7-2023

# Examining the influence of spectral envelope shape on pitch

Laura Reinert

*James Madison University*

## Recommended Citation

Reinert, Laura, "Examining the influence of spectral envelope shape on pitch" (2023). *Masters Theses, 2020-current*. 187.
https://commons.lib.jmu.edu/masters202029/187

# Examining the Influence of Spectral Envelope Shape on Pitch

Laura Reinert

A thesis submitted to the Graduate Faculty of

JAMES MADISON UNIVERSITY

In

Partial Fulfillment of the Requirements

for the degree of

Master of Arts

Department of Graduate Psychology

July 2023

---

FACULTY COMMITTEE:

Committee Chair: Michael D. Hall

Committee Members/ Readers:

Jeff Dyche

Lincoln Gray

## Table of Contents

# Acknowledgements

I would like to thank my family for their continued support and encouragement as I pursued a graduate degree. I am only involved in research because of my parents and psychology because of my grandmother. I hope to continue their legacies in science and dedication to learning.

Special thanks to my colleagues, Dominique Akesse, Steph Osborn, Sam Harmon, Ben Stone, and Sydney Iwanski. My program experience was made better for knowing you.

Finally, thanks to my advisor, Dr. Michael Hall. There are not enough words to express how grateful I am, but I know I am a better researcher and scientist because of you.

## List of Tables

## List of Figures

# Abstract

In the tritone paradox, there are many questions surrounding how listeners can make pitch judgments since Shepard tones are comprised of all octaves and this makes pitch ambiguous. The current study examined the influences from spectral envelope shape, spectral centroid, chroma, and musical training to identify how timbre and pitch interactions impacted pitch judgments to different tone types, including Shepard tones. Each trial consisted of a standard and a comparison tone differing by spectral envelope shape. Listeners were presented with these tone pairs and asked to judge whether the tone pairs were going up or down in pitch. For Shepard tones, sensitivity varied across centroid and chroma while acoustic analyses of the Shepard tones supported that pitch judgment performance was predicted by an aspect of spectral envelope shape, the relative amplitude of the $F_0$. The current study suggests that listeners first try to use the $F_0$ to make a pitch judgment, and when that component is not resolvable, systematically process the next components until they can judge pitch. This idea of a shared pitch processing mechanism is applicable to all tone types, including Shepard tones, and provides an explanation for the observed pattern of results seen in the tritone paradox. Future research should aim to confirm the relative amplitude of the $F_0$ predicts pitch judgment using the tritone paradox procedure to determine how listeners process pitch for Shepard tones.

# Examining the Influence of Spectral Envelope Shape on Pitch

From conveying warning through alarms to emotional intent in vocal intonation, pitch can convey complex information. The ability to process pitch is a significant skill as people are able to accurately synthesize a wide range of frequencies. However, pitch perception can be misled because of other auditory factors, such as timbral characteristics which are known to interfere with pitch accuracy as demonstrated through various auditory illusions.

Pitch is defined by "how high or low a sound is perceived, which mainly depends on frequency” (ANSI, 1994). In musical terms, how high or low the sound is can reference the tone chroma, the placement on the staff (e.g. a C vs an A), or the tone height, the octave in which the tone is present. For example, $C_4$ is an octave lower than $C_5$, being different in tone height though both encompass the same tone chroma. Examining the waveform of a higher frequency would show more cycles per second compared to a low frequency. Within this distribution, the first harmonic in the series, the fundamental frequency ($F_0$), often corresponds with reported pitch. In naturally occurring tones, the greatest amount of energy will often be found at the $F_0$ in the lower frequency region. From that peak, the amplitude height of each harmonic will decrease at a rate of 12 dB per octave until the signal is fully dissipated, forming what will henceforth be referred to as a *slope*-like shape.

Traditional theories state that pitch processing relates to basilar membrane sensitivity (place theory) and sensory neuron activity (temporal theory; Shamma, 2004). However, both of these theories cannot explain how low-frequency processing occurs. Relating to place theory, listeners can still differentiate low frequency tones beyond the peak of maximum displacement on the basilar membrane. For temporal theory, the same is true even when exceeding the minimum rate of neuronal firing. The ability to process pitch could also be attributed to being able to use the largest amount of energy in low frequency regions of the spectrum. However, listeners can also discern pitch of slope-shaped tones with little to no energy in the low frequency region or when the $F_0$ is missing, especially if they have a greater amount of musical training (see the missing $F_0$ demonstration or Seither-Preisler et al., 2007). It was thought this kind of pitch processing could be explained by the notion of periodicity pitch, which states all harmonics move in phase at the rate of the $F_0$. However, listeners can still perceive pitch in tones without periodicity, such as in narrow band noise (Eggermont, 2015).

Theories about pitch perception become even more complex because pitch is susceptible to influences from timbral characteristics. The nature of these interactions will provide the basis for the focus of the proposed investigation. Musical timbre refers to tone quality. For instance, a flute has a “light and airy” sound compared to the “darker” sound a clarinet produces. Specifically, “musical timbre concerns perceived tone quality that characterizes different sounds matched in pitch, loudness, and duration” (ANSI, 1973). This definition of timbre only identifies what it is not and is insufficient because timbre is multidimensional, comprised of the temporal, spectral, and spectrotemporal dimensions.

The temporal dimension refers to how the (intensity of the) tone changes over time. This includes the “rise time,” the start of the signal to the highest peak, “decay,” how the signal diminishes over time to a consistent amplitude, “sustain,” where the signal produces a consistent amplitude, and “release”, the end of the signal. Multidimensional scaling studies showed how a specific measure of rise time, log rise time, determined how similar or dissimilar instrument tones were perceived (see Grey, 1977; Grey & Gordon, 1978; McAdams, 1995; McAdams & Siedenburg, 2019).

The spectral dimension of instrument timbre is associated with spectral envelope shape, or how amplitudes change as a function of frequency. One acoustic measure that often captures the impact of differential weightings of the harmonics is the spectral centroid, which refers to the center frequency of the energy distribution of a signal (McAdams, 2019). There is some debate whether spectral envelope shape (e.g., Krumhansl, 1989) or spectral centroid (McAdams, 1995) is more pertinent to timbre. For instance, models comparing various groupings of spectral characteristics showed that spectral centroid best explained musical instrument timbre and was highly correlated to perceived similarity judgments as opposed to spectral shape (Grey & Gordon, 1978). Alternatively, the spectral envelope shape could be argued to be more useful when identifying spectral dimensions of timbre, at least for some particular instruments. For example, when musical instrument centroid was independently manipulated from formants (peaks in the sound spectrum), pitch perception changed as a function of spectral envelope shape determined by the center frequencies of formants, not the spectral centroid (Hall & Beauchamp, 2009).

