Effects of Harmonicity and Musical Training on the Loudness of Two-tone Complexes

by

Daniel Scott Jochelson

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Master of Engineering in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2001

© Daniel Scott Jochelson, MMI. All rights reserved.

The author hereby grants to MIT permission to reproduce and distribute publicly paper and electronic copies of this thesis document in whole or in part.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
JUL 1 1 2001
LIBRARIES

BARKER

Author .....
Department of Electrical Engineering and Computer Science
May 23, 2001

Certified by .....
Louis D. Braida
Henry Ellis Warren Professor Of Electrical Engineering
Thesis Supervisor

Accepted by .....
Arthur C. Smith
Chairman, Department Committee on Graduate Students

Effects of Harmonicity and Musical Training on the Loudness of Two-tone Complexes

by

Daniel Scott Jochelson

Submitted to the Department of Electrical Engineering and Computer Science
on May 23, 2001, in partial fulfillment of the
requirements for the degree of
Master of Engineering in Electrical Engineering and Computer Science

Abstract

The effects of harmonicity and musical training on the perception of loudness for two-tone complexes was studied by performing human auditory experiments. These experiments included two-interval and four-interval experiments that were performed by six subjects, with three of these subjects designated as musical subjects, and the other three as non-musical subjects. The results of these experiments suggested that harmonicity does not play a major role in loudness perception, but more study is needed before a definitive conclusion is reached. Musical subjects seemed to perceive consonant intervals a little softer than non-musicians for shorter stimuli, but, since the magnitude of this effect was not highly pronounced, more experiments are needed to support this further. The results yielded no other significant differences between musical and non-musical subjects.

Thesis Supervisor: Louis D. Braida
Title: Henry Ellis Warren Professor Of Electrical Engineering

Acknowledgments

I would like to thank my advisor, Professor Louis Braida, for his help and guidance throughout this project. He was always willing to listen and advise, and I enjoyed learning about the field of psychoacoustics from him during the last couple years. I also greatly appreciate the input of Andrew Oxenham as I wrote the thesis paper. I thank my colleagues in the Sensory Communication Lab for all their support and encouragement during the last few months. I greatly appreciate the support of Jeanie Krause and my office-mate Michael "Q" Qin. I have thoroughly enjoyed the long and entertaining conversations we have had over the last few months.

I am indebted to my subjects AW, DB, ST, BE, GY, and PV for their time and patience as they endured long hours in the isolation of a soundproof booth. I also thank subjects DC and CT for helping me to refine the experimental procedure.

Lastly, and most importantly, I would like to thank my friends and family, without whom I could not have survived my studies at MIT. If I were to try to list everyone in this category, this section would be longer than the entire thesis, but there are certainly some people I would like to thank specifically. To my father, Neil, and my brothers David and Adam, I thank you for always being there for me, through good and bad times. To Uncle Doug and Grandma Ruth in Cleveland, I have always enjoyed spending holidays with you all, and your support throughout my time at MIT has been wonderful. To Grandma Lilian in South Africa, I remember the day that you held my acceptance letter from MIT in your hand, and I hold you in my heart as I complete my journey here. To all my MIT friends (too many to list), you all are the most amazing people I have ever met; I feel fortunate to have been surrounded by such talented and kind friends for the last five years. To Zhe, I do not believe that I could have completed this year without you, and I hope to be just as supportive as you look forward to medical school in the future.

In memory of my mother, Karen Ruth Jochelson (1941-1996).

Contents

1 Introduction **13**

1.1 Motivation 13

1.2 Psychoacoustic Principles 14

1.2.1 Cochlear Auditory Processing 14

1.2.2 Critical Bands 15

1.2.3 Loudness 16

1.2.4 Combination Tones 17

1.2.5 Harmonicity 17

1.2.6 Pitch 20

1.3 Thesis Background 21

1.3.1 Loudness Summation 21

1.3.2 The Role of Harmonicity 25

1.4 Project Goals 27

2 Experiments **29**

2.1 Golub's experiment 29

2.1.1 Protocol 29

2.1.2 Stimuli 30

2.1.3 Subjects 31

2.2 Adaptive Loudness Estimation Methods 31

2.2.1 Psychometric functions 32

2.2.2 Levitt up-down procedure 32

2.2.3 Jesteadt up-down procedure 33
2.2.4 Calibration Experiments 34
2.3 Thesis Experiments 35
2.3.1 Protocol 36
2.3.2 Stimuli 37
2.3.3 Experiment Creation 38
2.4 Subjects 38

3 Results **41**
3.1 Calibration Data 41
3.1.1 Variation 41
3.1.2 Bias 45
3.2 Experimental Data 46

4 Discussion **61**
4.1 Presentation: Diotic versus Dichotic 61
4.2 Standards: Consonant versus Dissonant 62
4.3 The Consonant Dip 63
4.3.1 Two-Interval Experiments 63
4.3.2 Four-Interval Experiments 63
4.3.3 Standard Deviations 64
4.4 Musicality 65

5 Resolution with Golub's work **67**

6 Conclusions **71**

A Experiment ESPUD Code **75**

List of Figures

1-1 Spreading of Masking 16

1-2 Growth of Loudness in relation to Critical Band 22

1-3 Case for Comparison between theories of S. S. Stevens and Zwicker 23

1-4 Dynamic Range based upon Spread of Spectrum 24

1-5 Inharmonicity: Unaccounted for by Loudness Summation Models 25

1-6 Loudness versus $\Delta F = F_2 - F_1$ 27

2-1 Consonant and Dissonant Standards for Experiments 30

2-2 Psychometric Function 33

2-3 Levitt Up-Down Procedure 34

2-4 Jesteadt Up-Down Procedure 35

3-1 Calibration Experiments Across All Sessions, Matching 1000-Hz Tone to Fixed 1000-Hz Tone 42

3-2 Calibration Experiments Across All Sessions, Matching 1000-Hz Tone to Fixed 500-Hz Tone 43

3-3 Calibration Experiments Across All Sessions, Matching 500-Hz Tone to Fixed 1000-Hz Tone 44

3-4 Experiment 1. Summary of Two-Interval Experiment Data: Diotic Presentation, Dissonant Standard 48

3-5 Experiment 2. Summary of Two-Interval Experiment Data: Diotic Presentation, Consonant Standard 49

3-6 Experiment 3. Summary of Two-Interval Experiment Data: Dichotic Presentation, Dissonant Standard 50
3-7 Experiment 4. Summary of Two-Interval Experiment Data: Dichotic Presentation, Consonant Standard 51
3-8 Summary: 2-Interval Data, Average Across Subjects for all Stimuli Conditions 52
3-9 Experiment 5. Summary of Four-Interval Experiment Data: Diotic Presentation, Dissonant Standard 53
3-10 Experiment 6. Summary of Four-Interval Experiment Data: Diotic Presentation, Consonant Standard 54
3-11 Experiment 7. Summary of Four-Interval Experiment Data: Dichotic Presentation, Dissonant Standard 55
3-12 Experiment 8. Summary of Four-Interval Experiment Data: Dichotic Presentation, Consonant Standard 56
3-13 Summary: 4-Interval Data, Average Across Subjects for all Stimuli Conditions 57
3-14 Summary of Two-Interval Experiment Data: Standard Deviations 58
3-15 Summary of Four-Interval Experiment Data: Standard Deviations 59

List of Tables

1.1 Examples of Critical Bandwidth 15
1.2 Partials and Overtones 18
1.3 Combination and Difference Tones for Consonant and Dissonant Complexes . 26

2.1 Calibration Tests: Delta/Variance and Bias for Matching 1000-Hz tone to Another 1000-Hz tone 36
2.2 Subject Information 39

3.1 Calibration: Biases Across All Sessions 46

Chapter 1

Introduction

This thesis looks for a relationship between two topics in the hearing sciences, as it looks at the possible connections between loudness and harmonicity. In this chapter, we first discuss motivation behind this project. Next, many important psychoacoustic principles are introduced to ensure that the reader understands the basic ideas needed to comprehend the purpose, ideas, and experiments in this project. Then, background that is specifically vital to the ideas explored in this thesis are presented, including specific work by Howard Golub upon which this project is based. Lastly, the project goals are set forth.

1.1 Motivation

The loudness of complex tones plays an important role in our daily lives. Noise pollution is a very serious problem in today's world, and understanding the human perception of tone complexes allows us to manipulate our sound environment to lessen the effect of noise pollution. Also, understanding the perception of "musical" tone complexes allows composers to create more intriguing works that entertain us. When listening to music, people hear some distinct timbres more easily than others. For instance, brass instrument players can exert much less energy than woodwinds and string players and still be heard fairly easily. In musical terms, some instrument sounds, such as trumpet notes, "cut through the orchestra" better than other instruments. Why this is so is not very well understood, and composers and

conductors, who are constantly concerned with what would be easily heard by the audience, would be very interested in any advances in this area.

The different timbres are differentiated by temporal and spectral characteristics, with temporal characteristics more prevalent toward the beginning of the tone [10]. When a trumpet plays a note, the tone continues to "cut through the orchestra" throughout the tone, even when the attack transient has subsided and the sound reached a steady state spectrum. This suggests that a sound's spectral envelope plays an important role in loudness perception. Because harmonics play an important role in defining an instrument's sound spectrum, understanding loudness as related to harmonicity is very important if we are to ever understand the perceived loudness of musical instruments. This thesis is a step in that direction.

1.2 Psychoacoustic Principles

In order to understand the analysis and ideas in this thesis, a brief introduction to some principles of psychoacoustics is necessary for complex tone stimuli. First, the basic principles of cochlear transduction are presented, and this will lead to an understanding of critical bands and simultaneous masking. Then, the concept of loudness will be presented, followed by a discussion of combination tones and harmonicity, which play vital roles in this research endeavor. Lastly, perception of pitch will be discussed, since pitch ambiguity may play an important role in this experiment.

1.2.1 Cochlear Auditory Processing

The transduction of physical waves to electrical impulses occurs in the cochlea, which is located in the inner ear.[1] Waves in the cochlear fluids are caused by the stapes acting on the oval window of the cochlea, and the region of hair cells excited on the basilar membrane within the cochlea depends on the frequency content of the input signal. In other words, there exists a direct mapping of frequency to location on the basilar membrane. Scientists

[1] For a more detailed discussion of auditory transduction and physiology, see Zwicker/Fastl [20, p.23-28].

Table 1.1: Examples of Critical Bandwidth

Number	Center Frequency (Hz)	Critical Band Width (Hz)	Lower Cut-off Frequency (Hz)	Upper Cut-off Frequency (Hz)
1	50	–	–	100
2	150	100	100	200
3	250	100	200	300
4	350	100	300	400
5	450	110	400	510
9	1000	160	920	1080
10	1170	190	1080	1270
11	1370	210	1270	1480
12	1600	240	1480	1720
24	13500	3500	12000	15500

have studied the characteristics and consequences of this mapping extensively, and many of the concepts of psychoacoustics grow out of this understanding of the auditory physiology.

1.2.2 Critical Bands

A natural question that arises from the frequency-to-location mapping is how the particular locations on the basilar membranes are grouped when processed by the brain. Because humans have limited frequency resolution, it is clear that areas on the basilar membrane correspond to particular frequency *bands*, much like a standard multi-channel filterbank. Each channel of this filterbank is labelled a *critical band*, and the width of these critical bands increases as the frequency increases. Table 1.1 shows a typical example of characteristics of some critical bands. It is important to note that these are not fixed frequency bands, but rather depend on the stimuli. However, the important aspect of this chart to note is the increase in the relative widths of the critical bands.

Masking

The summation of complex auditory stimuli by the peripheral auditory system is not a linear process. Stimuli within critical bands can greatly *mask* each other; that is, they can raise the hearing threshold of the neighboring frequencies. Figure 1-1 shows this pictorially in the

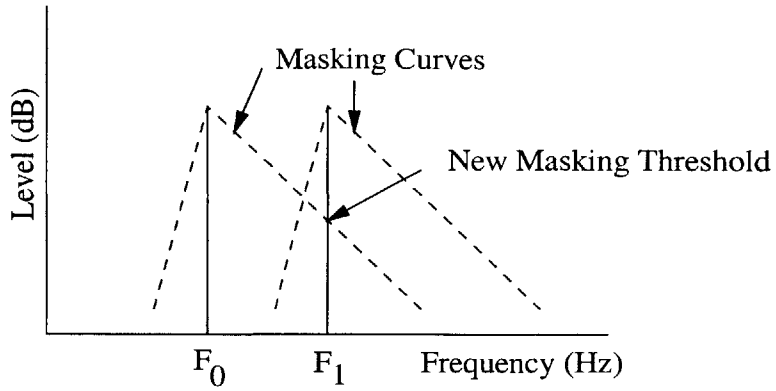


Figure 1-1: Spreading of Masking

frequency domain. For a two-tone complex, each tone has a spread-of-masking curve under which other tones are inaudible. This masking curve is not symmetric about the tone, as the figure shows that higher frequencies are masked more easily by a tone than frequencies lower than that tone. Because the masking curve of F_0 overlaps the tone at F_1, F_0 masks F_1 and, thus, makes F_1 harder to hear (softer).

1.2.3 Loudness

While location on the basilar membrane corresponds to frequency resolution, loudness can roughly be characterized by the number of nerve impulses generated. More specifically, louder sounds cause the fibers of a particular critical band to fire more, while softer sounds hardly increase the "firing count" above its normal (zero-input state) firing count. The details of the nerve firings can be thought of as a Poisson process where the rate parameter is determined by the intensity of the sound. Also, measurement on auditory nerves [8] has shown that there exists a saturation point where the number of nerve impulses cannot exceed. It has also been shown that the pulse can lock onto the phase of a sinusoidal signal for frequencies less than 6 kHz.

Since nerve firings can be studied carefully by electrically probing cat ears, their relation to loudness are understood relatively well. It is the processing by the brain which is a much more difficult area to study, since examining how a human is thinking is a hard task. For

instance, how the brain sums the information from the auditory fibers to perceive loudness must be studied by psychoacoustic experiments. Work by Colburn [1] and Siebert [17] have proposed models for the summation of the counts from all the filterbanks. However, these models do not consider the frequency content in a signal in these loudness models. The brain's analysis of the frequency content of a signal with respect to loudness may be important, as will be seen in the discussion of Golub's work later in this chapter.

1.2.4 Combination Tones

While the cochlea can, to a first approximation, be thought of as a filterbank where the ear simply takes the acoustic input and filters it linearly, there do exist some important nonlinearities. One consequence of such nonlinearity is the generation of *combination tones*, and work done by J.L. Goldstein [3] in the 1960's brought the importance of these tones to the attention of the loudness research community. These tones can be described as tones that are generated by the ear itself in response to the stimuli. Thus, if we consider a two-tone complex, the actual filterbank input is the original stimulus, plus sums and differences of integer multiples of the tone frequencies. Thus, the percept must also take into account these extra tones generated by the ear. The importance of these tones is discussed more later in the discussion of experiments by Howard Golub.

1.2.5 Harmonicity

The sounds in the natural world are not single-tone entities. Because many sounds, like notes played by a musical instrument, involve the excitation of a resonant structure, many tones are produced simultaneously. The tones excited by resonant structures are often characterized by the *overtone series*. This series consists of all integer multiples of the lowest frequency, which is termed the *fundamental frequency*. The frequency spectrum of a sound can then be characterized by its *partials*. The fundamental frequency is also known as the first partial, and the rest of the partials of the sound may be either *harmonic* or *inharmonic*. Harmonic partials are defined as integer multiples of the fundamental frequency, while inharmonic partials are not integer multiples. Since the harmonic partials are often the most prevalent

Table 1.2: Partials and Overtones

Partial Number	Frequency of Partial (Hz)	Name of Nearest Note	Frequency of Nearest Note (Hz)
1	110	A	110
2	220	A	220
3	330	E	330
4	440	A	440
5	550	C#	554
6	660	E	659
7	770	G	784
8	880	A	880
9	990	B	988
10	1100	C#	1109
11	1210	D#	1245
12	1320	E	1318
13	1430	F	1397
14	1540	G	1568
15	1650	G#	1661
16	1760	A	1760

and important, much of the literature often just uses partials to refer to harmonic partials.

Table 1.2[2] summarizes the overtones of the A string on the guitar. The table shows the integer multiples of the fundamental frequency, along with the nearest note for each partial. For the lower partials, the partials align with the nearest note exactly; however, in the higher partials, the nearest note tends to stray farther from the integer multiple.

Consonance and Dissonance

Music composers rely heavily on the properties of the harmonics. All these intervals have some measure of *consonance* and *dissonance*. These two terms lie on opposite ends of a spectrum, where more harmonicity usually means a consonant sound, and inharmonicity yields dissonant sounds. To the average listener, consonant sounds are deemed as more "pleasing and soothing", while dissonant sounds are "harsh and grating" to the ear. The harmonics have their own levels of consonance relative to each other. We now shall define a

[2]Reproduced from Sethares [16, p.22].

couple musical intervals in terms of the harmonics in Table 1.2 and discuss the level of their consonance when these pure tones are played simultaneously. Octaves are defined as partials where

$$f = (2^N) f_0, \tag{1.1}$$

where f is the partial frequency, f_0 is the fundamental frequency, and N is a positive integer. The octave is the most consonant interval, and this means that, when played simultaneously, tones are grouped more easily by the brain than more dissonant tones. An important concept here is the idea of *octave equivalency*. Because all octaves are perceived similarly by the brain, a tone can be transposed up or down an octave easily by a listener. Thus, if the note E were played at 165 Hz instead of 330 Hz, it can usually be resolved and associated with the E at 330 Hz. This becomes important when discussing the musical fifth. A musical fifth can be defined as

$$f = (2^N) 3 f_0, \tag{1.2}$$

where f is the partial frequency, f_0 is the fundamental frequency, and N is defined as an integer not equal to zero. This musical fifth (the note E in Table 1.2) is the second most consonant interval. Octave equivalence becomes important here because, since the note E at 165 Hz is perceived equivalently to the note E at 330 Hz, it is heard as a consonant interval, even though it is not an integer multiple of the fundamental frequency ($f = \frac{3}{2} f_0$). This fact is important in the selection of stimuli for the experiment discussed later in this thesis.

Timbre

As mentioned in the motivation section, harmonics play an important role in defining the timbral characteristics of a sound. The strength of each harmonic defines the spectral envelope, which is an important cue in differentiating timbres. For instance, the clarinet essentially only resonates at its odd harmonics, and the relative strength of these harmonics changes in different registers of the instrument [13, p.60]. Thus, a clarinet playing high notes has a different spectral envelope than a clarinet in the bottom of its register. For string instru-

ments, the amplitude of the nth partial relative to the first is approximately $\frac{1}{n}$, and there can also exist many inharmonic sounds in the spectrum due to aperiodic string scrapings by the bow [13, p.55-56]. It is clear that the amplitude and location of the partials for musical sounds play an important role in defining the sound, and this project aims to look at how timbre and loudness may be related. Timbre is discussed further in the context of work by Miskiewicz and Rakowski [14] later in this chapter.

1.2.6 Pitch

While critical bands and loudness are highly dependent on the peripheral auditory system, man's perception of pitch is mainly a cognitive process that takes place in the mind. In other words, the perception of pitch is essentially the brain's interpretation of a complex sound. The mind groups related frequencies, forming a Gestalt representation of the complex. When subjects are asked to match a single tone to have the same "pitch" as a two-tone complex, they often select a tone near the fundamental frequency of the complex [20]. The fundamental frequency can be defined as the difference of the harmonic components. For instance, in the case of a harmonic complex with tones at 500 Hz and 1000 Hz, the pitch will be matched to a single 500-Hz tone.

The fundamental frequency does not have to be present in order for the brain to extrapolate the same pitch from the complex. For instance, the sounds produced by orchestral chimes contain the second, third, and fourth harmonics without the fundamental frequency, and yet the pitch perceived is equivalent to the fundamental frequency [2, p.59]. This phenomenon, called virtual pitch, emphasizes the Gestalt nature of pitch perception. For example, consider the case of a harmonic tone complex with tones at 1060 Hz and 1590 Hz. The fundamental frequency of this complex would be $(1590 - 1060) = 530$ Hz. No properties of masking excitation on the basilar membrane can explain why a pitch of 530 Hz would be perceived. A 1060-Hz tone's spread of masking curve does not even reach 530 Hz, so the mind must extrapolate 530 Hz from the 1060-Hz and 1590-Hz excitations.

For inharmonic tone complexes, pitch becomes much more ambiguous, as the relationship between tones is much less clear, since they are no longer rational multiples of one another.

