reduced from seven to four, 0.0, 3.0, 5.0 and 7.0 dB, and the number of motifs was reduced from three to two.

Subjects were presented two signals at a time, Signal A and Signal B, using one of two test methods. Test Method 1, which was the same one as used in the previous study (3), required subjects to listen to all of Signal A and then all of Signal B before answering if the signals were the same or different. Test Method 2, based on Bradley *et al*'s work (2), allows subjects to freely switch between Signals A and B in real-time with the use of a custom interface box, as shown in Figure 1. Half of the subjects completed the test first using Test Method 1 (*no switch)*, then Test Method 2 (*with switch*), while the other half were given the opposite order.

Figure 1: Custom subject-interface control that allows subjects to toggle between Signals A and B in real-time.

3.2 Signal Generation

The target C80 values for each base case were set based on preferred C80 and T30 values for large concert halls and small chamber music halls (4), where Base Case 1 is for the former with T30 of 2.1 s and C80 of -3.0 dB at 1 kHz, and Base Case 2 is for the latter with T30 of 1.6 s and C80 of +1.0 dB (4). The spectral shape of the C80 signals was determined in the previous study (3) by averaging an extensive set of concert hall measurement data from (4). The target values for each base case are shown in Table 1.

Table 1: Target C80 values for each base case

Octave-Band Center Frequency (Hz) :	**125**	**250**	**500**	**1000**	**2000**	**4000**	**8000**
Base Case 1 C80 values (dB) ***(approximates large concert hall)***	-6.0	-5.0	-4.0	**-3.0**	-2.5	-2.0	-1.5
Base Case 2 C80 values (dB) ***(approximates small chamber music hall)***	-2.0	-1.0	0.0	**+1.0**	+1.5	+2.0	+2.5

Signals were generated using a Yamaha Digital Mixing Engine, Type DME64n, and the accompanying software program Yamaha Designer, and played through eight spatially arranged loudspeakers in the anechoic chamber (see Figure 2 and Figure 3).

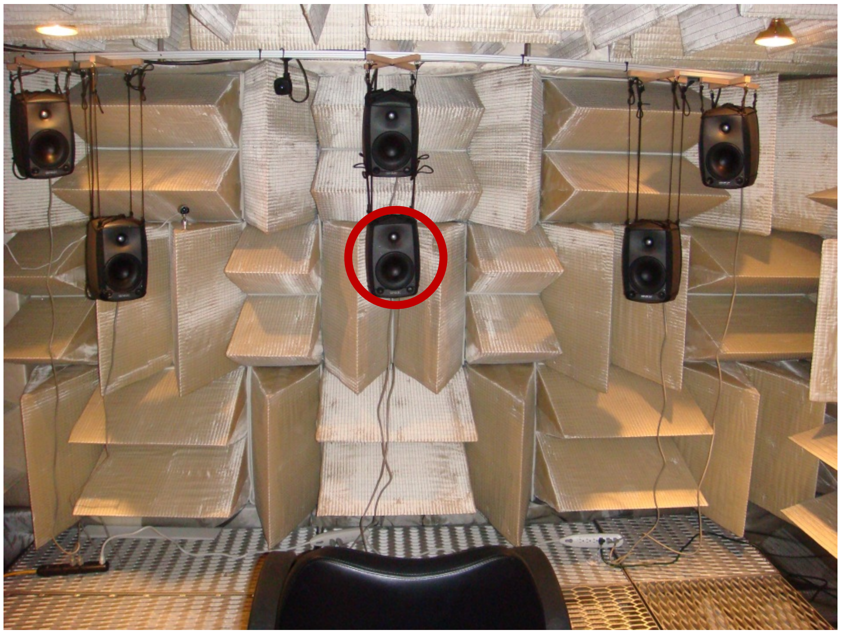

Figure 2: Six front loudspeakers as viewed from behind the listener's position.

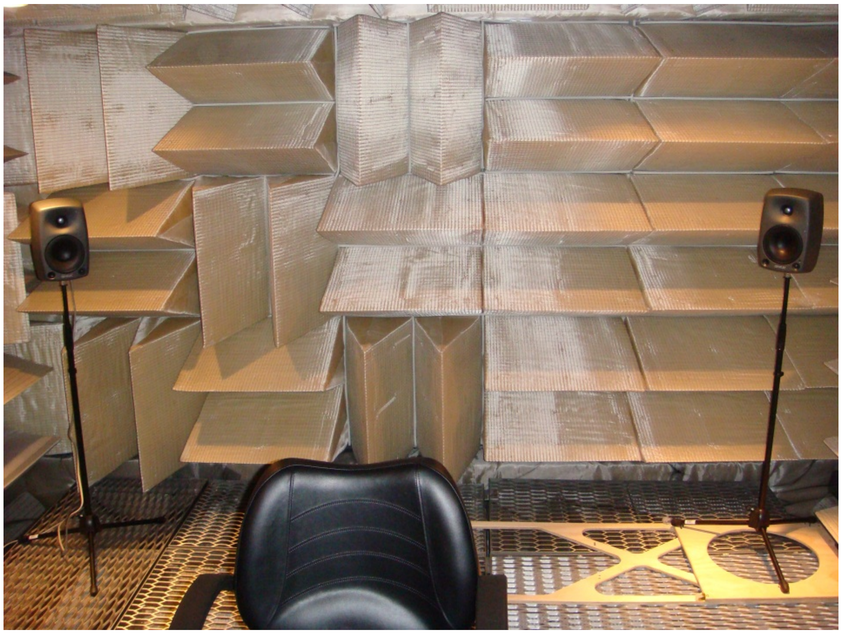

Figure 3: Two rear loudspeakers as viewed from in front of the listener's position.

A typical room impulse response (RIR) consists of the direct sound, a series of early reflections that are separated in time and then a series of late reflections overlapping in time. In order to generate a realistic RIR, the simulated sound field consisted of three components: 1) the direct sound, which was supplied to the central loudspeaker, and 2) a series distinct early reflections and 3) late reflections in the form of artificial reverberation were supplied to all eight loudspeakers (see Figure 4).

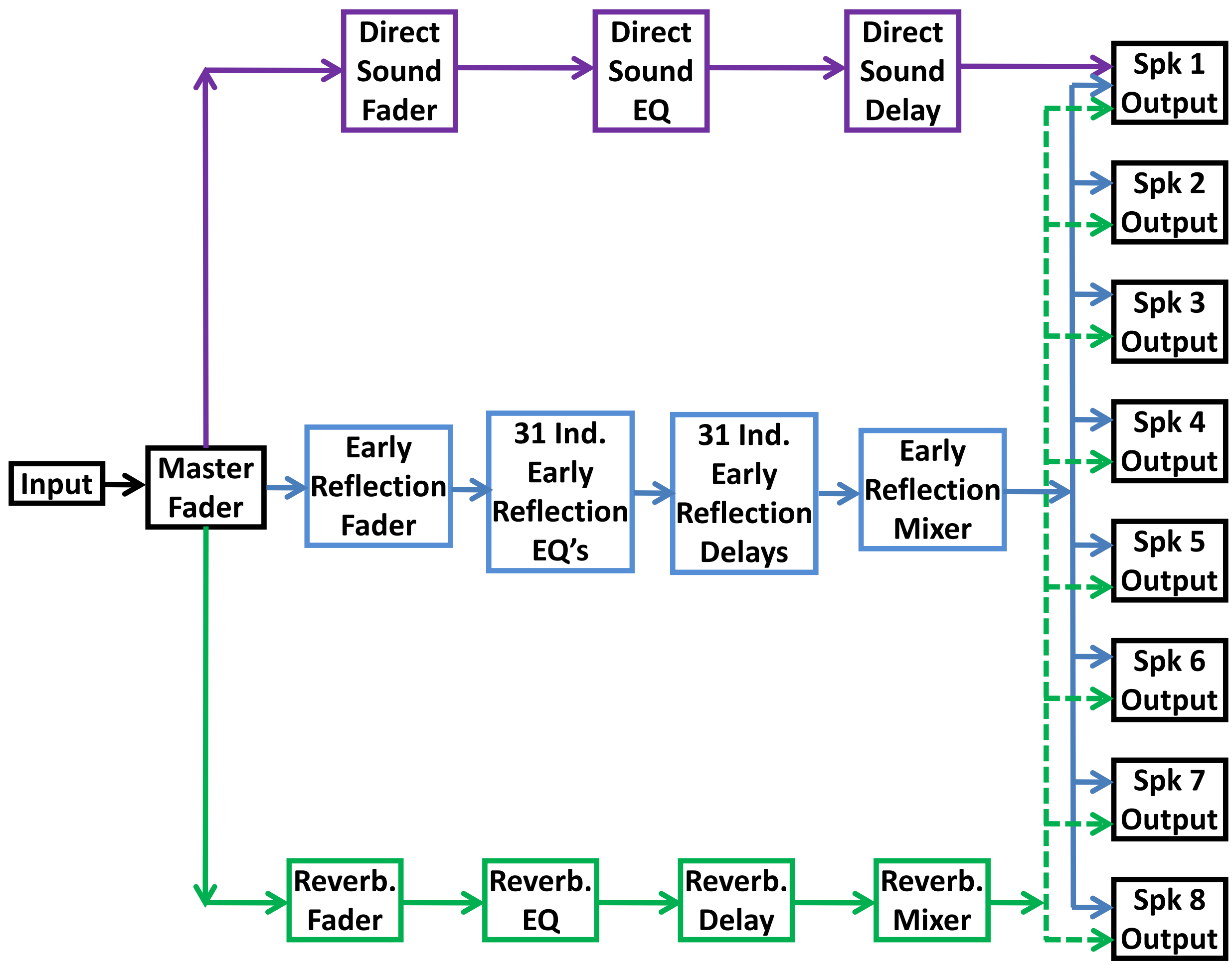

Figure 4: Overview of room impulse response signal generation for simulated sound fields.

The signals for the two base cases were initially created using the experimental set-up by adjusting the early reflection levels and pre-set reverberation time. The early reflections had to be adjusted for each octave band using a built-in equalizer, i.e. by increasing the early energy level C80 increases and vice versa. Once a signal had been programed, the resulting RIR was measured using the sine-sweep method with software WinMLS 2004. The signal generation process is iterative since it requires fine tuning of the early reflections equalizer in order to achieve the precise target C80 values in each octave band.

Once the base case signals were finalized, the signals with differences of +3, +5 and +7 dB were created using the same iterative process. The measured C80 values of the generated signals are shown in Figure 5 for Base Case 1 and in Figure 6 for Base Case 2. The signals were convolved with short anechoic recordings and presented in a series of pairs, Signal A and Signal B. Signal A was one of the base case signals and Signal B was one of the following: the same signal again or one of the three other signals with differences of +3, +5, and +7 dB.

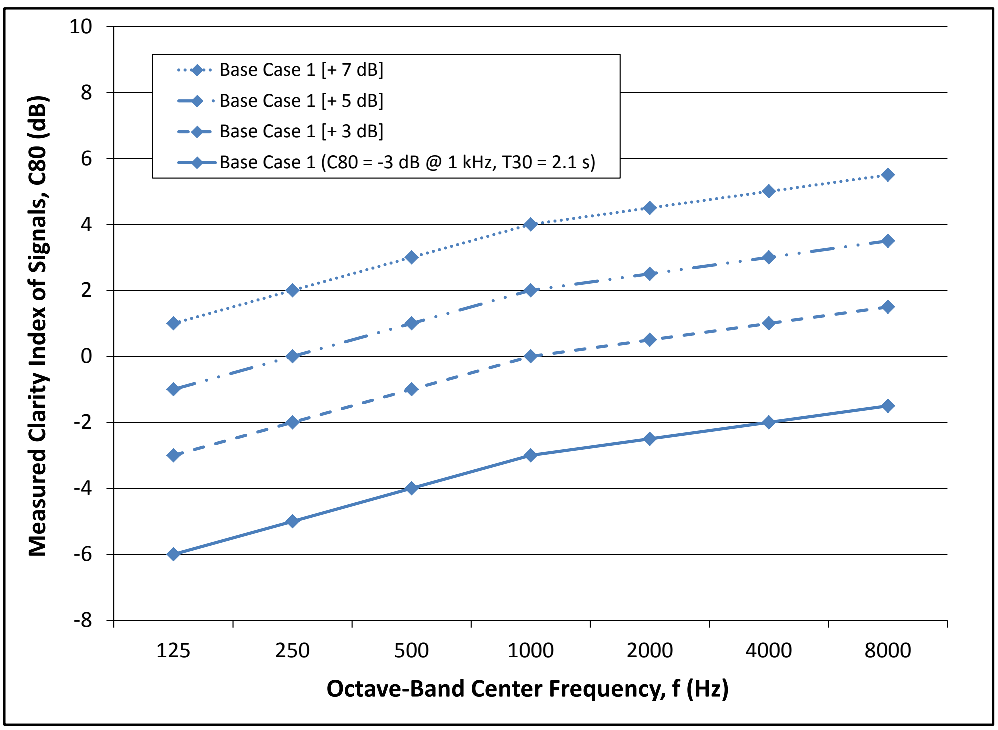

Figure 5: Measured clarity index values (C80) for 125 – 8000 Hz center octave bands for Base Case 1.

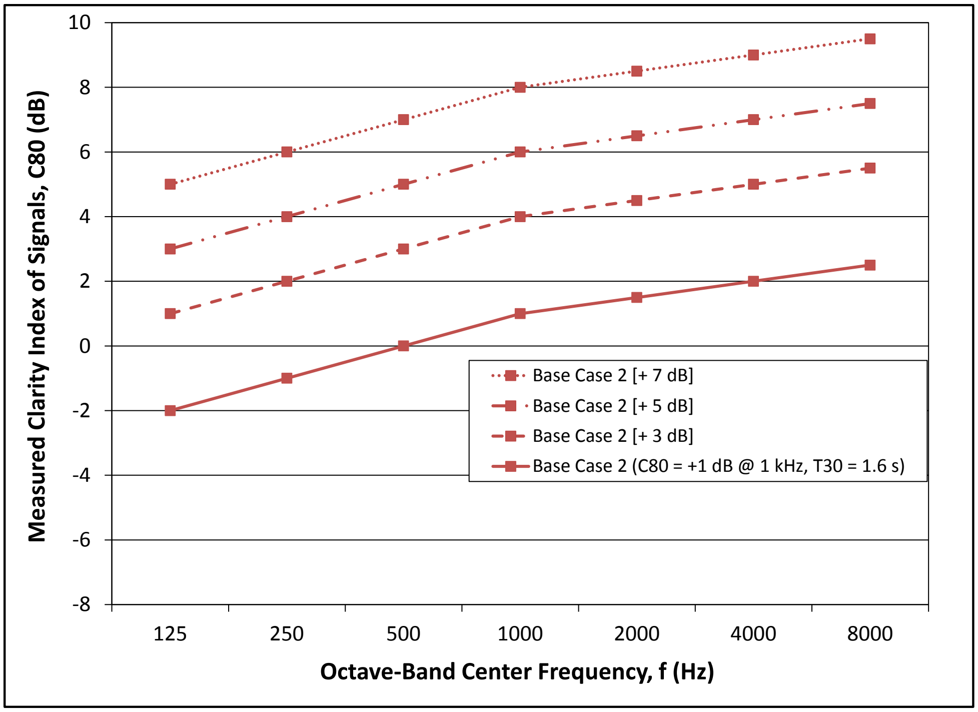

Figure 6: Measured clarity index values (C80) for 125 – 8000 Hz center octave bands for Base Case 2.