Finally, the spectrotemporal dimension refers primarily/commonly to spectral flux, the degree to which changes in the spectral envelope occur over time (e.g., see Krumhansl, 1989; McAdams, 1995). Spectral flux was highly correlated with three dimensions of the six-dimension BIC timbre space so McAdams (1995) concluded a three-dimension timbre space, consisting of log rise time, spectral centroid, and spectral flux, was sufficient to classify various synthetic instrument timbres. This model provided the basis for creating accurate, complex perceptual representations of musical instruments (McAdams, 1995).

## Timbre and Pitch Interactions

Pitch judgments can be detrimentally affected by influences from timbre. The current investigation will examine several of those factors and focus on the relevant mechanisms to pitch processing. Before this can occur, the existing evidence for how pitch and timbre interact, as well as the contribution of specific timbre dimensions to pitch, needs to be reviewed.

There are many demonstrations that timbre and pitch interact. For example, one demonstration showed that timbre and pitch are perceptually integral characteristics through the Garner speeded classification task (Garner & Felfoldy, 1970). Pitt (1994) examined how listeners responded to pitch changes while keeping timbre constant, timbre changes while keeping pitch constant, and a combination of both pitch and timbre changes alike when presented with musical tones. Combined timbre and pitch changes took the most time to identify, but changes to one dimension impacted judgment for the other. For example, when timbre was varied, listeners were slower to identify pitch changes. In addition to being slower at identifying timbre changes, listeners categorized as non-musicians performed less accurately compared to musicians, often confusing timbre for pitch (Pitt, 1994).

Additional evidence of timbre and pitch interactions can be seen in the case of the missing $F_0$ (Seither-Preisler et al., 2007). To create pitch ambiguity, the $F_0$ was removed from individual tones, shifting the entire tone spectrum higher. Tones were paired together that would typically be perceived as having a falling $F_0$ (if the $F_0$ was not removed) and were defined by a rise in spectrum (because of the missing $F_0$). Listeners classified as non-musicians were unable to differentiate pitch due to the removal of the $F_0$ because they relied on that information to discern if tone pairs were rising or falling in pitch (Seither-Preisler et al., 2007). Changes in spectral envelope shape led to inaccurate pitch judgments. Becker and Hall (2014) replicated the missing $F_0$ experiment, but also examined the possible effects from centroid. Removing the $F_0$

shifted the centroid higher, along with the shift in spectrum, resulting in listeners perceiving tone pairs as rising in pitch (Becker & Hall, 2014). It is possible the de-emphasized lower frequency region negatively impacted pitch judgments and the interaction between spectral envelope shape and $F_0$ dictated the perceived pitch direction. For instance, when chroma tone pairs matched in $F_0$, spectral envelope shape dictated how listeners perceived pitch direction. Meanwhile, a redundancy gain occurred for different chroma tone pairs because of the missing $F_0$ shifting the spectrum higher in addition to the tone pair actually rising in pitch.

The experiments discussed previously have demonstrated the effect of timbre on pitch. It is noteworthy that the reverse, the effect of pitch on timbre, has also been observed. For example, we can observe how changes in $F_0$ affects judgments of timbre, specifically with the spectral envelope shape. Listeners are worse at vowel identification for both male and female voices with higher $F_0$ compared to the average, or low, $F_0$ (Ryalls & Lieberman, 1982). A possible explanation of these results pertains to how listeners were using the available harmonic information provided. Since the vowels were higher in pitch, the harmonics were becoming more spread out and there was less information within the spectrum, thus, the listener had less information to use to judge pitch. Therefore, both spectral envelope shape and spectral centroid should be considered when evaluating the pitch of complex tones because of the many contributing factors to pitch perception.

Timbre and pitch interactions can be further examined through the tritone paradox (Deutsch, 1986). Shepard tones are artificially made tones that are comprised of a series of octaves with the same tone chroma. These tones were created so amplitudes of the low and high frequency regions of the spectrum are de-emphasized within a bell-shaped spectral envelope

(Shepard, 1964). It was originally thought the spectral characteristics of Shepard tones created pitch ambiguity, but it is unclear which attributes or combination of attributes were the cause.

In the tritone paradox, listeners have to decide if pairs of Shepard tones are ascending or descending in pitch. These pairs are presented six semitones apart in distance to ensure an equidistant change in perceived pitch direction. Additionally, tone pairs include all chroma within an octave (e.g. C-$F^{\#}$, $C^{\#}$-G, D-$G^{\#}$, etc.) to account for influences from (fundamental) frequency. Listeners reported a cluster of successive tones pairs ascending in pitch in one direction while the other following of of tone pairs were judged as descending. The location around the chroma circle where tone pairs were judged from ascending to descending, or vice versa, was different across groups of people, but consistent with individuals’ data (Deutsch, 1986). Deutsch concluded that listeners have an absolute pitch ability based on chroma and could identify the tone height for each chroma position (Deutsch, 1991). However, listeners with absolute pitch should have been confused when presented with Shepard tones because they should have heard all octaves within the spectrum simultaneously and not able to make a pitch judgment at all. This explanation focuses on the pitch accuracy of the listener but fails to address the contributions of the physical aspects of Shepard tones. Another explanation of the individual variation seen in the tritone paradox has correlated the individual differences seen in pitch judgment accuracy with listeners’ native language (e.g. Deutsch, 1991). However, these linguistic effects could not be reproduced.

