Memory and Production of Standard Frequencies in College-Level Musicians

A Thesis Presented

by

SARAH WEBER

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

MASTER OF MUSIC

September 2013

Music Theory

Memory and Production of Standard Frequencies in College-Level Musicians

A Thesis Presented

by

SARAH WEBER

Gary S. Karpinski, Chair

Andrew Cohen, Member

Brent Auerbach, Member

Jeff Cox, Department Head

Department of Music and Dance

DEDICATION

For my parents and Grandma.

ACKNOWLEDGEMENTS

I would like to thank Kristen Wallentinsen for her help with experimental logistics, Renée Morgan for giving me her speakers, and Nathaniel Liberty for his unwavering support, problem-solving skills, and voice-over help. This project would have been impossible without guidance from Dr. Andrew Cohen, who introduced me to statistical analysis and helped with the interpretation of my experimental results. I would also like to thank my professors at Ithaca College who modeled great teaching and thinking.

ABSTRACT

WITHIN-SEMITONE ACCURACY OF THE LONG-TERM MEMORY OF ABSOLUTE PITCH IN NON-ABSOLUTE PITCH POSSESSORS

SEPTEMBER 2013

SARAH ELIZABETH WEBER, B.M., ITHACA COLLEGE

M.M., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Dr. Gary S. Karpinski

This thesis examines the nature of long-term absolute pitch memory—an ability traditionally assumed to belong only to absolute pitch (AP) possessors—by testing for evidence of this memory for "standard" frequencies in musicians without AP. Standard frequencies, those based on the equally tempered system with A = 440 Hz, are common in the sonic environment of the Western college musical education, and thus could have the opportunity to penetrate listeners' long-term memories. Through four experimental tasks, this thesis examines musicians' ability to recognize and produce frequencies from the set of equally tempered frequencies based on A = 440 Hz, without regard to those musicians' pitch-labeling abilities. The experimental tasks also compare freshmen with seniors to test if exposure to standard frequencies during a college musical education engrains standard frequencies in long-term memory. The results suggest that musicians without AP cannot distinguish between standard and nonstandard frequencies during listening tasks, but they may be able to recall them without prompting when singing familiar folk songs. However, musical training during the college years does not seem to improve these abilities. Further experimentation is needed to corroborate the results, including modifications to the current tasks and methodology, as well as a larger subject size.

TABLE OF CONTENTS

CHAPTER

APPENDICES

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

A. Background and Related Work

Empirical studies provide evidence that listeners without the pitch-labeling abilities traditionally associated with absolute pitch still exhibit long-term auditory memory for absolute pitch (Levitin 1994, Deutsch 1991, Schellenberg and Trehub 2003).[1] Forty percent of Daniel Levitin's subjects—a mix of musicians and nonmusicians—reproduced the pitch levels of songs with which they were familiar "without error" on at least one trial (418). Diana Deutsch found that listeners use some form of absolute memory for pitch in deciding which of two Shepard tones presented in the interval of a tritone sounds higher and which sounds lower; subjects consistently put the same pitch classes into one category or another. Schellenberg and Trehub found that adults with no musical training successfully identify the original pitch levels of familiar instrumental television theme songs in a forced-choice task. They claim that their results indicate that "ordinary listeners retain fine-grained information about pitch level over extended periods" (262).

"Fine-grained," in these studies, however, is not precisely quantified, and usually refers only to the difference between adjacent semitones. No studies to date have explored the within-semitone precision with which musicians remember pitch levels of music. This is

---

[1] David Butler defines absolute pitch as the ability to "accurately and immediately label (or in some cases produce on demand) pitches without having to consult a reference tone" (48). McLachlan and Wilson state, "In some individuals, finely tuned identities for musical notes (absolute pitch) may be learned if their pitches are consistently presented with verbal labels at a young age when association hierarchies are sufficiently flexible to create branches dedicated to this pitch information" (McLachlan and Wilson, 179). They point out that non-AP possessors rely on comparing pitch heights stored in short-term memory in pitch processing tasks (relative pitch).

the central topic explored by this thesis. The issue may affect choral ensembles, aural skills teaching, and student practice. Any musical pursuit that involves checking the voice with a piano or other instrument of fixed pitch level is affected by the ability of participants to match the set of pitches available on that instrument in the absence of accompaniment. A student practicing a sight-singing excerpt may decide to check a challenging leap with a piano. If the student is singing at a pitch level in between a level reflected on the keyboard, hearing the frequency of the piano might dislodge the student's tonal center and cause the student to adjust to the piano, thereby destroying the relationships between the pitches he or she was singing and ruining the student's chances of learning to find the correct pitch on his or her own.[2]

Absolute pitch (AP) listeners are generally acknowledged to be able to label pitches with fine-grained accuracy, discriminating between a "slightly flat E," and a standard E, for example, or judging that a tone is "about halfway between B and C". McLachlan and Wilson (2010) state that "absolute pitch possessors can often report the pitch name as well as a microtonal pitch variation" (189). Most trained musicians who are exposed to tonal music develop *relative pitch,* or the ability to identify the pitch of a tone when given a reference pitch (Butler 1992, 234). They are called non-absolute pitch listeners or NAP listeners.[3]

---

[2] Students may sing sight-singing excerpts seemingly in tune, only to find that their pitch center has "drifted" a fraction of a semitone up or down by the end. The drift may not be discovered prior to checking a fixed pitch reference such as a piano. Even though pitch drift is technically a relative pitch issue because it involves changing the size of intervals gradually so that the pitch center changes slightly, the error can be so small that even skilled instructors may not notice. Students who check notes on a piano afterward can be mystified at the mismatch and remain unable to identify where the problem(s) occurred.

[3] For this study, I will assume that all subjects have developed their relative pitch to a significant degree, which allows them to recognize transposed melodies as equivalent, and also allows them to produce a tonal center and sing in tune.

Scholars agree that NAP listeners cannot label isolated pitches by letter-name, but it is unclear from a reading of Levitin, Deutsch, and Schellenberg and Trehub whether NAP listeners have the ability to identify the presence of microtonal pitch variations from standard frequencies based on A = 440 Hz when pitches are presented in isolation.

It is significant that not all scholars agree on what skills comprise "absolute pitch." Tests designed to identify AP possessors vary in terms of the complexity of stimuli, and therefore test a range of skills that might be considered to be AP. Ross et al. (2004) assess the works of scholars that posit the existence of absolute pitch memory in NAP listeners—Levitin and others—writing that "paradigms used by those groups test the ability to evoke the memory of a specific, spectrally complex stimulus that has accumulated across many repeated presentations." The authors describe the form of AP found by Levitin as "latent AP." In contrast, Ross et al.'s own paradigm tests AP by "explicitly [isolating] the ability to encode an immediate representation of stimulus fundamental frequency without the presence of any extrinsic cues" (1798). The current experiment uses methods from both camps, providing both isolated frequencies and spectrally complex stimuli to subjects in order to test their memory of standard frequencies.

Scholarship neither confirms or disproves the tendency of non-absolute pitch possessors to store in long-term memory the tones from the chromatic set based on A4 = 440 Hz, as opposed to the set based on A4 = 427.5 Hz, for instance, whose pitches are approximately a quarter tone below those of the A = 440 Hz set.[4] Levitin proposes that "perhaps everybody does have AP to some extent" (1994, 414). His claim is supported by his

---

[4] Octave designations in this paper conform to those put forth by the Acoustical Society of America.

experimental results in which most musicians and nonmusicians vocally produced familiar popular songs and rock songs from memory within two semitones of the pitch level at which the songs were recorded. However, Levitin does not address whether this form of AP allows listeners to reproduce frequencies at semitonal increments away from the recorded pitch level.[5] One must assume that some or all subjects did not render the songs at standard levels. Levitin documented his subjects' pitch levels—they could sing, hum, or whistle their responses—after quantizing their responses to the nearest semitone (416). Neither his hypothesis nor his method of data reporting acknowledge the question of whether listeners can reproduce the set of precise frequencies of standard Western tuning.

Most experiments dealing with long-term pitch memory in NAP possessors quantize the stimuli to the nearest semitone, thereby forcing subjects to interact with only standard frequencies. Deutsch's subjects were asked to listen to the 12 standard pitch classes; they were not asked to respond to standard *and* nonstandard frequencies. Similarly, Schellenberg and Trehub had their subjects listen to standard pitch levels only when judging "correctness" of pitch level of familiar television theme songs.

The literature addresses perception of standard versus nonstandard pitch levels only in anecdotal observation. For example, Geringer (2010) describes the performance of one AP subject who adjusted the overall pitch of a recorded excerpt of orchestral music on a smooth pitch continuum until he found the pitch level that he preferred. Geringer reports that the subject's responses were within 4 cents of the original pitch level when the subject knew the key used by the composer, and within a few cents of 100 cents, 200 cents, and 300 cents away from the original pitch level used by the composer when the subject was not familiar

---

[5] Levitin also does not address whether the recorded pitch levels of the songs used in the experiment were standard or not.

with the key in which the excerpt was written. “No one else exhibited this ability,” writes Geringer. “Perhaps this might be a method with which to test purported possessors of absolute pitch; both musicians and nonmusicians could be tested with such a procedure” (303).

A problem arises in Geringer’s hasty use of the word “ability.” His observation shows only that when this AP subject was asked to adjust the pitch level of recordings to his preferred level, he relied on his long-term memory of the collection of standard Western frequencies. The experiment does not prove that all AP possessors prefer frequencies from standard Western tuning. More relevant to this thesis, it also does not disprove that non-AP (NAP) listeners store the collection of standard frequencies in long-term memory, only that when asked to adjust the pitch to a preferred level, they tend not to settle at a standard level. Further testing is needed to separate the issue of preference from ability in an adjustment task such as Geringer’s.

It seems hasty to assume that NAP listeners automatically snap to an *a prioi* semitonal template while processing or producing tones if they have not “learned” the set of standard pitches through extensive repetition.[6] The human ear is capable of perceiving many discrete pitches that fall between adjacent semitones, an ability we’ve been aware of for more than a century. In his seminal book, *On the Sensations of Tone,* Hermann von Helmholtz (1863) wrote, “According to Waldemeyer there are about 4,500 outer arch [fibers] in the human cochlea. If we deduct 300 for the simple tones which lie beyond musical limits…there

---

[6] Learning to recognize or expect a musical event over prolonged musical training is one consequence of *statistical learning*. This process can be as informal as constant exposure to the music of a particular culture. “Auditory learning is dominated by statistical exposure,” writes David Huron (2006, 72). Perhaps musicians (and even nonmusicians) learn to recognize standard frequencies if they have been exposed to them enough.

remain 4,200 for the seven octaves of musical instruments, that is, 600 for every Octave, 50 for every semitone; certainly quite enough to explain the power of distinguishing small parts of a semitone" (147).

The smallest pitch difference that humans can perceive is known as *just-noticeable difference (JND),* and varies according to register (Butler 1992, 40). Butler writes, "For tones with frequencies up to about 1,000 Hz, the JND for inequal pitches…is 3 Hz" (40).[7] This means that within the practical music range, including that of the human singing range, listeners are capable of perceiving changes in frequency that divide the semitone from approximately three parts through twelve parts. It is conceivable that NAP listeners could therefore develop memory for a set of nonstandard frequencies because they are easily perceived as different pitches than standard frequencies.

Burns and Campbell (1994) are quick to dismiss the idea that pitches from standard Western tuning exploit natural human physiological preferences. According to these authors, "There is a complete lack of physiological evidence which would support the existence of regions of natural sensitivity, separated by semitones and consistent with A-440 tuning, along either pure- or complex-tone frequency continua" (2717). The authors add that "the perception of sequential frequency ratios in relative pitch is also characterized by a similar lack of physiological evidence for natural boundaries" (2717). This suggests that recognition of intervals by NAP listeners does not depend on absolute pitch level. Indeed, A4 has not always been performed at 440 Hz, or even at whole numbers of semitones away from 440 Hz, but has varied along a spectrum of frequencies over the last few centuries.

---

[7] This is a generalization; others may disagree. Kollmeier, Brand, and Meyer (2008) state that JND is about 3 Hz for frequencies below 500 Hz only (65).

Just how precise is memory for absolute pitch among NAP listeners? The studies by Levitin, Geringer, Schellenberg and Trehub, and Deutsch cloud the answer to this question. As mentioned before, Levitin's study does not take within-semitone pitch levels into account, leaving the reader to wonder whether listeners choose standard levels or close-to-standard levels when reproducing familiar songs. Geringer implies that NAP listeners do not care at what level they hear pitch stimuli, but, as already noted, subject preference of nonstandard pitch levels does not prove that NAP listeners cannot distinguish between standard and nonstandard levels. Schellenberg and Trehub and Deutsch do not even give their subjects a chance to respond to stimuli from both standard and nonstandard pitch levels.

The issue of recognition versus recall, two distinct types of memory tasks, also suggests that the previous experiments do not tell the whole story of long-term pitch memory in NAP listeners. Except for Levitin, all of the experimenters cited thus far tested only recognition of pitch levels, not recall. Deutsch, for example, tested recognition of certain pitch classes that were given as experimental stimuli. No subjects in her study or in similar studies of absolute pitch have ever been asked to recall or produce pitches with the aim of measuring their conformance to standard versus nonstandard pitch levels. George Mandler (1980) describes the mental task of recognition as a search for whether an item has been previously encountered. According to Mandler, the recognizer does not need to be able to identify the item; identification is a subsequent process that places the item "within the relational network of long-term memory" (253).[8] Recall, a process in which "the context is present and a missing event is sought," can be more rigorous, especially when the subject is tasked with conjuring the context without external cues, such as being asked to sing a familiar song given no aural stimuli.