For instance, consider a two-tone complex with tones at 1060 Hz and 1108 Hz. The natural choice for fundamental frequency would be $(1108 - 1060) = 48$ Hz, but this would mean that all the harmonics from 48 Hz to 1012 Hz are missing. With all these harmonics missing, the mind's ability to extract virtual pitch greatly diminishes. Thus, for inharmonic tone complexes, there is much pitch ambiguity.

1.3 Thesis Background

Loudness and harmonicity of complex tones have often been researched in separate arenas. Loudness studies have focused on masking principles and critical band summations of complex spectra; on the other hand, most studies in harmonics have involved pitch perception. Since both of these areas are important for this thesis, background information for both subjects are presented here. Section 1.3.1 presents the loudness summation models of S. S. Stevens and Eberhard Zwicker, and it then discusses them in the context of musical instrument sounds in work by Miskiewicz and Rakowski. Section 1.3.2 then looks at the possible role of harmonicity in loudness, focusing on the work done by Howard Golub.

1.3.1 Loudness Summation

Much research has been done on the loudness of complex tones, and many of these investigations have involved broadband signals to simulate noise conditions. In 1957, Zwicker, Flottorp, and Stevens [21] showed that, as the total SPL is kept constant, the loudness of noise is fairly constant when the bandwidth is less than a critical band, but rises as the bandwidth is increased beyond this. Figure 1-2[3] shows their results explicitly. For a 1000-Hz tone, the critical band in this area is a little less than 200 Hz wide (approximately 160 Hz from Table 1.1), and the loudness begins to rise as the separation between tone components in this 4-tone complex increases beyond a critical bandwidth.

These results suggest that critical bands are very important in perceived loudness. Based on experiments like this, loudness models for complex signals have been developed which

[3] Adapted from Zwicker, Flottorp, and Stevens [21]

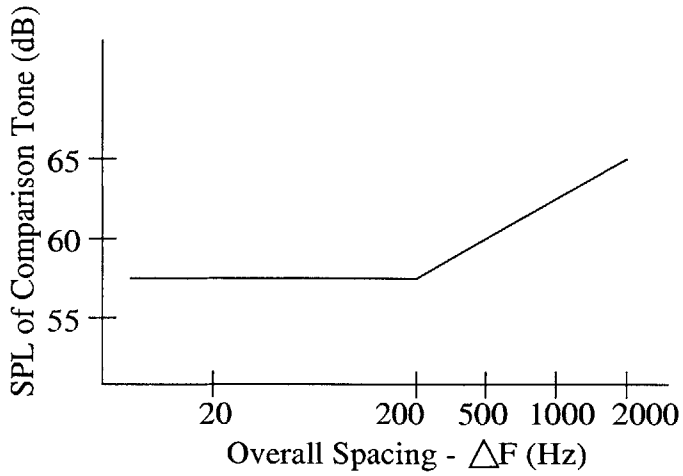


Figure 1-2: Growth of Loudness in relation to Critical Band

incorporate psychoacoustic principles such as masking and auditory critical bands. In particular, S.S. Stevens and Eberhard Zwicker created the most prominent theories of "loudness summation" of these critical bands. It has been found that the perceived loudness is not a simple addition of the independent energies in each auditory critical band, but rather that the channels affect each other through masking.

Stevens's Mark VI and Mark VII Loudness Summation Models

In the 1950's and 1960's, S. S. Stevens conducted loudness experiments at Harvard University in order to develop a theory for loudness of complex sounds. His research was greatly motivated by and oriented toward noise spectra, as opposed to harmonic tone complexes. His theory can be described by his loudness summation equation:

$$S_t = S_m + F(\Sigma S - S_m) \tag{1.3}$$

where S_t is the total perceived loudness, S_m is the loudness of the loudest band (in sones), ΣS is the sum of the loudnesses of the remaining bands, and F is a factor determined by the size of the bands used and the value of S_m [19]. In the previous version of this theory (Mark VI), Stevens made F simply a function of the size of the bands used; more specifically, F had

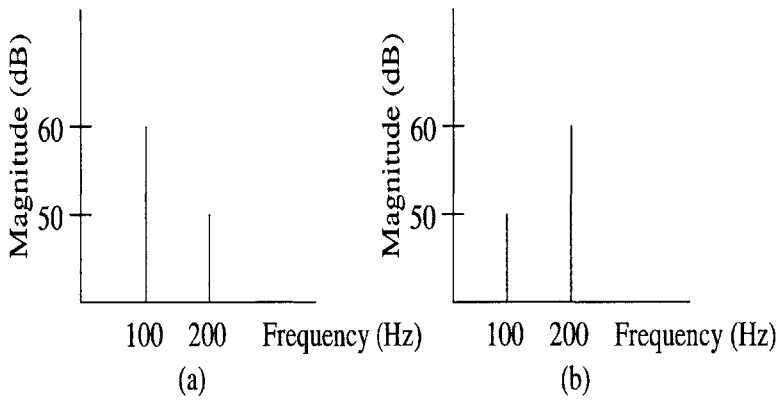


Figure 1-3: Case for Comparison between theories of S. S. Stevens and Zwicker

the values of 0.3, 0.2, and 0.15 for octave, half-octave, and third-octave bands [18]. In Mark VII, Stevens makes F dependent also on the value of S_m because he noticed that energies located outside the most prevalent critical band contribute more to loudness perception at midlevels than at low or high levels (also known as the "midlevel bulge").

Zwicker's Loudness Summation Model

Eberhard Zwicker's theory models the excitation level within each critical band from a particular stimulus. This excitation level is dependent on the energy within the particular critical band, as well the neighboring bands that may be masking it. Then, this loudness level is transformed to a specific loudness for that band. The total loudness can then be determined by integrating the specific loudness over all the critical bands. In contrast to Stevens's theory, Zwicker treats upward masking differently than the masking of lower critical bands, as his excitation level incorporates this asymmetry of masking. Figure 1-3 demonstrates a situation where the predictions of the two theories would be vastly different. As computed by Mark VII, the tone complex in (a) would be the same loudness as in (b); on the other hand, Zwicker's model would predict that (b) would be louder because masking is less below a tone component than above it.

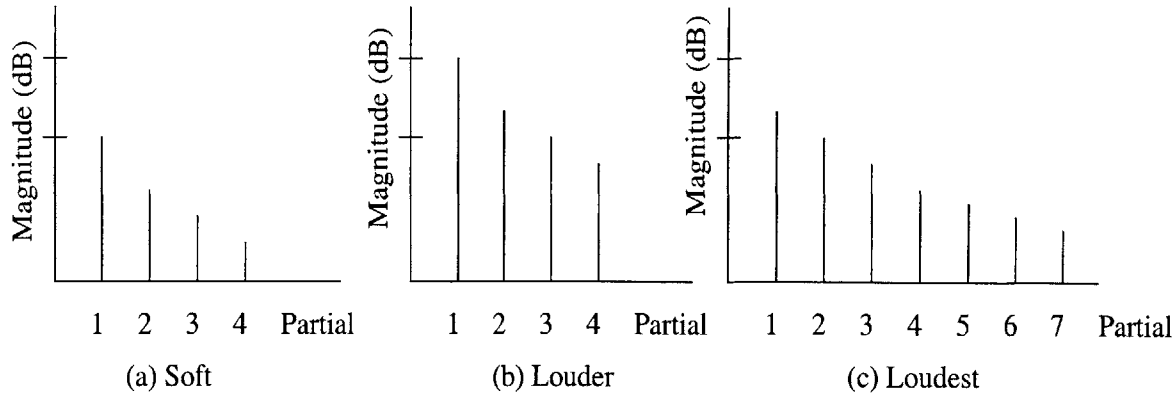


Figure 1-4: Dynamic Range based upon Spread of Spectrum

Miskiewicz and Rakowski - Dynamic range of musical instruments

Using Stevens's Mark VI model and Zwicker's specific loudness model, Miskiewicz and Rakowski [14] calculated the loudnesses of the spectra of several musical instruments. They sampled spectra from brass, woodwinds, and string instruments and then simulated these models to explain the dynamic range (i.e. loudest note divided by softest note) of them. They demonstrated the fact that, for some of the louder instruments, the louder notes utilized more of the upper harmonics than softer notes. In other words, to play louder, the key aspect is the spreading of the spectrum, as opposed to an increase in level of the lower harmonics. This is very common for brass instruments, as well as the clarinet; on the other hand, the spectral envelope of the strings do not change very much as the player tries to increase his volume. Figure 1-4 shows frequency spectra to elucidate this idea. In (a), we have a reference "soft" sound representing the lower end of the dynamic range. In (b), the partials are simply raised by the same amount with no spectral spreading; this is approximately the case for strings and some woodwinds playing louder. In (c), the upper part of the spectrum rises more than the lower part, causing a spreading of the spectrum, and this is approximately the case for brass instruments creating louder notes.

This work by Miskiewicz and Rakowski demonstrates the growth in loudness as the bandwidth of a signal exceeds a critical band, and it highlights the importance of the partials in loudness summation. However, it does not address the issue of the *location* of the partials.

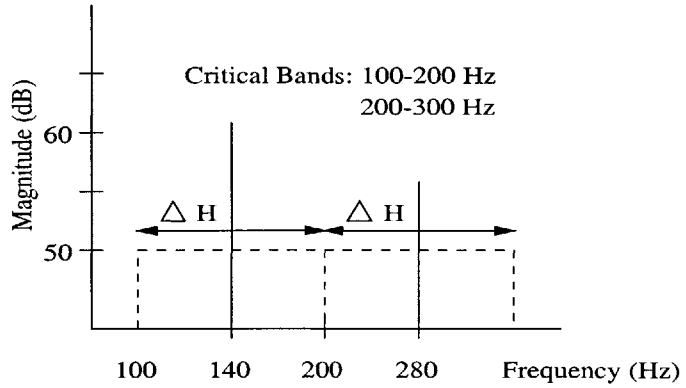


Figure 1-5: Inharmonicity: Unaccounted for by Loudness Summation Models

Inharmonicity does occur in musical sounds, especially in string instruments, and this could be a factor in perceptual loudness.

1.3.2 The Role of Harmonicity

While the theories by Stevens and Zwicker provide adequate models for loudness summation, they fail to take into account the harmonicity of tone complexes. For example, if the two components of a two-tone complex are located in different critical bands, they could be moved anywhere within those two critical bands, and neither model would account for this. More specifically, either tone component in Figure 1-5 could be moved anywhere within ΔH, while keeping the other component fixed, and current loudness models would show no difference in the loudness. Experiments by Howard Golub [4] point out that this could be, in fact, a very important issue.

Golub's Thesis

The work of Howard Golub is the foundation upon which this thesis will be based. In his M.I.T. Masters thesis in 1975, Golub reported his research on the loudness of harmonic and inharmonic tone complexes. His motivation was to find ways to reduce the noisiness of jet engines, and he was curious about why a jet engine should sound louder and more noisy than a symphony orchestra, even though the SPL levels of the two stimuli can be comparable at

Table 1.3: Combination and Difference Tones for Consonant and Dissonant Complexes

	Consonant Complex	Dissonant Complex
Stimuli Tones	f_1,f_2,f_3	f_1,f_2,f_3
Combination Tones	$2f_1 - f_2$, $2f_2 - f_3$	$2f_1 - f_2$, $2f_2 - f_3$
Difference Tones	$f_2 - f_1$, $f_3 - f_2$	$f_2 - f_1$, $f_3 - f_2$
Tone Relationships	$f_2 = 2f_1$, $f_3 = 3f_1$	$f_2 = \frac{B}{A}f_1$, $f_3 = \frac{C}{A}f_1$ where A, B, $C \neq$ integers
Overall Sound (Number of Frequencies Present)	f_1, f_2, f_3	f_1, $(2 - \frac{B}{A})f_1$, $(\frac{B}{A} - 1)f_1$, $\frac{B}{A}f_1$, $(2\frac{B}{A} - \frac{C}{A})f_1$, $(\frac{C}{A})f_1$, $\frac{C}{A}f_1$

times. According to Golub, the main difference between the two stimuli is that sounds from symphonies tend to contain more harmonicity than jet noise. Therefore, Golub decided to conduct experiments on the loudness of two-tone complexes where the criterion was the harmonicity of the complex. His experiment involved adjusting a standard complex until it matched the loudness of a comparison complex. Golub used both a dissonant standard and a consonant standard in separate runs of his experiment.

When he presents his initial theory for these two-tone complexes, Golub explains that more combination and difference tones are detectable for inharmonic complexes presented to the same ear, so this could account for the increase in loudness for inharmonicity. For example, for three-tone complexes, Table 1.3 combination and difference tones for consonant and dissonant complexes. Because the consonant harmonics are integer multiples of one another, the combination and difference tones lie at the same frequencies as those in the input stimulus. Thus, the overall sound perceived by the ear contains still only three tones, where the amplitudes of the tones interact. On the other hand, the dissonant complex forms combination and difference tones that differ from the input stimulus tones. Because of this, there are more tones present in the ear; because the energies of these tones add (when outside a critical band), they should be perceived louder than the consonant case.

Golub's first experiments support this idea, as he presents both tone components to the same ear so that combination and difference tones are created. The data for these experiments show a prominent dip in loudness where $F_2 = \frac{3}{2}F_1$, which is equivalent to the

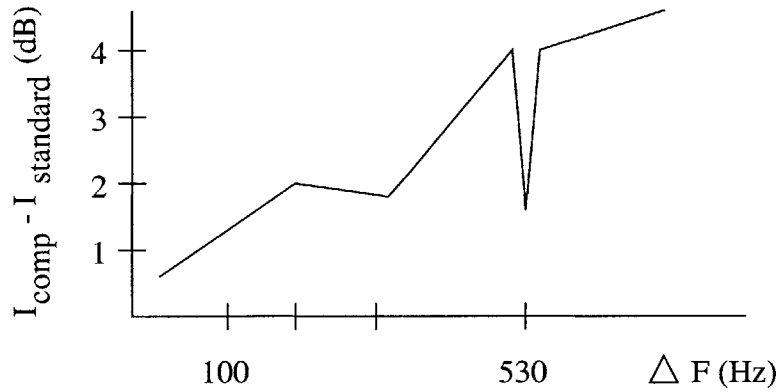


Figure 1-6: Loudness versus $\Delta F = F_2 - F_1$

consonant musical interval of a fifth. This occurs for both the consonant and the dissonant standard. Figure 1-6 shows the loudness as a function of the difference between the two tone components, where the lower tone component is at 1060 Hz. The dip occurs near 530 Hz, which corresponds to the musical fifth interval.

Another fascinating result of the experiment is that, when the two tone components are presented to separate ears, the dip still occurs in some experiments! This suggests that the increase in loudness for inharmonic tones depends also on a cognitive function instead of only an increase in combination and difference tones at the basilar membrane. When a consonant standard is used in this case, the dip becomes much broader, and Golub explains this by saying that, with the consonant standard presented dichotically, the comparison of pitch ambiguity is the overriding issue. With the consonant standard's minimal pitch ambiguity, the subject is cognitively comparing pitch ambiguity instead of purely loudness.

The most important finding by this thesis is the "consonant dip" that occurs for two-tone complexes, and the work of this thesis focuses on this area.

1.4 Project Goals

This thesis compares the loudness of harmonic two-tone complexes versus inharmonic two-tone complexes. It aims to provide at least partial answers to the following questions:

1. Do inharmonic tones sound louder than harmonic tones?

2. Do musicians estimate the loudness of tone complexes differently than non-musicians?

These questions are explored through human auditory experiments, which are described in detail later in this paper.

The remainder of this paper will describe the details of this project. First, chapter two will describe Golub's procedures and the experiments of this project. Then, chapter three will present the results, and chapter four will discuss the implications of this data. Lastly, chapter five will present the conclusions and suggest future work.

Chapter 2

Experiments

This chapter introduces the project experiments and explains the reasoning behind the methods employed for this study. We first describe Golub's experiment in detail, since we are investigating the reproducability of his data with more currect methods. We then present newer methods for estimating equal loudness. Next, the experimental protocol for this project is described; lastly, information on the experimental subjects is presented.

2.1 Golub's experiment

In order to understand the basis of the experiments for this thesis, it is important to understand the details of Golub's experiments. The information in this section is based on Golub's thesis [4], as well as personal communication with Howard Golub [5].

2.1.1 Protocol

Golub's experiment consisted of a loudness matching paradigm. In other words, the subject was asked to adjust a "comparison" tone complex until it was as loud as the "standard" tone complex. When the subject felt that the two complexes were equally loud, he would depress a button that would end the trial and move on to the next "comparison" tone complex. Doing this for a set of comparison complexes traces out a loudness curve in terms of the standard tone complex. In other words, it shows the level of the comparison complex *relative to* the

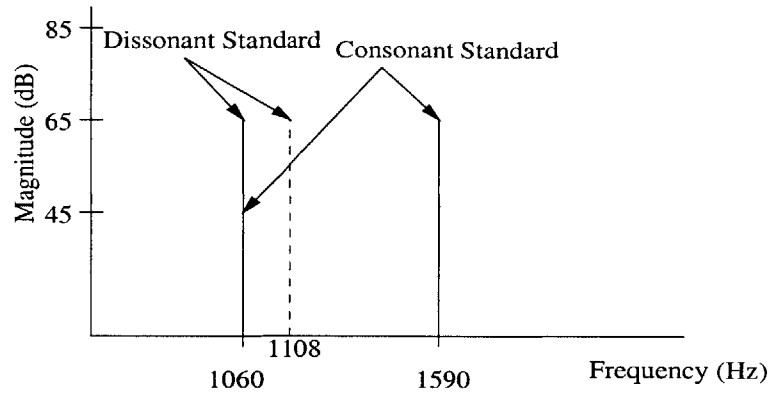


Figure 2-1: Consonant and Dissonant Standards for Experiments

standard complex.

In order to give the subjects as much time as they needed to compare the loudnesses of the tone complexes, Golub allowed the subjects to freely alternate between the comparison and standard complexes. He also allowed them to control the length of the stimuli. These two traits could be important aspects in the reproducibility of his experimental data. Because both the length of the stimuli and the number of presentations are highly variable, it is very difficult to redo his experiment exactly. Thus, because there are more uncontrolled variables in the stimuli, his conclusions may not be as strongly tied to consonance and dissonance as he expected. These variables are removed in the experiments for this thesis, which are discussed later in this chapter.

2.1.2 Stimuli

The stimuli of Golub's thesis were two-tone complexes, where F_1 is the lower tone component, and F_2 is the upper component. For the comparison sound, F_1 was held constant at 1060 Hz, and F_2 was tested at eight different values: 1108 Hz, 1225 Hz, 1265 Hz, 1333 Hz, 1540 Hz, 1590 Hz, 1650 Hz, and 1800 Hz. The most important of these points was at $F_2 = 1590$ Hz, since, for this set of frequencies, this point yields the most consonant interval with 1060 Hz (1590 Hz $= \frac{3}{2}$1060 Hz).

Two different standards were used in his experiments: a consonant standard and a dis-

sonant standard. These standards are displayed in Figure 2-1. The consonant standard had an F_1 of 1060 Hz and an F_2 of 1590 Hz; the dissonant standard had an F_1 of 1060 Hz and an F_2 of 1108 Hz. The frequency of F_2 for the consonant standard was selected so that it lies outside the critical band for 1060 Hz, but not too far away to dramatically increase the loudness. For the dissonant standard, Golub selected an F_2 of 1108 Hz, which is one fourth of a critical band away from 1060 Hz, based on work by Plomp and Smoorenberg [15] that defines maximum dissonance as a function of critcal bands. Both components in each standard were fixed at 65 dB SPL, while the comparison complex level is adjusted by the subject.

Also, the stimuli were presented diotically, as well as dichotically. In other words, in the diotic case, both tone components were presented to both ears; in the dichotic case, only one tone component was presented to each ear.

2.1.3 Subjects

Golub's subjects were essentially friends and collegues from lab, and he was unaware of their musical background. He tested about eighteen subjects once, and sixteen of these exhibited the consonant dip. Golub also tested four subjects 9-11 times for the diotic conditions, and he tested three subjects 9-10 times for the dichotic conditions. Two subjects completed all four test conditions (diotic presentation, dissonant standard; diotic presentation, consonant standard; dichotic presentation, dissonant standard; and dichotic presentation, consonant standard). However, it is unclear whether three completed all four tests. In any case, Golub's project does not have a larger number of subjects completing all tests multiple times. This aspect is improved in this project.