The motifs were selected from the limited available high-quality orchestral anechoic recordings with the criteria of having either a relatively slow or relatively fast tempo and to be relatively short. Four different musical motifs were selected from DENON *Anechoic Orchestral Music Recordings* CD (5) and all motifs were approximately 10 s in length. Motifs from Strauss's *Pizzicate-Polka* and the overture to Mozart's *Le Nozze di Figaro* were selected for the training sets to provide subjects with some variety to reduce fatigue and boredom. Two motifs were selected for the actual test sets. Motif 1 was the same as used in the previous study (3), Bizet's *L'Arlésienne Suite No. 2* and Motif 2 was Debussy's *Prelude to the Afternoon of a Faun*. Motif 2 was changed from the previous study to provide more contrast to Motif 1, as it has a flowing string passage as compared to the Bizet, which has a nicely articulated woodwind passage.

3.3 Summary of Test Participants

Eleven subjects were used who had two different training backgrounds in terms of musical and critical listening experience. Seven subjects were from the Music Production and Technology (MPT) program at the University of Hartford, which included both students and faculty, and the remaining four were music performance majors at the Hartt School of Music. The subjects' ages ranged from 19 to 36 with the average age being 24. All involved have taken and/or taught courses in critical listening at the University of Hartford, and all have been involved in recording or mixing live sound. The MPT subjects averaged 11.5 hours of critical listening per week and 12.5 years of formal instruction on their instrument. The performance major subjects had no weekly critical listening instruction or experience and averaged 10 years of formal instruction on their instrument. Of the 11 subjects, four were women and seven were men. All subjects attend live classical music concerts at least once a month.

Test participants were required to be at least 18 years old, to have a threshold of hearing at or below 15 dB HL across the frequency spectrum (from 250 to 8000 Hz) and to have at least five years of formal musical training.

3.4 Testing Procedure

Each subject was presented with a total of four sets of questions. Sets 1 and 3 were training sets with six questions each, while Sets 2 and 4 were the actual test sets with 24 questions each (see Table 2). The first eight signal pairs presented in Sets 2 and 4 were actually additional practice questions unbeknownst to the subjects. Half of the subjects were given Testing Order 1, Test Method 1 "no switch" first and then Test Method 2 "with switch", and the remaining half were given Testing Order 2, as shown in Table 2. After having completed the first two sets, the subjects were required to take a 5 minute break before carrying onto the remaining two sets. The subjects were instructed to determine if the two signals sounded the same in terms of clarity, which was defined as how clear each individual note sounds and also how clear each note sounds relative to the subsequent note.

Table 2: Overview of Testing Procedure

<table>
<tr><th></th><th colspan="2">First Half of Test</th><th colspan="2">Second Half of Test</th></tr>
<tr><td rowspan="3">Testing Order 1</td><td colspan="2">Test Method 1, no switch</td><td colspan="2">Test Method 2, with switch</td></tr>
<tr><td>Set 1</td><td>Set 2</td><td>Set 3</td><td>Set 4</td></tr>
<tr><td>Training (6)</td><td>Practice (8) & Test (16)</td><td>Training (6)</td><td>Practice (8) & Test (16)</td></tr>
<tr><td rowspan="3">Testing Order 2</td><td colspan="2">Test Method 2, with switch</td><td colspan="2">Test Method 1, no switch</td></tr>
<tr><td>Set 1</td><td>Set 2</td><td>Set 3</td><td>Set 4</td></tr>
<tr><td>Training (6)</td><td>Practice (8) & Test (16)</td><td>Training (6)</td><td>Practice (8) & Test (16)</td></tr>
</table>

The signals included in the training sets are shown in Table 3 and the signals for the actual test sets are shown in Table 4. In the practice set, all subjects heard the combination of specific signals and motifs in the same order. In the actual test sets, the presentation order of the signal pairs was randomized to eliminate the possibility of the effect of the question order on the test results.

Table 3: Signals for Training Sets 1 and 3.

Signal Pair	Motif	Signal A C80 value (dB) at 1kHz	Signal B C80 value (dB) at 1 kHz	C80 Difference between Signal A and Signal B (dB)
1	Strauss	-3.0	4.0	+ 7.0
2	Strauss	1.0	6.0	+ 5.0
3	Strauss	-3.0	-3.0	0.0 (same signal)
4	Mozart	1.0	8.0	+ 7.0
5	Mozart	-3.0	2.0	+ 5.0
6	Mozart	1.0	1.0	0.0 (same signal)

Table 4: Signals for Test Sets 2 and 4 (first eight signal pairs were for additional practice).

Test	Motif	Signal A C80 value (dB) at 1kHz	Signal B C80 value (dB) at 1 kHz	C80 Difference between Signal A and Signal B (dB)
1	Debussy	-3.0	2.0	+5.0
2	Debussy	1.0	6.0	+5.0
3	Debussy	-3.0	4.0	+7.0
4	Debussy	1.0	8.0	+7.0
5	Bizet	-3.0	2.0	+5.0
6	Bizet	1.0	6.0	+5.0
7	Bizet	-3.0	4.0	+7.0
8	Bizet	1.0	8.0	+7.0
9	Debussy	-3.0	-3.0	0.0 (same signal)
10	Debussy	1.0	1.0	0.0 (same signal)
11	Debussy	-3.0	0.0	+3.0
12	Debussy	1.0	4.0	+3.0
13	Debussy	-3.0	2.0	+5.0
14	Debussy	1.0	6.0	+5.0
15	Debussy	-3.0	4.0	+7.0
16	Debussy	1.0	8.0	+7.0
17	Bizet	-3.0	-3.0	0.0 (same signal)
18	Bizet	1.0	1.0	0.0 (same signal)
19	Bizet	-3.0	0.0	+3.0
20	Bizet	1.0	4.0	+3.0
21	Bizet	-3.0	2.0	+5.0
22	Bizet	1.0	6.0	+5.0
23	Bizet	-3.0	4.0	+7.0
24	Bizet	1.0	8.0	+7.0

4.0 RESULTS AND DISCUSSION

4.1 Overall C80 JND Results

The data were averaged over both base cases and both motifs to provide an overall C80 JND result, as shown in Figure 7. The results are plotted as the percentage of subjects who reported hearing a difference versus the difference in the presented signal pairs. When there was no difference, approximately 20% of the subjects reported they heard a difference while above 70% reported a difference when the difference was 7 dB. These results at the extreme cases are in line with the hypothesis that the percentage should be low when there's no difference and relatively high at the maximum difference. However, it was expected that the percentage who reported hearing a difference at the maximum difference of 7 dB would be closer to 90%. A linear regression was applied to this data and by taking the intersection of this curve at the 50% difference, these results give a C80 JND of 3.8 dB. Despite the high correlation coefficient of the trend-line (about 98%), the significance level of the correlation is $p < 0.06$. The high p-value is likely due to the small sample size used in this study.

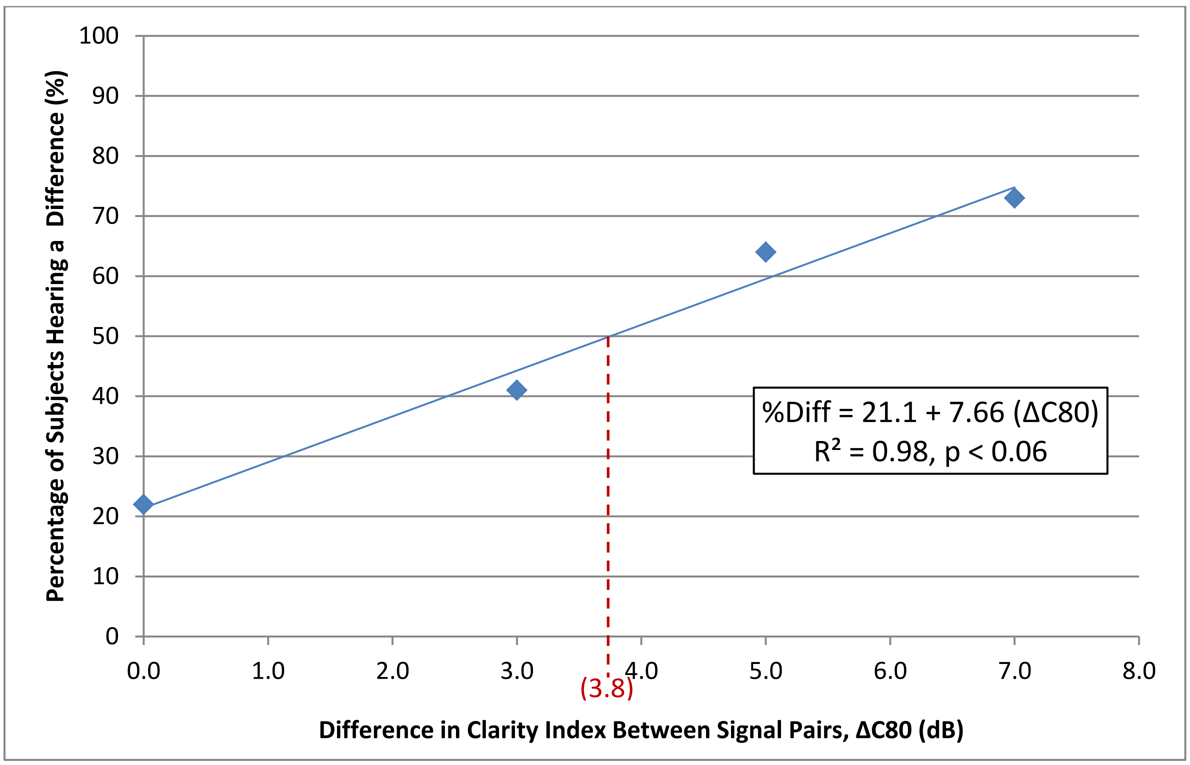

Figure 7: Overall C80 JND results.

4.2 Effect of Base Case on C80 JND Results

Figure 8 shows the results for each base case (Base Case 1 with C80 @ 1kHz = -3 dB and Base Case 2 with C80 @ 1kHz = +1 dB). The results are similar for both base cases at all differences except for the maximum difference of 7 dB. For this difference, 10% more subjects reported hearing a difference with Base Case 2, which had the higher C80, than with Base Case 1. This result is in agreement with the subject feedback obtained post-test in which subjects remarked it was easier to hear differences for the base case with less reverberance/higher clarity. The y-intercepts are similar for both regression lines around 20%, but due to the differences for the upper two points, there is a distinct difference in the slopes, which results in a C80 JND of 4.2 dB for Base Case 1 and C80 JND of 3.4 for Base Case 2.

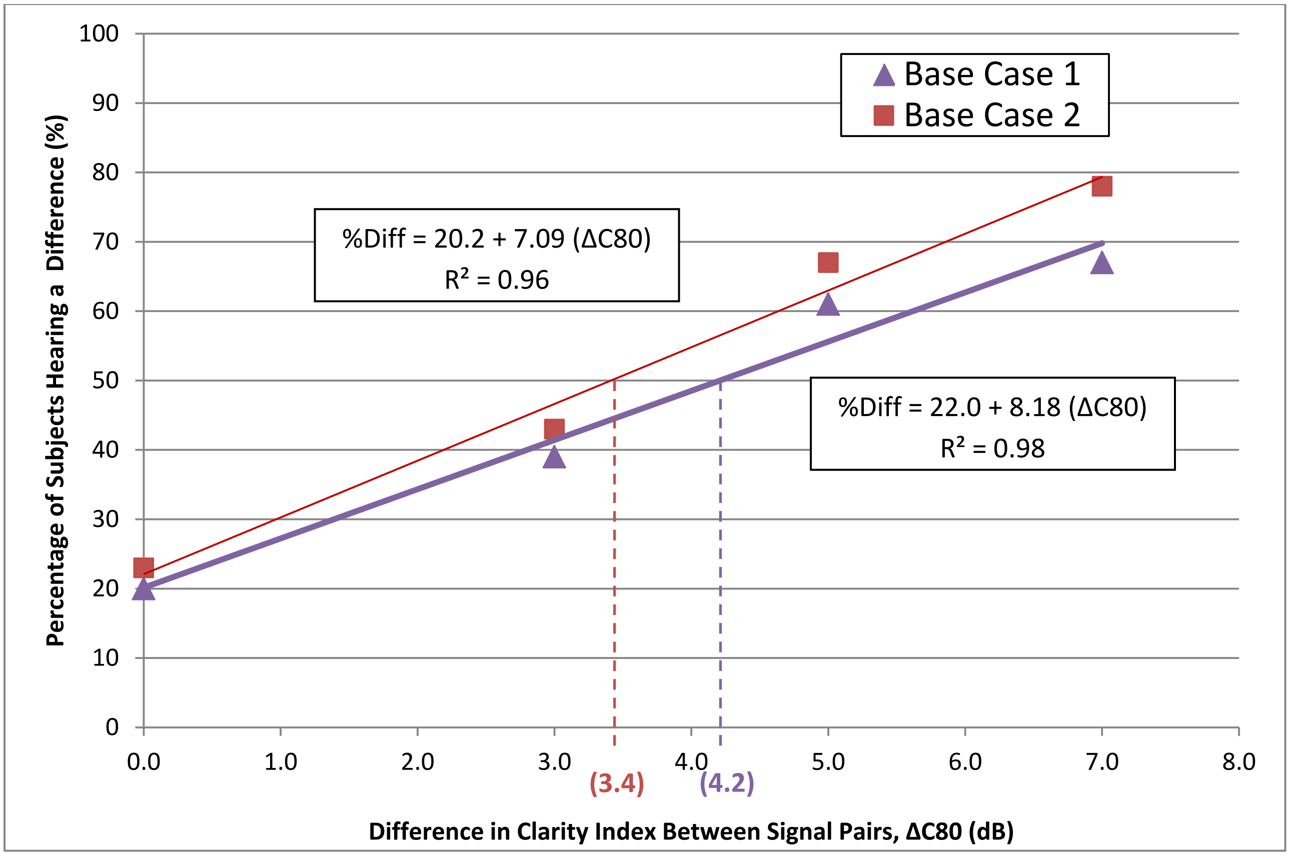

Figure 8: C80 JND results separated by base case.

4.3 Effect of Motif on C80 JND Results

Two motifs were used in this experiment; the Menuet of Bizet's l'Arlesienne and the first 15 seconds of Debussy's Prelude to the Afternoon of a Faun. The first motif (Bizet) was more staccato and had more room between notes whereas Motif 2 (Debussy) was more fluid and legato. The graph below shows that with Motif 1, less than 20% of the subjects reported they heard a difference when there was none and about 80% reported a difference when there was a 7 dB difference, which are in line with the general hypothesized results for the trend lines. The resulting C80 JND is 3.5 dB. For motif 2 however, about 25% reported a difference with a difference of 0 dB and only about 65% reported they heard a difference when there was a higher difference, resulting in a higher C80 JND of 4.3 dB. This results is to be expected as the two motifs had very difference musical content.