Vanzella and Schellenberg (2010) claim that AP is only evident among those with musical training; they posit that listeners who have not learned to associate note names with specific pitches will not exhibit signs of AP (1). Ross et al., (2003) found an exception by which a nonmusician was determined to possess AP using a test that did not involve naming pitches. This ability involves representing absolute pitches in long-term memory, a task that Levitin suggests all listeners can do (Levitin, 415). This musician most likely recognized pitch classes but had not learned to link names to them.

This sort of absolute pitch memory comprises the first part of Levitin's two-component model of absolute pitch. Levitin refers to this ability when he posits that "perhaps everybody does have AP to some extent." The second ability in Levitin's model is verbal labeling of pitch. To explain the phenomenon, McLachlan and Wilson posit that "[reporting the pitch name as well as a microtonal pitch variation] could occur by [comparing] in auditory short-term memory…the pitch information streamed through the auditory core with recalled pitch information for a given pitch class" (189). It seems that the first component of the model may rely on the ability to compare pitch information stored in short-term memory

---

8. D. A. Norman (1968) holds that a subject must recognize the object that is being encountered, and cannot simply recognize that it has been encountered before. Most psychological research on recognition and recall involves words, and may not have direct application to pitches, but the lack of agreement on the precise mechanism involved in recognition may suggest that in order to recognize a pitch as standard, one needs to be able to identify the note, i.e., assign it a pitch-class label.

with an acquired mental template of the set of standard frequencies. Instead of identifying heard pitch classes by name, putting each pitch into one of 12 categories, perhaps listeners without AP can identify the standard vs. nonstandard quality of pitches, putting each pitch into one of 2 categories.

Further support for Levitin's two-component model is found in Deutsch's 1991 experimental study. Her results show that during an aural illusion known as the tritone paradox, listeners tend to describe the direction of pitch changes according to the relative placement of the starting pitch on the pitch-class circle. She attributes these results to "the language or dialect to which the listener has been exposed, particularly in childhood" (2006, 2). Although listeners hear tones in one region of the pitch-class circle as the higher tone in the pair and those in the complimentary region as lower, they cannot necessarily identify the tones by name. According to Deutsch, this indicates that the difference between AP and NAP listeners lies in verbal labeling abilities, not in long-term memory, corroborating Levitin's two-component model of AP (2006, 2), and suggesting that NAP listeners exhibit the first component.

All this evidence suggests that NAP listeners might be able to tell when a given frequency matches one from the standard set of frequencies and when it does not. Yet, NAP listeners are generally assumed not to display tuning preferences for isolated frequencies or overall pitch levels. One might argue that since NAP listeners lack the frame of reference for pitches that allows AP listeners to label them or reproduce them accurately, NAP listeners cannot discriminate between "in-tune" and "out-of-tune" frequencies when they are presented out of context, as an individual tuning note, for example. However, because long-term memory for pitch and labeling ability are separate skills, as demonstrated by Deutsch and Levitin, further testing is needed to investigate the within-semitone precision of that pitch memory in NAP listeners.

B. Thesis Statement

The literature review confirms that AP listeners can associate specific frequencies with consistent labels better than NAP listeners. It further reports that NAP listeners do possess long-term memory for absolute pitch on the order of one or two semitones. However, it does not address the within-semitone accuracy of that memory. This thesis tests whether NAP musicians retain the collection of pitches of standard tuning in long-term memory, rather than close approximations of frequencies. The hypothesis of this thesis is twofold: that within-semitone, or "standard" pitch memory is common in trained musicians without AP *and* that repeated exposure to these standard frequencies engrains pitches into the long-term auditory memories of trained musicians who engage in musical activities that conform to this tuning system. The frequencies that will be referred to as "standard" in this thesis are those from the collection of 12 chromatic pitch classes in equal temperament based on A4 = 440 Hz. Most trained musicians who have worked through a university musical education have interacted with this set extensively through both listening and production tasks, especially because of the ubiquitous use of the keyboard in classroom instruction.

Through a four-part experiment, this thesis aims to determine if musicians without AP remember—through recognition and production, not labeling—the precise frequencies of standard Western tuning.[9] It will also examine the effect of the amount of exposure to these standard frequencies by testing both freshmen and seniors at the University of Massachusetts Amherst. By comparing students in their first year of collegiate-level musical training with those in their fourth year, the study investigates whether the three years of concentrated exposure to standard frequencies that separates these groups enables seniors to remember the frequencies from standard Western tuning better than freshmen. This exposure includes both listening and singing activities that students do in their music theory, aural skills, and history classes as well as their ensemble work and private study. The decision to test both freshmen and seniors in this experiment was predicated on the assumption that the majority of the frequencies they listen to during their college years are standard frequencies or very close to standard.[10]

Although a marked difference between freshmen and seniors likely would indicate that the amount of exposure that a musician receives does improve their performance on the four tasks, the *absence* of such a difference would not disprove the hypothesis that exposure to certain frequencies improves memory of them. It is possible that the formative period for a set of frequencies to be embedded in long-term memory occurs in young people before they

---

[9] To meaningfully compare AP listeners with NAP listeners would require more subjects than this study was able to recruit. The results of the three AP subjects tested will be reported incidentally, but should not be used as a statistically sound baseline to which to compare NAP subjects.

[10] It would be impossible to survey the pitch levels of all frequencies to which the subjects are exposed during college. Some information is known, however. The director of bands at the University of Massachusetts Amherst confirmed that the Wind Ensemble and Symphony Bands tune to a standard level, but that the pitch level at which those ensembles rehearse is often "slightly higher." The 2012–2013 director of the University Orchestra reported that that ensemble tuned to A = 440 Hz in that year. The University piano tuner confirmed that pianos in the Fine Arts Center are tuned once or twice a semester to A = 440 Hz. The marching band director reported tuning that ensemble to A = 442 Hz.

reach college age. If this is this case, it is also possible that subjects have received significant exposure to standard frequencies before arriving at college and that further exposure during their college years does not have any effect on their performance during this experiment.

CHAPTER 2

THE PRESENT STUDY

A. Experimental Tasks

All experimental procedures were approved by the Institutional Review Board of the University of Massachusetts Amherst and subjects gave written informed consent before participating. Before starting any of the tasks, subjects watched and listened to a four-minute slide presentation created for this experiment. It was designed to help subjects conceive of the continuous pitch spectrum and the precise definitions of the terms “standard” and “nonstandard” frequencies as they were used in the experiment. The only sounds presented to subjects during this presentation were a speaking voice and a sine tone that swept upward in pitch like a glissando.

The first two tasks tested subjects’ abilities to recognize standard and nonstandard frequencies, explicitly in a rating test, and implicitly in a just-noticeable difference test. In the folk-song task, subjects converted recalled pitch into sound, singing a familiar tonal melody at a pitch level of their choosing. No aural stimuli were presented before subjects sang. The adjustment task tested subjects’ abilities to discriminate between standard and nonstandard pitch levels in a musical context.

Frequencies were categorized as standard or nonstandard, standard being the equal tempered frequencies based on A = 440 Hz. A small range of frequencies clustered around each standard semitone were considered to be “standard” pitches for the folk-song and adjustment tasks; this range varied from subject to subject according to his or her individual just-noticeable difference for pitch, which was obtained for standard and nonstandard frequencies collectively in the second task.

Frequency must be distinguished from *pitch* in order to study cognitive processing of frequency. Frequency is a physical measurement of the number of vibrations of an object over a specified time interval. Pitch is a reflection of a listener's perception of frequency. Butler notes that "Backus (1977), p. 127, states that pitch perception is 'essentially nonexistent' above 7,000 or 8,000 Hz, although W. D. Ward…has found some indication that listeners can identify octaves extending up to 10,000 Hz" (205). Frequency is an objective value corresponding directly to a physical property of sound, whereas pitch is subjective, and depends on the auditory and cognitive abilities of the listener. A more striking example of subjectivity in pitch is JND. If a listener has a JND of 3 Hz for the frequency 440 Hz, then the two frequencies 440 Hz and 442 Hz, when played melodically, will seem to be the same pitch for that listener. Another listener might be able to perceive the difference between the two frequencies, hearing those tones as discrete pitches.

B. Subjects

The experimental subjects included seven seniors and five freshmen, all music majors at the University of Massachusetts Amherst.[11] Subjects who completed all four tasks were entered in a drawing to win a gift card. Two seniors and one freshman reported having AP. Subjects were drawn from various instrumental and vocal majors. At the conclusion of their participation in the study, each subject filled in a questionnaire which asked for country of birth, how long subjects lived there, first language, the country in which they received most of their musical training, and the instruments or vocal type on which they had trained.

---

[11] One "senior" was an undergraduate in her sixth year of study at the University of Massachusetts Amherst.

Subjects were asked to report any significant exposure to tunings other than A = 440 Hz that they had used in their studies, such as Baroque or microtonal tuning. Appendix A presents subject profiles based on the information they provided. No formal analysis was done in order to draw correlations between the experimental results and subjects' musical backgrounds. However, a future experiment might rely on collecting as detailed information as possible that would indicate if subjects had repeated exposure to the collection of standard Western frequencies or some other collection, thereby facilitating statistical learning of the frequencies heard most often.

CHAPTER 3

RATING TASK

A. Method

The rating task tested subjects' ability to decide if tones were from standard or nonstandard pitch levels. Subjects listened to sine tones of both standard and nonstandard frequencies and rated how certain they felt that each frequency was standard or nonstandard. The tones presented were 48 pitch-classes in eighth-tone increments within a two-octave range above and including G3, totaling 12 standard frequencies and 36 nonstandard frequencies. The order of tones was randomized, with the restrictions that successive tones were not in an octave or compound-octave relationship, nor were successive tones within one semitone of each other. Additionally, pitch classes were not repeated during the 48 trials, that is, no two tones shared octave equivalence. The restrictions on randomization of the order of tones were designed to make it difficult for subjects to use relative pitch strategies to make their decisions.

Each tone sounded for two seconds, and was followed by a seven-second pause. About two seconds before the next tone was to begin, the number of the next tone was announced to help subjects keep their place. Subjects were told to rate their confidence in whether each tone was a standard or nonstandard frequency during the silence following each tone. They indicated their decision using a scale of 1 to 7 on a response paper. A response of "1" indicated that they were certain that a tone was a nonstandard frequency, and a response of "7" indicated their certainty that the tone was standard.

B. Data Interpretation

The data from the rating task was analyzed according to a method from psychology known as *signal detection theory.* Subjects' confidence ratings were used to determine their skills in discriminating between standard and nonstandard frequencies. The number of times each subject used each response category (how often they responded "1," "2," "3," etc.) was tabulated separately for standard tones and for nonstandard tones. The totals in each category were then converted into conditional probabilities by dividing standard totals by 12, the number of standard tones presented, and nonstandard totals by 36, the number of nonstandard tones presented. Next, seven separate cumulative probabilities were calculated to show how often a subject selected a response category equal to or lower than each response category. For example, the cumulative probability for the response "3" for standard tones is equal to how often a subject responded within the range of 1–3.

If a subject did not use the outer response categories, such as 1 or 7, for either standard or nonstandard frequencies, that category was collapsed inward and added to the next inner category. For instance, if a subject did not respond "1" to any standard tones but did respond "1" to at least one nonstandard tone, the total of "1s" for nonstandard tones was added to the total for the "2" category, and the "1" category was discarded entirely. This led to varied numbers of data points among subjects; Subject 3 had only 3 data points because she did not use the response "1" at all, nor did she use "2," "6," or "7" for nonstandard tones. Most subjects had seven data points, the maximum possible.

Each cumulative probability was transformed into a z-score using the NORM.S.INV function in Microsoft Excel 2011, which assumes normal distribution, a mean of zero, and a standard deviation of 1. The resulting data were then plotted on z-coordinates as a receiver operating characteristic (ROC), a type of graph used to show subjects' accuracy in judging the attributes of stimuli. For this experiment, the x-axis represents standard response probabilities, and the y-axis, nonstandard response probabilities, and correct detection of nonstandard tones was treated as a "correct" response. "False alarms" were graphed on the x-axis; this was when a standard tone was incorrectly identified as nonstandard. "Hits" were graphed on the y-axis; this was when a nonstandard tone was correctly identified as such. This method of graphing tracked the number of true positive answers against the number of false positive answers, taking the full range of response categories into account, thus showing how well each subject performed.

According to Stanislaw and Todorov (1999), the false alarm and hit rates reflect both response bias and sensitivity. Bias is a subject's tendency to respond either way to a stimulus—in this case, either that a tone was a standard or nonstandard frequency. They define sensitivity as "the degree of overlap between the signal and the noise distributions" (139). Subject sensitivity—a measure of the success of subjects in detecting the "signal," or in this case, nonstandard tones—can be represented by various statistics in signal detection theory. One such commonly used index is d', which compares the means of the signal and noise distributions, and assumes equal variance in those distributions (Macmillan and Creelman, 7). An alternative index is the area under each subject's ROC, also known as A_z. Macmillan and Creelman suggest using this index when there is unequal variance in the signal and noise distributions. A_z was calculated using Excel's NORM.S.DIST function and that subject's D_{YN}.[12] A_z is a typical index of subject performance in rating experiments: this

---

13. A_z is geometrically verifiable as $\mathbf{\Phi}(d_a/\sqrt{2})$, where $d_a = D_{YN}\sqrt{2}$. D_{YN} is the shortest distance between the zROC and the origin.

numerical value increases from .5 to 1.0 as subject sensitivity increases (MacMillan and Creelman, 2005, 63). This index allowed for comparison of the performance of freshmen and seniors and of individual subjects.[13] A one-tailed t-test was performed using the TTEST function in Excel. This type of t-test test can be used in signal detection theory when one of two groups of subjects is expected to perform better than the other. This difference is considered statistically significant when $p \leq .05$ (Windsor, 2004, 213–214). In this case, the t-test calculated the likelihood that the difference between the two groups—freshmen and seniors—was due to actual difference between the two groups and not due to chance.