2.2 Adaptive Loudness Estimation Methods

Golub's experiments ask difficult, subjective questions of his subjects when he requires them to equate the loudness of the comparison to that of the standard. To remedy this situation, adaptive methods allow the subject answer an easier question: Which sound was louder? This

allows the subject to hold his criterion for decision-making more stable as the experiment is run. We first introduce the concept of psychometric functions to facilitate the understanding of these adaptive procedures. Then, the two procedures considered for this experiment, the Levitt up-down procedure and the Jesteadt up-down procedure, are introduced in the context of loudness matching. Lastly, the reasoning behind the choice of experimental procedure is developed.

2.2.1 Psychometric functions

Psychometric functions are functions that relate responses on a binary task to a physical stimulus attribute, such as level. If the experiment were deciding when Sound 1 is louder than Sound 2, then the psychometric function simply maps the percentage of "Sound 1" choices to the necessary level to attain this percentage. Figure 2-2[1] shows a typical psychometric function. If an experiment aims for 50 percent positive responses, then this maps to a level at X_{50}, which is the point of subjective equality [9]. If the experiment aims for the 75-percent point, then this gives us X_{75}, which is defined as the difference limen or just-noticable difference (JND) level. Each of these points are important in the two procedures considered for this experiment.

2.2.2 Levitt up-down procedure

The Levitt up-down procedure [9] is an adaptive procedure that alters the level up or down based on the subject's previous responses. Figure 2-3 shows how one sound (Sound B) is varied to obtain a loudness level equal to that of the other sound (Sound A). In this example, Sound B starts above Sound A, but it can also start below it. The subject is instructed to select what he perceives to be the louder sound in each trial. If the subject selects Sound B on a trial, then the level of Sound B on the next trial will be lowered by a certain step size. If Sound B is selected again, its sound level on the next trial will be lowered again, but by a smaller step size. If Sound A is selected, the level of Sound B will be raised. Continuing in this manner will trace out an "up-down sequence" for Sound B, and each change in direction

[1]Reproduced from Levitt[9].

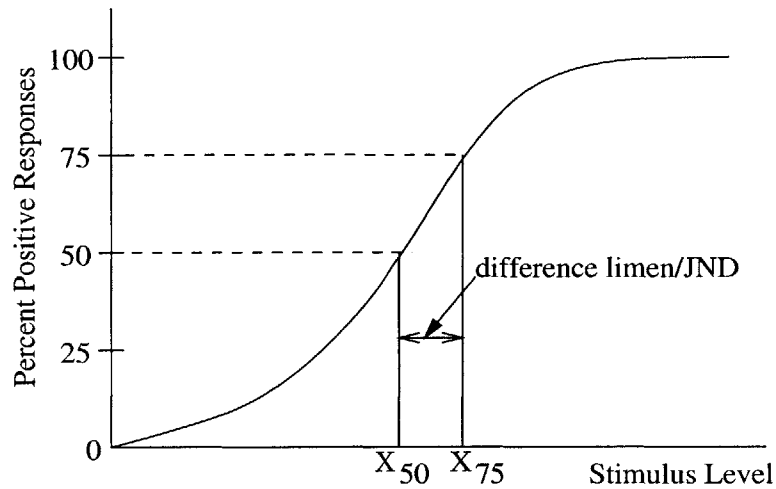


Figure 2-2: Psychometric Function

for Sound B is labelled a "turnpoint". The equal loudness point (X_{50} on the psychometric function, since Sound B is selected about 50% of the time) is then obtained by averaging all the turnpoints in the sequence. Two Levitt sequences (one starting below and one starting above) can be interleaved to gain more data and discount the starting point of Sound B as being an important issue.

2.2.3 Jesteadt up-down procedure

Another procedure that can be used to estimate the loudness levels for the comparison tones is the Jesteadt up-down procedure [7]. This procedure essentially consists of interleaving two Levitt up-down sequences in the same experimental run. By this interleaving, it makes it difficult for the subject to guess the form of the experiment, so he focuses only on his loudness perceptions.

Figure 2-4 displays the details of the Jesteadt procedure. Each Levitt sequence here is estimating the difference limen/JND by changing the rules for level changes compared to the X_{50} Levitt procedure described in the previous section. One of the up-down sequences starts above the loudness of Sound A and aims for the subject to respond "louder" 71% of the time. If the subject states that this sound (Sound B) is louder than the standard sound twice in a row, then Sound B's level is reduced. If the subject responds that it is softer, the

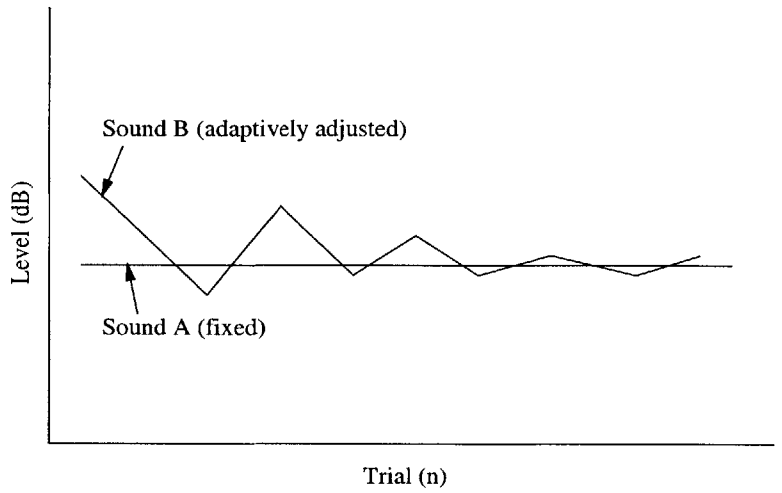


Figure 2-3: Levitt Up-Down Procedure

level of the Sound B is increased. The other up-down sequence begins below Sound A and aims for the subject to respond "softer" 71% of the time. If the subject responds that this sound (Sound C) is softer than the standard twice in a row, the level of Sound C is increased. If the subject says that Sound C is louder once, its level is reduced. Upon completion of these interleaved up-down sequences, the two estimates are then averaged to determine a final loudness estimate. It is important to note that even though Sounds B and C converge symmetrically around Sound A in Figure 2-4, this is not a necessary condition, especially when we are comparing vastly different stimuli.

2.2.4 Calibration Experiments

To determine which up-down experiment to use, all the subjects first participated in a calibration session where both up-down experiments were performed with single-tone stimuli. More specifcally, the subjects matched a 1000-Hz tone to another 1000-Hz tone, a 1000-Hz tone to a 500-Hz tone, and a 500-Hz tone to a 1000-Hz tone. For all these experiments, a statistically "normal" subject would match these tones at equal SL levels on the equal loudness contours, and the purpose of these calibration routines was to see if the up-down methods produced approximately the same level estimates. Table 2.1 summarizes the results of these tests. Comparing the average deltas, where delta is the difference between two

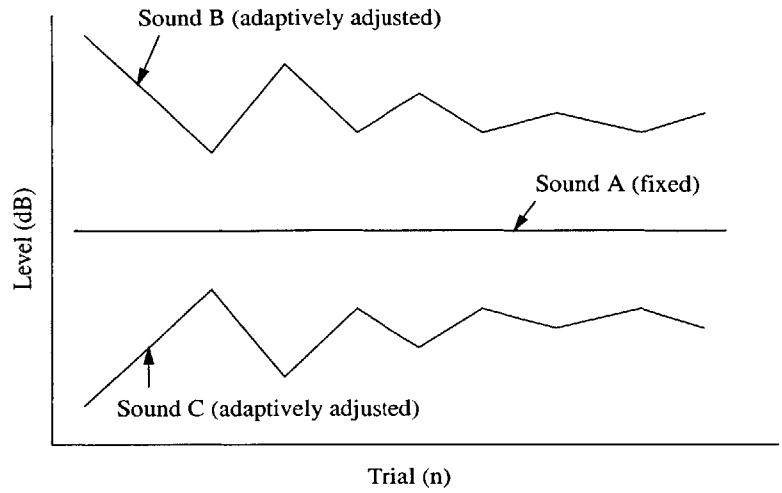


Figure 2-4: Jesteadt Up-Down Procedure

different trials (essentially like a standard deviation) for the 1000-Hz to 1000-Hz comparison experiment, we see that there is only about 0.2 dB difference between the two procedures. The biases, i.e. the differences between the perceived equal loudness and the equal SPL level, only differed by about 0.2 dB as well. Therefore, these two procedures obtain fairly equal estimates of loudness. The Levitt procedure has an advantage over the Jesteadt procedure, however, because it takes much less time to complete, even when two Levitt sequences are used. Thus, the subject encounters less fatigue when completing the experiments. Because of these time savings, the X_{50} Levitt procedure was selected for all the experiments in this project.

2.3 Thesis Experiments

The experiments took place in the laboratory of the Sensory Communication group in the Research Laboratory of Electronics, from January until April 2001. This section describes the protocol, the stimuli, and the details of creating the experiment.

Table 2.1: Calibration Tests: Delta/Variance and Bias for Matching 1000-Hz tone to Another 1000-Hz tone

Subject	Average Delta (Jesteadt)	Average Delta (Levitt)	Average Bias (Jesteadt)	Average Bias (Levitt)
AW	0.1553	0.8255	0.8024	0.4891
DB	0.3794	0.6440	0.9659	1.6944
ST	0.5095	0.2143	1.5300	1.2763
BE	0.2072	0.8974	0.8419	0.0289
GY	1.7651	0.3563	-0.2482	-0.6121
PV	1.1933	0.0964	2.9798	2.9820
Average	0.7016	0.5056	1.1453	0.9766

2.3.1 Protocol

The experimental procedures were similar to Golub's experiments (since we are trying to replicate his data), but there are a few important differences. First, using the two alternative, forced-choice, Levitt procedure means that the subject has no explicit control over the stimuli. He is forced to simply choose which sound is louder based on his perception. Also, the stimuli are only presented once; the subject is not able to freely switch back and forth between the stimuli and then make a decision. While this may seem more difficult for the subject, this way of testing is much more like real life. In the natural world, people are not able to rewind sounds and play them again to decipher which one was louder. They hear the sounds and then make a judgement. This emphasizes that this project's results may be more accurate and useful than Golub's experiments. Lastly, in this project, the "standard" sound's level is adjusted, and the "comparison" complex for which we are getting a loudness level is held constant. This difference should not present a problem, since the overall loudness contours should not be affected. Essentially, estimating the loudness of a fixed Sound B in terms of Sound A should yield about the same relative results as comparing the loudness of Sound B to a fixed Sound A.

2.3.2 Stimuli

The stimuli included all four of Golub's tests. The duration of the tone complexes, as well as the time between sounds, was fixed at 250 msec; these times were uncontrolled in Golub's project. For the dichotic case, F_1 was presented to the left ear, and F_2 presented to the right ear. Feedback (i.e. telling the subject if they are correct or not) could not be used since there is no "correct" answer in these subjective tests.

The stimuli for the four experiment conditions can be summarized as follows:

- Experiment 1: diotic presentation, dissonant standard, 8 ΔF's
- Experiment 2: diotic presentation, consonant standard, 8 ΔF's
- Experiment 3: dichotic presentation, dissonant standard, 8 ΔF's
- Experiment 4: dichotic presentation, consonant standard, 8 ΔF's

Each experiment was run 3 times, non-consecutively, so that all four experiments were performed before any experiments were repeated. The experiments were separated into 4 to 6 session of 2 hours each. In each trial, the fixed comparison tone complex was always presented first, followed by the moving standard complex. The subject was instructed on a computer monitor to select the interval he perceived as the louder interval. They were told that guessing was allowed (since most of the level differences were less than a JND), as long as their perceptions guided their guesses. In the two interleaved Levitt procedures, the variable that is changed on the moving standard is the level of both tones in the standard complex. For instance, for the consonant standard, the level of the 1060-Hz component and the 1590-Hz component are always equal and moved together.

In addition, to simulate Golub's experiment more closely to see the effects, all the conditions were tested in a 4-interval experiment, with each interval 500 msec in length and 50 msec between intervals. The subject was asked to decipher whether Interval 1 and 3 were louder, or Interval 2 and 4. This was done to see if allowing more time by switching back and forth between the stimuli made a difference in the consonant dip.

The stimuli for these experiment conditions can be summarized as follows:

- Experiment 5: 4 intervals, diotic presentation, dissonant standard, 3 ΔF's
- Experiment 6: 4 intervals, diotic presentation, consonant standard, 3 ΔF's
- Experiment 7: 4 intervals, dichotic presentation, dissonant standard, 3 ΔF's
- Experiment 8: 4 intervals, dichotic presentation, consonant standard, 3 ΔF's

2.3.3 Experiment Creation

The stimuli were created using an Ariel DSP-96 DSP board, with Motorola DSP96002 DSP chip with 16-bit precision, and they were presented the stimuli via Telephonics TDH-39P (296D000-2, serial #063026) headphones, whose frequency response is fairly flat from 50 Hz to 4 kHz [12, p.30].

The stimuli and experiments were designed in a LISP-like interpreter named ESPUD[2]. This interpreter allows the creation of tones and band-limited noise in controlled time frames, and all programs for this interpreter were written in Scheme. Appendix A shows some of the vital code in the creation of these experiments.

2.4 Subjects

Six subjects were used in these experiments. All of these subjects reported having normal hearing. They were undergraduate or graduate students at MIT, and they were paid for their time. Each subject went through a calibration procedure at the beginning of each session to establish subject characteristics, as well as look for consistency among the sessions.

Because not all of Golub's subjects performed all of his experiments, his conclusions are not as strong as they could be. For this thesis, six subjects performed all the experiments. Also, in order to investigate how musical background affects auditory perception, half of the subjects were musicians, while the other half were people with little musical training. Table 2.2 summarizes this background information for the six subjects. The table shows that subjects BE, GY, and PV can be classified as non-musicians because they have performed

[2]For more information on ESPUD, please see Lum [11].

Table 2.2: Subject Information

Subject	Music Performance Experience	Music Composition Experience	Music History Experience	Music Preferences
AW	violin	study in school	brief introduction to most musical eras	Popular, Rock and Roll, Classical
DB	euphonium	some study	none	Classical, Popular
ST	trumpet, tuba, string bass, percussion	study in school and composes actively	extensive study of musical eras	Popular, Rock and Roll, Classical
BE	none	none	none	Instrumental, Popular, Some Classical, New Age
GY	none	none	none	Rock and Roll
PV	very little	none	none	Popular, Rock and Roll

musically very little in their lifetimes and have not studied music composition and history. On the other hand, subjects AW, DB, and ST are the musical subjects in this study, since they are actively performing and studied music some in the past. Musical preferences are also presented here to see if what subjects normally listen to affects their perception of consonance and dissonance. From the table, it is clear that the non-musical subjects hardly listen to classical music, while the musicians listen to it regularly. Otherwise, all the subjects listen to popular music and rock and roll. This background information may allow us to see if musicality affects the perception of consonance and dissonance.

Chapter 3

Results

This chapter displays the data from the experiments of this project. First, the data from the single-tone loudness comparisons are presented to get an idea of the natural biases of the subjects. Then, the data from all the experiments are presented, and these data will be analyzed in detail in the discussion of the next chapter. Lastly, the standard deviations for the data are presented, for this information is important in Golub's conclusions. In all graphs, the musicians and non-musicians are plotted on separate graphs to make analysis easier and clearer.

3.1 Calibration Data

3.1.1 Variation

Figure 3-1 through Figure 3-3 show the calibration data for all subjects across all the sessions. Looking at Figure 3-1, we see that all the subjects match the 1000-Hz tone to another 1000-Hz tone without much volatility across sessions, as each subject varies by at most 1.5 dB. In Figure 3-2, where the subject matches a 1000-Hz tone to a 500-Hz fixed tone, we notice much more variability across the sessions, as most subjects vary by about 3 to 4 dB. Also, the musical subjects seem to vary more across sessions, and we shall look at data variance in the musical context more closely later in this paper. For Figure 3-3, the volatility is fairly equal for the two groups, as each subject varies by 2 to 3 dB.

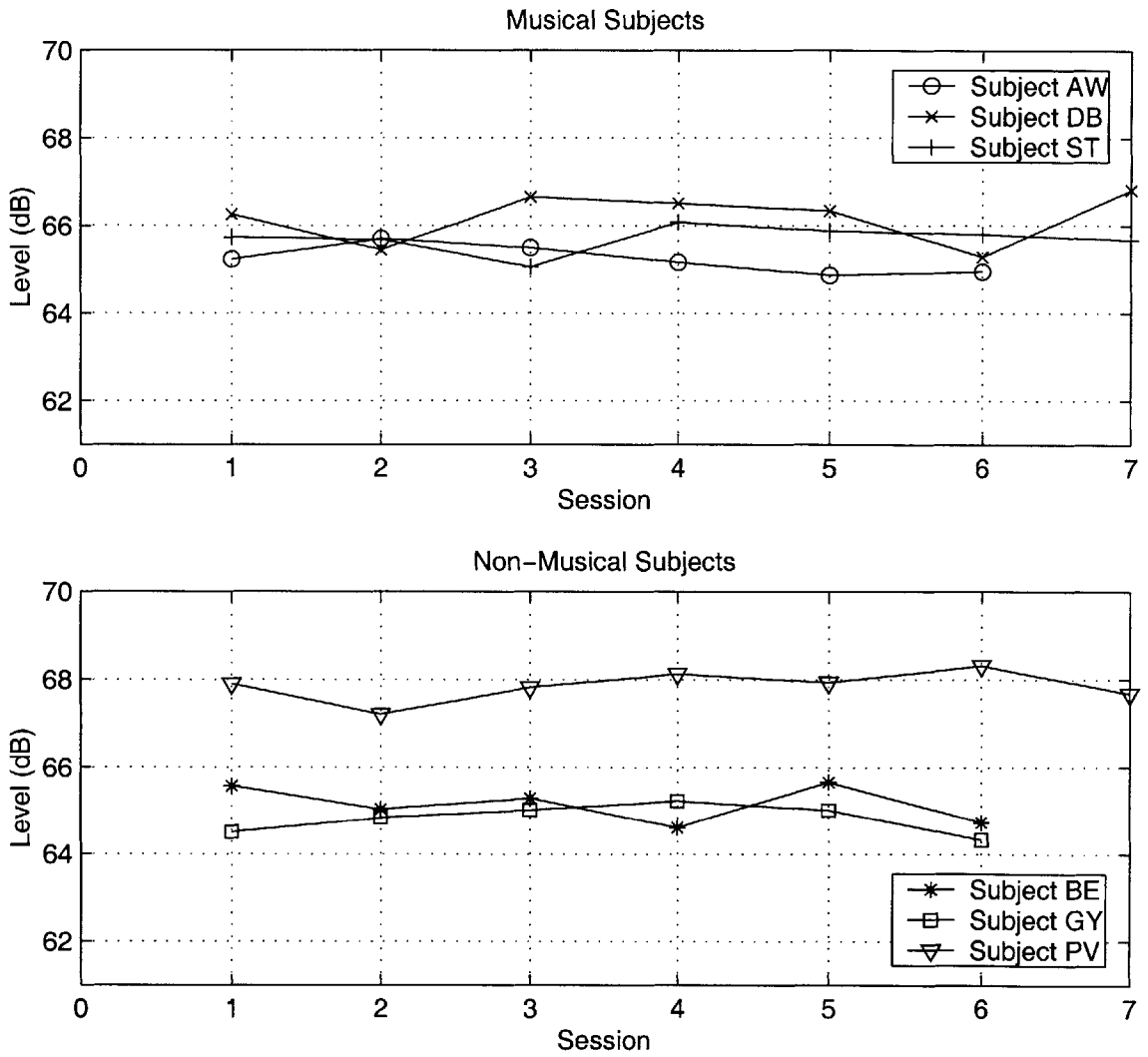


Figure 3-1: Calibration Experiments Across All Sessions, Matching 1000-Hz Tone to Fixed 1000-Hz Tone