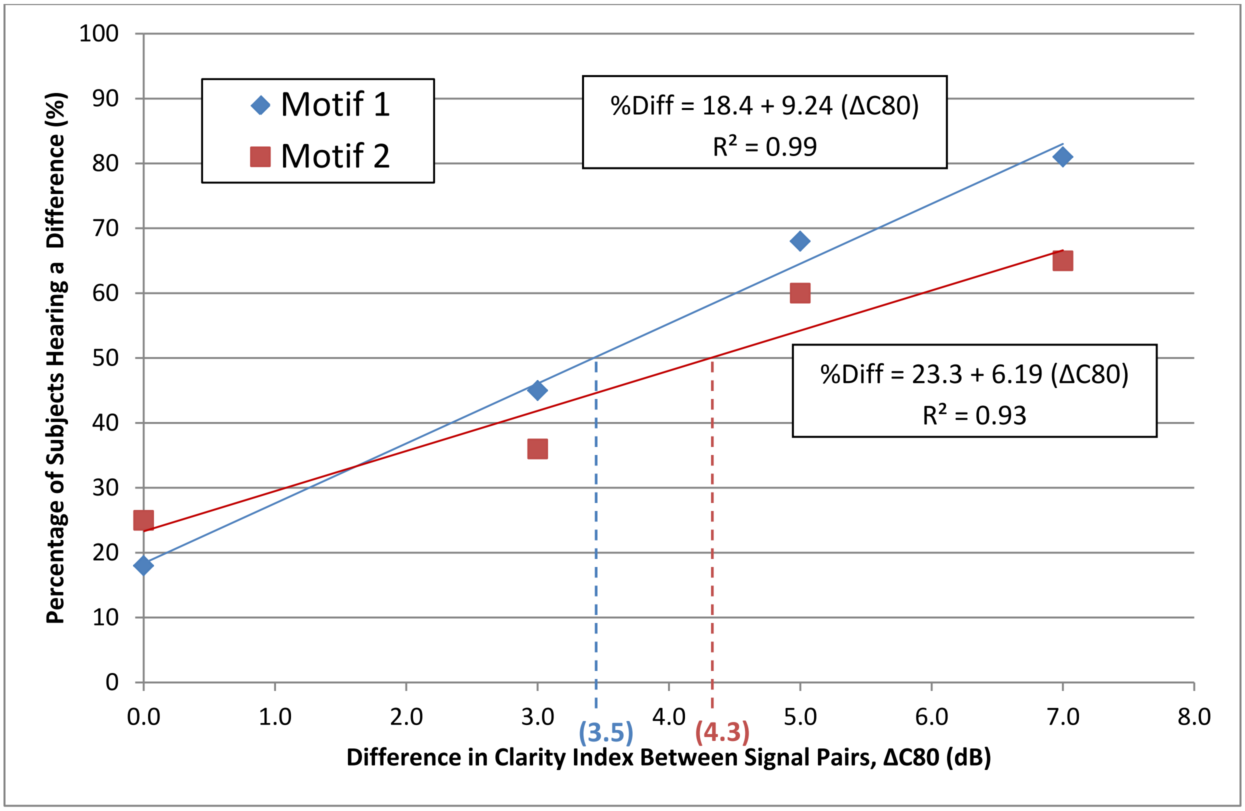

Figure 9: C80 JND results separated by motif.

4.4 Interaction Effect of Base Case and Motif on C80 JND Results

Figure 10 shows the effect that the base case and motif have on the subjects' responses. Motif 1 (blue and green lines) yielded higher positive responses at the higher end (around 80%) and relatively lower positive responses at the low end (around 10% for base case 1 and 30% for base case 2). As is expected, the worst results were obtained from the second motif (Debussy) and the first base case (-3 dB); with over 30% giving positive responses when the signals were the same and below 60% recognizing a difference when there was a high one. The two trendlines that are most consistent with each other and the previous results are Motif 1 & Base Case 1 and Motif 2 & Base Case 2 and yield a C80 JND of approximately 4 dB, as compared to the other two lines which result in C80 JNDs of 2.8 and 4.7 dB.

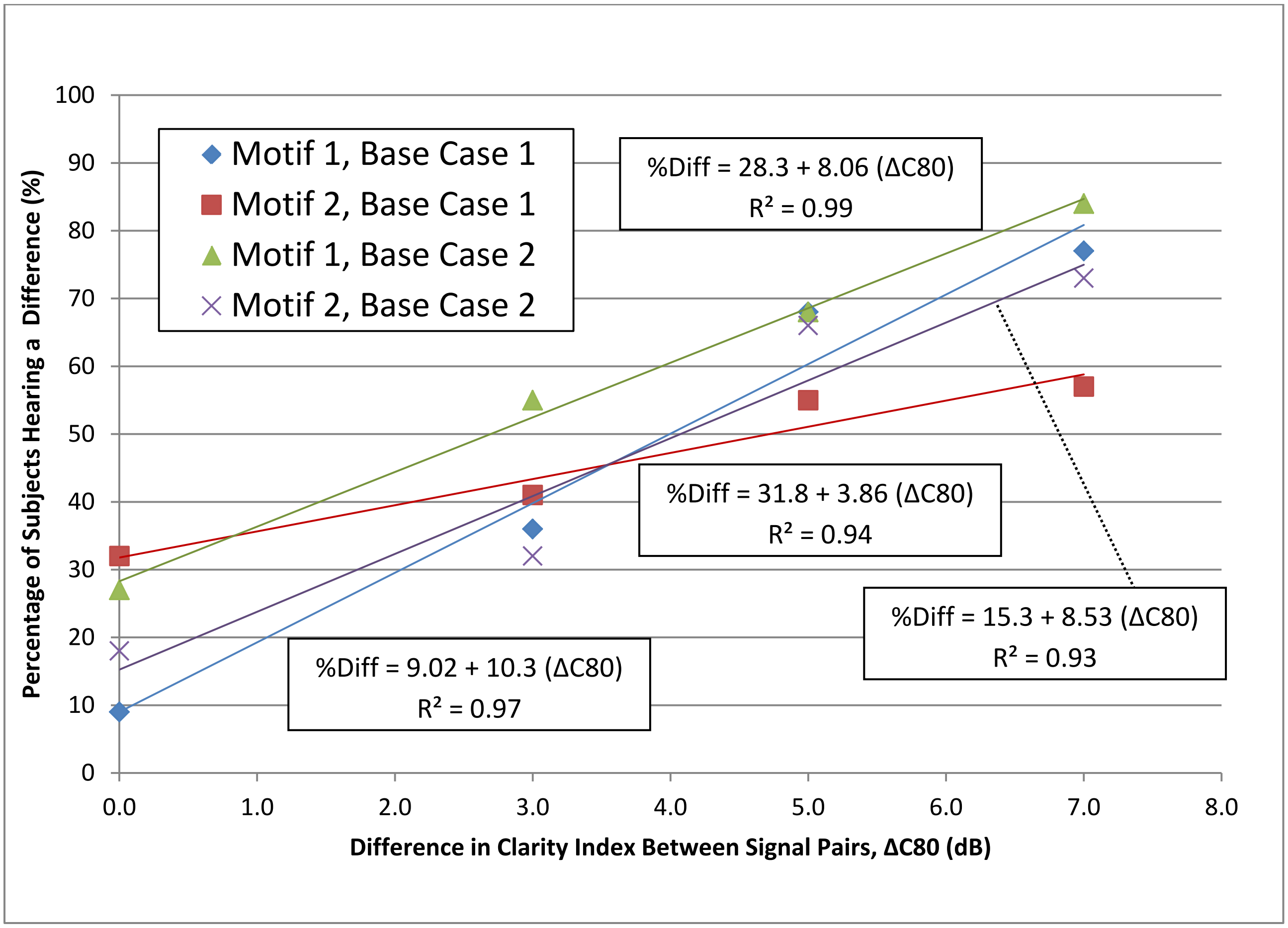

Figure 10: C80 JND results separated by the four combinations of motif & base case.

4.5 Effect of Subject Type on C80 JND Results

Two subject types were tested in this experiment; those majoring in or teaching Music Production and Technology (Subject Type 1), and those majoring in Music Performance (Subject Type 2) both at the University of Hartford. Students and teachers in Music Production and Technology have a core curriculum of music study (private instruction and ensemble performance), and are trained in music production, recording engineering, and have considerable experience listening critically to live and recorded music. Music Performance majors at the University of Hartford study a curriculum of music-focused courses, involving private lessons, ensemble performance, ear training, music theory, music history, and are see and conduct live performances regularly. As can been seen in the following graph, there is very little difference between the responses of the two subject types. The results of the Type 1 subjects yielded a C80 JND of 3.7 dB, and the results of the Type 2 subjects gave a similar C80 JND of 3.9. Given the similar trend lines and resulting C80 JND, no difference was found between the two subject types, and since there are much higher numbers of music performance majors than MPT majors, future studies will continue to recruit from the large population of music students.

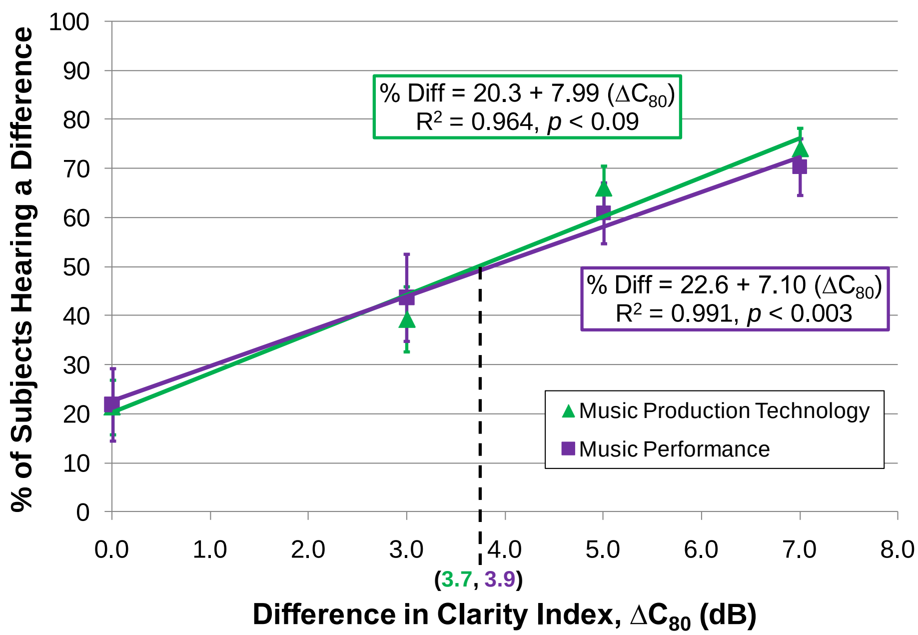

Figure 11: C80 JND results separated by subject type.

4.6 Interaction Effect of Base Case and Subject Type on C80 JND Results

Figure 12 shows the combined effects of base case and subject type on the C80 JND results. The blue and red curves are the results for the low base case (-3 dB clarity), and the green and purple curves are the results for the high base case (+1 dB clarity). Within a given base case, the JND for each subject type was very similar, however the JND for the two subject types was noticeably higher for the lower base case. While the Base Case 1 shows very little difference between the subject types' responses, when the higher base case signal is used (Base Case 2), differences between the two slopes are more apparent. This suggests that Subject Type 1 (MPT students and teachers) was better able to notice changes in clarity as the clarity increases, while Subject Type 2 (music majors) had more trouble identifying the changes for Base Case 2.

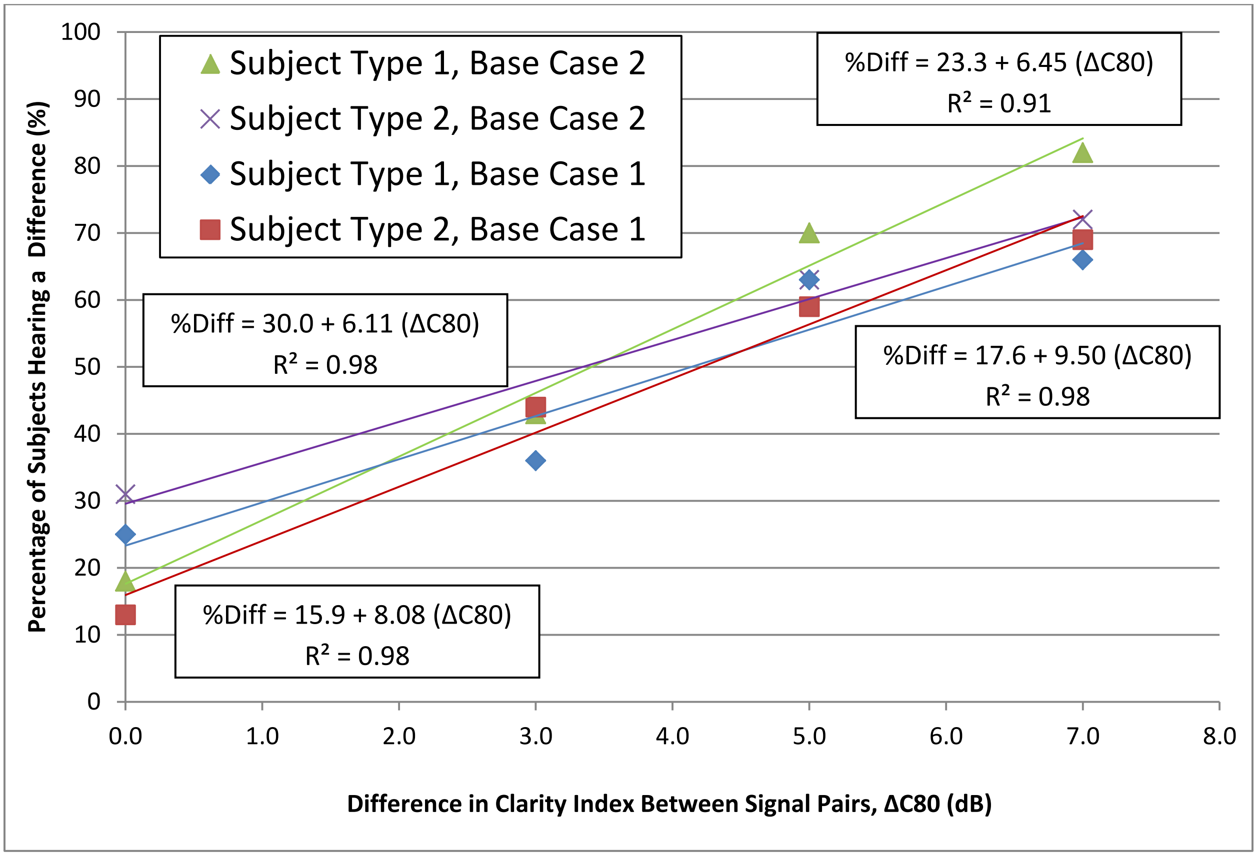

Figure 12: C80 JND results separated by the four combinations of subject type & base case.

4.7 Effect of Test Method on C80 JND Results

Two testing methods were used in this experiment; one which required the subject to listen to the entirety of a motif with one signal, then compare to the entirety of the next signal (Test Method 1) and one which enabled the subject to switch the track being played in real time, at any time during a motif (Test Method 2. The results are shown in Figure 13. It is notable that while the slopes of the percentage of subjects answering ‘different’ against the clarity difference are very similar for both methods, Test Method 2 has a much higher correlation, and also has a much lower value of people answering ‘different’ when the clarity difference is zero dB. This suggests that the Test Method 2, which allows the subjects to switch between the signals allows for more accurate answers by the subjects than the other method. It is also notable, however, that the JND for both methods was found to be drastically different. For Test Method 1 the C80 JND was found to be 2.8 dB, while for Test Method 2, the C80 JND was found to be 4.6 dB. These results suggest further research may be needed to investigate the differences in test method with a larger subject pool.