C. Results

The ROCs plotted on z-coordinates for each subject are shown in Figure 1a for seniors and in Figure 1b for freshmen. The false-alarm rate is shown on the x-axis, and the hit rate on the y-axis. Figure 2 compares A_z scores for freshmen and seniors: $t(7) = .48$, $p = .32$. The t-test shows no evidence that the experience gained during collegiate study made a difference in subjects' ability to discriminate between standard and nonstandard pitches in the rating task. A_z ranged from .53 to .85. The three AP possessors who completed this task all scored above .70; one NAP possessor in this task scored .67, and the remaining NAP possessors all scored below .59, indicating that NAP possessors have a relatively poor ability to discriminate between standard and nonstandard tones.

---

14. Because this experiment only recruited 3 AP possessors, it is not statistically sound to compare them to NAP possessors in these four tasks. Trends are suggested by the data, however. These will be discussed last in this chapter and in Chapter 7.

Figure 1a. zROCs for seniors.

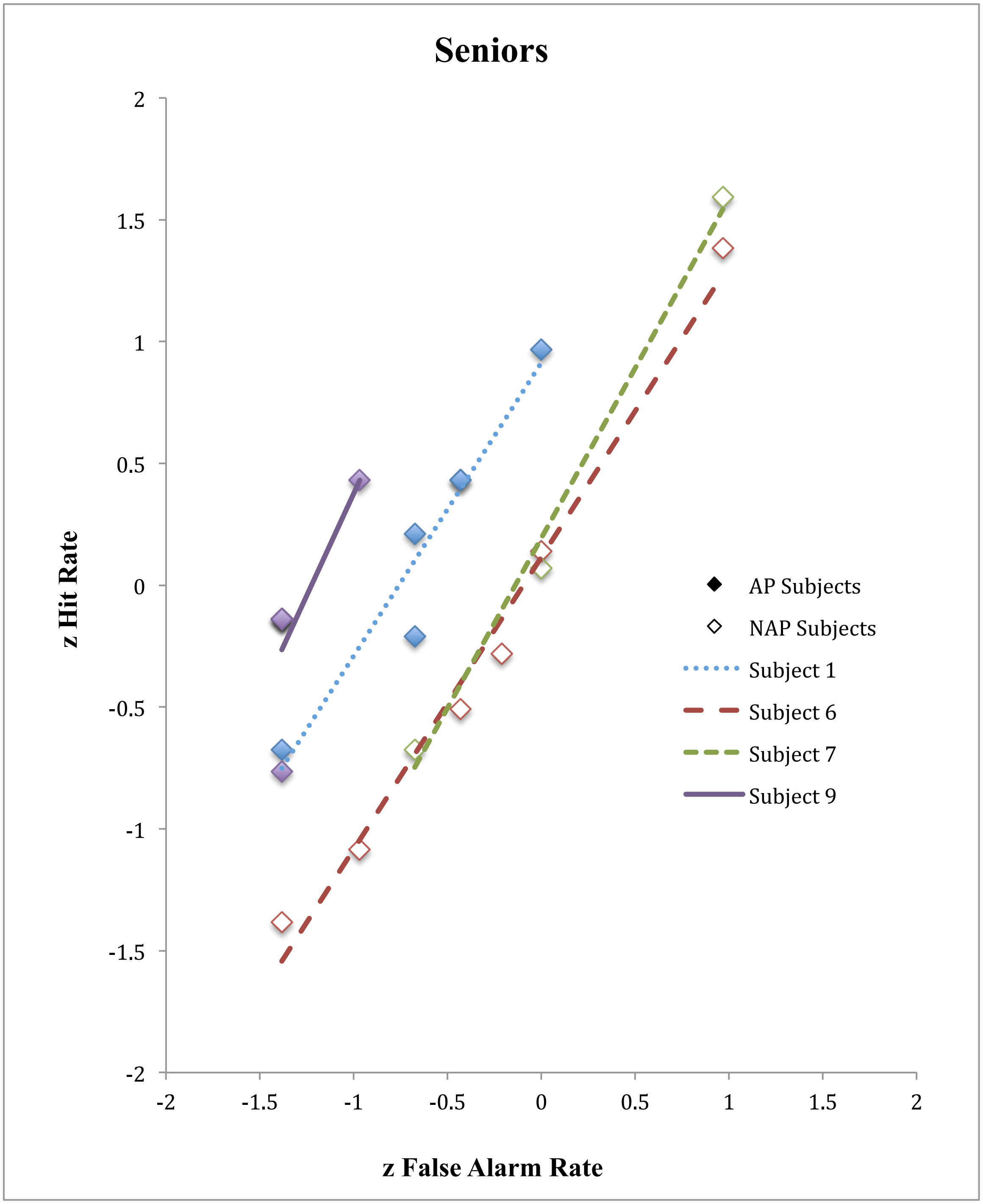


Figure 1b. zROCs for freshmen.

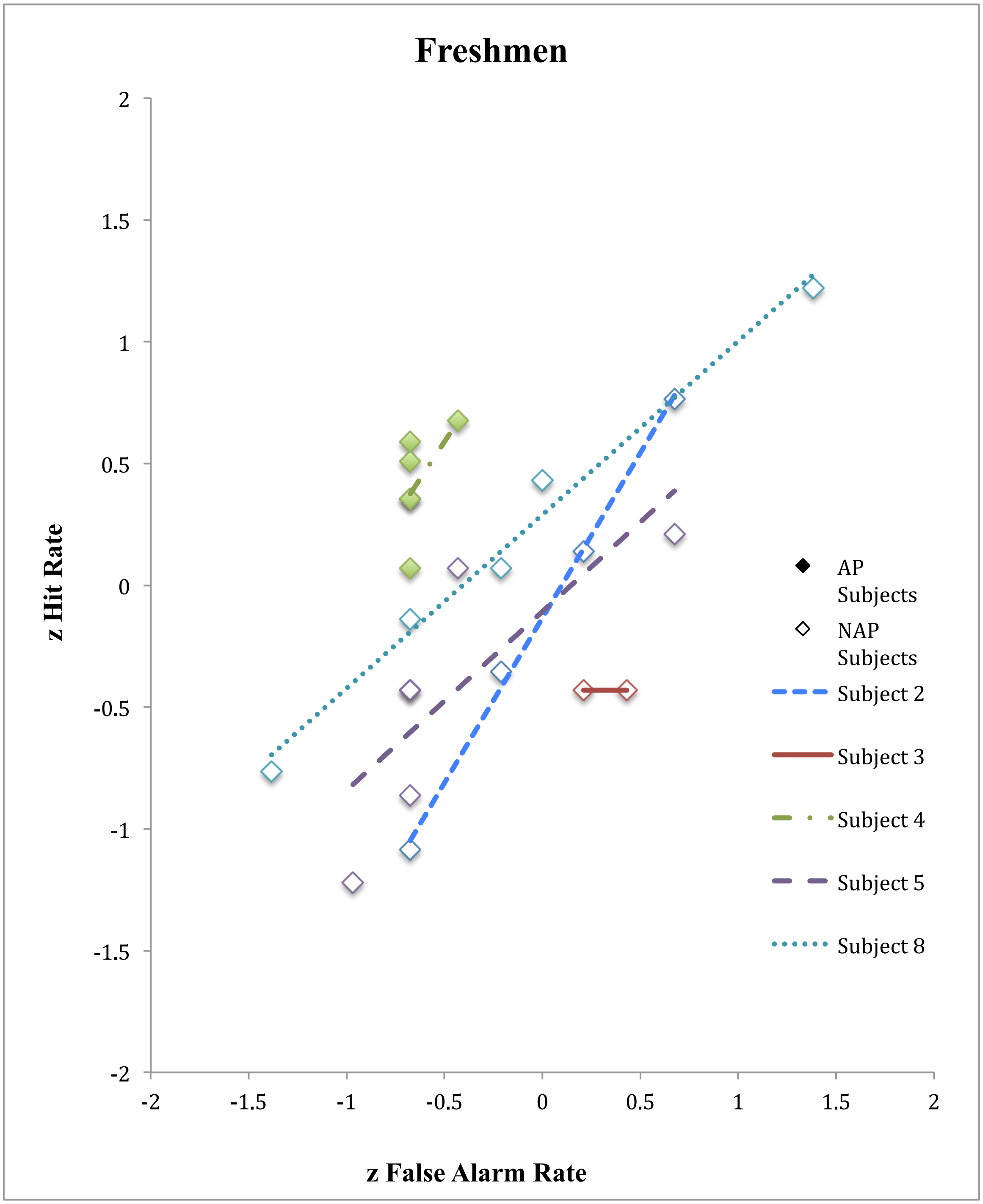
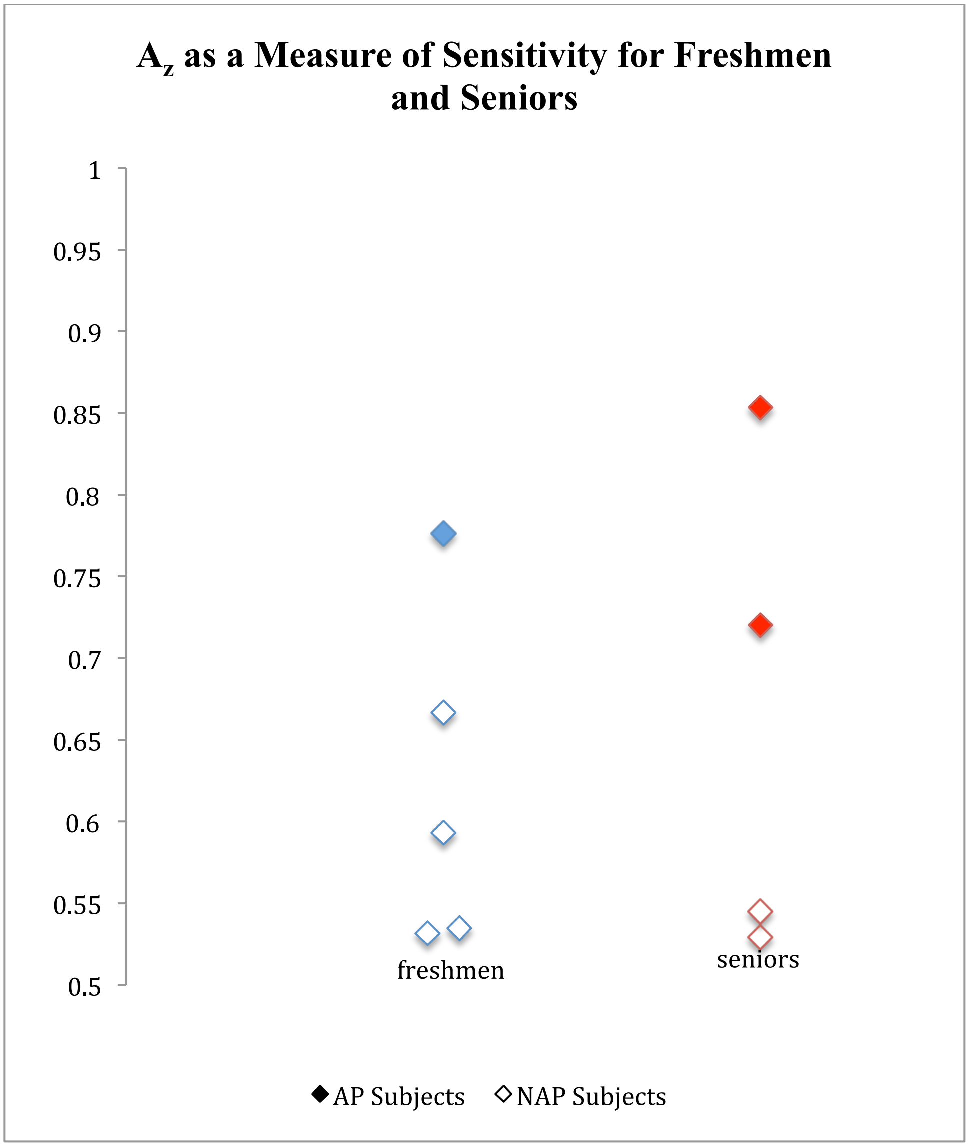


Figure 2. A_Z **scores as a measure of subject detection of standard and nonstandard tones, freshmen and seniors compared.**



CHAPTER 4

JND TASK

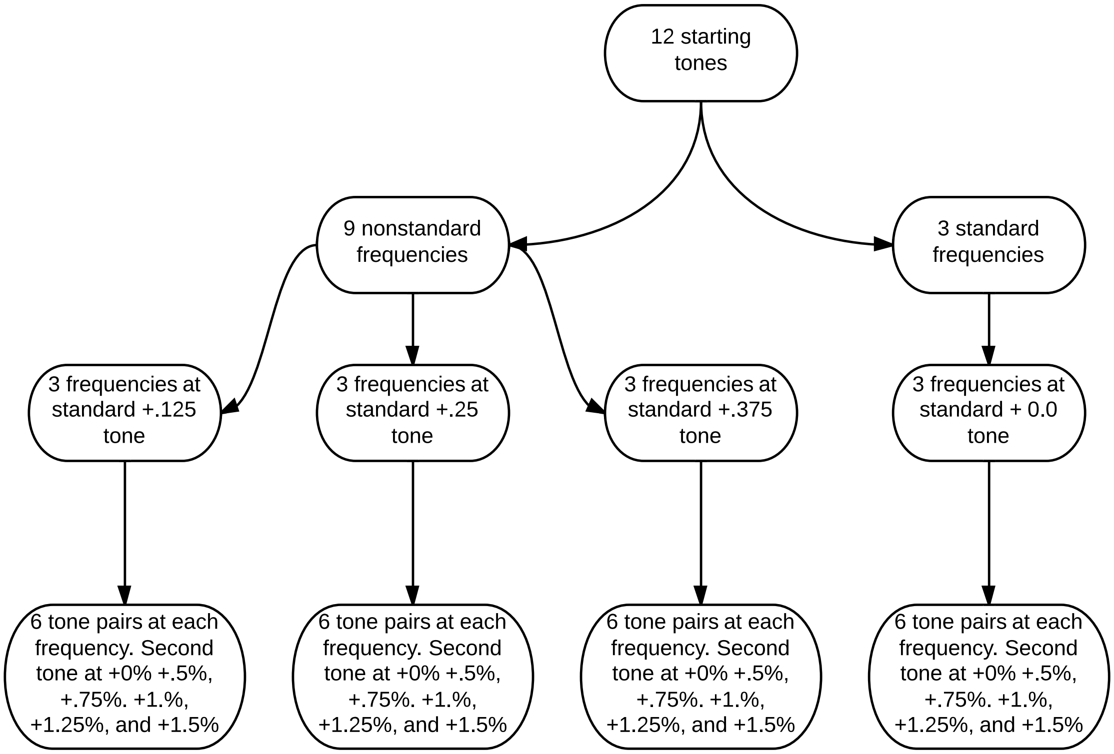
A. Method

The rating task asked subjects to consciously decide whether they thought tones were standard or nonstandard, whereas the just-noticeable difference (JND) task tested whether the standardness of frequencies affected their ability to detect small pitch changes. In the JND task, subjects listened to 72 pairs of tones. The second tone in each pair was usually slightly higher in frequency than the first. The difference in frequency between the two tones in each pair varied among 6 different percentages. In order to test whether subjects' JND was different for standard frequencies than for nonstandard frequencies, the first tone in each pair was either standard or nonstandard. If subjects were better able to detect small pitch changes when the first tone was standard, for instance, this would indicate that subjects discriminate between standard and nonstandard frequencies when performing a JND task.