The large individual variation seen with the tritone paradox could be explained by factors not explored by Deutsch (1986), the first being related to the reliance on Shepard tones as stimuli and the degree to which the spectral characteristics of Shepard tones can be used to create pitch ambiguity. It was originally thought that Shepard tones could be perceived as a never-ending circle because they were equal in pitch height. However, pitch circularity has also been accomplished using harmonic tones (Deutsch, 2008) and modified sawtooth waves (Fugiel, 2011), bringing into question the significance of chroma for Shepard tones. Therefore, it cannot be that a singular characteristic of the Shepard tone is solely responsible for achieving pitch circularity and does not create pitch ambiguity. In fact, Shepard tones may be perceived in similar ways to naturally occurring tones, or even tones with more unusual spectral envelope shapes. For example, pitch judgment accuracy was comparable when listeners were presented with Shepard tones and with tones missing the $F_0$ (Hall & Becker, 2010). It is possible the shared characteristics of the spectral envelope shape, specifically, the de-emphasized lower frequency region, could explain the comparable results because certain listeners rely on the spectral envelope shape to make pitch judgments. When regions of the spectral envelope shape are missing or lessened, some listeners attempt to use the residual, incomplete information and come to inaccurate judgments because of that. Additionally, some listeners may hone onto timbral characteristics such as spectral centroid when it is not appropriate to use such a factor to process pitch, perhaps due to a lack of understanding or ability to differentiate timbre and pitch. This kind of confusion is not seen with those with prior musical training and pitch judgments often change as a function of musicianship. Spectral envelope shape, spectral centroid, and musical training could all contribute to pitch perception, but their separate influences needs to be examined together.

The second explanation of why we see individual variation in the tritone paradox can be attributed to how the contribution of spectral envelope shape, spectral centroid, and musical training is weighted differently across studies. For instance, Repp (1997) replicated the tritone paradox but standardized centroid placement. The data was analyzed to examine the effects of high or low centroid placement on pitch direction. Listeners reported pitch class based on centroid and pitch shifts could be as great 6 semitones (Repp, 1997). Spectral centroid explained the individual variation seen in the tritone paradox task, but only Repp (1997) attributed this as being significant.

Malek (2018) also replicated the tritone paradox, accounting for centroid effects, and found comparable results to Repp (1997) pertaining to the influence of centroid. However, Malek (2018) thought the individual variation seen in the tritone paradox was based on the degree to which the listener could detect the lowest frequency in the spectrum, acting as a mediating effect between pitch class and envelope shape. If listeners were truly unable to access amplitudes at lower octaves and unable to realize the $F_0$, their pitch judgment would have been skewed upwards in a comparable manner seen with the missing $F_0$ since the missing $F_0$ shifted the entire spectrum higher (Becker & Hall, 2014; Seither-Preisler et al., 2007). This would make tones with low $F_0$ seem perceptually higher due to their individual amplitude weighting of the $F_0$.

All these prior demonstrations of timbre and pitch interactions also change as a function of musical training. As a result, any study that involves timbre-pitch interaction, including the current investigation, should consider the potential relevance of musical training to the obtained data. All listeners are influenced by timbre-pitch interference effects. However, listeners with extensive musical training (often categorized as musicians) can still make accurate pitch judgments as opposed to those with less musical experience (non-musicians). For non-musicians, performance accuracy decreased when there was a pitch change and even more so when there was a timbre change. Changing both pitch and timbre disproportionately affected non-musicians while musicians remained highly accurate for all categories (> 92%; Pitt, 1994). Musicians can also make accurate pitch judgments when the $F_0$ is missing and are less susceptible to spectral changes (Seither-Preisler et al., 2007). It could be assumed that non-musicians focused on the physical sound attributes to realize the $F_0$, using $F_0$-pitch cues to make their pitch judgments.

In Repp's replication of the tritone paradox, listeners seemed to fall into one of two categories, synthetic or analytic, which was related to non-musicians and musicians, respectively. Synthetic listeners reported hearing multiple components and were uncertain in their response. Analytic listeners heard the pitch change as either ascending or descending and were certain. It is possible the analytic listeners were responding to both the pitch (frequency) and timbre (spectral centroid) information they were presented with (Repp, 1997). Neither Deutsch (1986) nor Malek (2018) accounted for the relationship of musicianship and pitch judgment accuracy or the influence of timbre in their tritone paradox research. Even though Repp (1997) showed how important spectral envelope shape was in explaining the individual variation seen in the tritone paradox, replications have ignored this contributing factor.

## The Current Investigation

The tritone paradox provides an opportunity to study the combined influence of spectral envelope shape, spectral centroid, tone chroma, and musical training, but interpretation of results from the original study (Deutsch, 1986) as well as replications (Malek, 2018; Repp, 1997) differ in which aspects of timbre and what other factors are significant to pitch perception. This controversy has contributed to the uncertainty surrounding why individual variation is seen in the tritone paradox as well as how Shepard tones create pitch ambiguity. We know that timbre and pitch are integral properties that can be confused with one another, in part because they influence each other (Pitt, 1994; Seither-Preisler et al., 2007). At the same time, some listeners are less affected by this relationship while others are detrimentally impacted (as seen in the Garner speeded classification, missing $F_0$, and tritone paradox). Therefore, parsing between the effects of timbre and pitch can be difficult because they are interconnected. An orthogonal examination of contributing factors would convey both their individual and combined influence on pitch, but prior research has mainly focused on timbre characteristics separately.

The current investigation will orthogonally examine the specific circumstances in which spectral envelope shape, spectral centroid placement, and tone chroma influences pitch. Results could determine the possible shared mechanisms of pitch perception that could apply across various/most types of tones. This will be accomplished by replicating the tritone paradox task, asking listeners to make a directional pitch judgment for tones differing in $F_0$ by six semitones varying in spectral envelope shape (e.g. slope-shaped, tones de-emphasizing the lower frequency region, and Shepard tones). For example, slope-shaped tones could be directly compared to Shepard tones or other tones with shared spectral characteristics to Shepard tones to determine which spectral characteristics contribute to pitch ambiguity. By replicating the tritone paradox, we will also be able to clarify the discrepancies between the original (Deutsch, 1968) and various replications (Malek, 2018; Repp, 1997) pertaining to the relationship of centroid and individual variation in the tritone paradox as well as with Shepard tones and pitch ambiguity.

Finally, musical training will be evaluated through correlations with pitch judgment sensitivity to determine how past musical experience influences pitch processing. Musicians should be accurate in their pitch judgments and able to overcome influences from timbre. Specifically, people with more musical training should be able to separate timbre and pitch influences (see Pitt, 1994) and realize spectral information surrounding the $F_0$ even when it is missing (Seither-Preissler et al., 2007) or that region is de-emphasized (Malek, 2018).

# Method

## Participants

Study participants consisted of 43 undergraduate students at James Madison University. All participants were compensated with 1.5 credits for an hour of their time in partial fulfillment of their course requirement through *SONA participant systems* (Sona Systems, n.d.). Age range was limited to a maximum of 40 years to avoid effects due to presbycusis, age-related high-frequency hearing loss (Bonfils, Bertrand, & Uziel, 1988). No participants self-disclosed hearing difficulties or were above the age of 40 years. A prior exclusion criteria determined seven participants' data were removed from the final analysis due to technical errors during the testing session. Additionally, there was one instance in which it was determined the participant was unable to accurately judge the slope tones because they had an average d' value of 0.