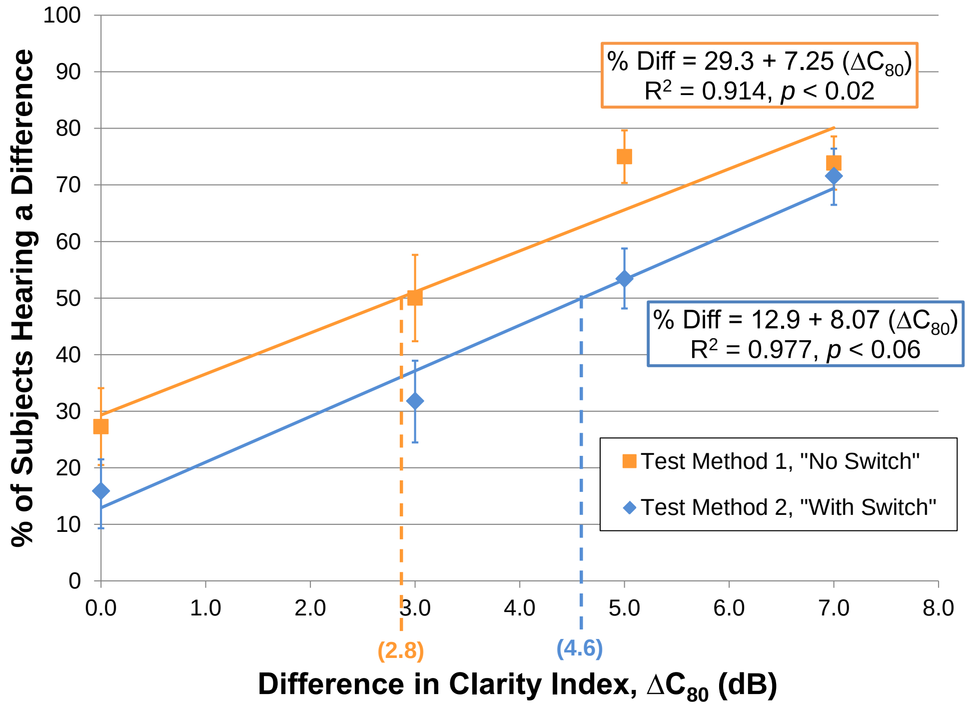

Figure 13: C80 JND results separated by test method.

4.8 Effect of Testing Order on C80 JND Results

Another variable affecting the data in this experiment was testing order, i.e. whether the subjects started the test using the "no switch" method or the "with switch" method. As shown in Figure 14, the lowest percentage of subjects responding 'different' to the 0.0 dB difference and the highest percentage of subjects responding 'different' to the 7.0 dB difference were found for the subjects when they used Test Method 2, "with switch", after having spent the first half of the testing using the other method, "no switch" (solid red line with red triangles). The C80 JND for this scenario was found to be 4.4 dB. This result suggests that the preferred testing method is to train test subjects using Test Method 1 and then obtain the actual test results using Test Method 2. It is also worth noting that the worst results were obtained when the subjects used Test Method 1 after having spent the first half of the test using Test Method 2 (dashed blue line with blue circles).

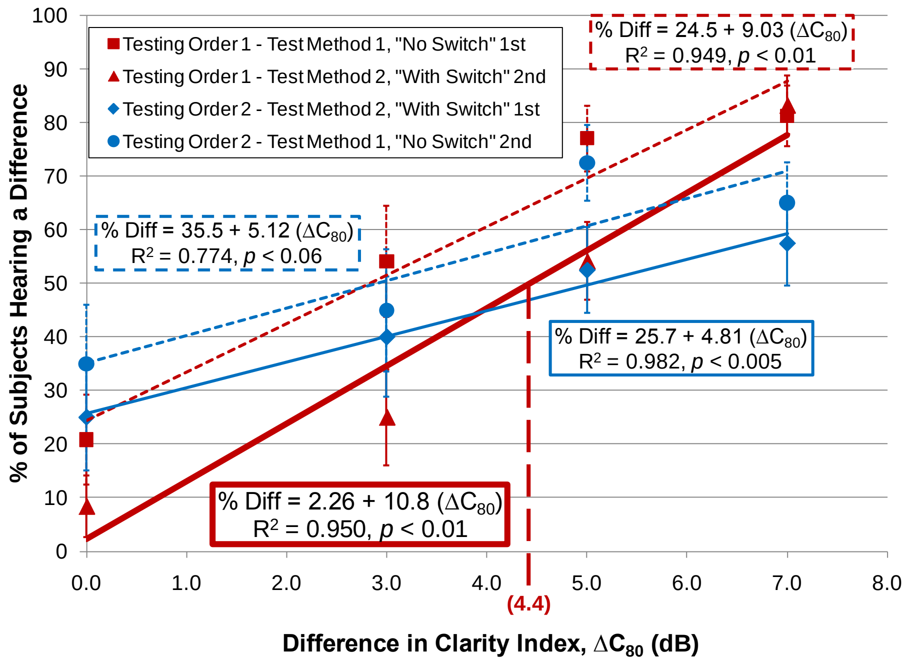

Figure 14: C80 JND results separated by the four combinations of testing order & test method.

5.0 Conclusions and Recommendations for Future Work

The test has been modified from previous research done at the University of Hartford. Subjects had previously reported a difficulty in hearing differences, so greater differences were implemented (up to 7.0 dB instead of 3.0 dB). The test subjects for this study reported fewer difficulties in hearing the changes in clarity. The number of test questions was also reduced from nine differences to four which made the test less tiring and repetitive.

The C80 JND determined by averaging all variables was 3.8 dB with a standard deviation of 0.7 dB. This JND is much higher than all of the prior investigations. The effect of Base Case and motif are similar to what was expected, with the high base case and the more staccato motif both resulting in more accurate responses. There appears to be little effect of subject type. It was unexpected that the use of the ‘no switch’ yielded the lower JND, rather than ‘with switch’, which was hypothesized to be an easier method of hearing differences. However, the ‘with switch’ method yielded greater overall accuracy. Overall, the use of the ‘with switch’ method in the second half of the experiment yielded the most accurate, and close to ideal results.

It is recommended that future work in this area involve an intensive training session using both the ‘switch’ and ‘no switch’ method, before the ‘actual’ test. The actual test should use the ‘with switch’ method, as it yielded the most accurate, and close to ideal results. Study 2c as described later in this report used this approach with a larger subject pool to further investigate the C80 JND.

Study 2b:

Effect of Motif Length in

Psychoacoustics Testing

by

Scott Edwards

Daniel Ignatiuk

with

Michelle Vigeant, Ph.D.

Robert D. Celmer, Ph.D., P.E.

Director, Acoustics Laboratory

University of Hartford

STUDY 2B TABLE OF CONTENTS

STUDY 2B LIST OF FIGURES

STUDY 2B LIST OF TABLES

1.0 INTRODUCTION

The purpose of this investigation, Study 2b, was to determine the effect of motif length in psychoacoustic tests requiring the use of musical samples. For psychoacoustic tests involving sound samples, there has been little previous research about how the length of samples can affect the results of testing. Sample lengths that are excessively long or short may test the auditory memory of the individual participant rather than a specific psychoacoustic quality which was meant to be tested. By determining which sample lengths allow the greatest accuracy in a subjective testing setting, future psychoacoustic research may benefit from these results.

2.0 BACKGROUND INFORMATION

In discussion of the performance of the auditory memory, Cowan (6) distinguishes between short auditory stores (lasting up to 300 ms) and long auditory stores (lasting 10–15 s). Research has shown that there was a significant drop in performance of memory recall tasks after 300 ms, and matching task performance decayed around 10–15 s suggesting that there are two distinct forms of auditory memory dependent upon task. Additional research showed that auditory distractions that were not to be recalled (termed “the suffix effect”) and silent gaps between matching samples of more than 1s led to significantly poorer performance of long memory stores (lasting only 2–4 s).

In his paper on the memory of noise, Kaernbach (7) distinguishes sensory memory from categorical memory by the trace code that is memorized. Kaernbach conducted a series of experiments in which subjects were required to listen for patterns and repetitions in noise. Subjects were tested for the maximum period that they could perceive a repeating cycle, detecting a single repetition (with a gap and without a gap in between), auditory memory capacity, and interference or distraction effects. Results showed that sensory memory and categorical memory have similar dynamics and interaction patterns. However, sensory memory is related more closely to short-term memory and cannot be stored or improved through rehearsal, while categorical memory can be stored and can be sustained through rehearsal techniques.

When determining the auditory memory persistence in European songbirds Zokoll *et al* (8) used artificial pure tones in Delayed Matching to Sample (DMTS) and Delayed Non-Matching to Sample (DNMTS) procedures. In these procedures, test tones were followed by a delay and several sample tones. In the DMTS procedure, the task is to identify the sample tone that matches the original test tone, while in the DNMTS procedure the task is to identify which of the sample tones did not match the original test tone. A similar DMTS procedure has also been used in order to test short-term memory in animals and humans (9).

The DMTS and DNMTS procedures are closely related to the ABX subjective testing method first described by Munson and Gardner used in the fields of noise and architectural acoustics (10). In the ABX method, signals are provided so that the first signal given is Signal A and a second Signal B is given different from Signal A. A final test Signal X is given that is either the same as A or B and the subject's task is to accurately match Signal X to A or B. The first formal application of this method was used by Stevens (11) in a study on frequency discrimination for damped waves. In Stevens' test every ABX group was balanced with an additional BAX group. Large differences between A and B were used for the first ten groups in order to introduce subjects to the test.

3.0 Experimental Procedure

3.1 Overview of Experimental Method

The goal of this study was to examine the effect of motif length on the ability of test participants to correctly match Signal X to either Signals A or B, with the hypothesis that a shorter motif would lead to the highest number of correct responses. The effect of motif type was also examined by using an excerpt from an orchestral piece and an excerpt from a purely percussive piece for the bongo drums. In order to simplify the test and reduce the chance of adding a confounding variable, the relatively simple variable of reverberation time (RT) was used to provide differences between the signals. Preliminary tests showed that a difference of 0.5 s between signals with 1.0 and 1.5 s RT's was clearly audible, while a difference of 0.3 s between signals with 1.0 and 1.3 s RT's was more difficult to detect. Three different signal lengths were used for each motif: 5 s, 7 s, and 10 s.

The testing took place in the anechoic chamber and the subjects were presented the signals over eight spatially distributed loudspeakers. The test began with a training set in which ΔRT was set to the maximum difference used in the test (0.5 s). Two test sets followed that included both motifs played back at one of the three signal lengths with the varying ΔRT's. A 5-minute break was given to subjects between the two test sets and the entire test lasted approximately 1 hour, including the hearing screening.

3.2 Musical Motifs

Four motifs were selected to use in the study, with two used for the actual test ("test motifs") and two used for a training set ("training motifs"). The test motifs varied between an orchestral excerpt from Mozart's Overture to *The Marriage of Figaro* and a sample of a solo percussionist playing bongos. The Mozart motif was chosen for its familiar orchestral timbre and dynamic contrast, while the percussion motif was chosen for its crisp rhythmic attacks and decays during rests. The test motifs varied in length between 5, 7, and 10 s. For both test motifs, the 5 s and 7 s lengths were made by reducing the length of the 10 s motif so that the same musical passage was used for each length. Two additional motifs were selected for a training set which included an excerpt from Beethoven's *Symphony No.9* and a Weber cello theme. The training motifs were fixed at 8 seconds for the Beethoven excerpt and 7 s for the Weber cello theme.

3.3 Signal Generation

The equipment and method for the signal generation was identical to the one used in Study 2a (see pages 9 – 11 for details.) A base case signal with a 1.0 s reverberation time was initially generated, with spectral shape as shown in Figure 1. Three additional signals were created for comparison to the base case with reverberation times of 1.3, 1.4, and 1.5 s; thus, when paired with the base case for the ABX tests one of three ΔRT values would be presented, i.e. 1) 0.3 s ΔRT (comparing 1.0 to 1.3 s); 2) 0.4 s ΔRT (comparing 1.0 to 1.4 s); and 3) 0.5 s ΔRT (comparing 1.0 to 1.5 s). The gain levels of the early reflections for each octave band were adjusted to keep the clarity index (C80) relatively constant, within 0.1 dB, for all of the reverberation time signals. The overall levels of each signal, which were measured using a Brüel and Kjær (B&K) Type 2260 sound level analyzer, were also adjusted to ensure that signals of equal length at the four different RT's were all within approximately 1 dBA (re 20 µPa). See for The sound pressure and sone levels of each signal are shown in Table 1 for the Mozart signals and in Table 2 for the percussion signals.

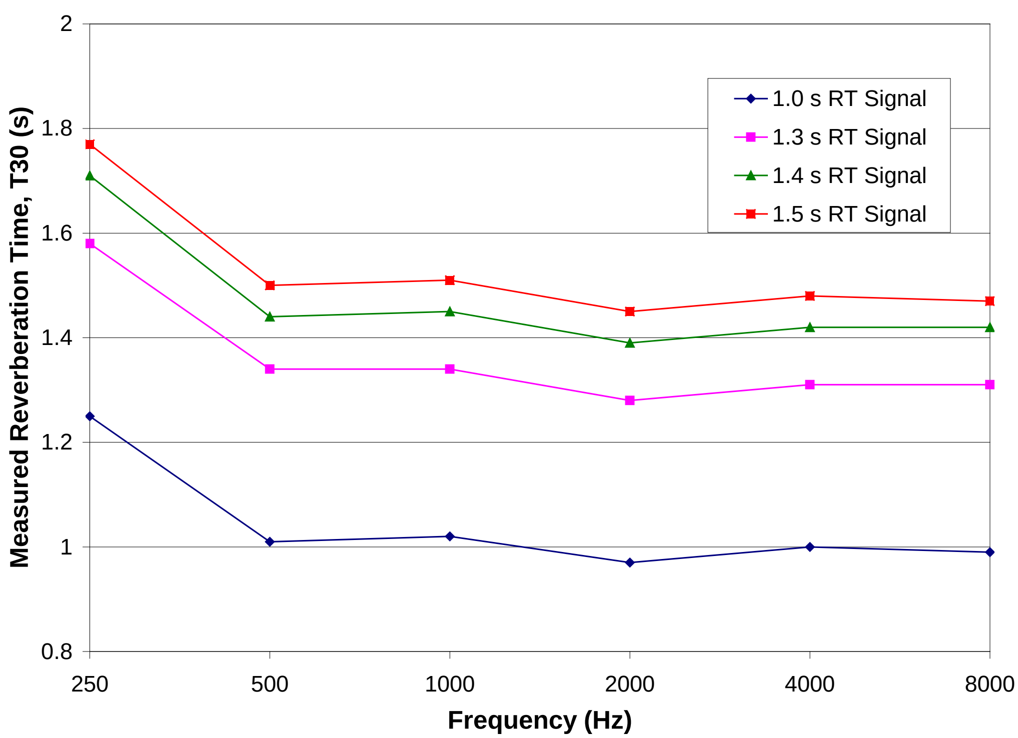

Figure 1: Measured Reverberation Time (T30) vs. Frequency

Table 1: Sound Pressure and Sone Levels for Mozart Test Motifs

	Sound Pressure Level (dBA re: 20 μPa)				Loudness Level (Sones)			
Motif Length (s)	RT = 1.0 s	RT = 1.3 s	RT = 1.4 s	RT = 1.5 s	RT = 1.0 s	RT = 1.3 s	RT = 1.4 s	RT = 1.5 s
5	68.8	69.9	69.7	69.9	22.0	23.7	23.4	23.7
7	69.1	69.0	69.5	69.7	22.6	22.5	23.2	23.5
10	68.3	69.4	68.5	68.2	21.5	22.9	21.9	21.4

Table 2: Sound Pressure and Sone Levels for Percussion Test Motifs

	Sound Pressure Level (dBA re: 20 μPa)				Loudness Level (Sones)			
Motif Length (s)	RT = 1.0 s	RT = 1.3 s	RT = 1.4 s	RT = 1.5 s	RT = 1.0 s	RT = 1.3 s	RT = 1.4 s	RT = 1.5 s
5	67.9	66.8	66.9	67.7	20.5	19.2	19.3	20.1
7	67.4	67.6	67.2	67.8	19.9	20.3	19.8	20.5
10	66.8	66.6	66.4	66.4	19.3	19.1	18.9	18.8

The signals were sent played back over eight Genelec loudspeakers in the anechoic chamber The speakers were arranged with six speakers in front of the listener (two on axis, four off axis) and two speakers off axis behind the listener, as shown in Figure 2 – Figure 4.