In order to select the stimuli for this task, the frequencies of all 24 standard semitones between G3 and F#4 as well as the frequencies in eighth-tone increments above each standard frequency in this range were assigned an integer label, 1 through 96. The integers increased as frequency increased. Twelve integers were then selected to represent so as to provide a near-even spread across the available range, and so as to include three instances of all four "levels" of pitch. These levels were standard, standard + .125 tone, standard + .25 tone, and standard + .375 tone. (The four instances of G3 will be indicated hereafter as *G3, G+3, G++3,* and *G+++3.*) These twelve starting pitches were each presented 6 times as the first of a tone pair throughout the task. The second tone in each pair differed from the first by various small percentages of frequency in order to test subjects' just-noticeable difference.

Each starting tone was paired either with its identity or a tone .5%, .75%, 1%, 1.25%, or 1.5% above it (measured in Hz). Figure 3 summarizes the organization of the stimuli.

Figure 3. Organization of the tone stimuli presented to subjects in the JND task.



Tones were one second in length, with one second between members of a pair, and six seconds between each pair. The order of tone pairs was randomized. Subjects indicated on a response paper whether they thought the two tones in each pair were the same pitch or different. By including both standard and nonstandard pitches as the first tone in each pair, this task determined if a subject's JND is different for standard and nonstandard tones.

B. Data Interpretation

For the JND task, three just-noticeable differences were calculated for each subject: one for all tones considered together, and one each for standard and nonstandard tones.[14] One graph was created for each of these conditions for each subject using an Excel file of the data and a script created by Andrew Cohen for the statistical computing software R.[15] The graphs show the percentage of trials for which subjects said that the two tones in a pair were different. Percentage of trials is plotted on the y-axis for each of the 6 pitch percent differences plotted on the x-axis.

To determine a subject's JND for the conditions of standard tones, nonstandard tones, or both, R computed the pitch percent difference at which that subject achieved 75% accuracy for that condition.[16] These percentages were compared between freshmen and seniors and within subjects for standard and nonstandard pitches. JNDs for each of the four levels of starting pitch (standard, standard + .125 tone, standard + .25 tone, and standard + .375 tone) were also compared for each subject.

---

15. The overall JND was set aside for use in the interpretation of the folk-song and adjustment data.

16. Andrew Cohen (University of Massachusetts Amherst) is a member of this thesis committee. R is a free "language and environment for statistical computing and graphics," according to its website. It was originally written by Robert Gentleman and Ross Ihaka.

17. A 75% threshold of accuracy in judgment was chosen to compare the percent pitch change necessary for subjects to achieve that threshold of accuracy for both standard and nonstandard tones. The *point of subjective equality* from psychometrics is the value of change in stimulus which subjects can detect 50% of the time (MacMillan and Creelman, 1991) and is often used in JND tasks. However, this threshold is susceptible to response bias, and a higher percentage was desirable to accurately reflect subjects' detection abilities.

C. Results

Each subject's overall just-noticeable difference for both standard and nonstandard frequencies is shown at the top of Table 1. As mentioned previously, a threshold of 75% percent correct was chosen in order to compare subjects' pitch discrimination abilities. JNDs for standard and nonstandard frequencies are also presented in Table 1. Note that Subjects 2 and 6 "maxed out" for standard tones, that is, they did not achieve 75% accuracy even when the difference between the two tones in a pair was increased to 1.5%. Subject 7 maxed out for nonstandard tones. The JNDs provided in Table 1 for these three subjects are extrapolated from the rest of their data. Whereas Subject 6's JND for standard and nonstandard tones was similar, Subject 2 and Subject 7 displayed a marked difference in JND for standard and nonstandard tones.

Table 1. Overall JNDs for standard and nonstandard tones combined, and for standard and nonstandard tones separately, by subject. "Nonstandard" refers to all levels of nonstandard frequencies. Values are given in percent Hz difference.

Subject	**1 (AP)**	**2**	**3**	**4 (AP)**	**5**	**6**	**7**	**8**	**9 (AP)**
Overall JND	1.03	1.50	1.31	1.26	1.37	1.50	1.44	1.08	1.10
Standard	0.77	2.13	1.31	1.38	1.05	1.54	0.82	1.00	1.21
Nonstandard	1.07	1.30	1.27	1.19	1.43	1.44	1.54	1.04	1.04

Figure 4a graphs the difference between each subject's JNDs for standard and nonstandard tones. This difference did not exceed chance levels; a one-tailed t-test yielded $t(8) = 0.11$, $p = 0.46$, indicating no statistically reliable difference between JND for standard and nonstandard tones. Figure 4b compares JNDs for standard and nonstandard tones for freshmen and seniors. For standard tones, $t(7) = 1.0765$, $p = .16$, and for nonstandard tones, t(7) = 0.2095, p = .42, indicating no significant difference between freshmen and seniors in either of these two categories.

Figure 4a. Difference between standard and nonstandard JNDs for each subject. Nonstandard JND was subtracted from standard JND.

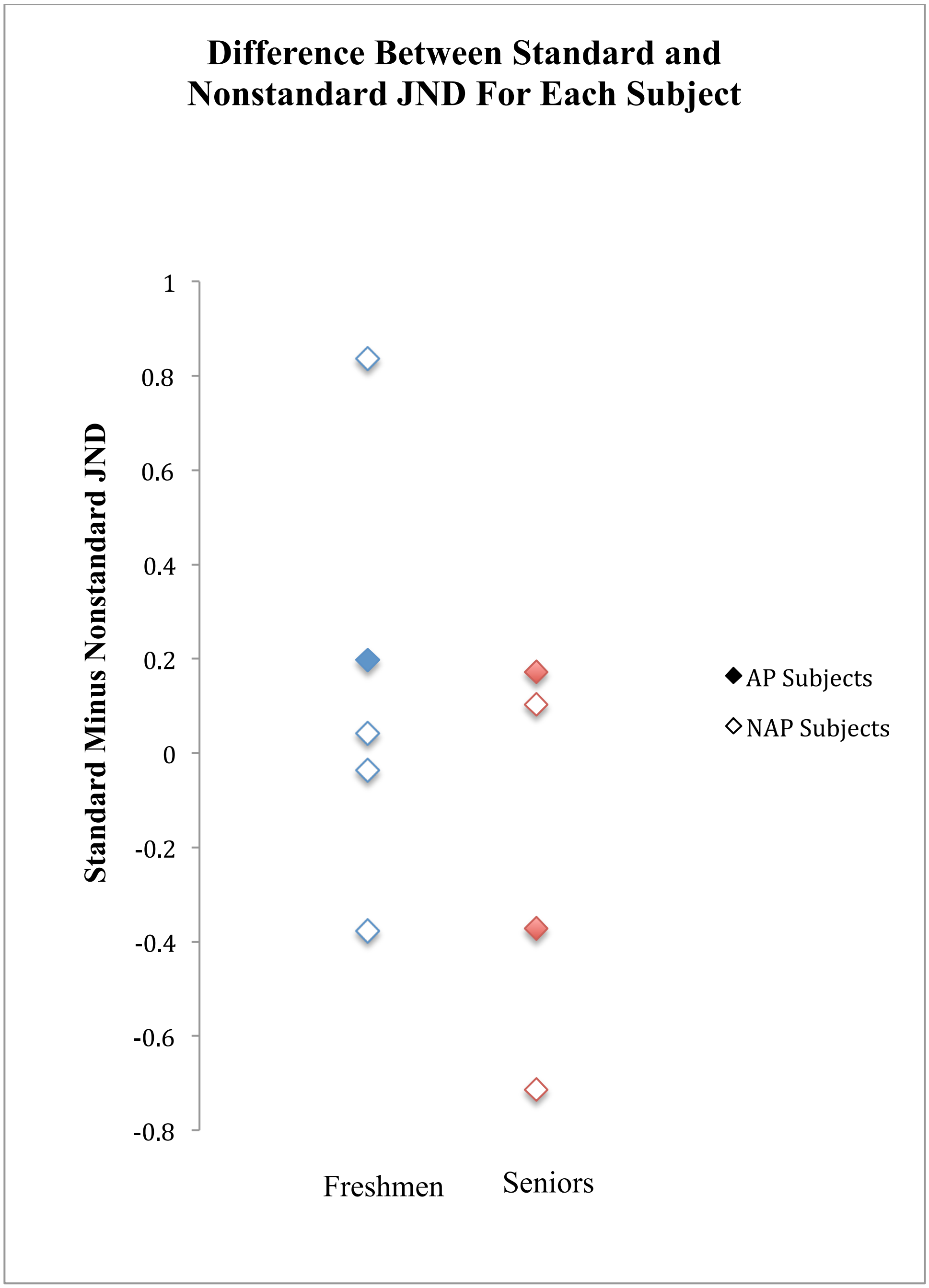
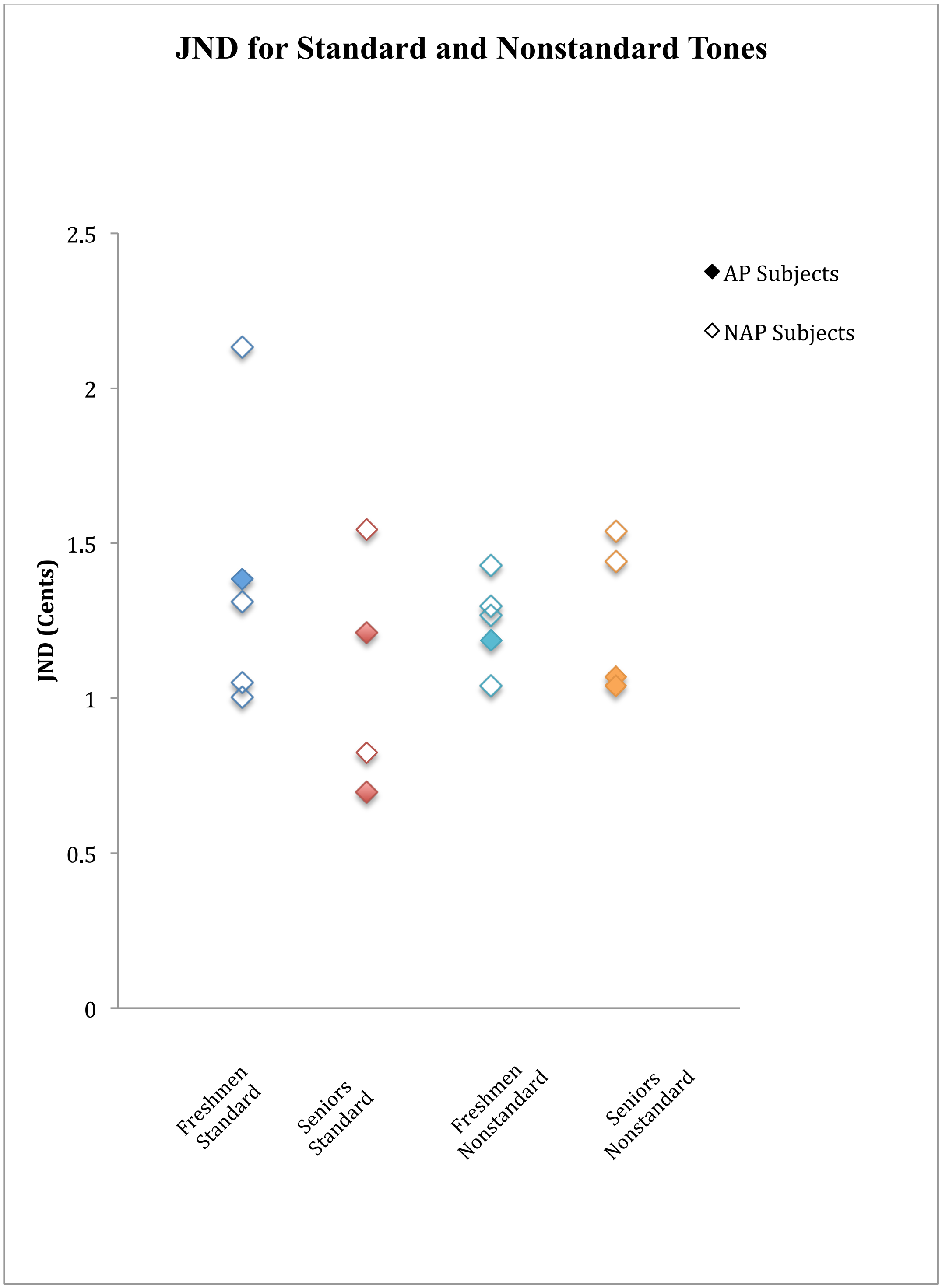


Figure 4b. Standard and nonstandard JNDs of freshmen and seniors compared.