Musical training was evaluated through self-report via a survey (discussed below). Students encompassed 0 - 41 cumulative years of instrumental study (*M* = 6.22, *SD* = 8.02), 0 – 14.42 years involved with music (*M* = 4.59, *SD* = 4.55), with 0 - 10 years of music lessons on an instrument (*M* = 1.26, *SD* = 2.21).

## Questionnaire

All participants completed the Musical Training and Experience (MuTE) survey (Daly & Hall, 2017) to permit an evaluation of possible correlations between amount of musical training and task performance. The full survey is included as Appendix A (Daly & Hall, 2017). The MuTE survey is a 20-item survey that focuses on the participant's musical training and prior experiences, both formal and informal, with music. Formal musical training includes music classes taken at school, from elementary to college level, and the type of class, such as composition or performance. Informal training includes experience with musical theater, dance, and musical video games. Items also pertain to amount of performance training with voice and instruments, as well as individually (e.g. private lessons/practice) and within groups (e.g. ensembles, bands).

Dichotomizing musicianship into musicians and non-musicians leads to decreased reliability and power as well as increased spuriousness. For example, using a 2-year cutoff point, people with 2 and 20 years of musical training could be categorized as musicians even though this reflects dissimilar musical experiences. From a data perspective, this leads to data loss and an overgeneralization of musicianship (Daly & Hall, 2017). A better measure of musicianship would be to look at musical training in years of experience as a continuous, rather than a categorical variable. The MuTE survey examines each aspect of musical training as a continuous measure (in years/months of experience), so we retain all significant information based upon the length of time spent on a particular type of training. These variables can be used in regression analyses to determine if task performance changes with the amount of musical training.

**Stimuli**

All tone stimuli were generated in Ableton *Live Suite 11* (Version 11.2, Ableton, 2021). All stimuli were 500*ms* in length with 20*ms* linear ramps in amplitude at onset and offset. The sampling rate was 44.1kHz (with 16-bit depth resolution). Tones were matched for average RMS amplitude, and the peak amplitude of all stimuli did not exceed 80 dB[A]. Sounds were presented in a sound-attenuated chamber over headphones (Sennheiser HD 25-SP II).

Twelve Shepard tones from the original samples used in Deutsch (1986) were extracted from *Musical Illusions and Paradoxes* (Deutsch, 1995) for several chroma ($C_3$ (130.81 Hz), $C^{\#}_3$ (138.59 Hz), $D_3$ (146.83 Hz), $D^{\#}_3$ (155.56 Hz), $F^{\#}_3$ (185.00 Hz), $G_3$ (196.00 Hz), $G^{\#}_3$ (207.65 Hz), and $A_3$ (220 Hz)) at two spectral centroids ($C_5$ (523.25 Hz) and $F^{\#}_5$ (740 Hz)). These particular chroma were chosen because they encompass the majority of the chroma circle (see Figure 1) while accounting for Deutsch's original stimuli and the spectral measurements of all spectral envelope shapes and corresponding centroid values. Using these chroma will provide the best chance at seeing the change in perceived pitch direction similar to Deutsch (1986).

The centroids $C_5$ and $F^{\#}_5$ were the highest values that Deutsch (1986) used. The higher centroid values allow for manipulation of amplitude slope to achieve the proper spectral envelope shapes while maintaining chroma position. By matching the centroid values of the Shepard tones to the other types of tones discussed below, we can determine if centroid could explain the individual differences seen in the tritone paradox. The centroid position will be referred to as centroid value while whether the centroid matched between tones on a trial or not will be referred to as "centroid match/mismatch," specifically referencing the comparison tone. Spectral centroid was calculated by summing all frequencies multiplied by their amplitudes, then dividing by the sum of amplitudes, as indicated in the formula below (Peeters, 2004):

$$= \frac{\sum_{k=1}^{N} kF[k]}{\sum_{k=1}^{N} F[k]} \tag{1}$$

Harmonic amplitudes were calculated to achieve the targeted centroid values and form the intended spectral envelope shapes. The individual amplitude values of each harmonic for the remaining types of tones were created through a *Max for Live* (Version 11.2, Ableton, 2021) plug-in called *FormAnt* (Hall & Redpath, 2016). *FormAnt* is able to create individual amplitudes for each harmonic and is capable of making amplitude adjustments to up to 50 harmonics. This software also prevents aliasing noise, distortion artifacts, by eliminating frequencies above 18kHz (for more on anti-aliasing, see Brahma, De Leon, & Kavasseri, 2009).

One subset of tones, referred to as *standard* tones, were used as the standard stimuli for perceptual comparisons to the remaining/*altered* tones. Standard tones lowered in amplitude as frequency increased by a user-defined spectral slope value in order to match the two spectral centroid values. There were sixteen standard tones, matching the combinations of the eight chroma of the Shepard tones with two possible centroid positions.

There were three sets of comparison tones defined by three distinct spectral envelope shapes. Based on the three spectral envelope shapes, eight chroma, and two centroid positions, there was a total of 48 comparison tones.

One set of comparison tones (referred to as “slope tones”) either matched or mismatched the centroid of the standard tones. Slope tones provided a baseline for participant performance because the spectral envelope shape of these tones formed a slope shape that approximated naturally produced tones. We also examined the effect of centroid on pitch judgments for mismatched centroid tone pairs. Since the lower frequency region of the slope tones were not de-emphasized, direct comparisons could be made to the other comparison tones with the de-emphasized lower frequency region to determine if individuals could realize the $F_0$ and determine the impact of the lower frequency region on pitch judgments.

The second set of comparison tones were the Shepard tones described previously. The use of Shepard tones helped determine the influence of low frequency prominence (relating to Malek, 2018 interpretation) since these tones are bell-shaped and de-emphasize both the low and high frequency regions.

The third/final set of comparison tones (henceforth referred to as “peak tones”) increased in amplitude for the first two harmonics and then decreased in amplitude while increasing in harmonic number. The spectral slope of these tones were adjusted in position to match each centroid value. These tones helped provide the basis for individual frequency region prominence since the spectral envelope shape de-emphasized the lowest frequency region without having the bell-shaped curve, a characteristic of the Shepard tone.