Figure 2 - Front Six Loudspeaker Location

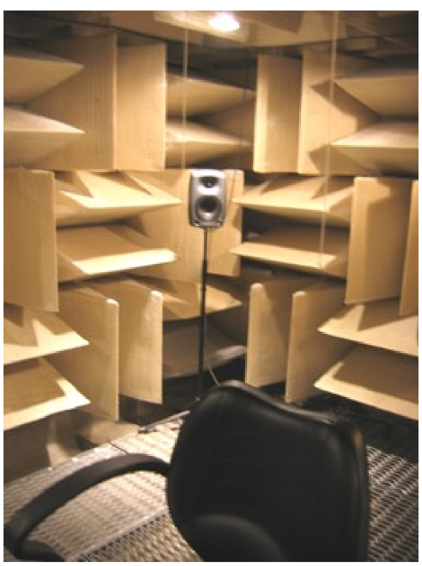

Figure 3 - Right Rear Loudspeaker Location

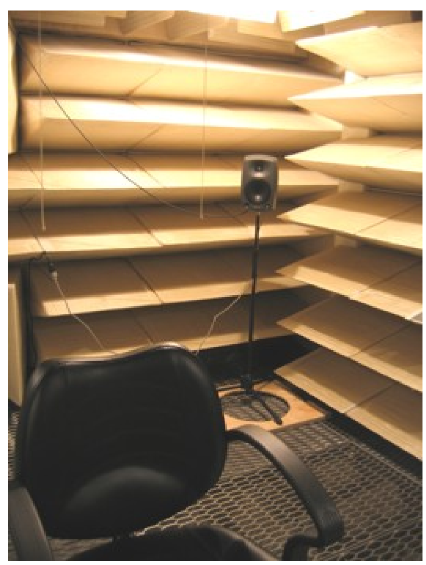

Figure 4 - Left Rear Loudspeaker Location

3.4 Summary of Test Participants

Test participants were required to have a minimum hearing threshold of 15dB HL in the 250 – 8000 Hz octave bands. All participants were required to be at least 18 years of age and have a minimum of five years of formal musical training. A total of 25 subjects participated in the study, 12 males and 13 females, with a majority of subjects being 18–22 years old. On average, test subjects had 12 years of music experience with 10 years of formal training. Ten of the test subjects participated in Study 2a, the C80 JND test method pilot study (3).

3.5 Testing Procedure

The testing procedure began with a training set which included eight ABX listening questions that consisted of two musical motifs played in a random order. The musical motifs used for the training were the Beethoven excerpt and Weber cello theme described in Section 3.2. Each question had a ΔRT of 0.5 s between samples A and B. Once this set was completed, subjects were notified that the test would begin. After the training set, the first six ABX questions were used as additional training unbeknownst to the subject and were not used in the experimental data. These additional questions included only ΔRT values 0.4 s and motifs with the shortest length of 5 s.

A total of 36 test questions were administered to each subject which included 2 motifs x 3 motif lengths x 3 ΔRT values x 2 possible orders ABX or BAX = 36 total questions. A random number generator produced a random ABX order of the three variables motif, motif length, and ΔRT for every test subject. Signal X in each ABX or BAX set was randomized in every case, so each subject received only two of the four possible combinations (ABA or ABB and BAA or BAB). The first part of the test included a set of 18 ABX questions which were followed by a five minute break and a second set containing 18 ABX questions. Subjects were notified when they were half way through each set.

All ABX questions in the test consisted of a musical motif played at a certain length, either 5, 7 or 10 s, with one of the three fixed ΔRT values between Signals A and B. The given motif, the motif length, and the difference in RT (ΔRT) between Signals A and B were random for each question.

4.0 Results and Discussion

The collected data for the 25 subjects was organized into contingency tables so that the test statistic χ^2 could be computed (12). If the χ^2 (chi-squared) test statistic is found to be significant, it means that subjects were actually able to determine if the test Signal X matched either Signal A or Signal B and weren't simply guessing. The data are grouped within tables to evaluate if differences exist between the different levels of the variables, i.e. if there is a statistically significant difference between the different motif lengths or the different ΔRT's. χ^2 was found to be statistically significant at $p < 0.05$ in all cases, indicating that subjects were able to correctly match Signal X to either Signal A or B. In order to compare the results beyond using the test of statistical significance, the numerical values of χ^2 were compared. A higher test statistic χ^2 correlates to a higher number of correct answers for a given combination of variables. The data were organized to first compare the effect of motif length within a given motif for each ΔRT (see Table 3– Table 4 and Figure 5 – Figure 6) and then the data were organized to compare the differences between each motif at the three different lengths for each ΔRT (see Table 5 – Table 7 and Figure 7 – Figure 9).

Table 3 and Figure 5 show how χ^2 varied between motif length and between ΔRT for the Mozart motif. The general trend shown in these results is that χ^2 (number of correct responses) increases with an increase in ΔRT, which was expected since a larger ΔRT should be easier to detect. An exception to this general trend occurs for the 7 s motif with the 0.4 s ΔRT, which resulted in higher correct responses than the lowest and highest ΔRT. When comparing differences in results with motif length, no differences are found between the 5, 7 and 10 s motifs for the smaller ΔRT's. However, for the condition with the largest difference in RT, or the easiest scenario, the shortest motif of 5 s yielded the highest subject accuracy. These results indicate that there is an interaction between these two variables, meaning that the motif length that will provide the highest number of correct responses will vary depending on the testing condition, in this case the ΔRT.

Table 3: Contingency Table of Chi-Square vs. ΔRT for Mozart Motif at Varying Lengths

Motif Length (s)	ΔRT = 0.3 s χ^2	ΔRT = 0.4 s χ^2	ΔRT = 0.5 s χ^2
5	21.5	21.8	44.6
7	18.9	24.5	18.1
10	12.8	19.8	31.6

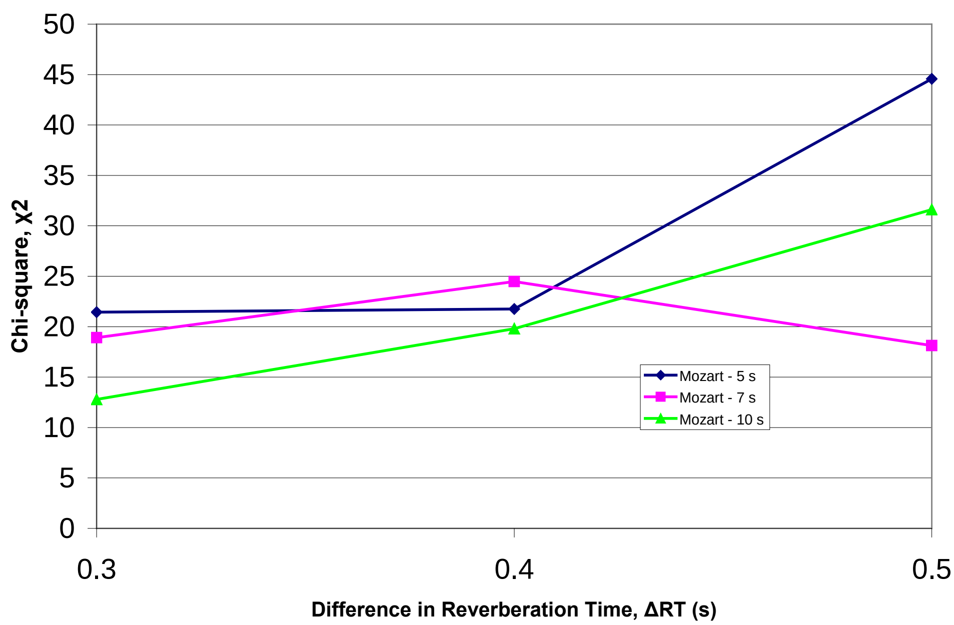

Figure 5: Plot of Chi-Square vs. ΔRT for Mozart Motif at Varying Lengths

The following table and figure provide the same comparison of differences between motif length and ΔRT for the percussion motif. The same general trend that was found with the Mozart motif is also shown in these results, in that the number of correct responses increases as ΔRT increases.

However, the 5 and 7 s motif yielded a higher number of answers for the most difficult case (ΔRT = 0.3 s), while the 10 s motif gave the highest number of correct answers for the easiest case (ΔRT = 0.5 s) These results again show the interaction effect between motif length and ΔRT.

Table 4: Contingency Table of Chi-Square vs. ΔRT for Percussion Motif at Varying Lengths

Motif Length (s)	ΔRT = 0.3 s χ^2	ΔRT = 0.4 s χ^2	ΔRT = 0.5 s χ^2
5	20.3	24.8	20.3
7	19.0	17.6	31.0
10	11.1	20.2	37.4

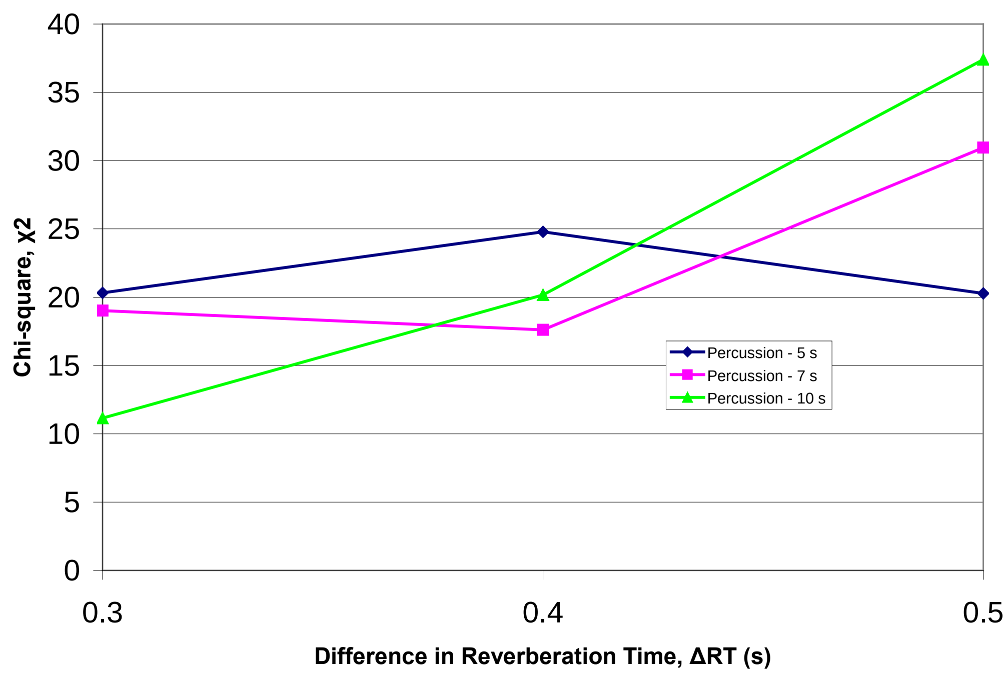

Figure 6: Plot of Chi-Square vs. ΔRT for Percussion Motif at Varying Lengths

The data were reorganized to provide a direct comparison of the two motifs as a function of the same motif length while holding ΔRT constant. Table 5 and Figure 7 show how the number of correct responses varied as the motif length increased when the ΔRT was held constant at ΔRT = 0.3 s, the smallest difference. For this case, the number of correct responses was found to decrease as the motif length increased and this trend was the same for both motifs. In other words, for this most difficult case, the shortest motif length was found to give the highest number of correct responses.

Table 5: Contingency Table of Chi-Square vs. Motif Length for both Motifs, ΔRT = 0.3 s

Motif	Motif Length = 5 s χ^2	Motif Length = 7 s χ^2	Motif Length = 10 s χ^2
Mozart	21.5	18.9	12.8
Percussion	20.3	19.0	11.1

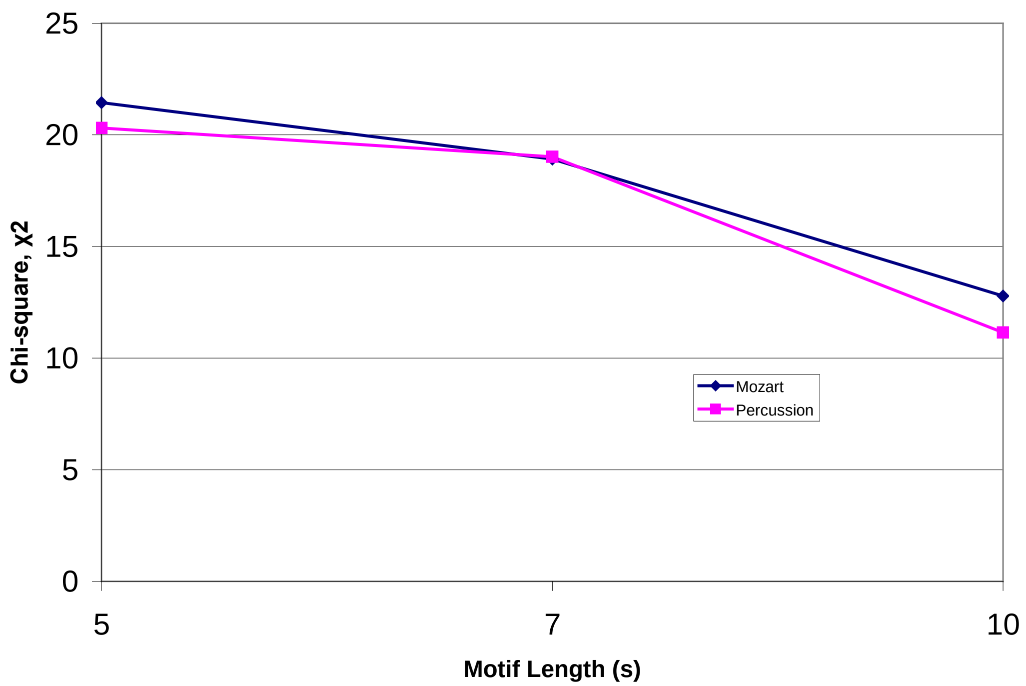

Figure 7: Plot of Chi-Square vs. Motif Length for both Motifs at ΔRT = 0.3 s

The following table and figure show how the number of correct responses varied as the motif length increased when the ΔRT was held constant at ΔRT = 0.4 s, the smallest difference. Unlike the previous case, no general trend is found as the motif length increases, except for the 7 s percussion motif. The ending of this motif as combined with the 1.4 s signal may have made it difficult to clearly hear the signal's decay making it difficult to match the test signal to the correct signal.