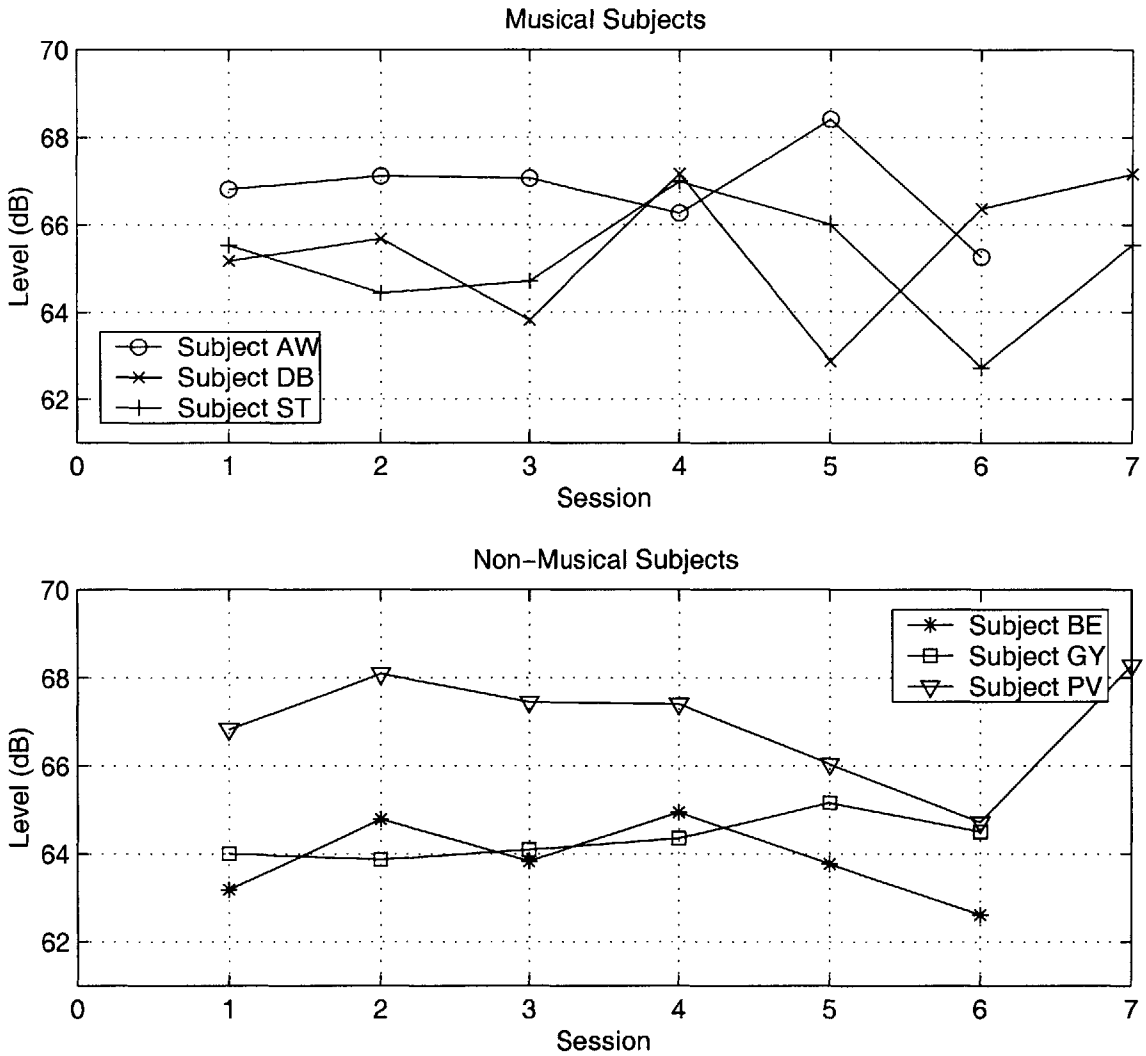


Figure 3-2: Calibration Experiments Across All Sessions, Matching 1000-Hz Tone to Fixed 500-Hz Tone

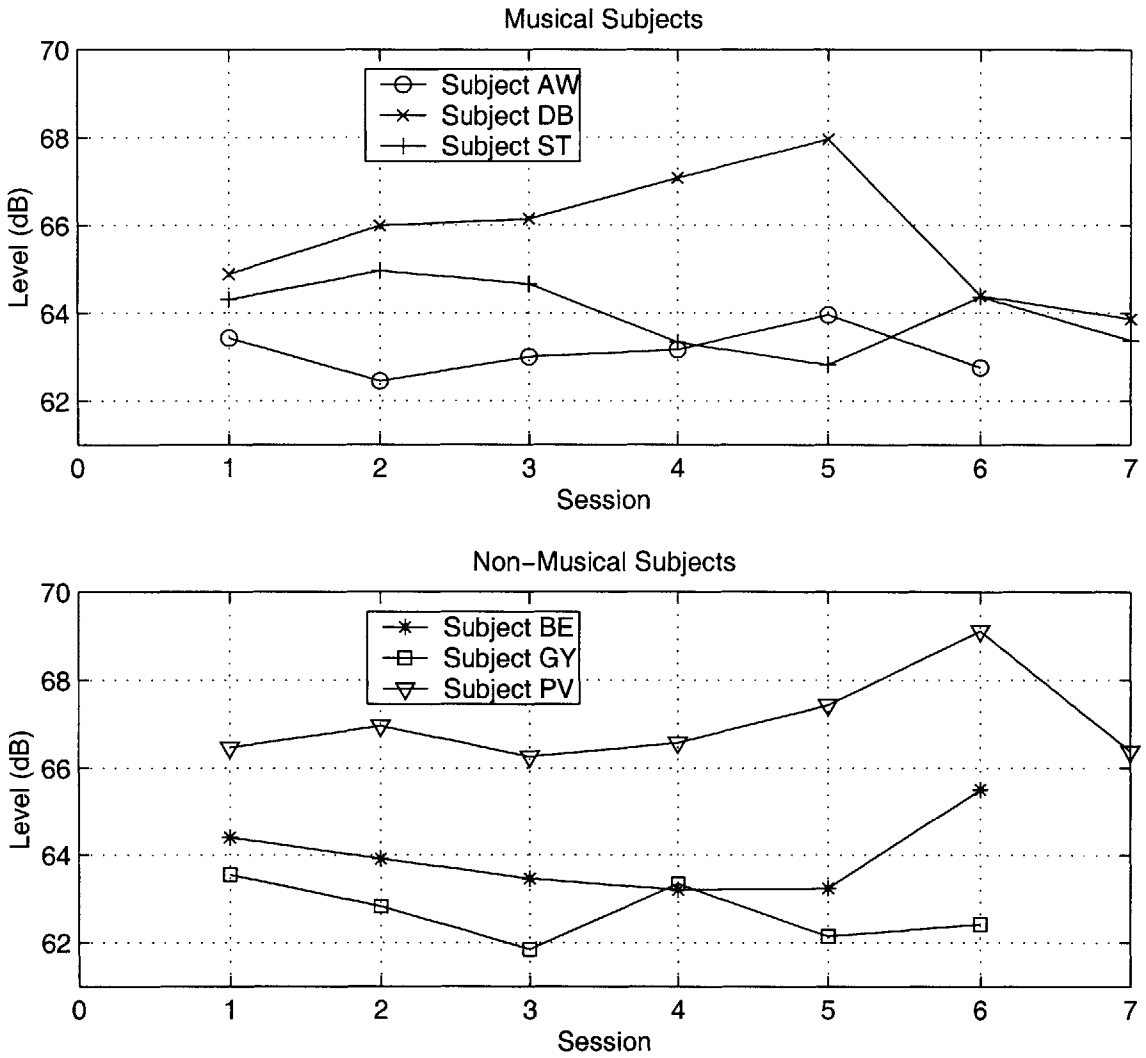


Figure 3-3: Calibration Experiments Across All Sessions, Matching 500-Hz Tone to Fixed 1000-Hz Tone

3.1.2 Bias

The biases, i.e. the difference between the perceived equiloudness level and the true equal SPL level, of the subjects is also an important aspect to extract from this calibration data. Table 3.1 summarizes the average biases across sessions for each subject. From this table, we can characterize the subject in two different ways: their estimated loudness difference and their estimated bias due to presentation order. Estimated loudness difference shows by how much the equal SL level for the subject differs from the equal SPL level; in terms of the 1000-Hz tone to the 500-Hz tone, this can be defined from the table as

$$LoudnessDifference = \frac{(Calib_2) - (Calib_3)}{2}. \tag{3.1}$$

Thus, a loudness difference of +1 dB would mean that the 1000-Hz tone needs to be 1 dB SPL louder than the 500-Hz tone in order to be heard equally loud by the subject; conversely, a -1 dB loudness difference would suggest that the 1000-Hz tone would need to be 1 dB SPL lower than the 1000-Hz tone.

The presentation order bias looks at whether the subject responds more readily to a particular interval based on when it is presented, and this can then be defined as

$$PresentationOrderBias = \frac{(Calib_2) + (Calib_3)}{2}. \tag{3.2}$$

A presentation order bias of +1 dB implies that the subject hears Interval 1 more readily than Interval 2 (since the fixed tone is always presented before the matching tone), while -1 dB suggests a bias toward Interval 2. Looking at the subjects in Table 3.1, we can characterize each subject. Subject AW hears the lower (500-Hz) tone more easily, since his estimated loudness difference is about +1.8 dB, and none of the other subjects exhibit this kind of bias. His presentation order bias is approximately 0 dB. Subject DB exhibits a very small estimated loudness difference and about 0.5 dB of presentation order bias. Subject ST exhibits about +0.5 dB of estimated loudness difference and about -0.4 dB of presentation bias, and both of these values are relatively small. Subject PV has a very small loudness difference, but has a presentation order bias of +2 dB, showing that he is greatly biased

Table 3.1: Calibration: Biases Across All Sessions

Subject	$Calib_1$ Average Bias (Match 1000-Hz to 1000-Hz)	$Calib_2$ Average Bias (Match 1000-Hz to 500-Hz)	$Calib_3$ Average Bias (Match 500-Hz to 1000-Hz)
AW	+0.2373 dB	+1.8211 dB	-1.8715 dB
DB	+1.1858 dB	+0.4578 dB	+0.7592 dB
ST	+0.6999 dB	+0.1306 dB	-1.0245 dB
BE	+0.1480 dB	-1.1476 dB	-1.0460 dB
GY	-0.1789 dB	-0.6726 dB	-2.3138 dB
PV	+2.8561 dB	+1.9707 dB	+2.0279 dB

to the first interval relative to the second interval. Subject BE displays a small loudness difference, while exhibiting -1 dB of presentation order bias. Subject GY has a loudness difference of +0.8 dB and a presentation order bias of -1.5 dB. This shows that Subjects BE and GY are biased to hear the second interval more readily than the first interval. These biases do not suggest any strong differences between musical subjects (Subjects AW, DB, ST) and non-musical subjects (Subjects BE, GY, PV).

3.2 Experimental Data

Figure 3-4 through Figure 3-12 show the main data of this project, to be discussed in detail in the next chapter.

Figure 3-4 through Figure 3-7 display the results of the 2-interval paradigm, described as Experiments 1-4 in the previous chapter. Each graph traces out the equiloudness curves as a function of the frequency of F_2, where the consonant interval is located at 1590 Hz. Figure 3-8 shows the averages across subjects for all the stimuli conditions in the 2-interval experiments.

These show the average equiloudness levels as the average of 3 runs per data point; thus, each data point is the average of 6 up-down sequences (since each trial interleaves two up-down sequences), where each up-down sequence is the average of 8 turnpoints.

Figure 3-9 through Figure 3-12 show the results of the 4-interval paradigm, described as

Experiments 5-8 in the previous chapter. Figure 3-13 shows the averages across subjects for all the stimuli conditions in the 4-interval experiments. Because these stimuli are longer, only the 3 points near the consonant dip were tested. This data is also the average of 3 trials per point, but each up-down sequence has 6 turnpoints in order to keep each run from taking too much time.

Lastly, the standard deviations for all these graphs are summarized in Figure 3-14 and Figure 3-15, since we will look at how this variation changes near the consonant dip, as well as variation differences based on the subject's musicality. In these graphs, the standard deviations of musical and non-musical subjects were averaged to look for these trends in musicality.

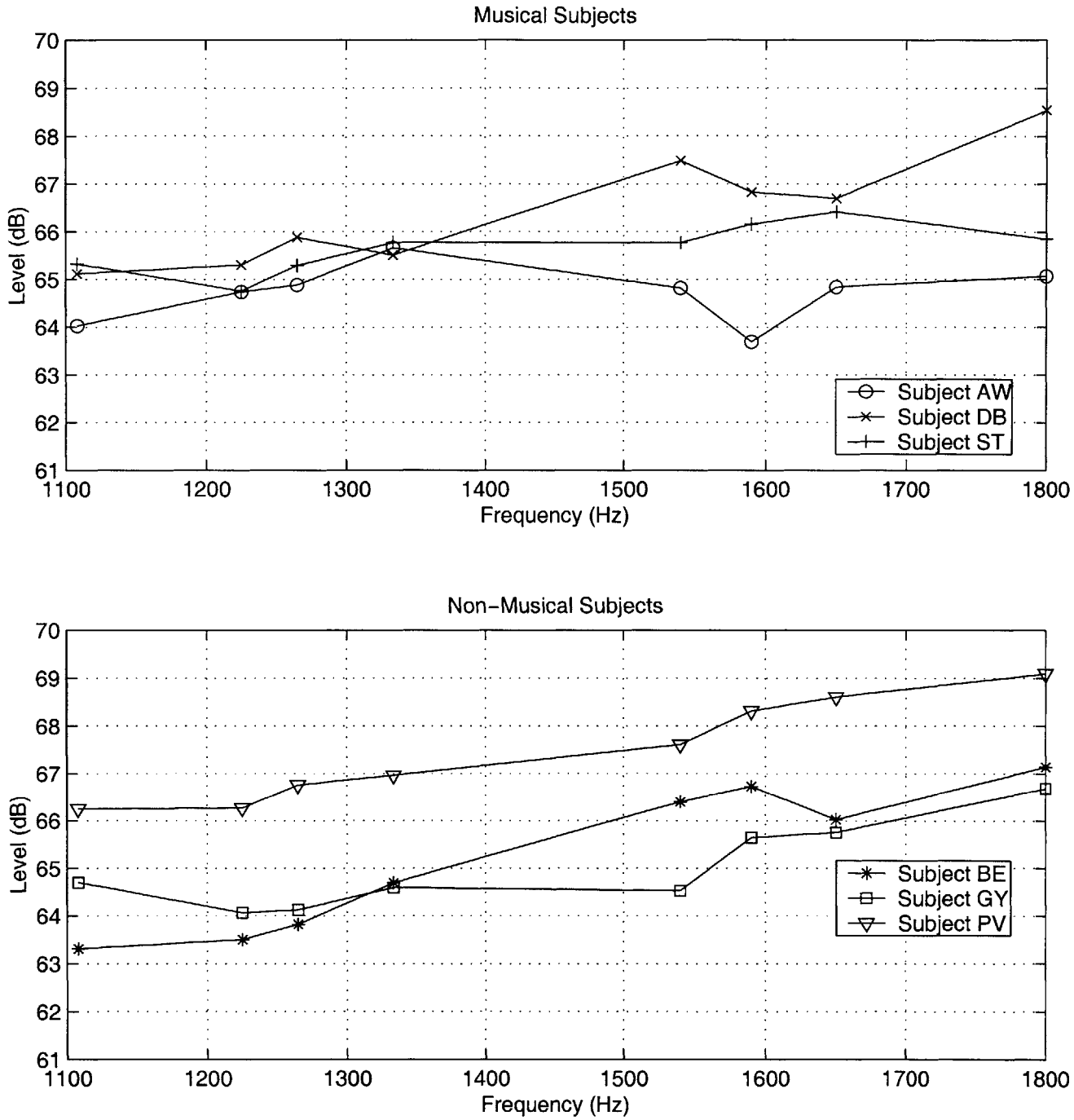


Figure 3-4: Experiment 1. Summary of Two-Interval Experiment Data: Diotic Presentation, Dissonant Standard

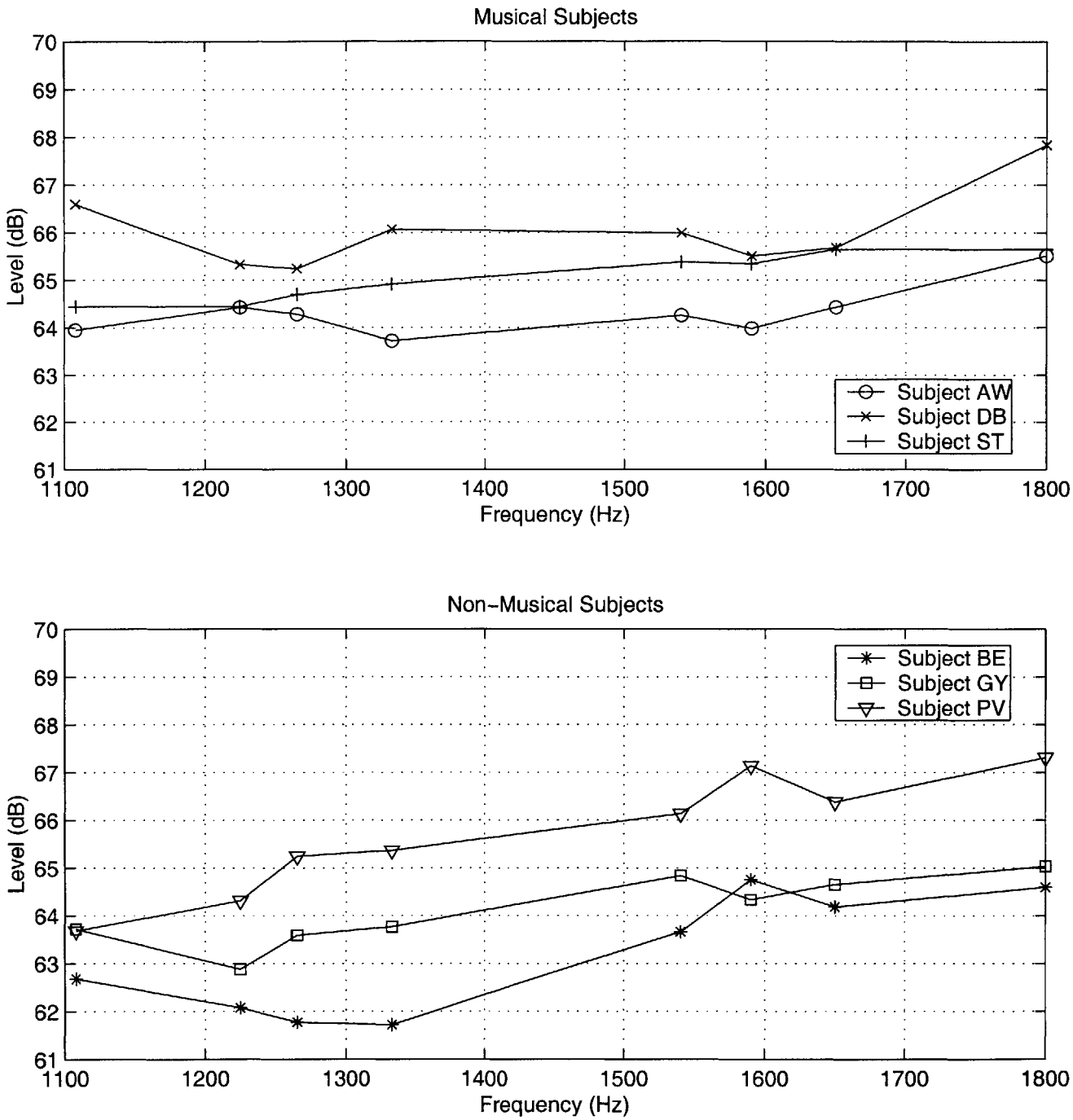


Figure 3-5: Experiment 2. Summary of Two-Interval Experiment Data: Diotic Presentation, Consonant Standard