## Procedure

First, informed consent was obtained. All participants were told about the nature of the task and what to expect during the testing session. Instructions used simple language so all participants could understand the task. It is critical that performance accuracy not be influenced by simple misunderstanding due to musical terminology since non-musicians can perform just as well as musicians on pitch judgment tasks if special terminology is avoided (Bigand & Poulin-Charronnat, 2006).

Participants completed a 2-alternative-forced-choice (2AFC) discrimination to determine pitch direction across a pair of tones on any given trial, replicating the tritone paradox task. This task was administered through Direct RT experiment and data collection software (Version 2018.1.117, Jarvis, 2018a). On each trial, a standard and comparison tone was presented in a sequence with a 300*ms* inter-stimulus interval. Listeners were asked to judge if the tone pairs were rising or falling in pitch. Responses were indicated by pressing one of two labeled buttons (i.e., “1” for “DOWN”, and “9” for “UP”).

There were four possible standard-comparison pairings of tone chroma (i.e., $C_3$-$F^{\#}_{3;}$ $C^{\#}_3$-$G_{3;}$ $D_3$-$G^{\#}_{3;}$ and $D^{\#}_3$- $A_3$). For each tone chroma pair, the difference in $F_0$ was always 6 semitones. Tone chroma ordering was counterbalanced (e.g. $C_3$ – $F^{\#}_3$ v. $F^{\#}_3$ – $C_3$), as well as the position of which tone represented the standard (i.e., standard-comparison; comparison standard). Additionally, the two possible centroid values ($C_5$ and $F^{\#}_5$) were either matched or mismatched between tone pairs. With four possible chroma pairs, two chroma pair orders, two standard positions, two centroid values, whether the centroid will match or mismatch, these variables provided 64 condition combinations for each chroma pair, totaling 192 possible combinations when including the three possible spectral envelope shapes for the comparison tones (see Table

1). Trials were block randomized across all conditions. There was a total of 960 trials. The 2AFC task took approximately 35 minutes but was self-paced. All participants had the opportunity to take a short break after each set of 192 trials. All combinations of stimulus conditions occurred after each block.

After the experimental task, participants completed the MuTE survey (Daly & Hall, 2017). All survey responses were collected through *MediaLab* survey administration and data collection software (Version 2018.1.109, Jarvis, 2018b). The MuTE survey took approximately 10 minutes. Finally, participants were debriefed to determine if they understood the experimental task or if any issues were encountered during testing. (See Appendix B for a summary of debriefing questions.) The entire testing session took approximately 60 minutes per participant.

## Hypotheses

All listeners (regardless of musical experience; discussed below) were expected to be most accurate in determining pitch direction at baseline for tones reflecting the same spectral envelope shape when there is a high amount of energy near the $F_0$ like in typical listening (standard-slope tones). Should this not be the case, these individuals' data were examined further (see "Results" section for more details).

According to prior tritone paradox results, since Deutsch (1986) showed tone pairs were perceived as changing in pitch direction somewhere across the chroma circle, it was expected the perceived pitch direction of Shepard tones would be based on the chroma as well. The position at which listeners judge where the pitch direction changes should vary across participants but remain consistent within an individuals' dataset as well. It was thought this individual variation should also relate to centroid position, similar to findings from Repp (1997) and Malek (2018).

A comparison of task performance for the Shepard and peak tones showed how spectral envelope shape influences pitch judgments separate from the emphasis of the low frequency region. If energy in the low frequency region significantly influenced pitch judgments, performance should be similar for Shepard and peak tones since they both de-emphasize the lower frequency region. If listeners' respond differently to the Shepard tones compared to the slope tones, the difference in spectral envelope shapes could explain this pattern of results.

If responses to Shepard tones are different compared to slope tones, another condition would be necessary to address why pitch circularity can occur with Shepard tones. Varying centroid will also support why we see individual differences in the tritone paradox. However, if performance is comparable between Shepard and slope comparison tones, this would suggest that the strange effects related to Shepard tones are not necessarily unique to Shepard tones even though they result in unusual pitch judgments.

If the pattern of responses comparing peak and slope tones are different, additional research will be required to examine how specific characteristics of peak tones influenced pitch judgments. For example, the largest amount of energy for the peak tones will be displaced slightly higher than the energy in the region surrounding the $F_0$ for the slope tone. This could potentially result in the peak tones being heard as being higher in frequency. If the pitch judgments for the peak and slope tones are the same, this could provide further evidence the de-emphasized lower frequency region is not the crucial element in the Shepard tone though it is unclear how listeners will respond to the peak tones.

Musical training should also influence pitch judgments. If musically experienced listeners respond to the slope tones and tones that de-emphasize the $F_0$ (Shepard and peak tones) the same, this could further support that musical training allows individuals to realize the $F_0$ (like with the missing $F_0$; Seither-Preisler et al., 2007) to make accurate pitch judgments and are not solely reliant on the physical characteristics of the signals. Listeners with more musical training should not exhibit any difficulty judging tones with spectral envelope shapes that de-emphasize the lower frequency region near the $F_0$ (Shepard and peak tones). However, if those with more musical experience judge the Shepard tones worse than peak tones, this could suggest there are elements not shared between the Shepard tones and peak tones that influence pitch judgments, such as how Shepard tones contain all octaves. Additional research pertaining to Shepard tones specifically addressing octave equivalence would be necessary while considering musical training. If musically experienced listeners judge the peak tones worse than the Shepard tones, there could be an additional characteristic about the peak tone that would need to be further examined.

Listeners with less musical training should struggle with both the Shepard and peak tones because they will be reliant on the spectral information left from the de-emphasized $F_0$, such as spectrum shift or centroid. There could be potential differences in pitch judgment accuracy between the peak and Shepard tones. If pitch judgment accuracy is worse for Shepard tones, this could also support there are additional unique spectral qualities about Shepard tones that contributes to pitch perception. If pitch judgment accuracy is worse for peak tones, additional conditions will be necessary to determine why this is so, but results such as this could potentially be explained by the unique spectral envelope shape. It is not expected that musically inexperienced listeners should perform just as well on the peak or Shepard tones compared to the slope tones.