Table 6: Contingency Table of Chi-Square vs. Motif Length for both Motifs, ΔRT = 0.4 s

Motif	Motif Length = 5 s χ^2	Motif Length = 7 s χ^2	Motif Length = 10 s χ^2
Mozart	21.8	24.5	19.8
Percussion	24.8	17.6	20.2

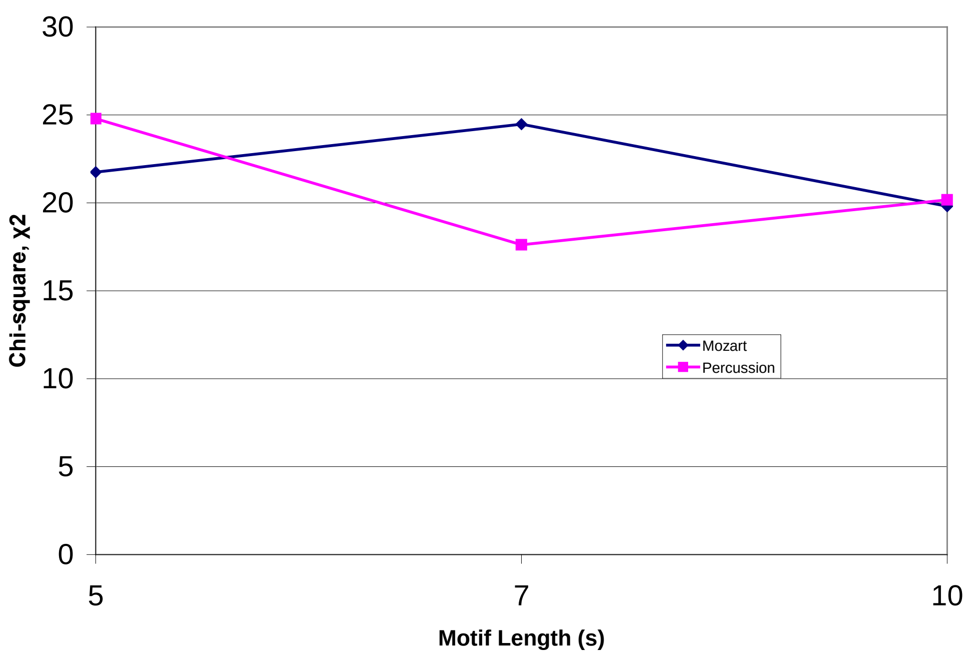

Figure 8: Plot of Chi-Square vs. Motif Length for both Motifs at ΔRT = 0.4 s

Finally, Table 7 and Figure 9 compare the results of the two motifs as a function of motif length with the third ΔRT difference of 0.5 s. For this ΔRT, an interaction effect is found between the motif type and motif length. In particular, the shortest motif length of 5 s yielded the highest number of correct responses for the Mozart motif and the lowest number for the percussion motif, while subject accuracy was found to increase for the percussion motif as the motif length increased. In addition, the number of correct responses was also higher for the longer percussion motifs than the longer Mozart motifs. The continuous nature of the percussion motif likely made it more difficult to hear differences when the motif was only 5 s, but became easier as the motif's length increased. The large decline in the correct responses between the 5 to 7 s Mozart motif may be accounted for by a difference in the actual Mozart motif between the 5 s and 7 s lengths. The 5 s Mozart motif length provided a constant musical passage without breaks in the music. The 7 s and 10 s lengths contained parts with shorter impulsive hits resulting in pauses within the motif.

Table 7: Contingency Table of Chi-Square vs. Motif Length for both Motifs, ΔRT = 0.5 s

Motif	Motif Length = 5 s χ^2	Motif Length = 7 s χ^2	Motif Length = 10 s χ^2
Mozart	44.6	18.1	31.6
Percussion	20.3	31.0	37.4

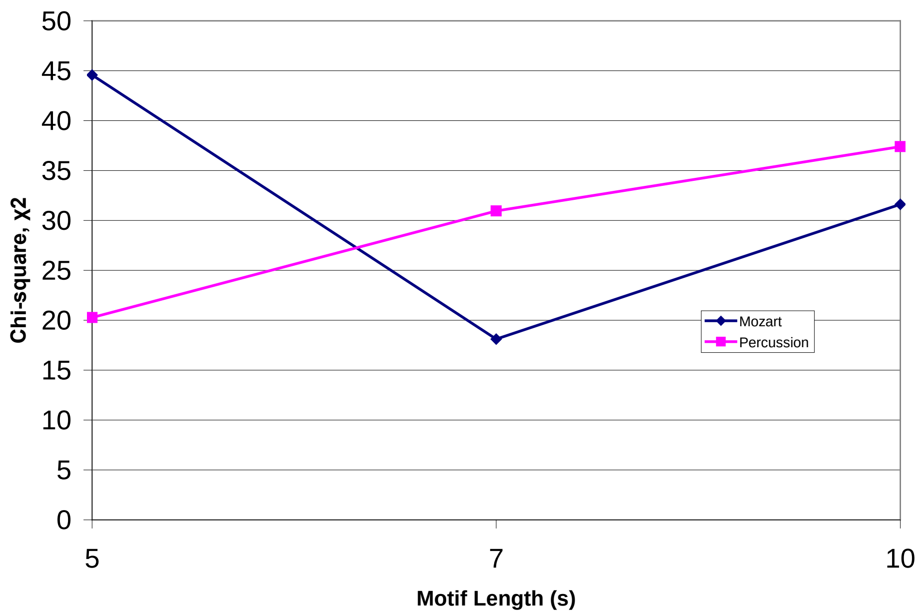

Figure 9: Plot of Chi-Square vs. Motif Length for both Motifs at ΔRT = 0.5 s

5.0 Conclusions and Recommendations for Future Work

Overall, the highest number of correct responses was found for the largest difference between signals of ΔRT = 0.5 s as was hypothesized. However, no clear trend was revealed in terms of an optimum motif length for this type of test. An interaction effect was found between motif type and ΔRT, and also between motif length and ΔRT. For the Mozart motif, no differences were found between the motif lengths of 5, 7 and 10 s for the cases of ΔRT = 0.3 and 0.4 s. Therefore, for those two scenarios any of the three motif lengths are suitable. For the largest difference of 0.5 s, no clear trend emerged, as the higher number of correct responses was obtained with the shortest motif and the lowest subject accuracy was for the middle motif length of 7 s. For the percussion motif, the shorter motifs yielded higher subject accuracy for ΔRT = 0.3 s, while the number of correct responses were found to increase as the motif length increased. These results are somewhat intuitive given the nature of the percussion motif, which was a continuous bongo drum beat with varying rhythms.

A general recommendation for an optimal motif length for this particular testing scenario cannot be specified due to the interaction effects found between variables. In other words, determining an ideal motif length for architectural-psychoacoustics tests is more complicated than initially hypothesized and further work is needed to investigate the interaction effects noted above. Additional studies should include a larger number ΔRT values with one to two additional motifs to determine if any larger overall trends exist in the data. In addition, RT could be held constant while other room acoustics variables are changed individually, such as clarity index, to determine if differences in motif length have a more significant effect for certain room acoustics characteristics over others.

Study 2c: Detailed Investigation of

Clarity Index (C80)

Just Noticeable Difference (JND)

by

Caitlin Ormsbee

Adam Wells

with

Michelle Vigeant, Ph.D.

Robert D. Celmer, Ph.D., P.E.

Director, Acoustics Laboratory

University of Hartford

STUDY 2C TABLE OF CONTENTS

STUDY 2C LIST OF FIGURES

STUDY 2C LIST OF TABLES

1.0 INTRODUCTION

The purpose of this study, Study 2c, was to conduct a detailed investigation of the clarity index (C80) just noticeable difference (JND) building upon the results obtained under the previous grant (3) and Studies 1 and 2 as described earlier in this report. The results from (3) revealed that the C80 JND is likely greater than 1.0 dB, since unclear results were obtained when the test was designed to include signals with differences in C80 around this value. The results from Study 2a indicate that the C80 JND might be closer to 4.0 dB and also showed that there is an effect of test method on the results, and these results were incorporated into the experimental design for this study. Study 2b gave inconclusive results about motif length, so a motif length of approximately 10 s as was used in previous studies was also employed in this study. Study 2c had two base cases, similar to the ones used in Study 2a, but six C80 differences, which was an increase from the number used in Study 2a (three differences). To reduce the overall testing time for each test subject, only a single motif was used in this final study.

2.0 BACKGROUND INFORMATION

2.1 Previous Studies

Three previous studies on C80 JND have been conducted: Cox, Lam, and Davies (1), Bradley, Reich and Norcross (2), and Ahearn, Schaeffler, Vigeant, and Celmer (3). See pages 4 – 7, within the report of Study 2a, for summaries of each of these studies. A brief summary of Study 2a is given below:

The purpose of Study 2a was to determine the effect of two different testing methodologies on the resulting C80 JND. Base cases with C80 values at 1 kHz of -3 dB and +1 dB were compared to signals with differences of 0, 3, and 7 dB. Test Method 1 required subjects to listen to all of Signal A and then all of Signal B before answering if the signals were the same or different and Test Method 2 allowed to freely switch between Signals A and B. A custom interface box was created for the subjects to use to both switch between the signals and also indicate their responses. The overall average C80 JND was found to be 3.8 $\pm$ 0.7 dB, but an effect of testing method was found. The testing order that gave the results most like those hypothesized, with a low number reporting hearing a difference at 0.0 dB and most hearing a difference at the maximum difference of 7.0 dB, was found for the subjects who used Test Method 1 for the first half of the test and then used Test Method 2. The results from latter portion of these subjects tests where they used the “with switch” method matched the hypothesized trend the best and resulted in a C80 JND of 4.4 dB. Due to the pilot nature of this study with only 11 test subjects, further work was conducted to investigate the C80 JND.

2.2 Helsinki Orchestral Anechoic Recordings

At Helsinki University of Technology, Lokki *et* al sought to create multi-channel anechoic recordings of individual instruments for several classical musical pieces for use research in acoustics research (13). A primary goal when obtaining the recordings was to maintain perfect signal separation between each instrument and to obtain directivity information for each part played. Signal separation was achieved by individually recording each instrumental part. In order for the individual parts to be synchronized in time when combined at the end, the musicians watched a video of a conductor and listened to the corresponding piano performance over headphones. The recordings were obtained in an anechoic chamber with a cubic dimension of 4.2 m (13.7 ft), with absorption wedge length of 80 cm (2.6 ft). With these large overall dimensions and wedge depths, the room is anechoic for frequencies above 10 Hz, which is needed for instruments noted to have significant frequency generation below 100 Hz (timpani, tuba, contrabass, and cello). The recordings were obtained using a total of 22 Rode NTI A-Type large-diaphragm microphones, which were positioned to form a dodecahedron as shown in Figure 1. A laser pointer was used to measure and align the angular orientations of the microphones.

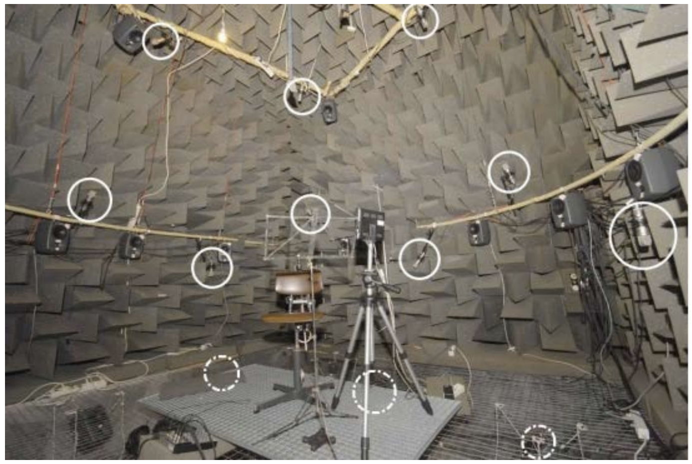

Figure 1: Dodecahedron Microphone placement

Additional equipment used in the study was the Motu 2408 mk3 audio interface and three 8-channel Presonus Digimax FS. To calibrate the system, a Genelec loudspeaker, Type 1032A was used to play back a sine sweep while a Brüel & Kjær (B&K) omni-directional microphone, Type 4191 was used to measure the impulse and frequency response. The response was similarly measured of the 22 Rode microphones at their final positions, with any variations in sensitivity between the microphones compensated for by de-convolving the Rode microphone in the recording room based on the measurement using the B&K microphone. A one-third octave band and low shelf filter were applied to reduce and smooth the low frequency response of the Rode microphones. It was noted that the additional equipment (i.e. the loudspeakers) may have been a cause for error, as well as the small offset of the microphone array to accommodate the room size and shape.

Lokki *et* al decided to record a number of musical styles in order to provide a range of musical motifs to use in architectural acoustics research, particularly in the study of auralization, and in the design process. Classical and Romantic styles were chosen for this experiment since both Baroque and Modern styles presented too many issues including copyright and instrumentation. In total, four pieces were chosen and each was then shortened to obtain passages under four minutes in length. Mozart's *Don Giovanni*, specifically the *Donna Elvira* soprano aria, was chosen from the Classical period. While it does feature a soloist, it requires a very small orchestra. The entire piece was recorded since it is less than four minutes. Next, Beethoven's *Symphony No. 7* from the late Classical periods was chosen so that when convolved with a signal, the reverberation would be quite audible due to the number of unified chords and pauses. Although the required woodwind and brass sections are much larger for this piece, the string section is very close to the size of that require for the Mozart. Bruckner's *Symphony No. 8* and Mahler's *Symphony No. 1* are both the late Romantic period. While these two pieces require large orchestras, the Mahler is considered more complex and a larger excerpt was played to give users a larger range of choices for possible motifs.

To reduce the number of musicians required to make the recordings, only one musician per instrument was used, i.e. only on string player for each instrument. For the woodwind and brass instruments, a musician for each instrument played all parts for that given instrument. This approach resulted in only needing 14 musicians to make the recordings. In order to re-create the effect of multiple string players playing simultaneously, multiple takes of each these instruments were taken for use in post-processing. Each musician was instructed to play how he or she would in a performance environment and only a single incidence resulted in a level that exceeded the microphone's range, one of tuba player's takes.

All of the clips from the musicians were compiled into full tracks of each individual instrument and then further edited to create the entire orchestra's track. Approximately two to three edits were made to each accepted take. It was noticed that when musicians were recording and the piano accompaniment began to play out of tempo, that the musician would follow the piano and not the conductor.

To reduce the amount of noise during the pauses in the recordings, when the musicians were resting, the tracks were edited to mute automatically if the signal level remained under a predetermine threshold. This was especially important for the string tracks as each was amplified and multiplied to recreate the effect of an entire string section.

2.3 Anechoic Orchestral Motif Used in Study

It was determined that for the test, a single anechoic recording would be used to reduce the amount of variables in the experiment. Mozart's *Don Givonanni* was chosen because it contained flowing passages without overly dense harmonies or rhythms that might complicate its musical evaluation. The particular motif used in the testing, with a total length of 10.4 s, was selected for its distinctive repeating musical pattern.