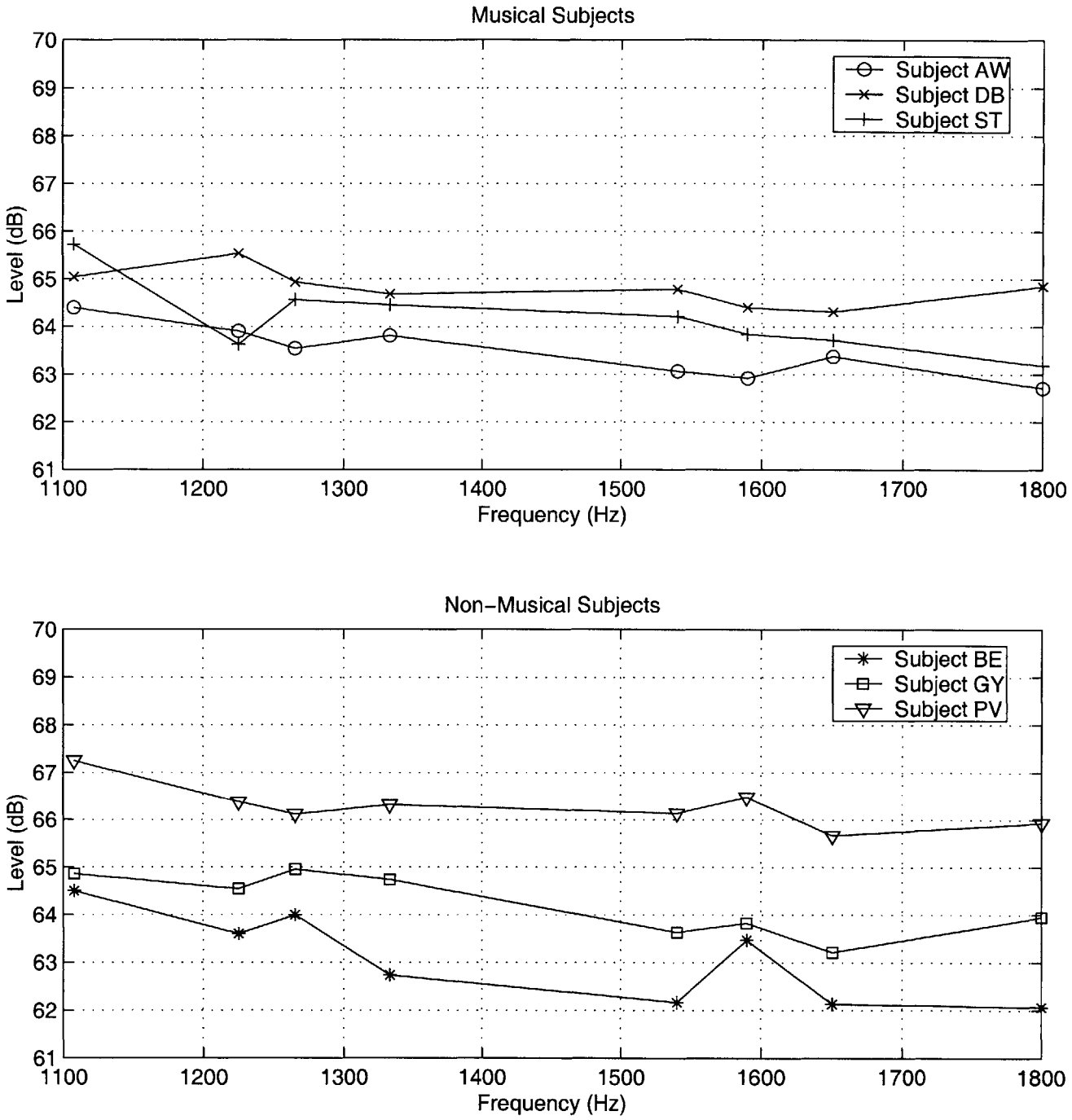


Figure 3-6: Experiment 3. Summary of Two-Interval Experiment Data: Dichotic Presentation, Dissonant Standard

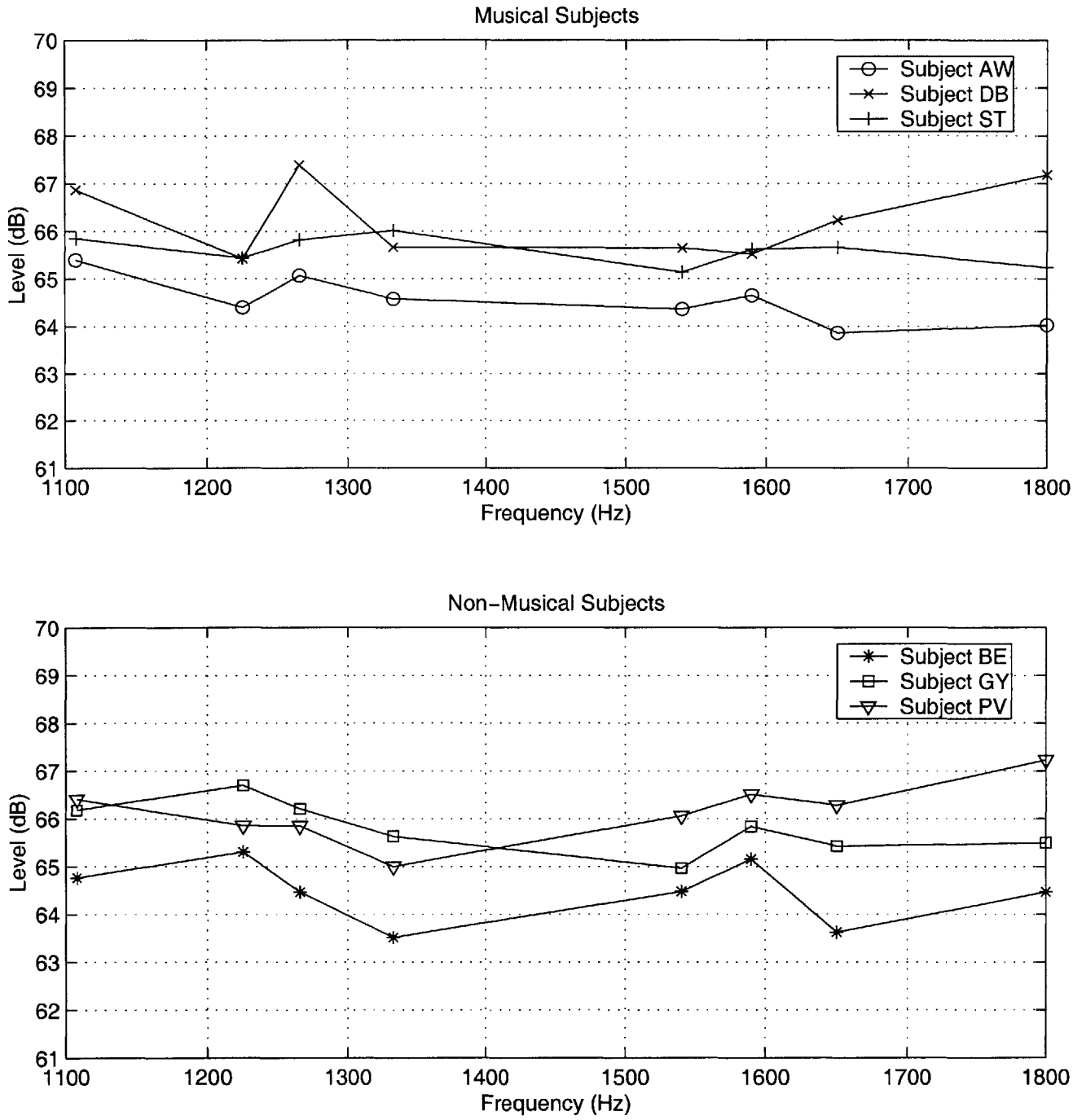
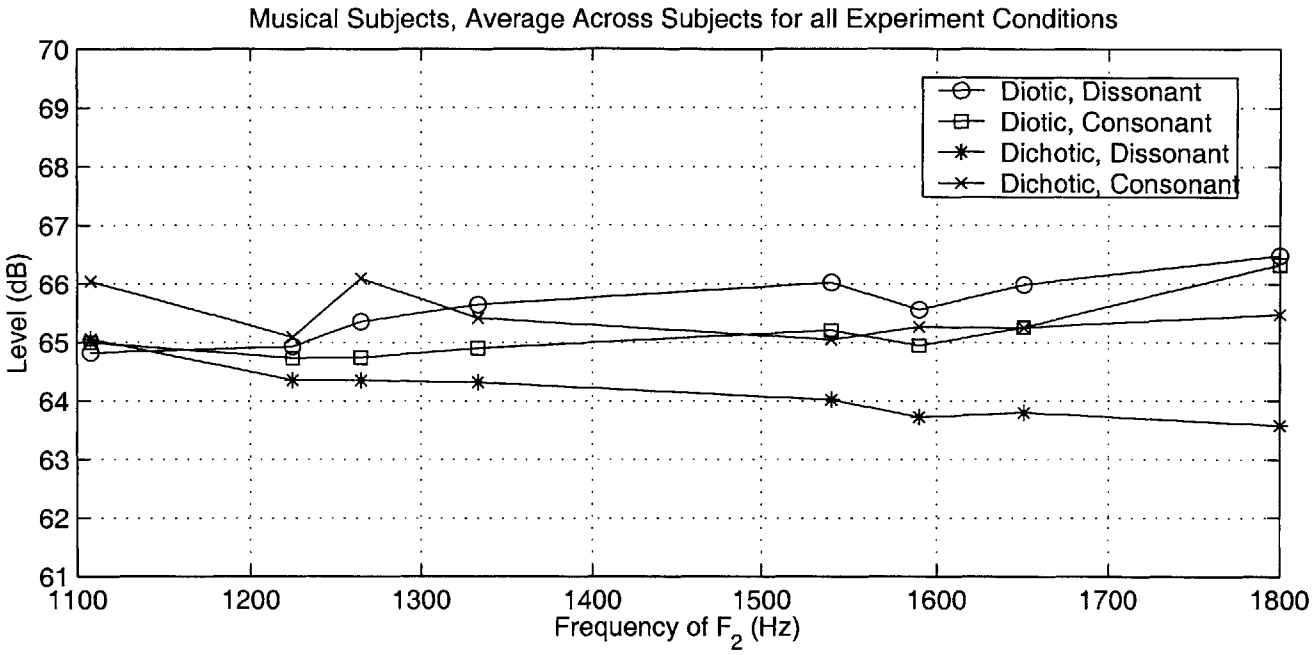


Figure 3-7: Experiment 4. Summary of Two-Interval Experiment Data: Dichotic Presentation, Consonant Standard



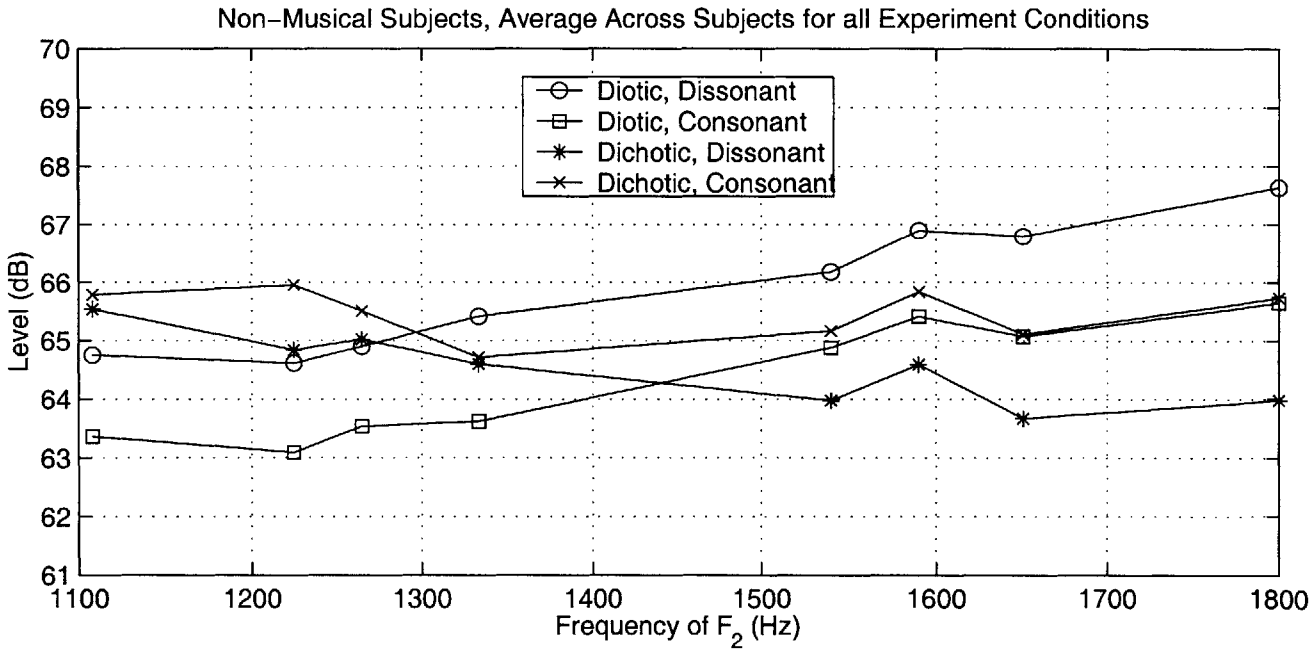


Figure 3-8: Summary: 2-Interval Data, Average Across Subjects for all Stimuli Conditions

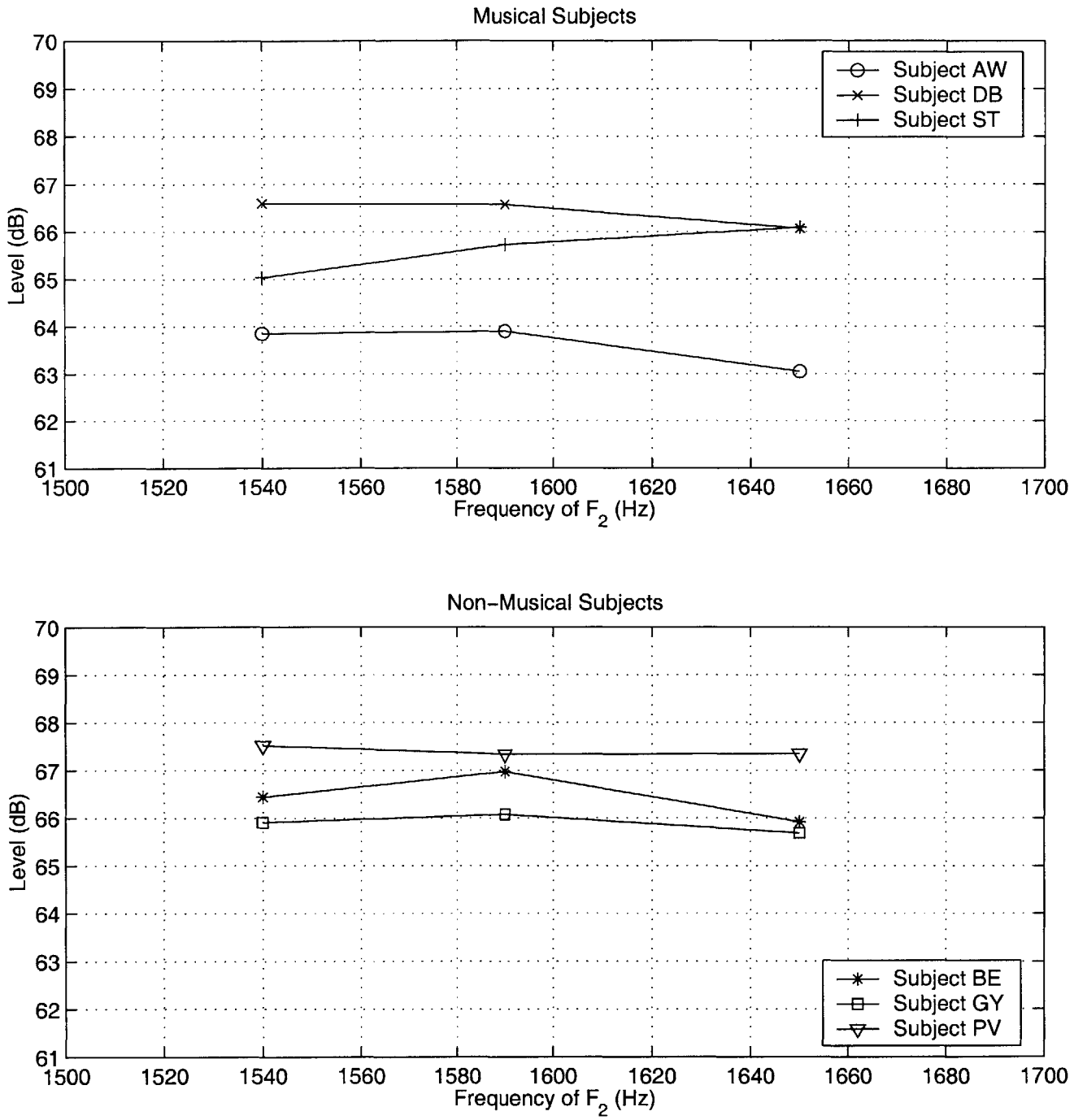


Figure 3-9: Experiment 5. Summary of Four-Interval Experiment Data: Diotic Presentation, Dissonant Standard

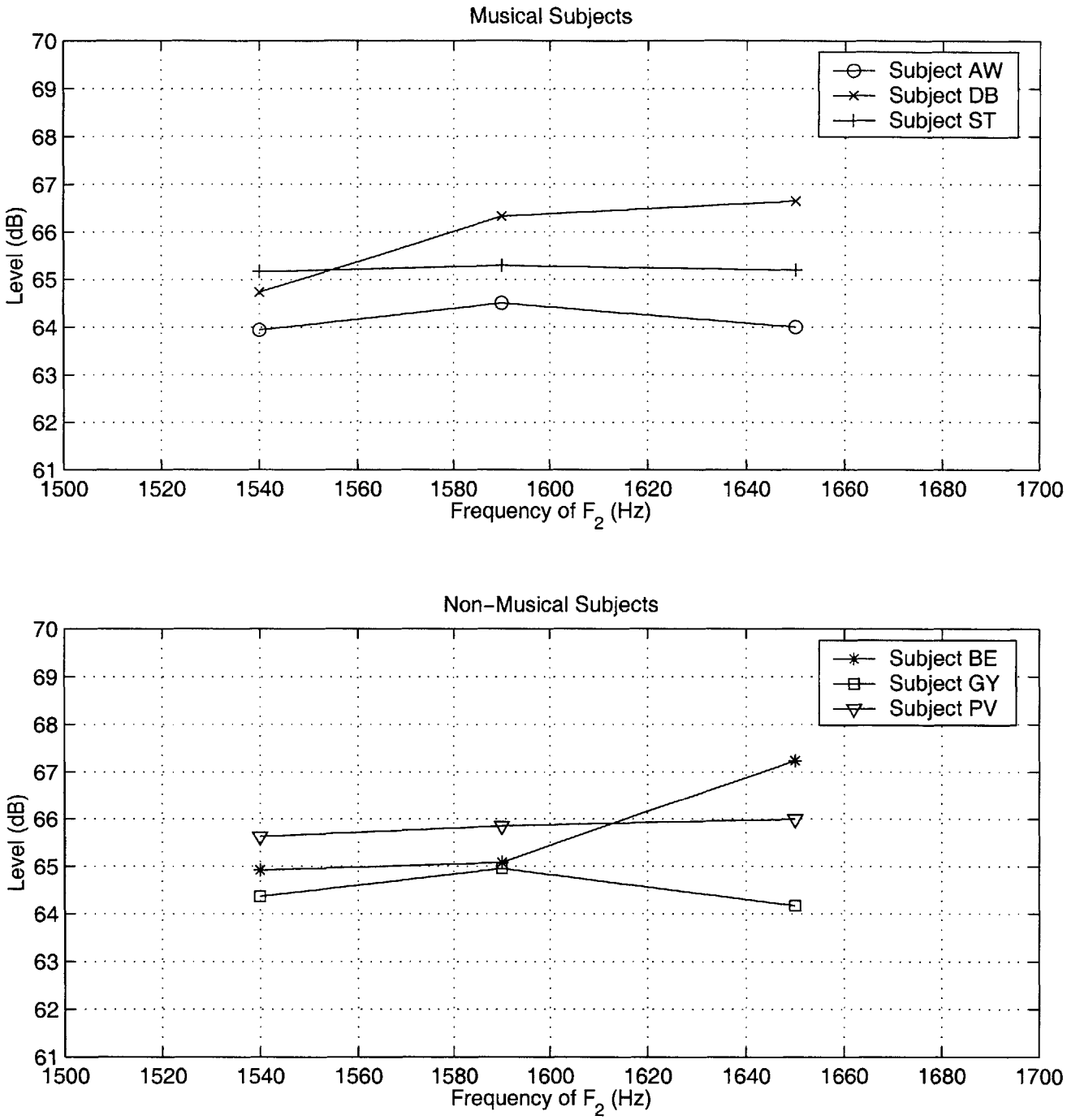


Figure 3-10: Experiment 6. Summary of Four-Interval Experiment Data: Diotic Presentation, Consonant Standard