CHAPTER 5

FOLK-SONG TASK

A. Method

The latter two tasks—the folk-song task and the adjustment task—used actual music instead of isolated tones. The folk-song experiment—a recall task—tested integration of the pitches of standard Western tuning in long-term memory. Upon entering the testing room, each subject was told to sing a simple folk tune from memory with no outside pitch stimulus.[17] They were instructed to pick a tune from a brief list or to choose one of their own in a similar style and to sing as much as they could remember using the words or the syllable “la.”[18] Their voices were recorded using a Zoom H2 Handy Portable Stereo Recorder and were then imported into Praat, a free speech analysis program (Boersma and Weenink 2012). The pitch levels were analyzed to determine if subjects chose standard or nonstandard frequencies.

B. Data Interpretation

The average frequency of the last pitch sung by each subject (the tonic pitch) was obtained using Praat’s “get pitch” function. Portions of the spectrograms shown in Praat were selected manually for frequency analysis through a point-and-highlight method. The span

---

18. This task and the subsequent adjustment task took place in the practice rooms at the University of Massachusetts Amherst, which are not soundproof. In an attempt to mask ambient instrumental or vocal sounds from adjacent rooms, brown noise was played through speakers as subjects entered the testing room and continued until subjects began singing in the folk-song task. Brown noise resumed before the start of the first excerpt of the adjustment task, as well as in between trials of the adjustment task.

19. All subjects chose a song from a list provided to them. This list is included in Appendix B, with the syllables on which subjects sang the songs.

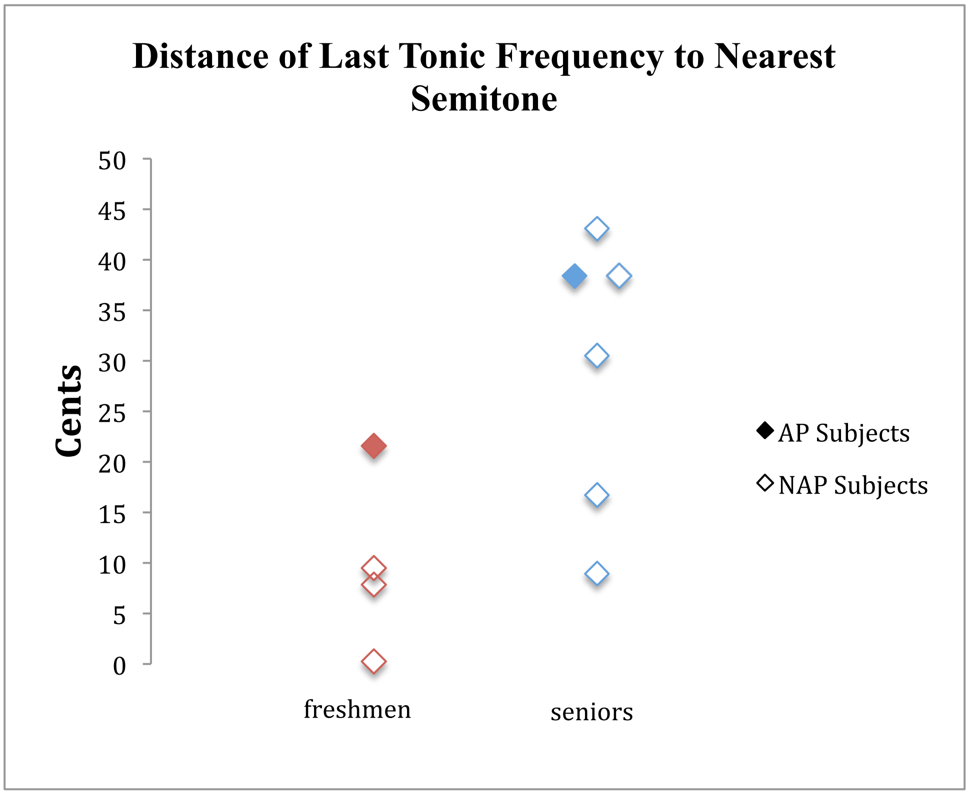
lengths were variable based on the length of the notes sung by the subjects. The distance of each subject's last pitch to the nearest standard semitone was calculated in cents, and the data is expressed in a scatter plot. Freshmen were then compared to seniors using a t-test to see if one group tended to sing closer to a standard pitch level.

C. Results

Nine subjects completed the folk-song task, but only seven of those subjects also completed the JND task. Four out of seven subjects for whom JND was known (Subjects 2, 3, 5, and 7) sang the last tonic within one JND of a standard semitone. None of these four subjects had AP. This is a significant result, given that the average JND of the seven subjects was just 17.43 cents, approximately 1/3 of the maximum 50-cent distance away from the nearest semitone. Although the JND of Subject 10 was not known, his ending frequency was 8.94 cents away from standard, likely within that subject's JND. Curiously, Subject 9, an AP listener, finished "Happy Birthday" 38.39 cents away from the nearest standard level. This could be due to unreliable vocalization technique, limitations of that subject's AP, or other factors that this experiment did not control.

All subjects' folk-song results are plotted in Figure 5, with freshmen contrasted with seniors. A two-tailed t-test (used for testing for a difference between two groups without prior idea about the direction of the difference) yielded $t(8) = 2.51$, $p < .05$. Freshmen were better than seniors at approximating standard frequency levels when they sang, again providing no evidence that collegiate training improves the long-term memory of standard pitches in a production task. Note the incredibly close-to-standard response of Subject 5, at just 0.24 cents away from standard.[19]

Figure 5. Distance of the last sung frequency to the nearest standard semitone.



Subjects 2, 3, 5, and 7 sang within one JND of standard, but they did not show particular ability to discriminate between standard and nonstandard tones in the tasks that did not involve singing. A possible exception is that Subjects 2 and 7 had considerably different JNDs for standard and nonstandard frequencies. Their A_z scores were 0.53, 0.67, 0.53, and 0.54, demonstrating poor detection abilities in the rating task. In the adjustment task, their

---

20. Such a small number may be sensitive to measuring biases like software precision and manual selection of the span of sound to be analyzed.

ending pitch levels were 0.31, 0.25, 0.34, and 0.28 semitones away from standard. (The results of this task will be discussed in more detail in the next section.) Refer to Table 1 to compare these subjects' JNDs for standard versus nonstandard levels; Subjects 2 and 7 were considerably better at detecting small pitch changes in one category.

Although subjects chose which syllables to use to sing the folk-song—seven used the words of the song and three used "la,"—syllable choice didn't create a statistically significant difference in the distance of the last sung pitch to the nearest semitone. An unpaired two-tailed, type-two t-test found that $t(8) = .5848$, $p = 0.57$.

CHAPTER 6

ADJUSTMENT TASK

A. Method

The adjustment task tested subjects' recognition of standard pitches in a musical context. Schellenberg and Trehub (2003) posit that "isolated tones are musically meaningless to all but AP possessors," and that "the absence of explicit memory for pitch level does not preclude relevant implicit knowledge" (262). Because Schellenberg and Trehub note that conventional AP tests use amusical stimuli (isolated tones), and therefore tend to fail to identify many subjects with good long-term absolute pitch memory, they presented subjects with familiar instrumental excerpts from television shows. The excerpts were shifted in pitch level by one or two semitones, with what the authors reported as "no discernible effect on tempo or overall sound quality" (263). Their experimental design motivated the adjustment task in part. The method of the current adjustment task is inspired by a 2010 study by John Geringer (298–300). His subjects listened to recordings whose overall pitch levels had been manipulated, and then had the opportunity to change the pitch levels on a smooth continuum until they were satisfied with the result.

Following Geringer's 2010 study incorporating manual pitch adjustment, the current subjects listened to four pieces of commercial-quality recordings of piano music with the aim of deciding if they were presented at a standard pitch level. Subjects were told to adjust the recordings as they played to standard if they were not. The recordings were Chopin's Etude No. 1 in A-flat Major, Brahms's Rhapsody in B Minor, Liszt's *Mephisto Waltz No. 1* (originally scored for orchestra), and Schumann's *Papillons.* A complete discography is presented in Appendix C.

Before they were presented to subjects, the excerpts had their overall pitch level changed by varying amounts, using Audacity 2.0.2, a free audio editor and recorder (Mazzoni and Dannenberg, 1999–2012). The overall pitch level of each recording was measured using the software Sonic Visualiser (Chris Cannam and Centre for Music at Queen Mary, University of London, 2005–2013). The Chopin was presented to subjects at 44.66 cents lower than E5, the nearest standard semitone level. The Schumann was presented at 19.38 cents higher than C3, but more than a semitone lower than the pitch level at which the piece was composed. The Liszt was presented at 29.93 cents lower than standard, and the Brahms at 26.85 cents lower. The Brahms recording was already at 26.85 cents below the nearest semitone; the overall pitch level of the recording was not manipulated. The Brahms recording had its pitch level changed twice—upward and then downward again—to protect against the biasing presence of sound artifacts only being in the "nonstandard" recordings.[20]

Recordings were played through the software Amazing Slow Downer (Roni Music), which enabled subjects to make their pitch manipulations. For each recording, subjects used a handheld remote control to change the pitch in cent increments until they thought the pitch level was standard. Subjects could use individual clicks of the up and down buttons to raise

---

22. The pitch level of the Brahms recording was raised approximately 30 cents up and then 30 back down, since the average pitch manipulations of the other three recordings was approximately 60 cents. Gary Karpinski notes that despite the effort to add sound "artifacts" to the untransposed Brahms recording (cues in the sound resulting from electronic manipulation), transposing the other recordings up or down from the original level shifts the formants present in the sound, a change which cannot be replicated in the Brahms recording which was presented at the original pitch level. Formants are frequency bands where certain overtones are amplified, depending on the shape of the vibrating air cavity that produces the sound (Butler, 70). Although the Brahms recording was manipulated up and down by the same amount in Audacity, Sonic Visualiser measured the end product as 12.29 cents lower than the original, indicating that there may be discrepancies between the two programs' pitch-measuring algorithms.

and lower the pitch by individual cents, or they could press and hold the buttons to affect the pitch change. They were not told by how much they were changing the pitch, nor were they told that a single click of the button equaled one cent; subjects completed the task entirely by ear.[21] Following Geringer's protocol, subjects performed a practice trial to get a feel for the remote and to explore the rate at which they could change the pitch level with the buttons. They were encouraged to take all the time they needed for each excerpt. Starting levels and final pitch levels of all trials were recorded in number of cents up or down from the original pitch level presented to subjects.

The order of the four recordings presented varied for each subject. Twelve orders were generated: every recording appeared three times in each of the four serial positions. Each recording also immediately preceded every other recording three times. Since only ten subjects completed this task, 10 of these 12 orders were used. These are listed in Appendix D.

B. Data Interpretation

The four ending pitch levels obtained in the adjustment task were converted into fractions of a semitone away from the nearest standard semitone. If a subject had also completed the JND task, that subject's four pitch levels were compared to that subject's JND, on the assumption that a subject's JND determines what range of frequencies he or she perceives as standard. The four trials of each subject were also averaged and plotted in order to compare freshmen with seniors.

---

23. One-cent increments are virtually undetectable to the human ear, meaning that subjects changed the pitch levels of the recordings on a perceptually smooth pitch continuum.

C. Results

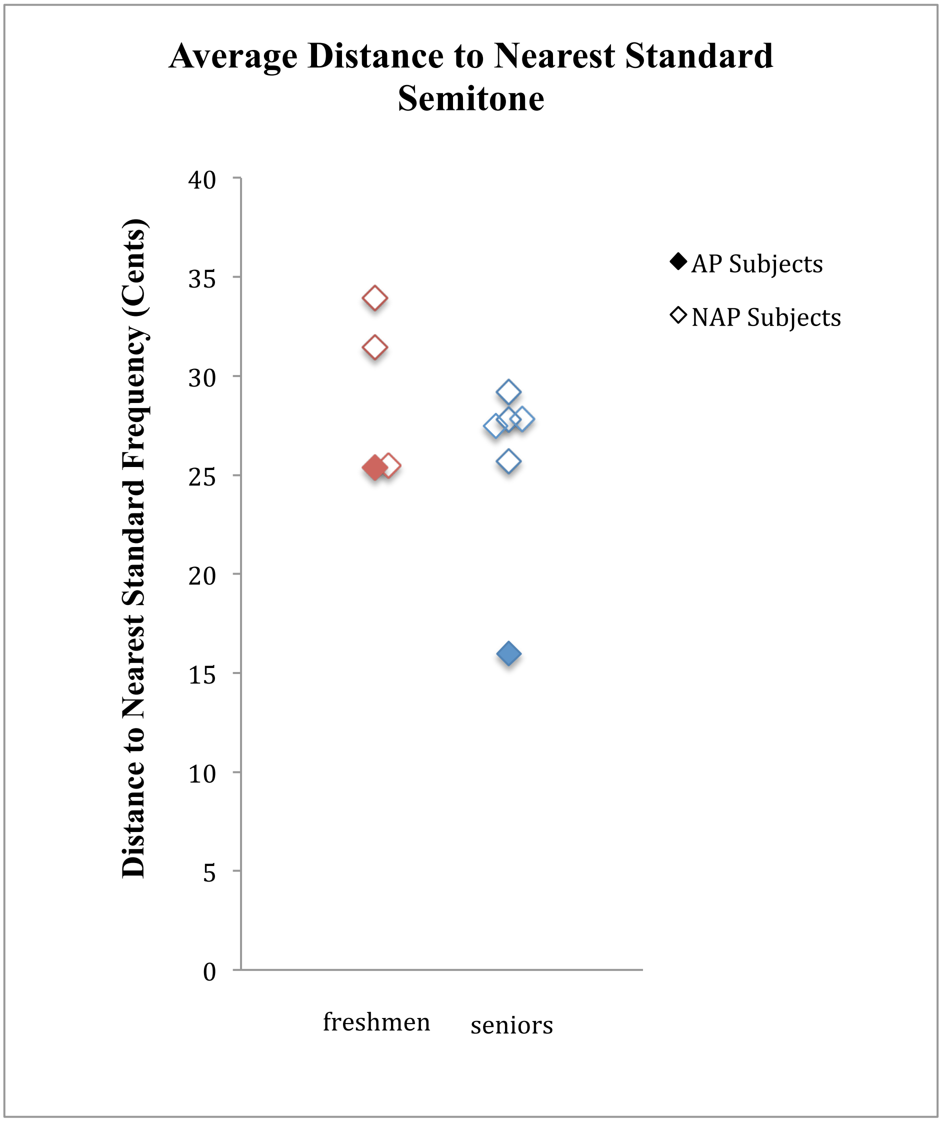
Table 2 summarizes the results of the adjustment task. Most subjects for whom JND was known adjusted one out of four recordings to a level within one JND of standard, indicating a poor ability to distinguish standard from nonstandard pitch levels. Although not enough AP subjects took part in this experiment to compare AP and NAP listeners in a statistically sound way, the results suggest that NAP listeners are not nearly as good as AP listeners at discriminating between standard and nonstandard frequencies. The three AP subjects had the highest A_z scores of all subjects in the rating task. The two AP subjects that completed the adjustment task adjusted the recordings the closest to standard on average.