The individual recorded instruments from the Helsinki orchestral anechoic recordings were compiled to create a single ensemble. For the string sections, the overall levels of each individual string recording were increased to represent the appropriate number of instruments for each section. In order to provide a somewhat spatially separated orchestra during playback over the front six loudspeakers, the instrumental recordings were panned to place each instrumental group in a representative position from left to right. The parts were panned to achieve separation and achieve a phantom location appropriate to their arrangement on stage as seen from the audience.

3.0 EXPERIMENTAL PROCEDURE

3.1 Overview of Procedure

The primary goal of this study was to determine the C80 JND building upon the results of Studies 2a and 2b. Given the results of Study 2a, which indicated that the C80 JND might be closer to 4 dB, rather than 1 dB, the differences between were set based on this result. The differences were also based on results found in (14), in which the authors studied the statistical validity of different sampling distributions. In particular, this work recommended having two differences below the expected C80 JND, and four differences well above this threshold. Using this approach, the resulting differences between signals were set to 0, 1.5, 3, 5, 7 and 9 dB. Similar base cases to Study 2a were used, with the following differences: 1) for Base Case 1, T30 was reduced to 1.9 s from 2.1 s for a more realistic sounding signal and 2) for Base Case 2, C80 at 1 kHz was decreased from +1.0 to 0.0 dB to have the C80 more in the middle of the preferred range for chamber music halls (4). Thus, Base Case 1 had a C80 value of -3.0 dB at 1 kHz and T30 of 1.9 s and Base Case 2 had a C80 value of 0.0 dB at 1 kHz and T30 of 1.6 s.

As with the previous studies, subjects were presented two signals at a time, Signal A and Signal B and were required to determine if the signals sounded the same or different. The first half of the test was an extensive training period for the test subjects, in which the participants used Test Method 1, where subjects had to listen to all of Signal A and then all of Signal B before giving their response (*no switch)*. The second half of the test, which was the actual test, required the subjects to use Test Method 2, which allows subjects to freely switch between Signals A and B in real-time (with *switch)*, with the use of a custom interface box (see Figure 2). This overall test procedure was used as it was found to produce the most consistent results in Study 2a.

Figure 2: Custom subject-interface control that allows subjects to toggle between Signals A and B in real-time.

3.2 Signal Generation

The procedure and equipment for the signal generation and playback was exactly the same as the one used in Studies 1 and 2 (see pages 9 – 11 for details.) The two base case signals were created first, as shown as the red lines in Figure 3 and Figure 4. Then, the overall early reflection gain levels and the parametric equalizer were adjusted to achieve the comparison signals with the same spectral shapes and differences of 1.5, 3.0, 5.0, 7.0 and 9.0 dB as shown in the figures.

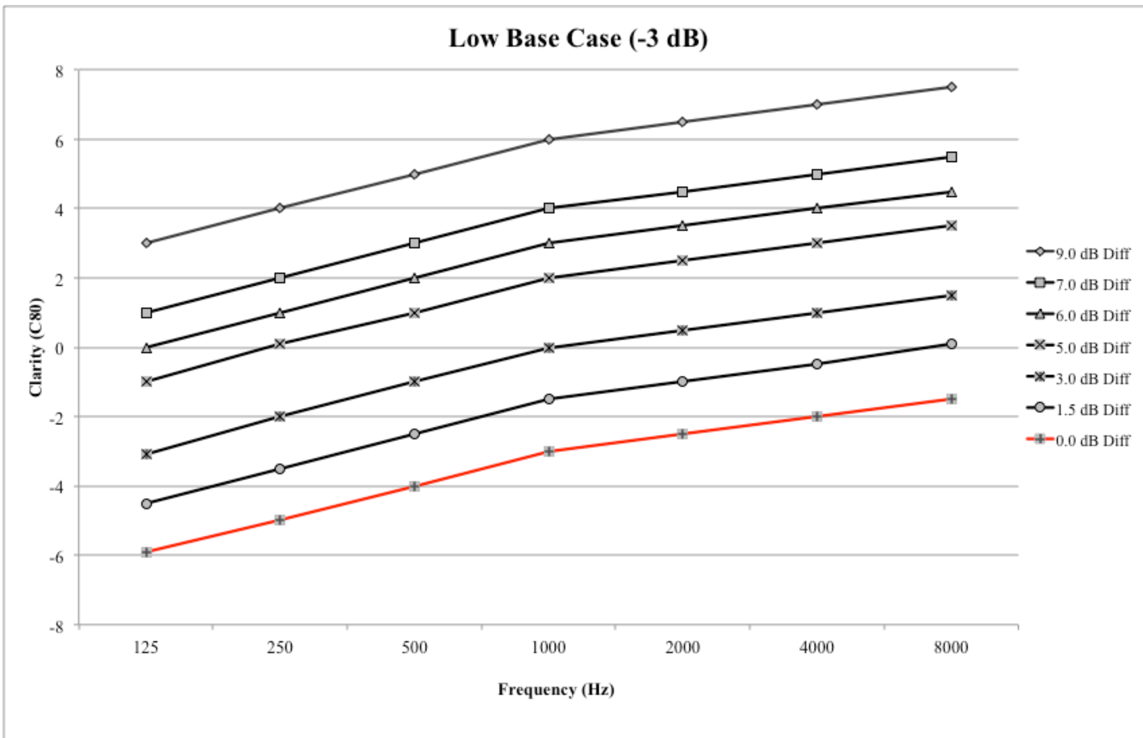

Figure 3: Clarity Values for Base Case 1 with C80 @ 1kHz = -3 dB and T30 of 1.9 s

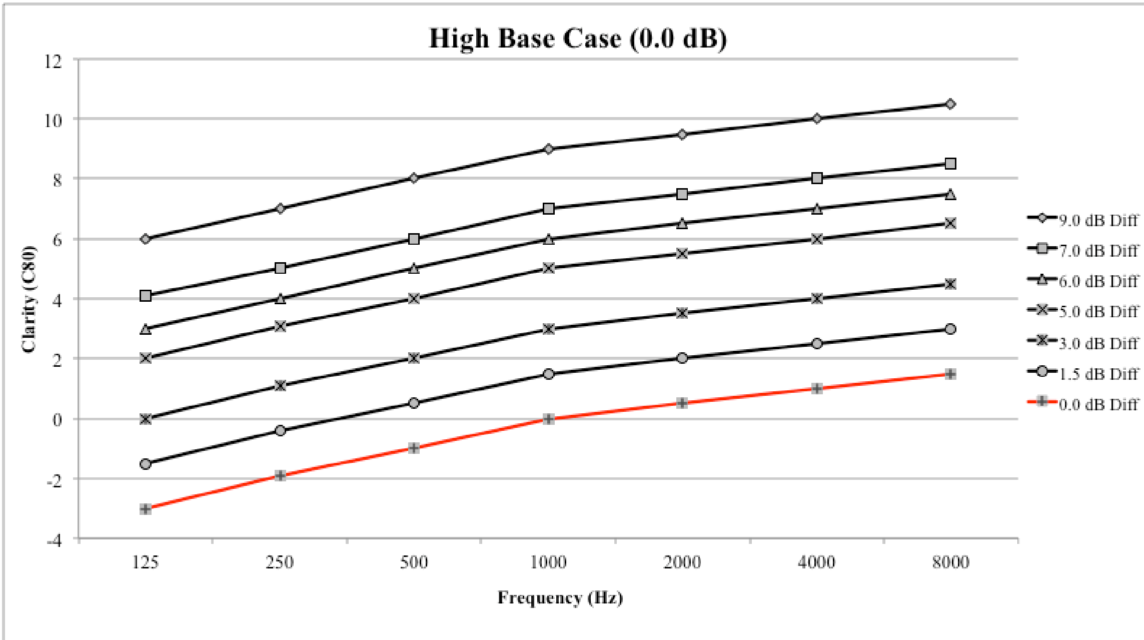

Figure 4: Clarity Values for Base Case 2 with C80 @ 1kHz = 0 dB and T30 of 1.6 s

3.3 Summary of Test Participants

A total of twenty-eight subjects from University of Hartford's The Hartt School, the university's music conservatory, were used for the experiment's sample pool. Subjects were required to have at least five years of formal musical training and in addition were required to also be currently taking music lessons. The sample pool consisted of 15 male and 13 female subjects who averaged 21 years old and had 10 years of musical instruction. The subjects were also required to pass a hearing screening, and have a hearing thresholds below 15 dB HL between 250 to 8000 Hz.

3.4 Summary of Testing Procedure

The results from Study 2a revealed that the most reliable results occurred when subjects used an extended training period using Test Method 1, "no switch", and then carrying out the actual test using Test Method 2, "with switch". The test divided into four sets with 16 questions each, with the first two sets entirely for training purposes and the second two sets the actual test, as shown in Table 1. For the first two training sets, the subjects were told that the first six questions were for training purposes and the remaining ten questions within the set were actual test questions, when in reality they were still training questions. Subjects used Test Method 1 in the first set and Test Method 2 in the remaining three sets. The final two sets contained 16 questions, all of which were actual test questions.

Table 1: Overview of Testing Procedure

	First Half of Test – Training		First Half of Test – Actual Test	
Subject Group 1	**Test Method 1, *no switch***	**Test Method 2, with *switch***	**Test Method 2, *with switch***	
	Set 1	Set 2	Set 3	Set 4
	Base Case 1 & 2	Base Case 1 & 2	Base Case 1	Base Case 1
	Training (6) & "Test" (10)	Training (6) & "Test" (10)	Actual Test (16)	Actual Test (16)
Subject Group 2	**Test Method 1, *no switch***	**Test Method 2, with *switch***	**Test Method 1, *no switch***	
	Set 1	Set 2	Set 3	Set 4
	Base Case 1 & 2	Base Case 1 & 2	Base Case 2	Base Case 2
	Training (6) & "Test" (10)	Training (6) & "Test" (10)	Actual Test (16)	Actual Test (16)

Before starting the test, the following formal instructions were given to the test subject, "*You are listening to differences in clarity between the two tracks. Consider both how **clear** the overall track sounds as well as how **clear** each note sounds relative to the subsequent note*." At the end of each section subjects were given a 2-3 minute break to counteract the fatigue of the listening test and being in the unusual environment of the anechoic chamber. During these breaks subjects remained mentally active by filling out a required musical history form, playing a logic-oriented computer game, or engaging in conversation with the test administrator(s).

The specific signal pairs given to each subject during the training sets, Sets 1 and 2, are shown in Table 2. For the first part of each set, part A, the subjects were given three signal pairs from each base case, where Signal A was always one of the two base cases and Signal B was either the same signal or one of the two signals with the largest differences of 7.0 and 9.0 dB, respectively. These three signals were chosen to have the subjects start the test with the most extreme cases. For the second part of each set, part B, two additional differences of 5.0 and 6.0 dB were included, while the smallest differences of 1.5 and 3.0 dB were not included, to again provide the subjects with the pairs with the larges differences in C80.

Table 2: Signals for Training Sets 1 and 2.

Type of Questions as Told to Test Subjects	Signal Pair	Signal A C80 value at 1kHz (dB)	Signal B C80 value at 1 kHz (dB)	C80 Difference between Signal A and Signal B (dB)
Sets 1A & 2A: Training	**1**	-3.0	-3.0	0.0 (same signal)
	2	-3.0	4.0	+ 7.0
	3	-3.0	6.0	+ 9.0
	4	0.0	0.0	0.0 (same signal)
	5	0.0	7.0	+ 7.0
	6	0.0	9.0	+ 9.0
Sets 1B & 2B: "Test" (additional training questions)	**7**	-3.0	-3.0	0.0 (same signal)
	8	-3.0	2.0	+ 5.0
	9	-3.0	3.0	+ 6.0
	10	-3.0	4.0	+ 7.0
	11	-3.0	6.0	+ 9.0
	12	0.0	0.0	0.0 (same signal)
	13	0.0	5.0	+ 5.0
	14	0.0	6.0	+ 6.0
	15	0.0	7.0	+ 7.0
	16	0.0	9.0	+ 9.0

The questions for the actual test sets, Sets 3 and 4, contained all C80 differences used in this study: 0.0, 1.5, 3.0, 5.0, 7.0 and 9.0 dB. Half of the subjects were presented Base Case 1 for both sets, while the other half were presented Base Case 2, as shown in Table 3 and Table 4, respectively. Each signal pair was presented twice for a total of 16 questions per set. The testing order was randomized for all sets for each test subject. For the training sets, the signal pairs for each base case were grouped together to be in either the first or second half of the set and this order was alternated for each subject.

Signal A was always the base case signal, with the exception of the final signal pair, where a different signal was presented as both Signals A and B for a second case of a 0.0 dB difference between signals.

Table 3: Signal Pairs for Test Sets 3 and 4, each repeated once – Base Case 1.

Signal Pair	Signal A C80 value at 1kHz (dB)	Signal B C80 value at 1 kHz (dB)	C80 Difference between Signal A and Signal B (dB)
1	-3.0	-3.0	0.0 (same signal)
2	-3.0	-1.5	+ 1.5
3	-3.0	0.0	+ 3.0
4	-3.0	2.0	+ 5.0
5	-3.0	3.0	+ 6.0
6	-3.0	4.0	+ 7.0
7	-3.0	6.0	+ 9.0
8	0.0	0.0	0.0 (same signal)

Table 4: Signal Pairs for Test Sets 3 and 4, each repeated once – Base Case 2.

Signal Pair	Signal A C80 value at 1kHz (dB)	Signal B C80 value at 1 kHz (dB)	C80 Difference between Signal A and Signal B (dB)
1	0.0	0.0	0.0 (same signal)
2	0.0	1.5	+ 1.5
3	0.0	3.0	+ 3.0
4	0.0	5.0	+ 5.0
5	0.0	6.0	+ 6.0
6	0.0	7.0	+ 7.0
7	0.0	9.0	+ 9.0
8	3.0	3.0	0.0 (same signal)

4.0 Results and Discussion

Results from the training sets, Sets 1 and 2, were analyzed by evaluating the number of questions correctly answered by each subject to determine how their performance might have changed between the two sets. They hypothesis, based on Study 2a, was that the subjects' performance would include from between Set 1, where they used the "no switch" method, to Set 2, where they used the "with switch" method. Only the ten "test" questions from these sets, part B of each set, were analyzed, as shown in Figure 5. None of the participants received scores of less than 50% during the training sets. However, on average the results showed that the subjects' performance actually got worse between the two sets, with the average result from the first set being 69% correct and for the second was 66% correct. When broken down by subject groups, a total of 11 subjects scores improved between the first set as hypothesized, with the scores increasing from 53% to 76%. A second group of 7 subjects stayed the same between both sets with an average score of 72%, while 10 subjects' performance got worse between the first and second sets with scores starting at 84% and then decreasing to 51%. These results indicate that the effect of test method may not be as clear as concluded in Study 2a, in that starting with Test Method 1 and then using Test Method 2 may not yield the most accurate results for all subjects.

Figure 5: Subject Results to Extended Training

For the actual test tests, Sets 3 and 4, half of the subjects were presented Base Case 1 for both sets (Table 3) and the remaining half were presented Base Case 2 for both sets (Table 4). This test design required the participants to compare the same pairs of signals a total of four times, two times in each set. Unlike in Studies 1 (PSVRF Grant 08.06) and 2a, only one major variable had two conditions, the base case, since the motif and subject type were kept constant.