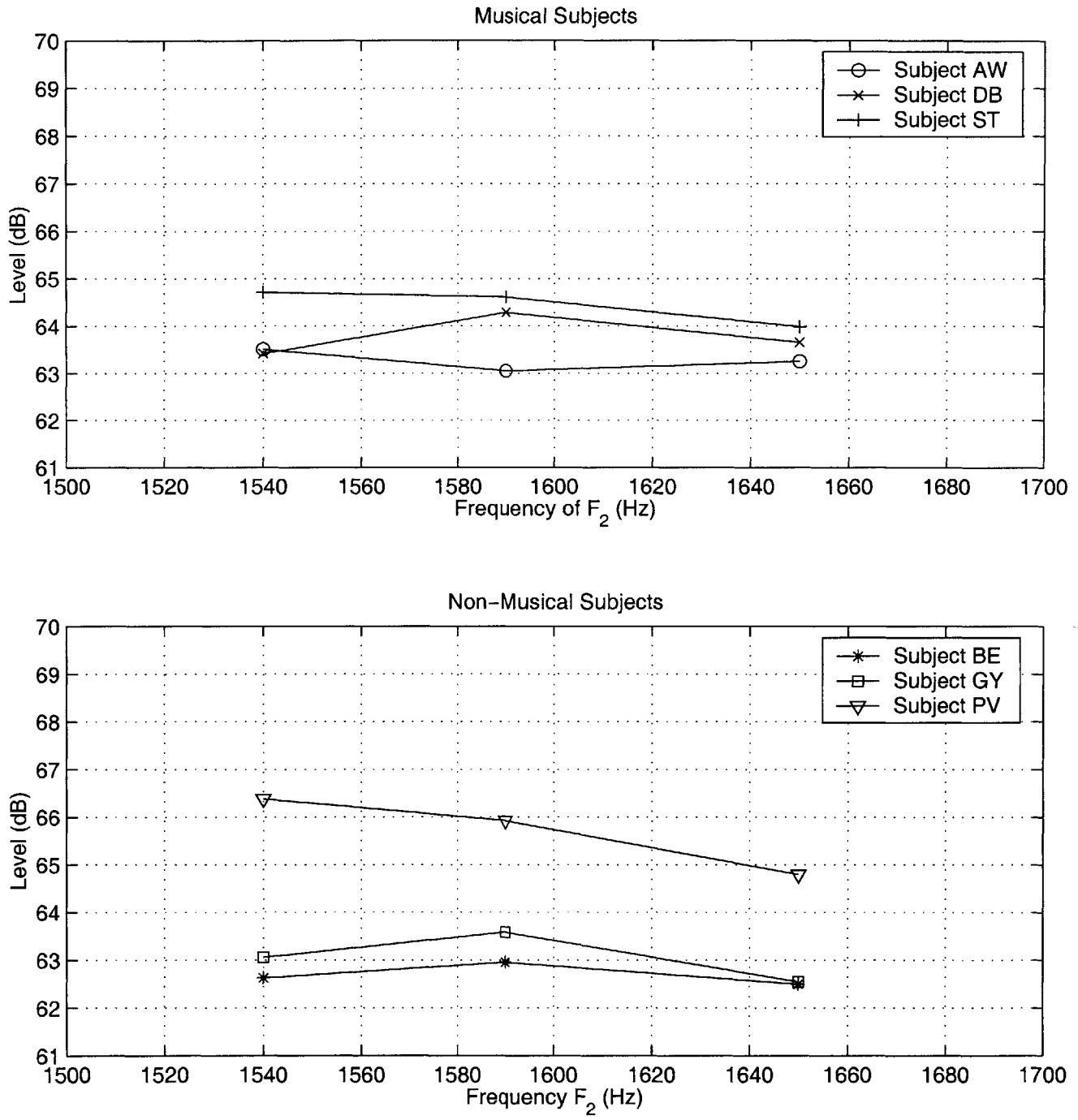


Figure 3-11: Experiment 7. Summary of Four-Interval Experiment Data: Dichotic Presentation, Dissonant Standard

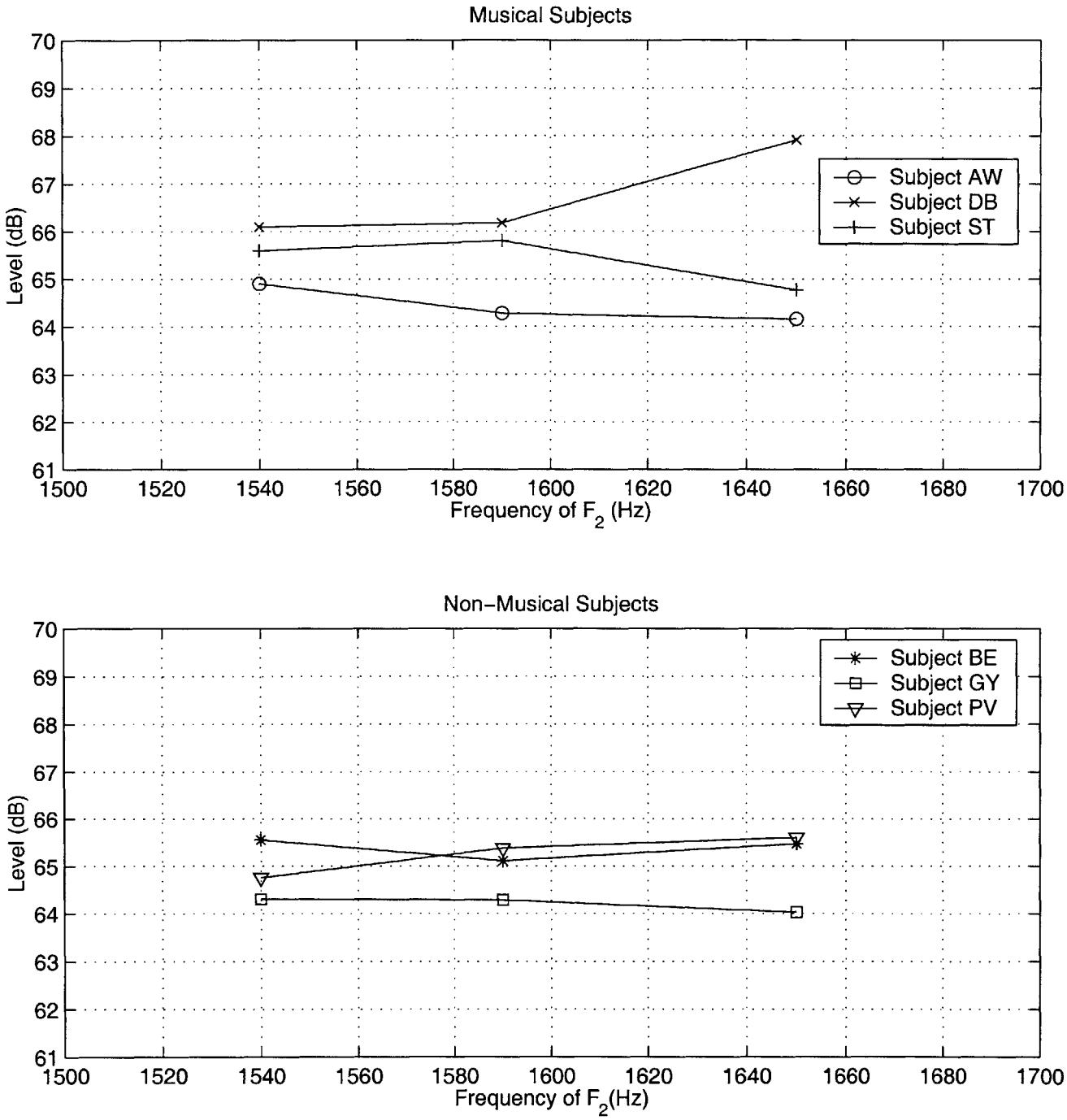
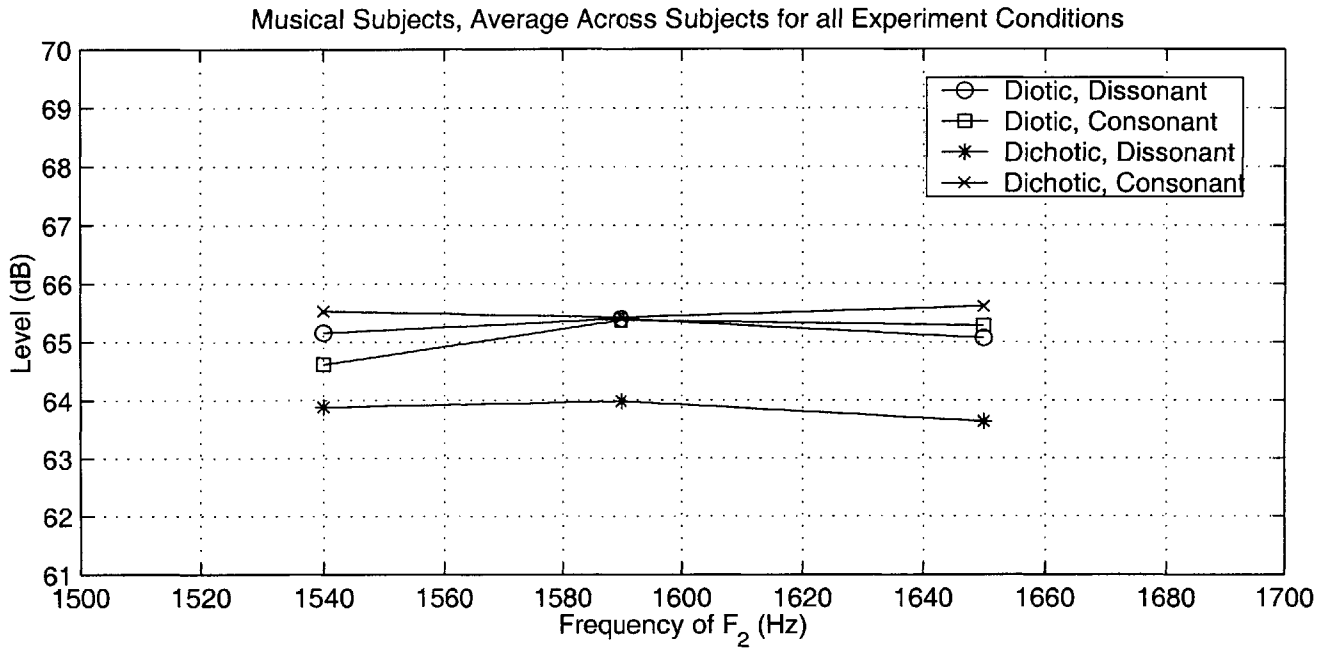


Figure 3-12: Experiment 8. Summary of Four-Interval Experiment Data: Dichotic Presentation, Consonant Standard



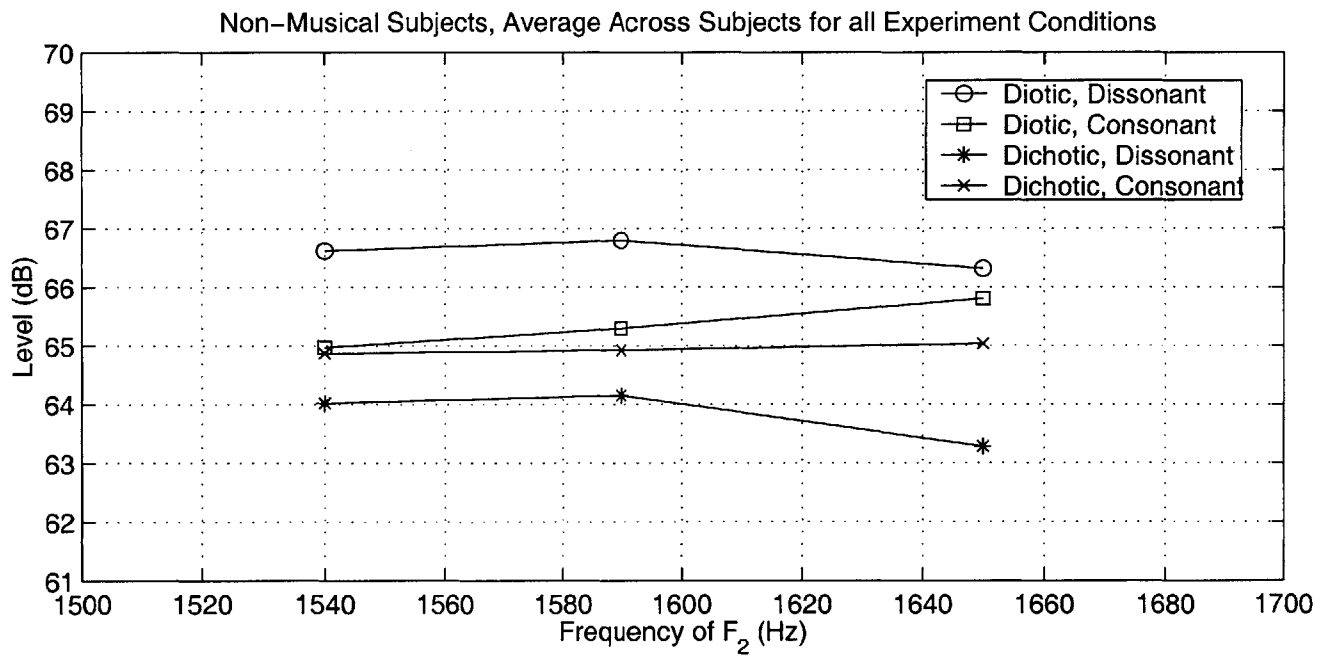


Figure 3-13: Summary: 4-Interval Data, Average Across Subjects for all Stimuli Conditions

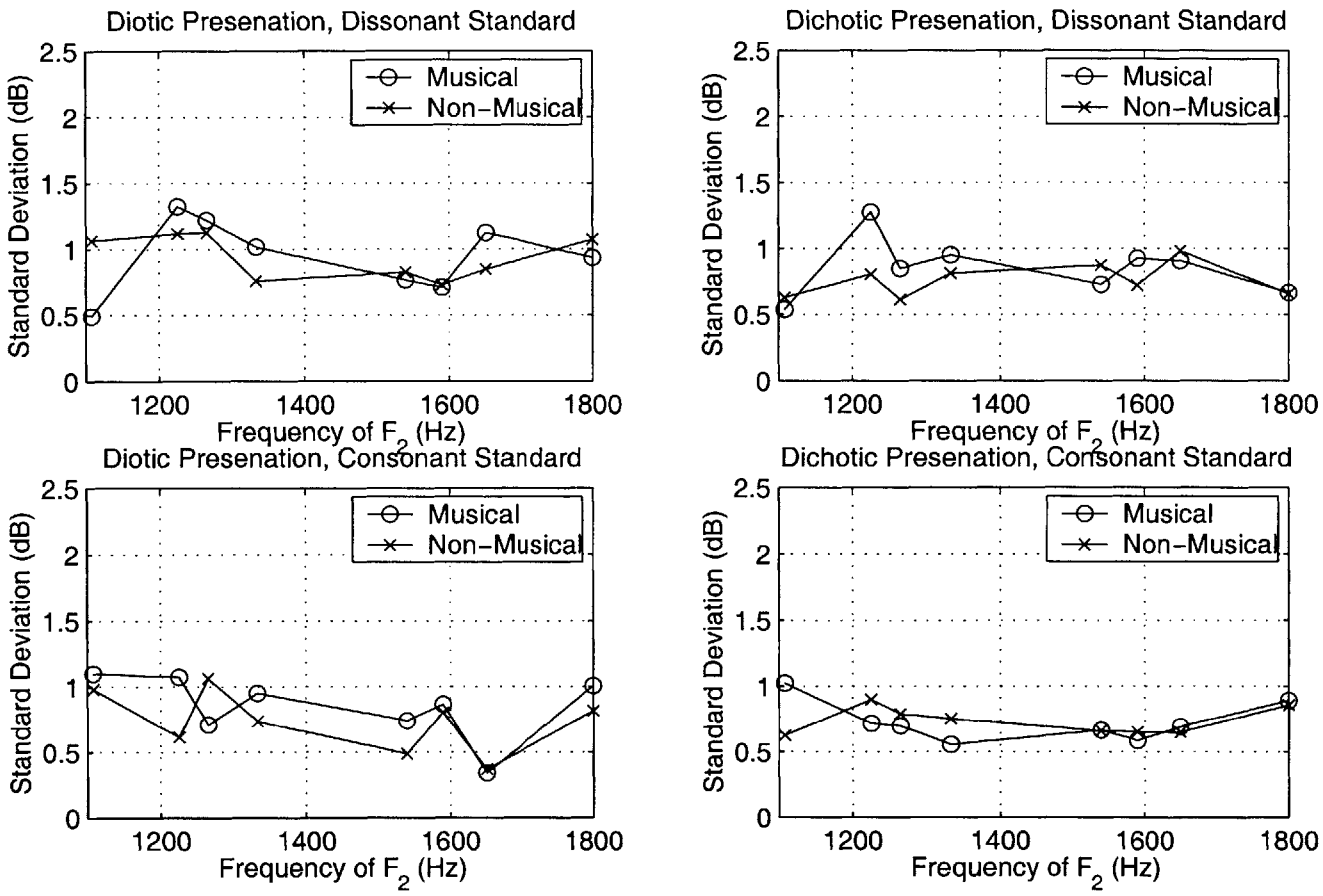


Figure 3-14: Summary of Two-Interval Experiment Data: Standard Deviations

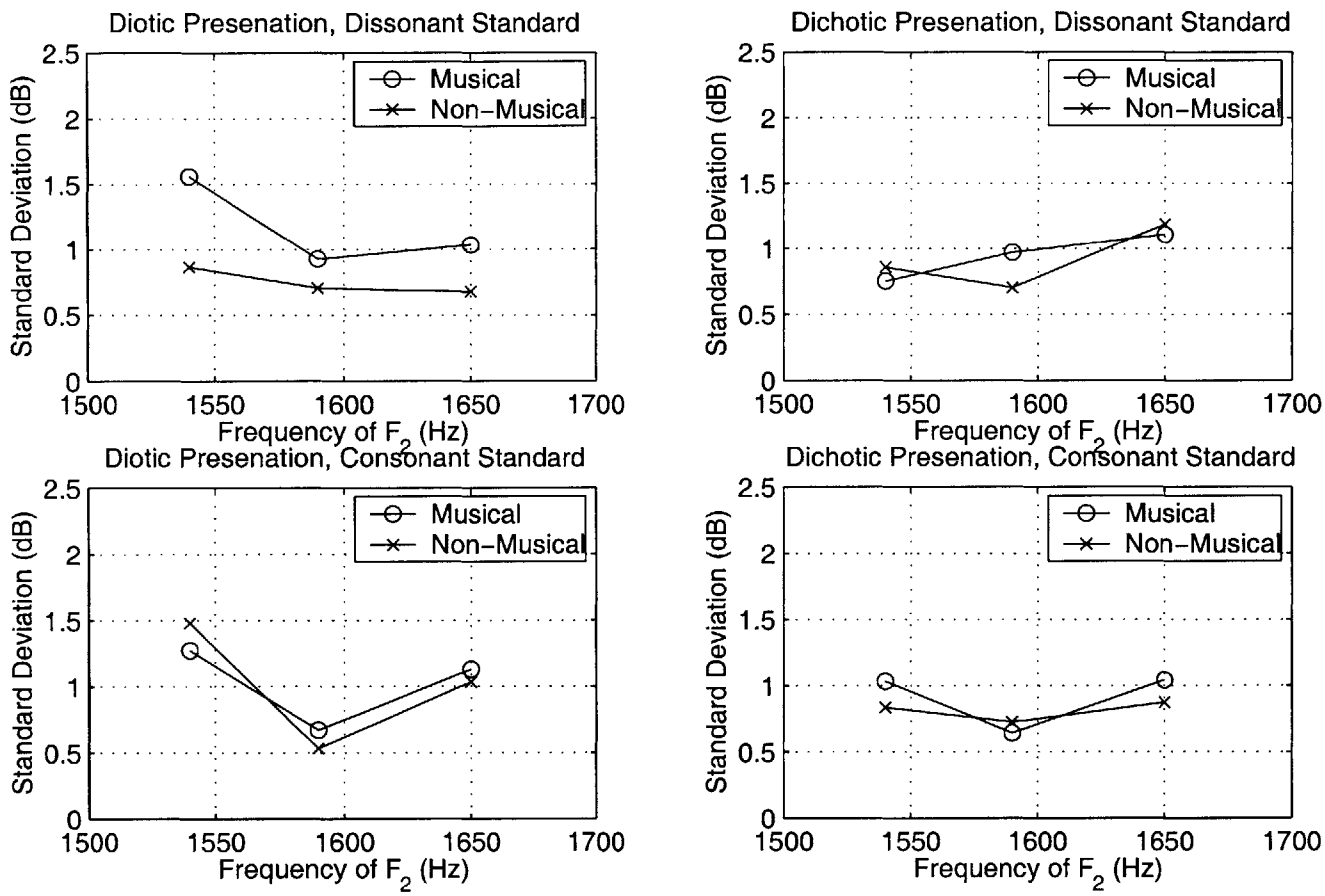


Figure 3-15: Summary of Four-Interval Experiment Data: Standard Deviations

Chapter 4

Discussion

This chapter discusses the data presented in the previous chapter and attempts to explain these results. First, we look at the method of stimuli presentation to examine the differences in the data for diotic and dichotic presentations. Then, we examine the consonant and dissonant standards to see how these factors affect the loudness estimates. Next, we focus on the region near the consonant interval to characterize the "consonant dip". We then analyze the data in the context of musicality to determine if musical background plays a role in loudness perception. Lastly, the data is resolved with the work of Howard Golub, highlighting differences in the experiments.