Table 2. Summary of the results of the adjustment task, including mean distance of each subjects' four trials, and the number of recordings that subjects adjusted to within one JND of a standard pitch level.

Subject	AP or NAP	Overall JND (Cents)	Mean Distance to Nearest Semitone (Cents)	Trials Within One JND of Standard
2	NAP	15.76	31.44	1
3	NAP	17.36	25.46	1
4	AP	17.29	25.36	2
5	NAP	19.29	34.94	0
6	NAP	21.97	28.79	1
7	NAP	18.04	28.82	2
9	AP	12.28	16.98	1
10	NAP	-	27.47	-
11	NAP	-	29.19	-
12	NAP	-	26.68	-

Figure 6 shows the average distance to the nearest semitone of each subject's ending pitches, with freshmen compared to seniors. None of these averages was within one JND of standard, and seniors did not outperform freshmen: t(8) = 1.1247, p = 0.15. The results cannot prove that NAP listeners, let alone seniors, use long-term memory of absolute pitch to make judgments regarding the standardness of piano recordings.

Figure 6. Average distance to the nearest standard semitone of all four trials.



The two AP subjects that completed the adjustment task adjusted the recordings the closest to standard on average, although a few NAP subjects' results were very close to these scores. Curiously, however, AP possessors did not sing as close to standard pitch levels as

NAP possessors in the folk-song task. The limitations of the folk-song task, as well as the possible ill-suitedness of the JND task to address the current hypothesis will be discussed in Chapter 7.

CHAPTER 7

DISCUSSION, LIMITATIONS OF THE EXPERIMENT, AND AVENUES FOR FURTHER RESEARCH

A. Discussion

The twofold hypothesis of this thesis from Chapter 1 can be expressed as follows: "within-semitone, or "standard" pitch memory is common in trained musicians without AP *and* repeated exposure to these standard frequencies engrains these specific pitches into the long-term auditory memories of trained musicians." The results of the rating, JND, and adjustment tasks indicate that NAP listeners do not distinguish between standard and nonstandard frequencies during listening tasks. The results of the folk-song task provide evidence that some memory for the set of standard frequencies does exist in NAP listeners. Therefore, NAP musicians' ability to call to mind the collection of pitches from standard tuning without an outside pitch reference may not be dependent on labeling ability. The results of the experiment, however, do not support the second part of the hypothesis because they do not disprove the null hypothesis that seniors' are no better at remembering the set of standard frequencies than freshmen.

This thesis assumes that if the ability of individual listeners to perceive two melodic frequencies as different depends on their JND, that JND also influences overall pitch level of a listener's singing. In the folk-song task, four out of seven subjects whose JND was known sang within one JND of a standard frequency. It is unlikely but conceivable that chance, and not absolute pitch memory, produced such results. Subject 3, for instance, had an overall JND of 17.36 cents. Dividing 17.36 by the 50 cents in half a semitone demonstrates that there is a 34.72% chance that Subject 3 would sing within one JND of standard even if that subject has no within-semitone absolute pitch memory. Again, chance is unlikely to have produced the results obtained by the folk-song task, but it still could confound the data interpretation. More subjects would be needed to eliminate the risk of this confound.

It remains unclear *how* subjects might have remembered the levels of standard tuning when singing if they could not recognize them in the other three tasks. Mandler writes, “An item is accessible in storage if it can be retrieved, and the accessibility of items, as determined by their retrieval in free recall, predicts the likelihood of their being recognized, independent of whether accessibility was determined before or after the recognition test” (257). How does one explain the better performance of subjects on the folk-song task, a recall test, over the other three recognition tasks? Furthermore, recognition is made possible when a subject “[combines] the details of each item,” according to Mandler (256). What are the details of a single frequency that might be familiar to listeners? Can NAP listeners learn to recognize frequency details without also displaying labeling ability?

Mandler describes recognition as the finding of a context for a given event, and recall as finding the event given a context. Recall is a product of “interrelating the items within a list” (256). When subjects were asked to sing a familiar song, the “list” might have been a subject’s mental record of the pitch levels of all the previous instances that song was sung. More likely, the list is a collection of possible frequencies from which to pick a starting pitch and the scale degrees of the song.

It is important to keep in mind that when Mandler writes about recognition and recall tests, he uses examples involving words, not musical tones. Mandler’s lists are comprised of words that are presented to subjects before they are asked to recognize or recall them. These words are likely already familiar to subjects in terms of spelling, meaning, sound when spoken, etc. In a study by Mandler, Pearlstone, and Koopmans (1969), “subjects sorted sets of unrelated words into a number of categories of their own choosing. Subsequently, they were given recall and/or recognition tests for the sorted items" (Mandler, 1980, 253). There was no set-up in the current experiment—that is, frequencies were presented fresh during the trials, not played for subjects before the experiment for sorting.

How much thought, conscious or subconscious, goes into selecting a pitch level for a song? Do listeners imagine the first pitch before they sing? Do they simply produce a sound and adjust it within milliseconds to reflect some preconceived grid of frequencies? Subjects were asked to say their name into the recording device immediately before they sang for identification purposes. Were the pitches sung by subjects influenced by the sound of their speaking voices? In general, speech is not comprised of discrete scale steps, but glides smoothly between bounding frequencies. Is it possible that the bounding frequencies of subjects who spoke immediately before singing influenced the starting pitches in the folk-song task by providing relative pitch cues? Deutsch, Henthorn, and Lapidis (2011) write that "the dichotomy between the physical characteristics of speech and nonspeech are not clearcut". Their study confirms the hypothesis that "the pitches forming [a spoken] phrase increase in perceptual salience [as a result of repetition] and that in addition they are perceptually distorted so as to conform to a tonal melody" (2246). It is possible for speech to sound like song, but it is difficult to say if this perceptual phenomenon occurred during the current folk-song task.

The inconsistent results of the JND task indicate that this type of task may not be effective at uncovering differences in the way listeners respond to standard and nonstandard tones. More AP and NAP subjects would be needed to test the validity of this method. Table 3 reproduces some of the contents of Table 1; it shows the overall JND scores of the three AP subjects for standard versus nonstandard frequencies. Subject 1 had a higher JND for nonstandard tones, and Subjects 4 and 9 had higher JNDs for standard tones.[22]

Table 3. Reproduction of the data from Table 1: JND for AP subjects.

Subject	**1 (AP)**	**4 (AP)**	**9 (AP)**
Standard	0.6972	1.3841	1.2124
Nonstandard	1.0692	1.1869	1.0401

The experimental results seem to suggest that there may a weaker link between the perception and production of the standard frequency collection than might previously be assumed, or at least reveal a need for further research into recognition and recall of musical frequencies. Do musicians really produce standard frequencies better than they can recognize them? In contrast to what Mandler writes about words, it seems that perception of frequencies is not a prerequisite for production.

The sine tones used in the rating and JND tasks are a far cry from the "spectrally complex stimul[i]" described by Ross et al. as necessary for possessors of latent AP to store in long-term memory after repeated hearings. All 12 subjects likely had extensive exposure to spectrally complex stimuli during their musical training. Perhaps timbre contributed to subjects' ability to recognize and recall absolute frequencies; Levitin's subjects likely called upon their memory of timbre as well as pitch and where the song lies in their vocal range in order to sing the songs they selected from memory. Schellenberg and Trehub's subjects

---

24. Scholars have not yet proven that the possession of AP influences one's JND for pitch. The only study that compares the JND for AP possessors with that of NAP possessors is by Pilko (2001). Her results were not able to show a difference between the two groups. She called for modifications of her experiment in order to better test for a difference.

listened to recordings of popular television programs that included “multiple instruments, each with multiple pure-tone components” (263)[23] Deutsch’s subjects listened to sinusoidal tones, however, and were still able to make implicit judgment about absolute pitch.[24]

B. Limitations of the Experiment

1. Experimental Design

The most limiting factor of the experiment is the small subject pool. Despite all other flaws in experimental design or methodology, an ample subject size is the first parameter that guarantees that the results of the experiment are generalizable to the population similar to those subjects who were tested. Although there were enough subjects in this experiment to show trends in performance, a subject size 5 to 6 times as large would be ideal for these tasks.

Given the complexity of designing a test for AP, the widespread scholarly disagreement about the validity of such tests, as well as the variety of concrete skills that AP possessors exhibit, this experiment takes for granted the word of subjects about whether or not they possess AP. Despite this limitation, the experimental methods and data interpretation presented here should be considered valid and can be implemented in future research with some adjustments. In order to better examine the abilities of NAP possessors and to uncover potential difference between AP and NAP possessors, subjects should be tested for AP. The test could ask subjects to label pitches presented as both sine tones and piano timbres,

---

25. The recordings used by Schellenberg and Trehub were excerpted from “E.R.,” “Friends,” “Jeopardy,” “Law & Order,” “The Simpsons,” and “X-Files” in the keys of B minor, A major, E-flat major, G minor, C-sharp major, and a minor (263).

26. The tone pairs were presented in such a way to control the effects of amplitude on subject perception (5–6).

replicating the sounds used in the rating, JND, and adjustment tasks. This would help identify subjects' individual strengths in relation to the tasks. The methods from the current experiment could then test balanced numbers of both AP and NAP possessors

It is conceivable that NAP musicians possess better memory for absolute pitches than nonmusicians. To truly test the first hypothesis of this thesis, that musicians do retain the pitches of standard tuning with fine-grained accuracy, it would be necessary to compare their performance with that of nonmusicians. To further test the effects of statistical learning on human's long-term memory for minute pitch shading, nonmusicians should be tested at regular intervals beginning in early childhood, as soon as relative pitch develops, and continuing up through freshman year of college, where the current study picks up.

Not all subjects participated in all 4 tasks. Only 7 out the 10 subjects who completed the folk-song or adjustment task also completed the JND tasks, and so the pitch level at which the remaining 3 subjects sang and the levels to which they manipulated the piano recordings could not be compared to their individual JNDs. As previously mentioned, Subject 10 sang only 8.94 cents away from the nearest semitone. This was likely within his JND, given that the average JND of the 7 subjects tested was 17.43 cents. Having all subjects complete all 4 tasks would have increased the power of each test to demonstrate skill or lack of skill in the task at hand, as it would have allowed for more thorough correlational analyses.

The rating experiment included three different "levels" of nonstandard pitches. The interpretation of the data in Figs. 1a and 1b, however, did not discriminate between varying degrees of nonstandard pitches; probabilities of responses were calculated for all nonstandard pitches as a group. To increase the statistical viability of the results, an equal number of standard pitches as nonstandard pitches should be presented to subjects.

Many years of singing and hearing the songs from the folk-song task in many different keys likely prevents there from being a “correct” starting pitch in the memories of most subjects. Performing these songs without standard-tuned instrumental accompaniment may contribute to the statistical learning of nonstandard frequencies, or at least, inhibit the learning of standard frequencies.

Subjects were told only to “sing as much of the song as you can remember from the beginning,” and to “try to sing in tune.” To fully separate the issue of preference from that of ability in the folk-song task, subjects could have been told to sing the folk song at a standard pitch level. To eliminate a confound from the possibility that certain syllables or vocal sounds are more difficult to produce at a constant pitch, all subjects could be asked to sing the songs on the syllable “la.” For instance, the diphthong in the word “row,” of “Row, Row, Row Your Boat’ caused some subjects to lower their pitch substantially as they changed the first part of the syllable to the second (no diphthongs occurred on the final tonic, the note whose pitch was analyzed in the folk-song task). Note that there was no statistically significant difference in the last pitch’s distance to the nearest semitone between those who sang on song lyrics and those who sang on “la.” However, uniformity of syllables is important to help control variables like these that might obscure the results in a future experiment.

Subject 9, an AP listener, sang 38.39 cents away from the nearest semitone. Her performance questions the power of the folk-song task to test for long-term memory of standard pitch levels. Several factors could lead to a response like this. One, already mentioned, is that subjects were not specifically told to sing at a standard pitch level. Another possibility is that Subject 9's JND is very high. This seems unlikely since the average JND of subjects who completed the JND task was 17.43 cents and the difference between the lowest and highest JND was only 9.69 cents. Measuring only the last pitch that subjects sing might not be the best method to determine the pitch level that subjects are consciously using. Alternatively, this subject might possess better pitch-labeling skills than vocal production skills.