## Results

Signal detection theory (Green & Swets, 1966; Macmillan & Creelman, 2004) was used to differentiate listeners' pitch judgment ability by calculating d' as a theoretically unbiased measure of sensitivity. A response consisted of a hit if the tone with the lower $F_0$ in the pair was present in the first interval and the participant said it was. When the tone with the lower $F_0$ was in the second interval and the participant said it was in the first interval, a false alarm was counted. The typical "square root of two" adjustment was used to account for any performance inflation that may be seen within a 2AFC task, which provides more conservative measures of d' (Macmillan & Creelman, 2004). When converting probabilities to z-scores, probabilities also were limited between a minimum of 0.001 and a maximum of 0.999 in to reduce any inflation that may occur when converting hits and false alarms (see Macmillan & Creelman, 2004). Values of d' were calculated using the equation below where z(H) represents the z-score for the hit rate and z(FA) represents the z-score for the false alarm rate:

$$= \frac{z(H)-z(FA)}{\sqrt{2}} \tag{2}$$

Respondents were expected to be most sensitive to slope tones since they are similar to naturally occurring tones by containing the largest amount of energy near/at the $F_0$ which reduces as component frequencies increase. Therefore, if a listener within the current investigation was unable to judge the pitch of slope tones based upon their respective fundamental frequencies, it similarly would not be expected that they would be able to accurately judge pitch for the other types of tones. If listeners were relatively insensitive in their standard-slope pitch judgments ($d' < 1$) across all combinations of conditions defined by centroid and chroma, then their data was discarded from analyses due to the fact that their data reflected negatively impacted pitch processing ability relative to the majority of listeners.

For each participant, d' value was separately calculated for each combination of spectral envelope shape, centroid value/match condition, and chroma value for the tone in the first interval. The resulting d' values were submitted to a 3 x 4 x 8 repeated measures ANOVA examining the interaction of spectral envelope shape (peak, Shepard, slope), centroid ($C_5$ match, $C_5$ mismatch, $F^{\#}_5$ match, $F^{\#}_5$ mismatch), and chroma (C, $C^{\#}$, D, $D^{\#}$, $F^{\#}$, G, $G^{\#}$, A). An additional analysis was conducted to further understand the influence of spectral centroid on pitch perception for Shepard tones. This analysis for the Shepard tone condition was a 2 x 2 x 8 repeated measures ANOVA was conducted with centroid value ($C_5$, $F^{\#}_5$), centroid match/mismatch, and chroma (C, $C^{\#}$, D, $D^{\#}$, $F^{\#}$, G, $G^{\#}$, A) as factors.

Separate sets of differences scores were calculated to further investigate the relative magnitude of the spectral shape effects for peak and Shepard tones. The resulting difference scores were submitted to a 2 x 2 x 8 repeated measures ANOVA with spectral envelope shape (peak, Shepard), centroid ($C_5$ match, $F^{\#}_5$ match), and chroma (C, $C^{\#}$, D, $D^{\#}$, $F^{\#}$, G, $G^{\#}$, A) as factors. Positive difference scores indicated that sensitivity to slope tones was higher than the alternative (peak or Shepard tone) comparison spectral envelope shape. Conversely, negative difference scores indicated that sensitivity to slope tones was lower than for the alternative comparison tones.

For all ANOVA analyses, Bonferroni-pairwise comparisons were conducted to clarify obtained effects. Since every analysis required multiple comparisons per factor due to the number of levels for each variable, the Bonferroni correction reduced the chance of a type I error, and its conservative measure ensured that any statistically significant findings were reliable. Greenhouse-Geisser corrections were used to address any violations of sphericity assumptions. Since all analyses contained within-subjects factors (e.g., spectral envelope shape, centroid, chroma), Greenhouse Geisser corrections were most appropriate because they account for the possible variance seen at each level of the independent variables by adjusting the *p*-value to 0.05.

**Spectral Characteristics**

Figure 2 depicts the average d' values and corresponding standard error bars for each spectral envelope shape according to centroid value/match and starting chroma. Looking at Figure 2, it can be seen that sensitivity remained high, and neither starting chroma nor centroid affected d' levels, for both the peak and slope tones. The greatest variation in sensitivity can be seen for the Shepard tones according to centroid (i.e. $C_5$ vs $F^{\#}_5$) as a function of starting chroma. Mean values of d' remained reliable across chroma for the matched and mismatched centroid conditions at $C_5$, while sensitivity varied across chroma for the $F^{\#}_5$ matched and mismatched centroid conditions. This pattern of results contributed to a significant interaction of centroid and starting chroma as a function of shape, $F(16.611, 697.651) = 9.603$, $p < 0.001$, $\eta^2 = 0.186$. The nature of this 3-way interaction can be understood by examining pairwise comparisons. For either peak or slope tones, sensitivity did not significantly/statistically vary as a function of centroid or chroma, $p \geq 0.08$. In contrast, for Shepard tones, sensitivity at the $C_5$ centroid was higher than sensitivity at the $F^{\#}_5$ centroid for chroma at C, $C^{\#}$, D, $D^{\#}$, $p \leq 0.024$; conversely, sensitivity at the $C_5$ centroid was lower than sensitivity at the $F^{\#}_5$ centroid for $F^{\#}$, G, $G^{\#}$, A, $p \leq 0.006$.

Since responses to Shepard tones showed the greatest variation in sensitivity, difference scores were calculated by subtracting corresponding d' for the peak/Shepard tone condition at each combination of centroid and starting chroma from the d' for the slope tone condition, which served as a baseline condition (i.e., d' for slope – d' for comparison). For each centroid-chroma combination, these difference scores represent the degree to which the critical spectral envelope shapes conditions impact pitch perception compared to their corresponding baseline conditions. Figure 3 shows the difference scores for peak and Shepard tones relative to the baseline/slope condition according to matched centroid condition and starting chroma. Within Figure 3, larger, positive difference scores for the Shepard tones indicated that sensitivity was higher for the slope tones compared to the Shepard tones across centroid and chroma. Specifically, the magnitude of the difference scores for the Shepard tones at the $F^{\#}_5$ centroid condition decreased examining sensitivity from tones with a starting chroma of C to A, while difference scores at the $C_5$ centroid remained at similar levels across chroma. In contrast, sensitivity for peak tones showed little deviation from performance for slope tones across all starting chroma values, as indicated by relatively small difference scores. This pattern contributed to a significant interaction of spectral envelope shape, centroid, and chroma, $F(5.028, 211.156) = 15.125$, $p < 0.001$, $\eta^2 = 0.265$.