4.1 Presentation: Diotic versus Dichotic

By comparing Figure 3-4 and Figure 3-5 to Figure 3-6 and Figure 3-7, we can see how the presentation of the stimulus affected the loudness estimates. The data shows that, for the diotic presentation, we have the expected "growth in loudness", predicted by Stevens, Zwicker, and Flottorp [21], as ΔF increases. However, for the dichotic case with dissonant standard, we see an average decline in loudness as ΔF increases. This decline can be explained by the subject's tendency to simplify the stimuli to one tone as ΔF increases. As the distance between the frequencies increases for the data points in the dichotic case, the subject separates the comparison complex into its two separate tones, but is not able to

separate the dissonant standard, since the frequencies are rather close together. Thus, the subject is comparing the two-tone entity (dissonant standard complex) to the single tones in the comparison complex, so the level of the standard decreases as the subject is able to split the two tones. In other words, when the comparison complex's frequencies are farther apart and heard as two separate one-tone entities, the subject must lower the level of the standard, which is heard as a two-tone entity.

4.2 Standards: Consonant versus Dissonant

By looking at Figure 3-8, we see an interesting trend when we consider the standard complexes used in these experiments. For the diotic tests, the average estimated level curve is higher for the dissonant standard than it is for the consonant standard, usually by about 1 dB. By comparing the dissonant and consonant standards in this diotic case, we can understand why this should be so. Essentially, the consonant standard covers a wider frequency range, so it is perceived louder than the dissonant standard at the same SPL [21]. Therefore, the dissonant standard must be higher in level in order to have the same SL as the consonant standard. In other words, the softer dissonant standard must be pushed higher in level as it is compared to the comparison complexes.

In the dichotic case, we see an opposite trend, in that the consonant standard traces out a higher curve than the dissonant standard. Since the two tones of each complex are presented to different ears, the critical band loudness growth explanation (Stevens, Zwicker, and Flottorp [21]) in the diotic case does not apply. Instead, we must consider how the subject would try to resolve these stimuli. In the consonant case, the subject can easily separate the tone complex into its constituent components and make judgements on these individual components. On the other hand, in the dissonant case, the tones a rather close in frequency, so separating them would be more difficult. Thus, the subject must use this louder, two-tone complex entity, as opposed to making judgements based on softer, single tones. This suggests that, in the consonant case, the subject uses a single tone, so its level needs to be pushed higher to achieve an equal SL. Conversely, the dissonant standard, perceived as a two-tone object, does not need to be as high in level to obtain an equal SL

compared to the comparison complex.

4.3 The Consonant Dip

One of the main purposes of this project was to determine if there is a dip in loudness at consonant intervals. We now look at the data carefully to determine if this dip occurs under these conditions.

4.3.1 Two-Interval Experiments

The data in Figure 3-4 through Figure 3-7 show results for all the subjects in all conditions for the two-interval experiment (Experiments 1-4). By focusing on the 1590-Hz point in each graph, we see a dip of greater than 1 dB in only one case: for subject ST in the diotic, dissonant case (Figure 3-4). Otherwise, we see no substantial dips in loudness at this point for any of the subjects. We look for dips greater than 1 dB in these graphs because, as Figure 3-14 shows, the standard deviations are between 0.5 dB to 1 dB in magnitude. In order to truly say that we have a consonant dip, we must exceed the standard deviation, which describes by how much the data varies from run to run. In fact, for some subjects (Subjects BE, GY, PV), we see more of a "bulge" at the consonant interval, and this is discussed more in the musical context later in this chapter.

Figure 3-8 summarizes the averages across subjects for all four experiments. This shows that, on average, any dip or bulge is less than 0.5 dB in amplitude; therefore, because it is less than the standard deviation, these effects cannot be considered significant unless they still remain after reducing the standard deviation through more future experiments.

4.3.2 Four-Interval Experiments

Because the two-interval data did not yield a dip in loudness at the consonant interval, the four-interval experiments were conducted to attempt to simulate Golub's experiments more closely. Because he allowed the subject to switch freely between the two stimuli in each trial and also allowed the subject to control the length of the stimuli, this allowed the

subject much more time to match the loudnesses. The two-interval experiment only gave the subjects 250 milliseconds for each interval, so perhaps the length of the stimuli would affect the loudness at the dip. Perhaps giving the subject more time and, thus, more cycles of the sinusoids would give the subject a better sense of the harmonicity, and then they would respond differently in the loudness comparisons. In the four-interval experiments, the subject is exposed to the comparison and standard complexes for 1 second each (since each interval is 500 milliseconds), so the subject should have plenty of time to judge the loudness and harmonicity of each stimulus.

Figure 3-9 through Figure 3-12 show the results of the four-interval experiments (Experiments 5-8). For Figure 3-9, we see that the loudness around the 1590-Hz point is essentially flat, with perhaps a small bulge here. In Figure 3-10, there is some growth in loudness as ΔF increase, but we do not see any drastic drop in the curves at the consonant point. Figure 3-11 displays a decline in loudness as ΔF increases, but we again see no dip, and, in some cases, we see a bulge (+0.5 dB for Subjects DB and GY). In the last figure for this part (Figure 3-12), we only see one slight dip (0.5 dB for Subject BE), and this is still less than the standard deviation at this point. The analysis of these four figures derives the conclusion that this four-interval experiment does not produce an appreciable dip in loudness at 1590 Hz. This conclusion is supported by the summary of the four-interval experiments in Figure 3-13 as well.

4.3.3 Standard Deviations

In his thesis, Golub plots his standard deviations for all four cases and notices that it decreases at the consonant interval by about 0.5 dB. He associates the standard deviation with pitch ambiguity, so he concludes that loudness is a function of pitch ambiguity, since his standard deviations resemble his loudness curves.

For the two-interval experiments, Figure 3-14 shows that the standard deviation at the consonant interval is relatively constant. The main appreciable drop in this case occurs for musical subjects at 1108 Hz, and this occurs because, at this data point, the subject is comparing two identical stimuli. For the four-interval experiments, Figure 3-15 shows

that we only obtain a dip (of about 0.5 dB) at 1590-Hz when we are using a consonant standard. Once again, this large decrease in the standard deviation occurs because the subject is comparing two identical stimuli for this data point. Therefore, we see no dip at the consonant interval due to a reduction in pitch ambiguity, but rather because the subjects compare identical stimuli, and this contradicts Golub's conclusions.

4.4 Musicality

By analyzing Figure 3-8, we can see if musicality plays a role in the loudness perceptions of two-tone complexes in the two-interval experiment. The most prominent difference between the musicians and non-musicians is that the musicians tend to have a slight depression (not large enough to be considered a dip) at the 1590 Hz, while the non-musicians tend to have a slight bulge. Looking at the individual averages in Figure 3-4 through Figure 3-7 also supports this trend, as most of the individuals in each group display it (i.e. it is not an artifact of taking the average over the group). More specifically, we in these figures that 11 out of 12 of the non-musician results have some sort of bulge, while 9 out of 12 of the results for the musicians have a slight depression. Figure 3-6 especially shows this difference, as the non-musicians exhibit prominent bulges, with Subject PV's bulge greater than 1 dB, and the musicians' curves are fairly flat with a little depression at the consonant interval. This seems to suggest that non-musicians hear consonant intervals a little bit louder than musicians do for these 250-msec tone complexes. This occurs for all four of the conditions tested, so this also suggests that the differences are in the brain and not in the peripheral processing of the combination tones, since, if it were differences in the peripheral auditory system, we would only see differences in the diotic cases.

While this difference is not a huge effect (slightly less than 1 dB), it is possible that using different experimental parameters could accentuate this difference. We now look at the four-interval experiments to see what effect the longer stimuli has in this context. Figure 3-13 shows the averages across subjects of all the four-interval experiments. It displays that these differences between musicians and non-musicians disappear in the four-interval experiments. All the subjects in this case exhibit either a flat loudness curve, or a slight bulge. This seems

to suggest that as length of the stimuli increases, musicians may increase their loudness estimates. However, this aspect of the study is certainly not conclusive and would need to be investigated further for a complete understanding. Except for this slight difference at the consonant interval in the two-interval experiments, there seem to be no other delineating factors between musicians and non-musicians.

Chapter 5

Resolution with Golub's work

While Golub concluded that consonant intervals are perceived softer than dissonant intervals, the data of this project does not support that conclusion. We now look at the differences between his 1975 experiments and this project in an attempt to understand why the result conflict. The following aspects are important differences:

1. Golub asks his subjects to adjust one stimulus until it is "just as loud" as another stimulus. These experiments simply ask to decide if one sound is louder than another.

2. Golub allows the subject to control the length of the stimuli. This project fixes the duration of every stimulus.

3. Golub allows the subject to freely switch back and forth between the two stimuli before making a decision. These experiments force the subject to make a decision after hearing the stimuli only once.

Difference #1 is very important, since this is essentially the main reason these experiments were done. Using the Levitt procedure in this project allows the subject to maintain his criterion more easily than in loudness matching experiments. Because the subject is able to maintain this criterion, the standard deviations in this experiment are much less than in Golub's experiments (0.5 dB - 1.5 dB, as opposed to 0.5 dB - 4 dB).

Difference #2 and #3 introduce many more variables in Golub's experiment, and these are eliminated here. By fixing the stimulus length and order, we can say whether the

consonant dip exists more confidently because there are fewer variables that come into play here. While this experiment has not eliminated all other possible variables (i.e. the subject listening to the tones separately in the dichotic case), it is more careful than Golub's protocol, since it eliminates length and order. Also, by presenting the subject the stimuli once per trial, this is a more accurate representation of auditory stimuli in the real world. When listening and judging loudness in everyday life, a person does not have a replay button so that they can hear the stimuli over and over, and then make a decision. They discern the loudness immediately, and often they do not need to actively and consciously think about the stimuli to perceive the loudness. Therefore, this study is more useful in the context of real-world stimuli. It is certainly possible that allowing the subject to switch between the stimuli for each trial may help cause this consonant dip, but even if this is so, it would carry no weight since we are usually concerned with how these studies relate to real life.

In addition, Golub's data actually does not support his final conclusion. In his experiments, he is allowing the subject to adjust the comparison complex (i.e. the complex of which he is estimating the loudness), but this, in fact, provides the *inverse* of the loudness of the comparison complex. For example, if we compare two separate comparison tone complex data points, one (Complex 1) with an equal-loudness estimate at +1 dB above the standard complex, and one (Complex 2) estimated at +2 dB above the standarad complex, we see that Complex 2 requires more SPL to be equally loud as the standard. Thus, if we were to make Complex 2 the same SPL as Complex 1, Complex 2 would, in fact, be perceived softer than Complex 1. Therefore, a more positive data point in Golub's thesis actually means that complex is *softer* in SL than lower points. Another way to think about this is that, for softer sounds, a higher SPL is needed such that the complex has the same SL as the standard. This seems to suggest that Golub's dip may in fact be an increase in loudness.

In this project, the comparison complex is held fixed, and the standard is adjusted to estimate the loudness of the comparison complex. Therefore, if the comparison is at a higher SL level, then the standard estimate must go up as well. If a dip occurs with this protocol, this suggests a decrease in the SL loudness. This would seem to suggest that the data of this project should be a vertical flip of Golub's data. This turns out to not be the case, however, so the discrepencies in the data are likely to do with some of the other protocol differences.

In any case, the data in this experiment does not bulge or dip with the same magnitude as Golub's experiments.

Thus, because there are more variables in Golub's project, and he is actually moving the complex that he is trying to estimate, the discrepancies between his data and the data of this project is to be expected, and we shall therefore derive conclusions that differ from his ideas.

Chapter 6

Conclusions

The focus of this chapter is to address the original goals of this project, summarize the important points that this study has discovered, and propose directions for future work in this area. In the introduction, we set forth to at least partially answer the following questions:

1. Do inharmonic tones sound louder than harmonic tones?

2. Do musicians estimate the loudness of tone complexes differently than non-musicians?

To answer these questions, two sets of experiments, the two-interval paradigm and the four-interval paradigm, were performed on six subjects, three of which had musical training, and the other three had no musical training. The standard deviations for these experiments were relatively low, and much lower than Golub's experiments; this makes the variability among runs less significant than in Golub's work. This smaller variability among runs makes this data more convincing than the data in Golub's study.

Resolving the Project Goals

The data of this project clearly shows that, for alternative forced-choice (AFC) experimental paradigms, there does not exist a large dip in loudness for the consonant interval. The two-interval experiments did not yield an appreciable dip, and performing the four-interval experiments produced even flatter loudness curves. This leads to the question of what experiments would result in harmonicity playing a factor in loudness judgements. It may

require the subject being allowed to switch back and forth between the stimuli, but then the usefulness of such an experiment in real life is diminished. It may involve a certain length of stimuli, but, since the two stimuli lengths (two-interval experiments and four-interval experiments) in this experiment did not yield a consonant dip, this option seems less likely. In any case, if a majority of experiments that simulate real life stimuli yield no loudness differences for dissonant and consonant sounds, then search for conditions that yield a consonant dip would net little useful information. A model for loudness perception that incorporates harmonicity was not developed here because this project casts some doubt on whether there even is a relationship. However, since very few studies have been done in this area, more experiments should be done before fully discounting Golub's ideas.

As for the differences between musicians and non-musicians, we see only a small difference at the consonant interval, since non-musicians have a slight bulge in loudness at 1590 Hz, while musicians lack this bulge for shorter stimuli. Focusing on this difference in various experimental conditions may prove to be useful in understanding how musical training affects how we hear. Because these differences occur in both diotic and dichotic conditions, this may suggest that the differences are in the brain's processing as opposed to peripheral coding.

Experiment Improvement

If additional experiments like these are performed in the future, then the main improvement needed is the design of standard complexes that are treated in the same manner throughout the experiment. For instance, as mentioned in the previous chapter, the dichotic standards were most likely treated differently by the subject based on the comparison complex. One possible standard that would avoid this is to use bandlimited noise as the standard instead, where the band includes all the comparison complex frequencies. It would be very difficult for the subject to extract the comparison complex frequencies from this presumedly white noise, so each data point comparison would be treated the same. While difficult cross-modal comparisons would be done by the subject in this case, these comparisons would be *the same* for each data point. This would eliminate the standard complex as a variable in the experiments.

Project Relevance to Practical Applications

In his thesis, Golub proposes that perhaps his findings could lead to making jet engine noises less loud by making the sound more consonant, yielding a sort of "musical jet". While this seems like an intriguing idea for an application, his conclusions are not supported strongly enough to suggest that this would work. The data in this project seems to suggest that harmonicity does not play an important role in the loudness of sounds. However, more study is still needed to completely discount harmonicity's role in loudness, since this study only used 6 subjects. For Golub's jet engine problem, the data suggests that, for right now, a more efficient way to approach the jet engine noise would be to attempt to bandlimit the spectrum and reduce the total energy of the sound, instead of looking at the harmonicity.

Future Work

Future work for this project could proceed in two main directions:

1. Dissonance/Consonance Loudness Studies

2. Musical Training Inquiries

For continued study in the area of harmonicity, many experiments would most likely provide more insight. One could look more carefully at how the length of the stimuli affected the loudness curves; while this study considered two different lengths, a more complete study could be done. One could repeat these experiments with more subjects so that the conclusions are more solidified and general. One could also investigate other consonant intervals to see if they result in a dip in loudness. Since natural stimuli often contain many overtones, this work could be extended to complexes with more than two tones. However, this should be done only after two-tone complexes are more thoroughly understood. If future experiments yield the consonant dip, then another useful experiment would be to look at the width of this dip in frequency. For continued study in how musical training affects perceptions, the experiments mentioned above could simply distinguish between musical and non-musical subjects when they are run.

While this project has discounted Golub's conclusions somewhat, it has highlighted the need for more general standards and continued inquiry in this area to obtain a more concrete understanding.

Appendix A

Experiment ESPUD Code

idisc.defs

This file sets all the necessary parameters for each experiment and is reset before each run.

```
(define ext_atten 0.0)

(define session 8)
(define subject "ST")
(define filename "ST8-dichcons7.dat")

; whenever calibrate? is true, set ear-mode to 'diotic
(define calibrate? #f)
(define tone-freq 1060.0)
(define tone-freq-calib 1000.0)
(define tone-freq-calib-roving 1000.0)

;(define ear-mode 'diotic)
(define ear-mode 'dichotic)
(define consonant-standard 1590.0)
(define dissonant-standard 1108.0)

(define data-1-freq 1108.0)
(define data-2-freq 1225.0)
(define data-3-freq 1265.0)
(define data-4-freq 1333.0)
(define data-5-freq 1540.0)
(define data-6-freq 1590.0)
(define data-7-freq 1650.0)
(define data-8-freq 1800.0)
```

```
;(define tone-freq-2-comparison data-5-freq)
(define tone-freq-2-comparison 1650.0)
(define tone-freq-2-standard consonant-standard)
;(define tone-freq-2-standard dissonant-standard)

(define tone-level-1 75.0)
(define calib-tone-level-1 75.0)
(define average-offset 10.0)
(define attenuation 0.0)
(define tone-noise-offset 1.5)

;(define tone-level-2 (+ tone-level-1 deltaI))
;(define tone-dur 250.0)
;(define interpulse 250.0)
(define intertrial 25.0)
(define the-delay 700)
(define feedback #f)
;(define target-prob 0.707)
(define target-prob 0.500)
(define init-level tone-level-1)
(define average-init-step 5.5)
(define skip-turns 1)
(define max-turns 6)
(define max-presentations 150)
(define min-step 0.10)
(define noiselevel -20.0)
(define rise-time 25)
(define fall-time 25)
(define num_intervals 2)
(define R 0.0)
(define 4-interval? #t)
(define tone-dur 500.0)
(define interpulse 50.0)
```

adapt.lsp

This program uses Levitt's transformed up-down method to find a subject's threshold in a psychophysical task. The file "user.lsp" sets everything up and contains the "updown" procedure called at the end of this file.

```
(define parmfile "idisc.defs")
(load parmfile)
(printf "parmfile loaded\n")
```

```
;debug.lsp contains debugging variable definitions
(load "debug.lsp")
(printf "debug.lsp loaded\n")
(load "response.lsp")
(printf "response.lsp loaded\n")
;utilities.lsp defines useful tools
(load "utilities.lsp")
(printf "utilities.lsp loaded\n")
;levitt.lsp contains procedures for the updown algorithm
(load "levitt.lsp")
(printf "levitt.lsp loaded\n")
; response.lsp defines procedures for handling subject responses
(load "response.lsp")
(printf "response.lsp loaded\n")
; fake-io.lsp fakes out i-o procedures
(if iodbg
    (begin
     (load "fake-io.lsp")
     (printf "%s\n" "fake-io.lsp loaded"))
    'ok)
;;; parmfile contains parameters for this run of trials
;;; and loads the experimenter specified file
(load "user.lsp")
(printf "user.lsp loaded\n")
(if calibrate?
    (calibrate-updown)
  (updown))
```