Schellenberg and Trehub suggest that familiarity with an excerpt presented to subjects for pitch level analysis influences subject's perception of the "correctness" of the pitch level of the presented excerpt (265). The adjustment task did not control for subjects' familiarity with the compositions or recordings employed. It is possible that subject familiarity with the excerpts may influence their recognition of the standard versus nonstandard frequencies, introducing a possible confound to the current experiment.

2. Methodology

Daniel Levitin suggests that perceived loudness should be removed as a variable when presenting varied frequencies to subjects (1999, 320). Fletcher and Munson (1993) demonstrated that the intensity of a sound in decibels is perceived as a function of frequency; in their system of equal-loudness curves, perceived loudness is equal only to decibels at 1,000 Hz. The perceived loudness of a sound decreases as frequency decreases from this benchmark. Therefore, assuming all experimental subjects had normal hearing, all of the tones presented to subjects in the rating and JND tasks should have had their loudness increased as frequency decreased. To compensate for the unevenness across the experimental frequency range would ensure that frequency is the only variable influencing subjects' perception of pitch.

During the rating and JND tasks, subjects could have relied on relative pitch when making judgments about whether a frequency was from standard tuning, even with strategic restrictions of randomization of the order of tones presented. If subjects made an incorrect judgment—or even guess—that the first frequency presented was from standard tuning, for example, it is conceivable that subjects with reliable relative pitch might have judge every subsequent tone incorrectly, relying solely on the intervallic distance between consecutive tones. This would demonstrate the power that the first tone had on establishing a pitch reference for subsequent tones. Similarly, even if subjects indicated that the final tone in a series was "mistuned," it might be possible to subsequently establish that pitch to seem firmly "in tune" for that subject by giving that tone tonal context—using it as the dominant scale degree in a tonal melody, for instance. It may be possible that context alone (or in the case of the first tone presented, *lack* of context) influenced listeners' decisions in the rating task or JND tasks.

For the adjustment task, if time and subject availability had allowed there to be at least one subject for each of the 24 possible orders of excerpts, it would have increased the reliability of the results task by reducing order effects. Alternatively, each subject could have listened to each order. Although subjects were told only to manipulate the recording to reflect a standard level (which was not necessarily the level at which the piece was written), subjects with piano experience or familiarity with the specific recordings used as stimuli might choose the original level of the recording instead of a standard level shifted from the original level. When testing excerpts at both standard and nonstandard levels, standard-level recordings could be divided into two classes: (1) those that replicate the original pitch intended by the composer, and (2) those that are shifted a whole number of semitones away from the composer-intended key. To test both could reveal differences in subjects' tendency to distinguish between original pitch level and a "standard-but-shifted" pitch level. Further study would be needed to determine whether subjects without AP hear whether familiar or unfamiliar recordings are in the right key. At least one NAP subject reported listening for the "brightness" of the piano sound at different pitch levels in order to make decisions.

Similar to order effects are learning effects, which are present if a subject's performance improves over the course of a task due to their increased understanding of the task in real time. Although learning effects in psychological tasks are generally unavoidable, they are worth noting, especially in the JND task. One subject asked after the task if there were varying degrees of pitch variance, indicating that she may have become more attuned to fine pitch changes as the task progressed. If subjects did not detect slight pitch differences early in the sequence of tone pairs, subsequent differences could prompt them to expect them, increasing their attention to small pitch changes and improving their detection ability as the task progressed. This is why it is important to distribute evenly throughout the trials the standard and nonstandard tones as well as the amount of pitch fluctuation between members of a pair. Alternatively, the results of the first few trials could have been discarded.

Finally, in order to better survey subjects' exposure to nonstandard frequencies, the subject questionnaire could have asked subjects to list in which ensembles they had participated while studying at the University of Massachusetts Amherst. Subjects may not be aware that their ensemble uses a tuning reference other than A = 440 Hz, in which case they would not have reported it as an alternate tuning system (the marching band regularly tunes to A = 442 Hz).

C. Avenues for Further Research

Evidence suggests that the sound of the human voice is processed by brain mechanisms specifically designed to interpret language and meaning (Vanzella and Schellenberg, 2010). Vanzella and Schellenberg's research concludes that "because the human voice is inextricably linked to language and meaning, it may be processed automatically by voice-specific mechanisms that interfere with note naming among AP possessors" (1). They found that AP possessors did not identify test tones presented in natural and synthesized voice timbres as well as they did piano and pure tones. Their results suggest that the production of tones using the voice might not activate long-term memory for absolute pitch. Vocal production tasks alone may not fully determine how musicians mentally represent various pitches; if this is the case, it could also explain the "poor" performance of the two AP subjects on the folk-song task in the current experiment. It may be inaccurate to conclude that the folk-song task alone is a good indicator of subjects' long-term memory for pitch.

An informal experiment conducted by the investigator in 2012 focused on how within-semitone memory for absolute pitch might affect the production of abstract pitch patterns, not learned melodies that might be associated with language. Subjects sang five-note scalar segments given various standard or nonstandard sine-tone starting pitches and their intonation was compared between standard and nonstandard trials. However, once an imagined pitch is translated into vocal sound, mechanisms other than long-term pitch memory may guide subjects' pitch. It is as if subjects were using the sound of their own voice as a stimulus, not the sine tones. Once subjects match the pure tone given at the start of each trial, subjects may find it easier to sing a scalar pattern from the timbre of their own voice.

Although the main purpose of the current experiment was to learn about the pitch-memory abilities of NAP listeners, the results inspire further investigation into what musical experiences account for observed differences between subjects within the AP and NAP groups, and within the freshmen and senior groups. It is possible that caveats described by scholars for AP possessors also apply to NAP possessors. "Some people with AP can only label tones produced by one particular instrument," write Levitin and Rogers, suggesting that labeling ability is bound up with the timbre of that instrument (28). Others have AP for only a single tone, such as a tuning note, or several notes, such as the white keys of the piano, which are often learned first during one's musical training. These phenomena offer further evidence that statistical learning facilitates the development of AP during a critical period of learning early in a child's life.

The adjustment task employed piano timbres. Some subjects may be better able to distinguish between standard and nonstandard tones when listening to piano music than when listening to sine tones. AP listeners are known to exhibit better pitch-labeling abilities with familiar, complex tones, such as piano or oboe. Perhaps NAP listeners who show an ability to discriminate between standard and nonstandard frequencies also perform better when listening to familiar timbres.

Researchers have increasingly linked aural imagery to motor memory. In their neuromusical studies, Halpern (2001) and Halpern and Zatorre (1999) found that the supplementary motor area of the brain is active during musical imagery, “especially during covert mental rehearsal” of notated music (Brodsky and Kessler et al., 429). Brodsky and Kessler et al. found that “notational audiation is a process engaging kinesthetic-like covert excitation of the vocal folds linked to phonatory resources” (427). More generally, when trained musicians imagine music in their heads in a process known as “auralization,” the brain behaves similarly as it would when actually perceiving music (427).[25] It is therefore possible that pitch memory is linked to muscle memory. Perhaps recognizing pitches is a different type of task than auralizing pitches or singing them without a pitch prompt. Both the rating and the JND task are recognition tasks, and might demand from students a different sort of mental activity than when they sing a familiar tune from scratch or auralize an unfamiliar sight-singing excerpt. Does statistical learning of the frequencies of standard tuning influence the pitch levels at which musicians (AP or NAP) auralize?

What aspects of music do listeners remember in order to approximate frequencies without a reference? Most of Levitin’s subjects produced the original pitch levels of familiar songs within 2 semitones. Levitin’s Figure 1 is reproduced here in Figure 7.

---

27. The history of the term *auralization* as applied to inner hearing is traced in a detailed footnote on page 49 of Karpinski’s *Aural Skills Acquisition* (2000).

Figure 7. Reproduction of Figure 1 from Levitin (1994).

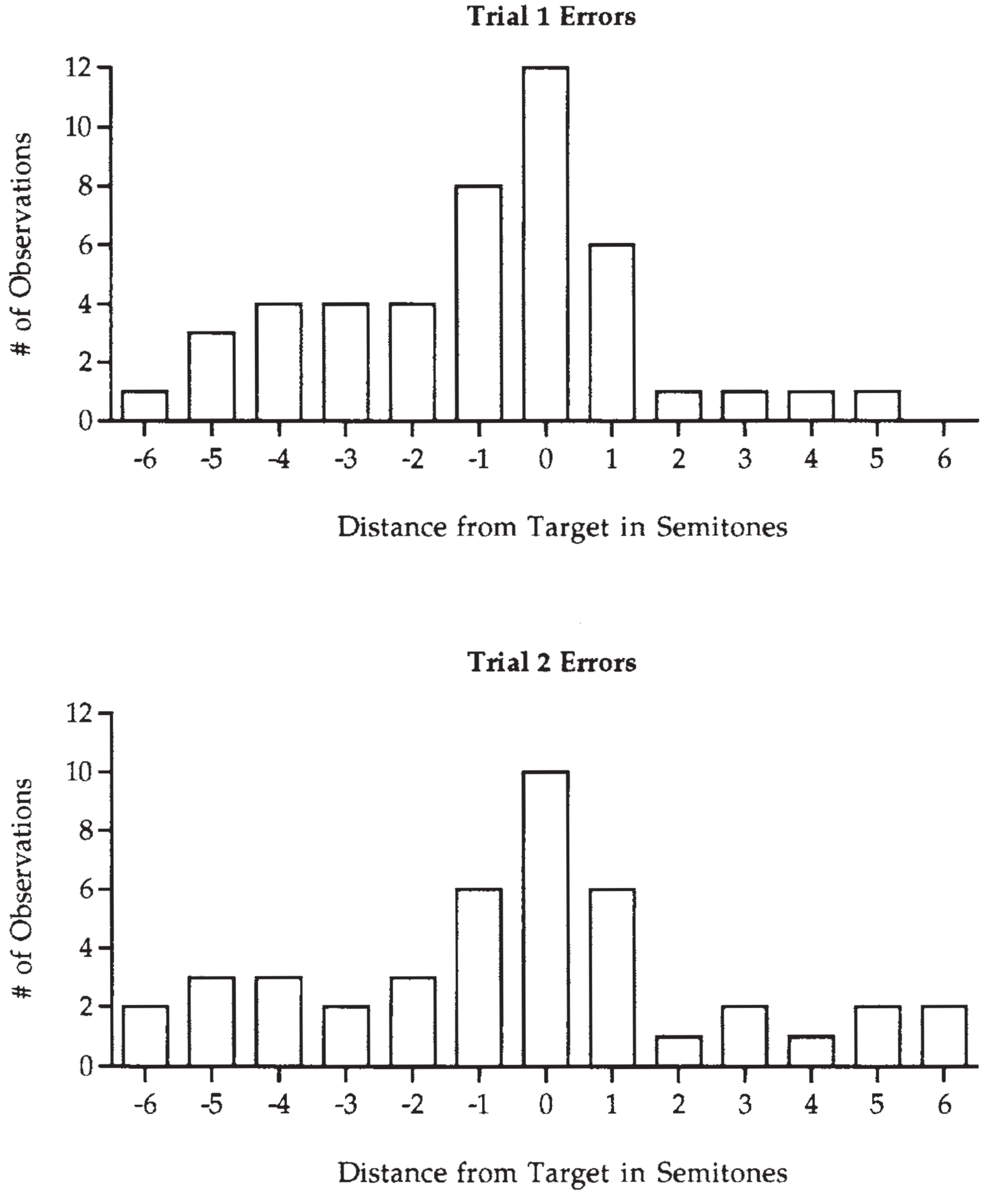


Figure 1. Subjects' errors in semitone deviations from the correct tone. Octave errors were not penalized. For Trial 1, mean $= -0.98$, $s = 2.36$. For Trial 2, mean $= -0.4$, $s = 3.05$.

The experiments of Levitin, Deutsch, and the current thesis raise a question about the perception of standard versus nonstandard tones. Does the smooth continuum of pitch class around the circle create an equally smooth memory of pitch class? In other words, is motion around the circle directly correlated to perception of pitch class height? Consider that Levitin does not divulge the exact pitch levels at which his subjects reproduced songs. Since he rounded subject responses to the nearest semitone, it can be assumed that at least some of his subjects rendered the songs at nonstandard pitch levels. Subjects therefore succeeded at *approximating* the original pitch levels of the recordings; most did so within two semitones. Subjects were told to imagine the song playing in their heads before singing, humming, or whistling it, but as Brodsky and Kessler demonstrated, auralization does involve covert motion of the vocal folds. In choosing a pitch level, subjects could have used kinesthetic memory of where the song lies in relation to their vocal or whistling range. Nevertheless, they must have recalled approximate aural information about the pitch level of the song before they produced it, that is, frequency or pitch height.[26]

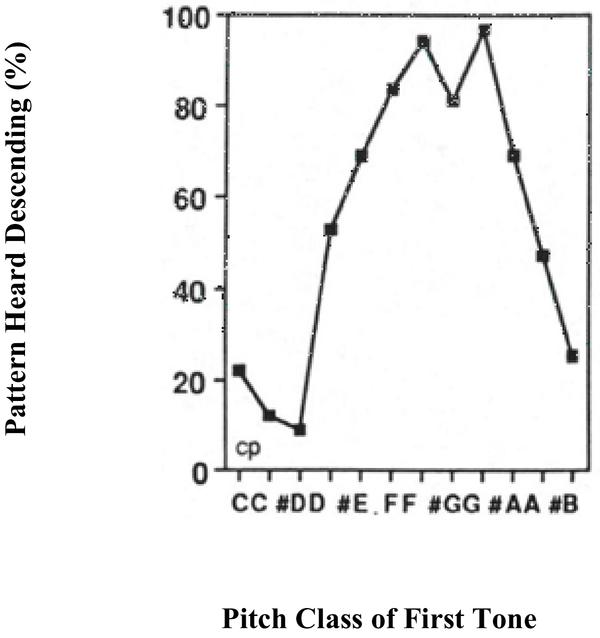
---

28. Music psychologists often use the term *chroma* to refer to a specific tonal quality that all members of a pitch class seem to share. Roger Shepard, author of the seminal 1964 article, “Circularity in Judgments of Relative Pitch,” proposes that pitch consists of two parameters: tone height and pitch class (chroma). However, he is careful to dismiss the notion that chroma is physically present in pitch, writing that “The main interest [in his experiment] is in the (psychological) problem of demonstrating circularity of judgments rather than in the (psychophysical) problem of systematically exploring the effects of physical parameters on auditory perception. Since the desired demonstration of circularity can be accomplished by an analysis of the pattern of judgments alone, it does not really depend upon any physical measurements of the stimuli” (2348). Similarly, I argue here that chroma is a psychological construct that relies on frequency (pitch height) alone, and that chroma is not an inherent property of sound. The frequencies 220 Hz, 440 Hz, and 880 Hz sound similar because of the 2:1 relationship of their fundamental frequencies. These frequencies produce the distinctive sound of the pitch class “A” because their frequencies are different than those of the pitch class “G,” or 196 Hz, 392 Hz, 784 Hz, etc.