Subsequent pairwise comparisons revealed that across Shepard tones difference scores at the $C_5$ centroid for C, $C^{\#}$, D, $D^{\#}$ were lower than scores at the $F^{\#}_5$ centroid for all starting chroma ($p \leq 0.015$) except $D^{\#}$ ($p = 0.058$). For tones with a starting chroma of $F^{\#}$, G, $G^{\#}$, and A, difference scores at the $C_5$ centroid were higher than scores at the $F^{\#}_5$ centroid across all starting chroma, $p \leq 0.021$. The uniform pattern of results for the peak condition in Figure 3 is further supported by pairwise comparisons which showed difference scores for the peak tones did not significantly vary as a function of either centroid or chroma, $p \geq 0.124$.

The pattern of results that contributed to the three-way interaction also yielded a significant interaction of spectral envelope shape (peak, Shepard, slope) and chroma (C, $C^{\#}$, D, $D^{\#}$, $F^{\#}$, G, $G^{\#}$, and A), $F(1.827, 76.743) = 11.056$, $p < 0.001$, $\eta^2 = 0.208$. Looking at Figure 2, it can be seen that sensitivity was less varied for the peak and slope tones across starting chroma and significantly more varied for Shepard tones. According to pairwise comparisons, overall performance across starting chroma for the peak and slope tones did not significantly differ from each other ($p \geq 0.091$), while sensitivity for both peak and slope tones was significantly higher than for Shepard tones ($p \leq 0.029$).

The greater variation in sensitivity for the Shepard tone condition is related to spectral centroid. Looking at Figure 2, it can be seen that sensitivity at the $F^{\#}_5$ centroid was significantly lower at C, $C^{\#}$, D, $D^{\#}$ relative to the $C_5$ centroid condition, while sensitivity at the $F^{\#}_5$ centroid was higher at $F^{\#}$, G, $G^{\#}$, and A. This pattern of results is supported by a significant interaction of centroid and starting chroma, $F(11.973, 502.863) = 10.050$, $p < 0.001$, $\eta^2 = 0.193$. Pairwise comparisons further revealed that sensitivity for the $C_5$ matched and mismatched centroid conditions was significantly higher than for the $F^{\#}_5$ matched and mismatched centroid conditions within tone pairs starting with C, $C^{\#}$, D, or $D^{\#}$ chroma, $p \leq 0.045$. For tone pairs with a starting chroma of $F^{\#}$, G, $G^{\#}$, and A, sensitivity for the $C_5$ centroid condition was frequently lower than for the $F^{\#}_5$ centroid condition. Specifically, sensitivity for the $C_5$ centroid matched condition was lower compared to the $F^{\#}_5$ matched centroid condition for the tone pair starting with $F^{\#}$, the $F^{\#}_5$ mismatched centroid condition for the $G^{\#}$ starting chroma, and for both the $F^{\#}_5$ match and mismatched centroid conditions for the $A^{\#}$ starting chroma, $p \leq 0.045$.

An additional analysis was conducted to further understand the contribution of the spectral centroid to pitch perception for Shepard tones, separating the effects of centroid value and centroid matching. This analysis for the Shepard tone condition was a 2 x 2 x 8 repeated measures ANOVA was conducted with centroid value ($C_5$, $F^{\#}_5$), centroid match/mismatch, and chroma (C, $C^{\#}$, D, $D^{\#}$, $F^{\#}$, G, $G^{\#}$, A) as factors. As evaluated by the aforementioned analyses, Figure 2 reveals sensitivity was lowest for C, $C^{\#}$, D, $D^{\#}$ and highest for $F^{\#}$, G, $G^{\#}$, A at the $F^{\#}_5$ matched and mismatched centroid conditions relative to the $C_5$ matched and mismatched centroid conditions. This pattern contributed to a significant interaction of centroid value and chroma, $F(3.590, 150.770) = 46.505$, $p < 0.001$, $\eta^2 = 0.525$. Pairwise comparisons further confirmed sensitivity for the $C_5$ centroid value was higher than the $F^{\#}_5$ centroid for C, $C^{\#}$, D, $D^{\#}$, $p < 0.001$. For $F^{\#}$, G, $G^{\#}$, A, sensitivity for the $C_5$ centroid value was lower than the $F^{\#}_5$ centroid, $p < 0.001$. While the choice of centroid placement did influence pitch judgments, the matched centroid condition had no effect, $F(1, 42) = 0.164$, $p = 0.687$, $\eta^2 = 0.004$.

## Relationships to Musical Training/Prior Experience

Finally, to assess the relationship between musicianship and task performance (d'), three musical training measures were derived from the MuTE survey. For a given listener, *cumulative years of instrumental study* were calculated by taking all of the instruments they studied and summing all years they studied each instrument. For instance, if a person studied flute for 19 years, piano for 2 years, and guitar for 4 years, the cumulative years of instrumental study would be 25 years. Additionally, *years involved with music* were calculated by subtracting the first date the listener started practicing music from the last date the listener practicing music. For example, if a listener started learning to play a musical instrument at 7 years of age and is 27 years old, then this person has 20 years of music involvement assuming they still are involved with music. Finally, listeners also directly indicated *years of lessons*, that is, how many years of formal music lessons they had received. All of these musical training measures were submitted to separate correlation analyses examining the relationship between musical training with the average d' according to shape (peak, Shepard, slope) and matched centroid ($C_5$ and $F^{\#}_5$). For these analyses, sensitivity was calculated by averaging across starting chroma for each spectral envelope shape and centroid value.

Figure 4 depicts the relationship between musical training and sensitivity for each spectral envelope shape. Panel 1 shows a scatterplot with d' as a function of the cumulative years of instrumental study, panel 2 depicts years involved with music, and panel 3 shows years of lessons. Figure 4 reveals that increased musical training was positively related to greater sensitivity for pitch judgments across all musical training measures and spectral envelope shape conditions. Statistically significant correlations were found for all spectral envelope shape conditions. Sensitivity in the Shepard tone condition was strongly, positively related to all musical training measures (years involved with music, $r = 0.714$, $p < 0.001$; cumulative years of instrumental study, $r = 0.571$, $p < 0.001$; years of lessons, $r = 0.454$, $p = 0.002$). While the corresponding obtained positive correlations were not as strong for performance in the peak and slope conditions, statistical significance also was achieved for each musical training measure (years involved with music: peak: $r = 0.525$; slope: $r = 0.587$, $p < 0.001$); cumulative years of study: peak: $r = 0.458$; slope: $r = 0.511$, $p < 0.002$; years of lessons: peak: $r = 0.351$; slope: $r = 0.373$, $p < 0.021$).