The results from Base Case 1 are shown in Figure, with the percentage of subjects reporting hearing a difference shown on the y-axis versus the difference in C80 between the presented signals. A linear regression was carried out on this data and the correlation coefficient was found to be 0.89. The percentage of subjects who reported hearing a difference when there wasn't a difference is about 26%, higher than would be expected, but the percentage who heard a difference at the extreme case of 9 dB was over 90 %. The resulting C80 JND from this curve fit is 4.3 dB, however, further work should be done to investigate other curve fits for an improved fit with the data.

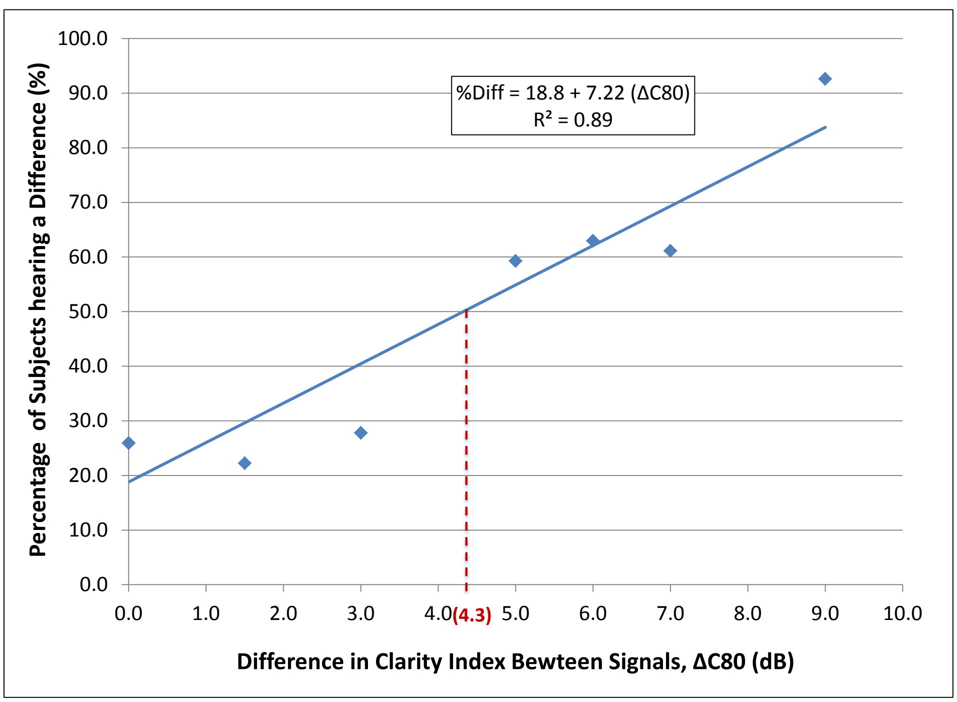

Figure 6: C80 JND Results for Base Case 1, C80 = -3.0 dB at 1 kHz and T30 = 1.9 s

The results for Base Case 2 are shown in Figure 7. A linear regression was also applied to these data and appears to be a better fit, with a correlation coefficient of 0.95. The percentage who reported hearing at 0.0 dB is lower than for Base Case 1 at 20%, and is similar at the high end with about 95% who reported hearing a difference. The resulting C80 JND for this base case is 3.6 dB. When the data are combined over both base cases, the linear regression appears to be a good fit with the data, as shown in Figure 8. The overall C80 JND is 3.9 dB, which is nearly identical to the result found in Study 2a, which was 3.8 dB, and are again much higher than the previously found quantity of approximately 1.0 dB in previous work (1-3).

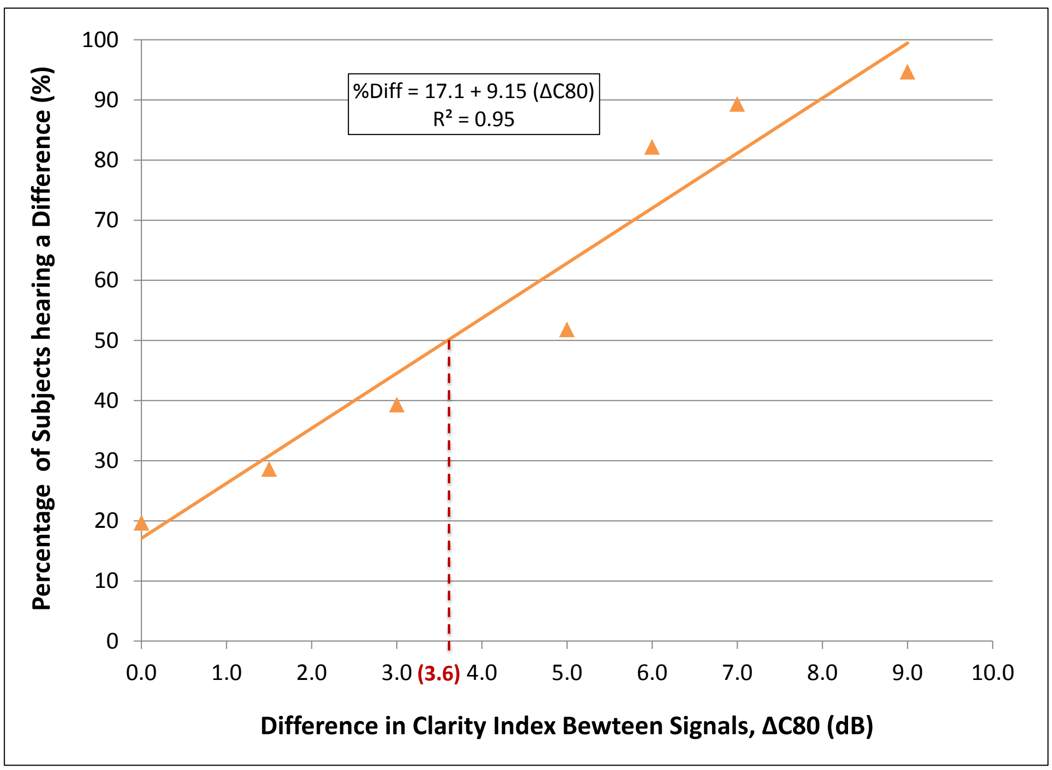

Figure 7: C80 JND Results for Base Case 2, C80 = 0.0 dB at 1 kHz and T30 = 1.6 s

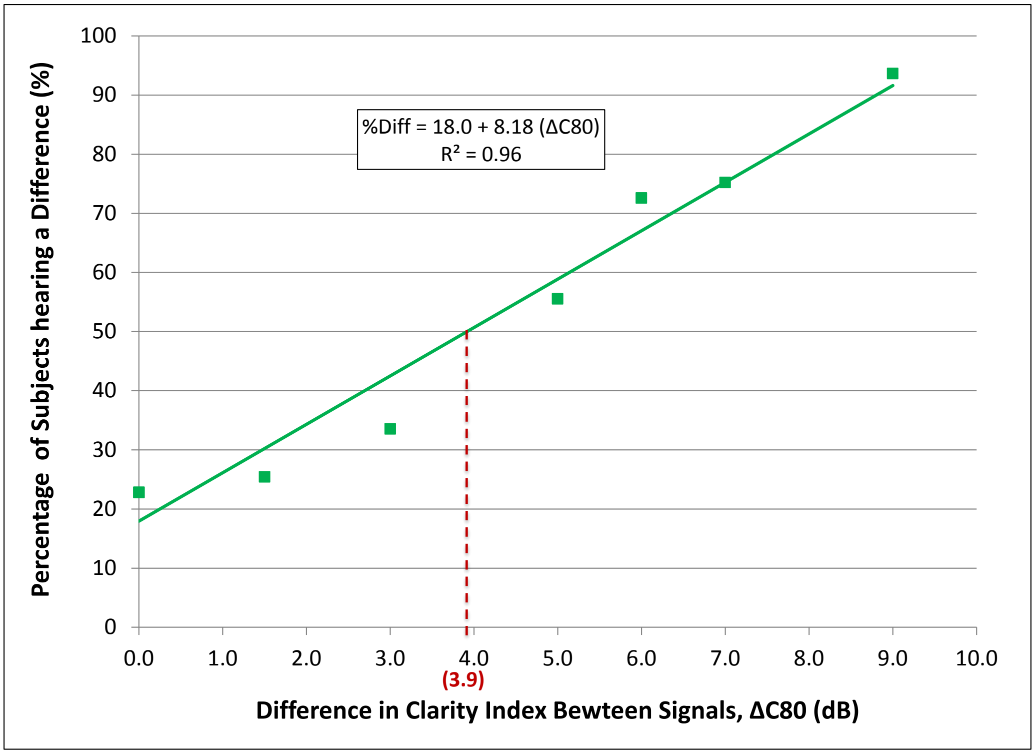

Figure 8: Overall C80 JND Results Averaged Over Both Base Cases

5.0 Conclusions and Recommendations for Future Work

In summary, a total of 28 subjects participated in the C80 JND listening test for this study, Study 2c. Subjects spent the first half of the test becoming familiar with the test and the second half actually completing the test. The training consisted of first requiring subjects to use Test Method 1 (no switch) and then use Test Method 2 (with switch), based on the results from Study 2a that showed the results were most consistent using this method. The subjects used Test Method 2 for the actual test which consisted of eight signal pairs repeated a total of four times. The C80 JND for Base Case 1 (C80 = -3.0 dB at 1 kHz and T30 = 1.9 s) was found to be 4.3 dB, while for Base Case 2 (C80 = 0.0 dB at 1 kHz and T30 = 1.6 s), it was found to be 3.6 dB. The overall C80 JND averaged over both base cases is 3.9 dB, which is in agreement with the results found in Study 2a. One recommendation for future work is to extend the training period using Test Method 1 to increase subject accuracy. In addition, a more complex statistical analysis could be carried out on the data using higher order curves to find an improved fit with the data.

Overall Report Conclusions

Overall Report Conclusions

The overall goal of this comprehensive study was to further investigate the C80 JND building on the work carried out in 2009 (Study 1), since this previous work indicated that the C80 JND was potentially greater than 1.0 dB as had been reported in the literature. An initial pilot study was conducted to determine the effect of testing method on the C80 JND results, Study 2a. In this study two test methods were compared, Test Method 1, where subjects had to listen to each signal in its entirety before answering, coined “no switch”, and Test Method 2, where subjects could switch between signals in real time, coined “with switch”. The results of this study showed an effect of both testing order and test method. In particular, the results that were most like those hypothesized, i.e. a low percentage reporting hearing a difference where there wasn’t one and high percentage reporting hearing a difference at the largest difference in C80 between signals, was as follows: subjects were first exposed to the *no switch* method and then the *with switch* method. These subjects’ results from the *with switch* method were the most consistent with the hypothesis and yielded a C80 JND of 3.8 dB.

The purpose of a second preliminary study, Study 2b, was to examine the effect of motif length on auditory memory in terms of architectural psychoacoustic testing. In this study, two motifs were selected, an orchestral excerpt and a percussion excerpt, and the lengths of these motifs was varied from 5 s to 7 s to 10 s. A total of 25 test subjects were presented pairs of signals with these two motifs at each length where the signals were varied in terms of reverberation time (T30) and subjects were asked if the signals sounded the same or different. For the orchestra piece, there was minimal variation in the number of correct answers for the 0.3 and 0.4 s differences, but for the largest difference of 0.5 s, the shortest motif was found to give the highest number of correct responses. For the percussion piece, there were also minimal variations for the first two differences between the three motif lengths, but for the 0.5 s difference, the longest motif was found to give the highest number of correct responses. As these results show, a number of interaction effects between the variables were found; thus a general conclusion cannot be drawn from these results. Further work with a larger number of differences in signals should be conducted to further investigate the topic of motif length in these types of studies.

Using the results of these two preliminary studies, the purpose of Study 2c was to investigate the C80 JND. A total of 28 test participants were presented a total of eight signal pairs with C80 differences ranging from 0.0 dB to 9.0 dB. The overall results of this final study led to a similar C80 JND to Study 2a of 3.9 dB. These results indicate that the C80 JND is therefore likely much higher than found in all previous studies and suggest that when all other variables are held constant, a change of approximately 4.0 dB in C80 is needed before subjects will be able to detect a difference. Further work will be conducted on the obtained data in terms of a more sophisticated statistical analysis than the simple linear regression method used in this study. Future work should also include a study of the JND of reverberation time since limited data exists to support the currently widely accepted value of 5%.

Acknowledgements

ACKNOWLEDGMENTS

The authors wish to gratefully acknowledge the support of The Paul S. Veneklasen Research Foundation (PSVRF) and in particular, Mr. John Lo Verde, PSVRF President. Continued support for this and past projects has been greatly beneficial to University of Hartford acoustics students. In particular, we are very pleased to have dedicated part of our laboratory facilities in honor of the foundation on October 15, 2010, through a naming ceremony for the *Paul S. Veneklasen Research Foundation Anechoic Chamber*.

References

REFERENCES

1. **Cox, T. J., Davies, W. J. and Lam, Y. M.** The sensitivity of listeners to early sound field changes in auditoriums. *Acustica.* 1993, Vol. 79, 1, pp. 27-41.

2. **Bradley, J. S., Reich, R. and Norcross, S. G.** A just noticeabe difference in C-50 for speech. *Applied Acoustics.* 1999, Vol. 52, 2, pp. 99-108.

3. **Ahearn, M., et al., et al.** *The just noticeable difference in the clarity index for music, C80.*: Engineering Applications Center Report #EAC-2009-11, June 2009.

4. **Beranek, L.** *Concert Halls and Opera Houses, 2nd Edition.* New York : Springer, 2003.

5. **Denon.** *Anechoic Orchestral Music Recording.* [Audio CD]: Osaka Philharmonic Orchestra, Nippon Columbia Co., Ltd., 1988.

6. **Cowan, N.** On short and long auditory stores. *Psychological Bulletin.* 1984, Vol. 96, pp. 341-370.

7. **Kaernbach, C.** The memory of noise. *Experimental Psychology.* 2004, Vol. 51, pp. 240-248.

8. **Zokoll, M. A., Klump, G. M. and Langemann, U.** Auditory short-term memory persistence for tonal signals in a songbird. *J. Acoust. Soc. Am.* 2007, Vol. 121, pp. 2842-2851.

9. **Paule, M. G., Bushnell, P. J., Maurissen, J. P. J., Wenger, G. R., Buccafusco, J. J., Chelonis, J. J. and Ellitott, R.** Symposium Overview: The use of delayed matching-to-sample procedures in studies of short-term memory in animals and humans. *Neurtoxicology and Teratology.* 1998, Vol. 20, pp. 493-502.

10. **Munson, W. A. and Gardner, M. B.** Loudness patterns - a new approach. *J. Acoust. Soc. Am.* 1949, Vol. 22.

11. **Stevens, K.N.** Frequency discrimination for damped waves. *J. Acoust. Soc. Am.* 1951, Vol. 24.

12. **Scheaffer, R. L. and McClave, J. T.** *Probability and Statistics for Engineers, 4th Edition.* Belmont, CA : Duxbury Press, 1995.

13. **Patynen, J., Pulkki, V. and Lokki, T.** Anechoic recording system for symphony orchestra. *Acta Acustica united with Acustica.* 2008, Vol. 94, pp. 856-865.

14. **Wichmann, F. A. and Hill, N. J.** The psychometric function: II. Bootstrap-based confidence intervals and sampling. *Perception & Psychophysics.* 2001, Vol. 63, 8, pp. 1314-1329.