AP possessors’ perception of frequency or pitch height is keener than NAP possessors. AP possessors generally recognize and/or recall pitch classes more accurately than NAP possessors, even those NAP possessors who display “good” absolute pitch memory, like the majority of Levitin’s subjects (Vanzella and Schellenberg, 1).

Perhaps the tritone paradox experiment could be used to test the keenness of a listener’s frequency perception. Recall that Deutsch’s subjects had no musical training. Since pitch memory seems only to increase with musical training, it is likely that Levitin’s subjects—a mix of musicians and nonmusicians without pitch labeling ability—would perform consistently in Deutsch’s tritone paradox experiment. That is, they would consistently rate the same pitch classes as higher and lower than their tritone-related pitch classes. A tritone paradox experiment that uses hundreds of standard and nonstandard pitch classes as stimuli would provide a more finely tuned “map” of how NAP possessors respond to pitches that are part of a smooth pitch spectrum, not just in scale-step increments. The results of such an experiment could be reported in a graph much like Deutsch’s Figure 3, the upper right-hand portion of which is reproduced here as Figure 8 with axis labels reset for formatting reasons.

Figure 8. Subject “CP,” results from Deutsch’s tritone paradox experiment (1991, 7). Percentages reflect average scores of how often a subject described pitch pairs as ascending or descending across multiple trials.



Deutsch’s subject “CP” produced a general up-down or down-up shape, as did each of her other 5 subjects whose results she reported in her 1991 article. Note that there are a few exceptions to this contour, however, at the pitch class G, for example, which was perceived on average as lower than the two surrounding pitch classes. The results show that subjects may not perceive pitch classes as changing smoothly around the circle. If CP had listened to hundreds of tone pairs, including standard and nonstandard tones from the whole pitch-class spectrum, perhaps more “bumps” or other interesting patterns would occur in the resultant graph. The graph’s contour could shed light on whether perception of pitch class is smooth as one “moves” around the pitch-class circle. This method might determine if NAP listeners describe the pitch class of standard tones as different than that of nonstandard tones. NAP subjects could then be compared to AP subjects.

D. Conclusion

The results of this experiment show that NAP possessors do not easily distinguish between standard and nonstandard frequencies during recognition tasks. However, the inconsistent performance of the few AP possessors tested suggest that testing within-semitone absolute pitch memory may not be a straightforward task. The JND task, especially, might not be suited for uncovering differences between listeners' perceptions of standard versus nonstandard frequencies. Further recognition tasks should meticulously control relative pitch cues, order effects, and learning effects. They could also test the effect that timbre has on subjects' ability to recognize the set of standard frequencies.

The statistically significant results of the folk-song task suggest further testing of NAP possessors' abilities to produce without reference the frequencies from standard tuning. Future experiments should isolate memory for timbre, auralization mechanisms, voice-brain interactions, as well as the separation of pitch class from pitch height in the brain, as factors that influence the pitch levels at which listeners sing. In order to better understand what experiences contribute to listeners' memories for within-semitone absolute pitch, children could be tested for standard-pitch production during various stages in their musical development and education. Absolute pitch is often used as an all-or-nothing label. By viewing absolute pitch memory as a complex set of disparate yet related abilities, researchers may eventually redefine the meaning of "absolute pitch," or at least shed light on the pitch-processing and production capabilities of our students.

APPENDIX A

INFORMATION PROVIDED BY SUBJECTS ON THEIR MUSICAL BACKGROUND

Subject	Year	AP/NAP	Country where most musical training occurred	Experience with nonstandard tuning levels or microtonal music
1	Senior	AP	USA	Baroque tuning
2	Freshman	NAP	USA	Tunes to A = 442 Hz, marching band at A = 442
3	Freshman	NAP	USA	None
4	Freshman	AP	USA	None
5	Freshman	NAP	USA	None
6	Senior	NAP	USA	Baroque tuning
7	Senior	NAP	USA	Early music tuning
8	Freshman	NAP	USA	None
9	Senior	AP	Taiwan and USA	A = 441
10	Senior	NAP	USA	None
11	Senior	NAP	USA	Early music tuning
12	Senior	NAP	USA	None

APPENDIX B

FOLK SONG LIST WITH SYLLABLES USED BY SUBJECTS

Song titles are presented in the order in which they were listed on the instruction sheet given to each subject. Subjects are listed under the song they chose, along with the syllables on which they performed the song.

Yankee Doodle
Subject 5: “la”

Old MacDonald

I’ve Been Working the Railroad

My Bonnie Lies Over the Ocean

Row, Row, Row Your Boat
Subject 3: words of the song
Subject 4: words of the song
Subject 11: “la”
Subject 12: words of the song

Jingle Bells

Joy to the World
Subject 7: words of the song

The Itsy, Bitsy Spider
Subject 6: words of the song

Frére Jacques
Subject 2: “la”

Happy Birthday
Subject 9: words of the song
Subject 10: words of the song

APPENDIX C

DISCOGRAPHY FOR ADJUSTMENT TASK

Brahms, Johannes. *Capriccio in B Minor,* Op. 76, no. 2. Artur Rubinstein, pianist. New York: RCA Red Seal, 1999, compact disc.

Schumann, Robert. *Papillons,* Op. 2. Jenö Jandó, pianist. Munich: Naxos, 1993, compact disc.

Liszt, Franz. *Mephisto Waltz* No. 1. London Philharmonic Orchestra, Bernhard Haitink, conductor. Baam: Philips, 1983, compact disc.

Chopin, Fryderyk. *Etude No. 1 in A-flat Major,* Op. 25, no. 1. Artur Rubinstein, pianist. Recital in Moscow. New York: RCA Red Seal, 1999, compact disc.

APPENDIX D

ADJUSTMENT TASK RECORDING ORDERS

Key:
C = Chopin S = Schumann
L = Liszt B = Brahms

Subject	Order
2	C, S, L, B
4	C, L, S, B
3	L, C, B, S
10	L, B, C, S
6	S, C, B, L
7	S, B, C, L
5	B, S, L, C
9	B, L, S, C
11	C, L, B, S
12	L, S, C, B

BIBLIOGRAPHY

Brodsky, Warren, Yoav Kessler, Jane Ginsborg, Bat-Sheva Rubinstein, and Avishai Henik. 2008. “The Mental Representation of Music Notation: Notational Audiation.” *Journal of Experimental Psychology* 34 (2): 427–445.

Burns, Edward, and Shari L. Campbell. 1994. “Frequency and Frequency-Ratio Resolution by Possessors of Absolute and Relative Pitch: Examples of Categorical Perception?” *Journal of the Acoustical Society of America* 96 (5): 2704–2719.

Butler, David. 1992. *Musician’s Guide to Perception and Cognition.* New York: Schirmer Books.

Butler, David, and W. Dixon Ward. 1988. “Effacing the Memory of Musical Pitch.” *Music Perception* 5 (3): 251–259.

Deutsch, Diana. 1991. “The Tritone Paradox: An Influence of Language on Music Perception.” *Music Perception* 8: 335–347.

Deutsch, Diana, Trevor Henthorn, and Rachael Lapidis. 2011. “Illusory Transformation from Speech to Song.” *Journal of the Acoustical Society of America* 129 (4): 2245–2252.

———. 2006. “The Enigma of Absolute Pitch.” *Acoustics Today* 2 (4): 11–19.

Fletcher, Harvey, and W. A. Munson. 1993. “Loudness, its Definition, Measurement and Calculation.” *Journal of the Acoustical Society of America* 5: 82–108.

Geringer, John M. 2010. “Musicians’ Preferences for Tempo and Pitch Levels in Recorded Orchestral Music,” *Journal of Research in Music Education* 58 (3): 294–308.

Halpern, Andrea R. 1989. “Memory for the Absolute Pitch of Familiar Songs.” *Memory and Cognition* 17(5): 372–381.

———. 2001. “Cerebral Substrates of Music Imagery.” In *Annals of the New York Academy of Science: Vol. 930,* edited by Robert. J. Zatorre and Isabelle. Peretz, 179–192. New York: New York Academy of Science.

Helmholtz, Hermann L. F. 1863. *On the Sensations of Tone.* 2nd ed. Translated by Alexander J. Ellis. New York: Dover Publications, Inc.

Huron, David. 2006. *Sweet Anticipation*. Cambridge, MA: Massachusetts Institute of Technology.

Karpinski, Gary. 2000. *Aural Skills Acquisition:* the Development of Listening, Reading, and Performing Skills in College-Level Musicians. New York: Oxford University Press.

Kollmeier, B., Brand, T., and Meyer, B. 2008. "Perception of Speech and Sound." In *Springer Handbook of Speech Processing,* edited by Jacob Benesty, M. Mohan Sondhi, and Yiteng Huang, 61–82. London: Springer.

Levitin, Daniel. 1994. "Absolute Memory for Musical Pitch: Evidence From the Production of Learned Melodies." *Perception and Psychophysics* 56 (4): 414–423.

——— 1999. "Experimental Design in Psychoacoustic Research." In *Music, Cognition, and Computerized Sound: An Introduction to Psychoacoustics,* edited by Perry R. Cook, 299–328. Cambridge, MA: MIT Press.

Levitin, Daniel, and Susan Rogers. 2005. "Absolute Pitch: Perception, Coding, and Controversies." *TRENDS in Cognitive Sciences."* 9 (1): 26–33.

MacMillan, Neal A., and C. Douglass Creelman. 1991. *Detection Theory: A User's Guide.*

Mandler, George. 1980. "Recognizing: The Judgment of Previous Occurrence." *Psychological Review* 87 (3): 252–271.

McLachlan, Neil, and Sarah Wilson. 2010. "The Central Role of Recognition in Auditory Perception: A Neurobiological Model." *Psychological Review* 117 (1): 175–196.

Pilko, Sara C. 2001. "Just Noticeable Difference Scores of Absolute Pitch and Non Absolute Pitch Possessors." Master's thesis, West Chester University of Pennsylvania.

Ross, David. A., Ingrid. R. Olson, and John C. Gore. 2003. "Absolute Pitch Does not Depend on Early Musical Training." *Annals of the New York Academy of Sciences* 999: 522–526.

———, Ingrid R. Olson, Lawrence E. Marks, and John C. Gore. 2004. "A Nonmusical Paradigm for Identifying Absolute Pitch Possessors." *Journal of the Acoustical Society of America* 116 (3): 1793–1799.

Schellenberg, E. Glenn, and Sandra E. Trehub. 2003. "Good Pitch Memory is Widespread." *Psychological Science* 14 (3): 262–266.

Shepard, Roger. 1964. "Circularity in Judgments of Relative Pitch." *The Journal of the Acoustical Society of America* 36 (12): 2346–2353.

Shepard, Roger N, and Daniel S. Jordan. 1984. "Demonstrating that Tones are Assimilated to an Internalized Musical Scale." *Science* 226 (4680): 1333–1334.

Stanislaw, Harold, and Natasha Todorov. 1999. "Calculation of Signal Detection Theory Measures." *Behavior Research Methods, Instruments, and Computers* 31 (1): 137–149.

Vanzella, Patricia, and E. Glenn Schellenberg. 2010. "Absolute Pitch: Effects of Timbre on Note-Naming Ability." *PloS One* 5 (11): e15449, 1–7.

Windhorst, Uwe, and Håkan Johannson. 1999. *Modern Techniques in Neuroscience Research.* Berlin: Springer.

Windsor, Luke. 2004. "Data Collection, Experimental Design, and Statistics in Musical Research." In *Empirical Musicology: Aims, Methods, Prospects,* edited by Eric F. Clarke and Nicholas Cook, 197–222. New York: Oxford University Press.

Online Resources

http://www.phy.mtu.edu/~suits/notefreqs.html

http://www.sengpielaudio.com/calculator-centsratio.htm

onlinetonegenerator.com

http://stattrek.com/statistics/random-number-generator.aspx

http://www.random.org/

http://www.alcula.com/calculators/statistics/linear-regression/

http://stattrek.com/online-calculator/normal.aspx

Downloads

Amazing Slow Downer is available for purchase here: www.ronimusic.com

Praat is available for download here: http://www.fon.hum.uva.nl/praat/

Sonic Visualiser is available for download here: http://www.sonicvisualiser.org

R is available for download here: http://www.r-project.org/

Audacity is available for download here: http://audacity.sourceforge.net/