levitt.lsp

This file contains the functions for the Levitt updown procedure.

```
(define offset (+ average-offset
  (get-rand -2 3)))
(define init-step
  (if (<= offset average-offset)
      (+ average-init-step
 (get-rand -1 1))
    (+ average-init-step
       (get-rand 0 2))))
(define active-exp-A #t)
(define active-exp-B #t)
(define print-tag #t)
(define (make-level-changer target)
```

```
  (define (pos? resp) resp)                ;;; responses are booleans
  (define (neg? resp) (not (pos? resp)))
  (define (p500 resp prev)
    (if (pos? resp) 'harder 'easier))
  (define (p707 resp prev)
    (cond ((neg? resp) 'easier)
  ((= prev 1) 'harder)
  (else 'same)))
  (define (p793 resp prev)
    (cond ((neg? resp) 'easier)
  ((= prev 2) 'harder)
  (else 'same)))
  (define (p841 resp prev)
    (cond ((neg? resp) 'easier)
  ((= prev 3) 'harder)
  (else 'same)))
  (lambda (resp prev)
    (cond ((= target 0.5000) (p500 resp prev))
  ((<= target 0.707) (p707 resp prev))
  ((<= target 0.793) (p793 resp prev))
  ((<= target 0.841) (p841 resp prev))
  (else
   (faltalf
    "MAKE-LEVEL-CHANGER illegal level %f\n"
    target)))))
;;; max-turns, max-trials, and min-step are global variables
;;; the first version is for real experiments
(define (stop? changes presentations step)
  (or (> changes (+ skip-turns max-turns))
      (>= presentations max-presentations)
      (<= (abs step) min-step)))

;;; make-levitt-maker hides many many things with internal definitions
;;; returns a procedure of one args that generates an adaptation procedure
;;; the adaptation procedure (no arguments) calls DO-TRIAL
;;; evaluating (make-level-maker) creates an environment in which several
;;;    variables and utility procedures are bound and hidden
(define (make-levitt-maker target)
  (define up +1)
  (define no-dir 0)
  (define down -1)
  (define (inc? dir) (> dir 0))
  (define (dec? dir) (< dir 0))
  (define (make-harder x y sequence)
    (cond ((eq? sequence 'low) (+ x y))
```

```
        ((eq? sequence 'high) (- x y))))
  (define (make-easier x y sequence)
    (cond ((eq? sequence 'low) (- x y))
          ((eq? sequence 'high) (+ x y))))
  (lambda (level-A level-B)
    (let ((presentations 0)
          (level-changes-A 0)
  (level-changes-B 0)
          (dir-A  no-dir)
  (dir-B  no-dir)
  (prev-dir-A no-dir)
  (prev-dir-B no-dir)
          (prev-A 0)
  (prev-B 0)
          (t-lst-A (list level-A))
  (t-lst-B (list level-B))
          (chng-proc (make-level-changer target)))
      (lambda ()
(define choose-sequence
  (if (= (get-rand 1 3) 1)
      'high
    'low))
(define next-change-A (+ level-changes-A 1))
(define next-step-A  (/ init-step next-change-A))
(define next-change-B (+ level-changes-B 1))
(define next-step-B  (/ init-step next-change-B))

(if (stop? (length t-lst-A) presentations next-step-A)
    (set! active-exp-A #f)
  'ok)
(if (stop? (length t-lst-B) presentations next-step-B)
    (set! active-exp-B #f)
  'ok)

(if (and dbg print-tag)
    (begin
     (printf "Upper sequence level changes = %d , " next-change-A)
     (printf "Lower sequence level changes = %d \n" next-change-B)
     (printf "Upper sequence next step = %f , "  next-step-A)
     (printf "Lower sequence next step = %f \n"  next-step-B)
     (printf "Upper sequence next direction change = %d , " dir-A)
     (printf "Lower sequence next direction change = %d \n" dir-B)
     (printf "Upper sequence # of turns done = %d , " (length t-lst-A))
     (printf "Lower sequence # of turns done = %d \n" (length t-lst-B))
     (set! print-tag #t))
```

```
  'ok)
(if (and (stop? (length t-lst-A) presentations next-step-A)
 (stop? (length t-lst-B) presentations next-step-B))
    (begin
     (list (list presentations level-changes-A next-step-A t-lst-A)
   (list presentations level-changes-B next-step-B t-lst-B)))
  (begin
   (cond
    ((eq? choose-sequence 'high)
     (let* ((rspo (do-trial level-A choose-sequence))
   (chng (chng-proc rspo prev-A)))
       (begin
(set! presentations (+ presentations 1))
(printf "Presentation Number %d \n" presentations))
       (cond
((and (eq? chng 'harder) active-exp-A)
 (set! t-lst-A (if (or (and (= dir-A no-dir) (inc? prev-dir-A))
       (inc? dir-A))
   (cons level-A t-lst-A) t-lst-A))
 (set! level-changes-A next-change-A)
 (set! level-A (make-harder level-A next-step-A 'high))
 (set! prev-A 0)
 (set! prev-dir-A no-dir)
 (set! dir-A down)
 (set! print-tag #t))
((and (eq? chng 'easier) active-exp-A)
 (set! t-lst-A (if (or (dec? dir-A)
       (and (= dir-A no-dir) (dec? prev-dir-A)))
   (cons level-A t-lst-A)
 t-lst-A))
 (set! level-changes-A next-change-A)
 (set! level-A (make-easier level-A next-step-A 'high))
 (set! prev-A 0)
 (set! prev-dir-A no-dir)
 (set! dir-A up))
(t
 (set! prev-A (+ prev-A 1))
 (set! prev-dir-A dir-A)
 (set! dir-A no-dir)
 (set! print-tag #t)
 ))
       '()))
    ((eq? choose-sequence 'low)
     (let* ((rspo (do-trial level-B choose-sequence))
    (chng (chng-proc rspo prev-B)))
```

```
       (begin
(set! presentations (+ presentations 1))
(printf "Presentation Number %d \n" presentations))
       (cond
((and (eq? chng 'harder) active-exp-B)
 (set! t-lst-B (if (or (and (= dir-B no-dir) (dec? prev-dir-B))
       (dec? dir-B))
   (cons level-B t-lst-B) t-lst-B))
 (set! level-changes-B next-change-B)
 (set! level-B (make-harder level-B next-step-B 'low))
 (set! prev-B 0)
 (set! prev-dir-B no-dir)
 (set! dir-B up)
 (set! print-tag #t))
((and (eq? chng 'easier) active-exp-B)
 (set! t-lst-B (if (or (inc? dir-B)
       (and (= dir-B no-dir) (inc? prev-dir-B)))
   (cons level-B t-lst-B)
 t-lst-B))
 (set! level-changes-B next-change-B)
 (set! level-B (make-easier level-B next-step-B 'low))
 (set! prev-B 0)
 (set! prev-dir-B no-dir)
 (set! dir-B down))
(t
 (set! prev-B (+ prev-B 1))
 (set! prev-dir-B dir-B)
 (set! dir-B no-dir)
 (set! print-tag #t)
 ))
       '()))
    (else
     (begin
      (set! print-tag #f))
     '())
    )))))))

;;; these procedures call REPORT (user-specified) with the arg
;;; (PRESENTATIONS (TRIALS TURNS) or
(define levitt-maker (make-levitt-maker target-prob))
(define (levitt ilvl report)
  (let ((turns '())
        (hl (levitt-maker (+ ilvl offset) (- ilvl offset))))
    (do ((count 0 (+ count 1)))
        ((not (null? turns))
```

```
        (report turns))
       (set! turns (hl)))))

;;; avg-tpts averages the turning points in a list to estimate the 50% point
;;; the final nskip points are omitted from the average
(define (avg-tpts tpts nskip)
  (let ((sum 0)
        (lst-len (- (length tpts) nskip)))
    (do ((len 1 (+ len 1)))
        ((or (null? tpts)
             (> len  lst-len))
         (list lst-len (/ sum lst-len)))
      (set! sum (+ sum (car tpts)))
      (set! tpts (cdr tpts)))))
```

user.lsp

This file sets up the 2-AFC experiment by setting all the ESPUD parameters, and then using the updown procedures. It calls functions in "response.lsp" to control interaction with the subject.

```
(define (calibrate-updown)
  (calibrate-setup)
  (if dbg (printf
   "\n%s\n"
   "CALIBRATION EXPERIMENT") 'ok)
  (printf "ready to wait for subject \n")
  (wait-for-subject-ready)
  (levitt init-level report-results))

(define (updown)
  (setup)
  (if dbg (printf
           "\n%s\n"
           "ADAPTIVE INTENSITY DISCRIMINATION EXPERIMENT") 'ok)
  (printf "ready to wait for subject \n")
  (wait-for-subject-ready)
  (levitt init-level report-results))

(define (calibrate-setup)
  (exp-reset-models)
  (exp-noise-inten 0.0 0 'both)
```

```
   (exp-tone-inten calib-tone-level-1 1 'both)
   (exp-tone-inten calib-tone-level-1 2 'both)
   (exp-car-freq tone-freq-calib 1 1)
   (exp-car-freq tone-freq-calib-roving 2 1)
   (exp-rt rise-time 0)
   (exp-ot tone-dur 0)
   (exp-ft fall-time 0)
   (exp-ip intertrial 1)
   (exp-ip interpulse 2)
   (exp-send-changes))

(define (setup)
  (exp-reset-models)
; <specify the values of things that don't change>
  (if (eq? ear-mode 'diotic)
    (begin
     (exp-noise-inten 0.0 0 'both)
     (exp-tone-inten tone-level-1 1 'both)
     (exp-tone-inten tone-level-1 2 'both)
     (exp-car-freq tone-freq 1 1)
     (exp-car-freq tone-freq-2-comparison 1 2)
     (exp-car-freq tone-freq 2 1)
     (exp-car-freq tone-freq-2-standard 2 2)
; ONLY for 4-interval experiment
     (if 4-interval?
 (begin
  (exp-tone-inten tone-level-1 3 'both)
  (exp-tone-inten tone-level-1 4 'both)
  (exp-car-freq tone-freq 3 1)
  (exp-car-freq tone-freq-2-comparison 3 2)
  (exp-car-freq tone-freq 4 1)
  (exp-car-freq tone-freq-2-standard 4 2))
       'ok))
    (begin
;this is the dichotic case
     (exp-noise-inten tone-level-1 1 'left)
     (exp-tone-inten tone-level-1 1 'right)
     (exp-noise-inten tone-level-1 2 'left)
     (exp-tone-inten tone-level-1 2 'right)
; ONLY for 4-interval experiment
     (if 4-interval?
 (begin
  (exp-noise-inten tone-level-1 3 'left)
  (exp-tone-inten tone-level-1 3 'right)
  (exp-noise-inten tone-level-1 4 'left)
```

```
  (exp-tone-inten tone-level-1 4 'right))
       'ok)
;     (exp-car-freq tone-freq 1 1)
     (exp-car-freq tone-freq-2-comparison 1 1)
;     (exp-car-freq tone-freq 2 1)
     (exp-car-freq tone-freq-2-standard 2 1)

     (if 4-interval?
 (begin
  (exp-car-freq tone-freq-2-comparison 3 1)
  (exp-car-freq tone-freq-2-standard 4 1))
       'ok)))
  (exp-rt rise-time 0)
  (exp-ot tone-dur 0)
  (exp-ft fall-time 0)
  (exp-ip intertrial 1)
  (exp-ip interpulse 2)
  (if 4-interval?
      (begin
       (exp-ip interpulse 3)
       (exp-ip interpulse 4))
    'ok)
  (exp-send-changes))

(define (do-trial lvl sequence)
  (if dbug
      (begin
       (printf "DO-TRIAL -- LVL: %6.2f\n" lvl)
       (if (eq? sequence 'high)
   (printf "SEQUENCE = UPPER\n")
 (printf "SEQUENCE = LOWER\n")))
    'ok)
;  (let ((alt (get-rand 1 3)))
  (let ((alt 1))
    (play-tones lvl alt)
    (let ((resp (get-response valid-responses)))
      (if dbug (printf "Resp. = %s;  Alt. = %d\n" resp alt) 'rsp)
      (if (eq? sequence 'low)
      (let ((ok? (correct? resp alt fbk-tbl)))
(give-feedback resp alt ok?)
ok?)
      (let ((ok? (not (correct? resp alt fbk-tbl))))
    (give-feedback resp alt ok?)
    ok?)))))
```

```
;;; miscellaneous definitions
(define (db2pwr dbs)
  (power. 10.0 (/ dbs 10.0)))
(define (pwr2db pwr)
  (* 10.0 (log. pwr)))

(define (play-tones level alt)
   (define roving
     (get-rand (/ (* -1 R) 2) (/ R 2)))
  (if dbg (printf "lvl: %f alt %d\n" level alt) 'ok)
  (delay the-delay)
  (if (= 1 alt)
      (begin
; comparison, then standard
       (printf "  Standard level = %f , " level)
       (printf "Standard attenuation = %f \n" (- tone-level-1 level))
       (printf "Comparison level = %f , " (- tone-level-1 attenuation))
       (printf "Comparison attenuation = %f \n" attenuation)
       (if (eq? ear-mode 'diotic)
   (begin
    (if calibrate?
(begin
 (exp-car-freq tone-freq-calib 1 1)
 (exp-car-freq tone-freq-calib-roving 2 1)
 (exp-tone-atten attenuation 1 1)
 (exp-tone-atten level 1 2)
 (exp-tone-atten (- calib-tone-level-1 level) 2 1)
 (exp-tone-atten level 2 2))
      (begin
       (exp-car-freq tone-freq 1 1)
       (exp-car-freq tone-freq-2-comparison 1 2)
       (exp-tone-atten attenuation 1 1)
       (exp-tone-atten attenuation 1 2)
       (exp-car-freq tone-freq 2 1)
       (exp-car-freq tone-freq-2-standard 2 2)
       (exp-tone-atten (- tone-level-1 level) 2 1)
       (exp-tone-atten (- tone-level-1 level) 2 2)
       (if 4-interval?
   (begin
    (exp-car-freq tone-freq 3 1)
    (exp-car-freq tone-freq-2-comparison 3 2)
    (exp-tone-atten attenuation 3 1)
    (exp-tone-atten attenuation 3 2)
    (exp-car-freq tone-freq 4 1)
    (exp-car-freq tone-freq-2-standard 4 2)
```

```
    (exp-tone-atten (- tone-level-1 level) 4 1)
    (exp-tone-atten (- tone-level-1 level) 4 2))
 'ok))))
 (begin
  (exp-car-freq tone-freq-2-comparison 1 1)
  (exp-car-freq tone-freq-2-standard 2 1)
  (exp-noise-inten (- tone-level-1 tone-noise-offset) 1 'left)
  (exp-tone-atten attenuation 1 1)
  (exp-noise-inten (- level tone-noise-offset) 2 'left)
  (exp-tone-atten (- tone-level-1 level) 2 1)
  (if 4-interval?
      (begin
       (exp-car-freq tone-freq-2-comparison 3 1)
       (exp-car-freq tone-freq-2-standard 4 1)
       (exp-noise-inten (- tone-level-1 tone-noise-offset) 3 'left)
       (exp-tone-atten attenuation 3 1)
       (exp-noise-inten (- level tone-noise-offset) 4 'left)
       (exp-tone-atten (- tone-level-1 level) 4 1))
    'ok))))
; standard, then comparison
    (begin
       (printf "  Standard level = %f , " level)
       (printf "Standard attenuation = %f \n" (- tone-level-1 level))
       (printf "Comparison level = %f , " (- tone-level-1 attenuation))
       (printf "Comparison attenuation = %f \n" attenuation)
       (if (eq? ear-mode 'diotic)
   (begin
    (if calibrate?
(begin
 (exp-car-freq tone-freq-calib-roving 1 1)
 (exp-car-freq tone-freq-calib 2 1)
 (exp-tone-atten (- calib-tone-level-1 level) 1 1)
 (exp-tone-atten level 1 2)
 (exp-tone-atten attenuation 2 1)
 (exp-tone-atten level 2 2))
      (begin
       (exp-car-freq tone-freq 1 1)
       (exp-car-freq tone-freq-2-standard 1 2)
       (exp-car-freq tone-freq 2 1)
       (exp-car-freq tone-freq-2-comparison 2 2)
       (exp-tone-atten (- tone-level-1 level) 1 1)
       (exp-tone-atten (- tone-level-1 level) 1 2)
       (exp-tone-atten attenuation 2 1)
       (exp-tone-atten attenuation 2 2)
       (if 4-interval?
```

```
  (begin
   (exp-car-freq tone-freq 3 1)
   (exp-car-freq tone-freq-2-standard 3 2)
   (exp-car-freq tone-freq 4 1)
   (exp-car-freq tone-freq-2-comparison 4 2)
   (exp-tone-atten (- tone-level-1 level) 3 1)
   (exp-tone-atten (- tone-level-1 level) 3 2)
     (exp-tone-atten attenuation 4 1)
   (exp-tone-atten attenuation 4 2))
'ok))))
(begin
 (exp-car-freq tone-freq-2-standard 1 1)
 (exp-car-freq tone-freq-2-comparison 2 1)
 (exp-noise-inten (- level tone-noise-offset) 1 'left)
 (exp-tone-atten (- tone-level-1 level) 1 1)
 (exp-noise-inten (- tone-level-1 tone-noise-offset) 2 'left)
 (exp-tone-atten attenuation 2 1)
 (if 4-interval?
     (begin
      (exp-car-freq tone-freq-2-standard 3 1)
      (exp-car-freq tone-freq-2-comparison 4 1)

      (exp-noise-inten (- level tone-noise-offset) 3 'left)
      (exp-tone-atten (- tone-level-1 level) 3 1)

      (exp-noise-inten (- tone-level-1 tone-noise-offset) 4 'left)
      (exp-tone-atten attenuation 4 1))
   'ok)))))
 (exp-send-changes)
 (exp-go))
```

Bibliography

[1] H. S. Colburn. Intensity perception. xv. relation of intensity discrimination to auditory-nerve firing patterns, 1982.

[2] Perry R. Cook, editor. *Music, Cognition, and Computerized Sound: An Introduction to Psychoacoustics.* The MIT Press, Cambridge, Massachusetts, 1999.

[3] J. L. Goldstein. Auditory nonlinearity. *The Journal of the Acoustical Society of America*, 41:679–689, 1967.

[4] H. L. Golub. Loudness of harmonic and inharmonic two-tone complexes. Master's thesis, Massachusetts Institute of Technology, 1975.

[5] H. L. Golub. Personal communication, March 2001.

[6] R. Hellman and E. Zwicker. Why can a decrease in dB(A) produce an increase in loudness? *The Journal of the Acoustical Society of America*, 82:1700–1705, 1987.

[7] Walt Jesteadt. An adaptive procedure for subjective judgements. *Perception & Psychophysics*, 28:85–88, 1980.

[8] N. Y. S. Kiang, T. Watanabe, E. C. Thomas, and L. S. Clark. *Discharge Patterns of Single Fibers in the Cat's Auditory Nerve.* MIT Press, Cambridge, Massachusetts, 1965.

[9] H. Levitt. Transformed up-down methods in psychoacoustics. *The Journal of the Acoustical Society of America*, 49:467–477, 1970.

[10] D. Luce and Melville Clark, Jr. Physical Correlates of Brass-Instrument Tones. *The Journal of the Acoustical Society of America*, 42:1232–1243, 1967.

[11] D. S. Lum. Runexp ESPUD Server Manual. MIT Research Laboratory of Electronics Sensory Communication Group, February 1997.

[12] David S. Lum. Evaluation of a hearing loss simulator. Advanced Undergraduate Project, Massachusetts Institute of Technology, Cambridge, Massachusetts, 1994.

[13] K. D. Martin. *Sound-Source Recognition: A Theory and Computational Model.* PhD thesis, Massachusetts Institute of Technology, 1999.

[14] A. Miskiewicz and A. Rakowski. Loudness level versus sound-pressure level: A comparison of musical instruments. *The Journal of the Acoustical Society of America*, 96:3375–3379, 1994.

[15] R. Plomp and G.F. Smoorenberg, editors. *Frequency Analysis and Periodicity Detection in Hearing.* A. W. Sijthoff Leiden, 1970.

[16] William A. Sethares. *Tuning, Timbre, Spectrum, Scale.* Springer-Verlag, Berlin, 1999.

[17] W. M. Siebert. *Recognizing Patterns*, chapter Stimulus Transformations in the Peripheral Auditory System, pages 104–133. MIT Press, Cambridge, Massachusetts, 1968.

[18] S. S. Stevens. Procedure for Calculating Loudness: Mark VI. *The Journal of the Acoustical Society of America*, 33:1577–1585, 1961.

[19] S. S. Stevens. Perceived Level of Noise by Mark VII and decibels (E). *The Journal of the Acoustical Society of America*, 51:575–601, 1971.

[20] E. Zwicker and H. Fastl. *Psychoacoustics: Facts and Models.* Springer-Verlag, Berlin, 1999.

[21] E. Zwicker, G. Flottorp, and S. S. Stevens. Critical bandwidth in loudness summation. *The Journal of the Acoustical Society of America*, 29:548–557, 1957.