## Discussion

### Spectral Envelope Shape and Chroma Interactions

If pitch judgment accuracy changed as a function of the amount of energy placed in the lower frequency region of the spectrum, then d' should be greater for slope tones relative to the other conditions because the frequencies including, and proximal to, the $F_0$ contained the highest amount of energy within the spectral envelope. Sensitivity for slope tones indeed reflected reliable task performance; d' remained above three, with little variability across centroid or chroma (see Figure 2). High d' values indicated that the slope tone condition could act as a reliable baseline for pitch performance.

Likewise, it was predicted that sensitivity/task performance for peak tones would suffer based on the tones' de-emphasized lower frequency region surrounding the $F_0$. Instead, sensitivity for peak tones remained high, comparable to that given slope tones (see Figure 2). This is despite the fact that the average relative amplitude of the $F_0$ for the peak tones was significantly reduced, rendered at -63.88 dB (*SD* = 0.08 dB). This suggests that listeners were still able to derive enough information from the $F_0$ to make accurate pitch judgments, that is, that the lowest frequency components in the peak tones were intense enough to be resolved. If this were the case, then the amplitude manipulation for the peak tones was not strong enough to manipulate pitch perception. In the current study, had the $F_0$ in the peak tones been reduced enough, pitch shifts should have occurred for listeners with less musical training, making sensitivity more variable, similar to what was seen in the case of the missing $F_0$ (Becker & Hall, 2014; Seither-Preisler et al., 2007). However, no such pitch shifts occurred for any respondents, and responses to peak tones more so resembled responses to that of slope tones. It was originally thought that the de-emphasized lower frequency region would result in inaccurate pitch judgments, and thus the peak tones would act as a comparison to the Shepard tones since both de-emphasized the lower frequency region around the $F_0$. Instead, sensitivity to the peak tones was consistently high while sensitivity to the Shepard tones was more variable. This pattern of results further supports that the peak and Shepard tones are not actually comparable.

If the $F_0$ was removed from the peak (or, for that matter, slope) tones, listeners with greater musical training would still be able to derive the $F_0$ because the remaining components would share periodicity at the rate of the $F_0$ (see Eggermont, 2015). This is not possible with Shepard tones since Shepard tones require octave-relationships between component frequencies (while the peak tones contain harmonic (i.e., integer multiple) frequencies) and have fewer component frequencies. Each component frequency within a Shepard tone provides equally important information pertaining to tone height, and multiple octaves are presented simultaneously. For Shepard tones with an $F_0$ that could not be resolved, it would be expected that these tones would be perceived as being an octave higher since all other components are harmonically related to the lowest remaining component frequency. If the Shepard tones containing octave components solely explains how listeners process Shepard tones, additional conditions exploring different tone types containing all octaves would be necessary. However, it is unlikely that a reliance on octaves explains the pattern of results in the tritone paradox since pitch circularity has been achieved with tones containing all octaves and for tones with harmonically-related components.

It was also hypothesized that sensitivity to Shepard tones would be lowest compared to the other spectral envelope shape conditions and d' would vary according to chroma because of the bell-shaped envelope of the Shepard tones de-emphasizing the lower and higher frequency region. As expected, sensitivity varied across chroma. However, as seen in Figure 2, sensitivity for tones with a starting chroma at $F^{\#}$, G, $G^{\#}$, and A remained above 1. Another characteristic of the Shepard tone must be the reason why such variability in sensitivity was seen.

As revealed by acoustic analyses, the spectral composition of all the Shepard tones was not standardized across chroma and centroid. According to the original specifications from Deutsch (1986), all Shepard tones were supposed to contain a total of six frequency components, each an octave above the next corresponding to the $F_0$. For example, the $C_3$ Shepard tone at the $C_5$ centroid contains six components starting at 130.81 Hz spaced by octaves as intended. However, not all Shepard tones from Deutsch (1986) used in the current study were rendered with six frequency components and the $F_0$ did not always correspond with the expected octave because the centroid value had to be maintained.

To achieve the proper centroid values, the spectral envelope shape was shifted, which influenced the components present within the spectrum. Figure 6 depicts the spectral slices of Shepard tones at each chroma for the low centroid value. Looking at the tones with a starting chroma of D and $D^{\#}$ at the $C_5$ centroid, the $F_0$ was located in the second octave frequency range instead of the third as intended. According to the acoustic analysis, there were also frequency components missing in the C and $C^{\#}$ tones at the $C_5$ centroid, resulting in the total spectrum containing five frequency components instead of six. Similar influences of centroid on the spectral envelope shape can be seen at the $F^{\#}_5$ centroid condition. Figure 7 shows the Shepard tones at each $F_0$ for the high centroid value. Adjusting the centroid value resulted in the inclusion of frequency components an octave lower than expected for the $G^{\#}$ and A chroma, as well as in the loss of a low component for the $F^{\#}$ and G chroma (see Figure 7).

Typically, a listener would rely upon $F_0$ to make pitch judgments. Table 2 provides information about the $F_0$ for each Shepard tone according to chroma as well as the corresponding obtained d' value and standard error for pitch judgments. Amplitudes at $F_0$ varied considerably across chroma, ranging from -53.9db to -18.4dB (*M* = 39.78, *SD* = 11.02, see Table 2). For Shepard tones ($D_3$ and $D^{\#}_3$ at $C_5$ centroid and $G^{\#}_3$ and $A^{\#}_3$ at $F^{\#}_5$ centroid) with an $F_0$ an octave lower than the intended frequency range, d' levels suggested listeners were not always able to resolve the $F_0$ because it was so low in amplitude. In these cases, listeners would try to use the second component, an octave above the previous component until they could make a pitch judgment. The listener’s range of hearing would then determine the accuracy of their pitch judgments based on their ability to resolve amplitude components in the lower frequency region.

This would then change their perception of pitch direction. For example, when presented with the $G^{\#}$ – D tone pair at the higher centroid, listeners who could resolve the lowest component would report this tone pair as ascending while listeners who could not resolve the lowest component would report this tone pair as descending because they were judging pitch based on $G_2$ and $G_3$ respectively. In these cases, the way in which d' was calculated would have to be changed to follow the pitch direction according to either the $F_0$ (when the $F_0$ could be resolved) or the frequency at the second component (when the $F_0$ could not be resolved).

Since Shepard tones are supposed to be ambiguous to pitch, some might argue that d' cannot accurately represent sensitivity to Shepard tones since hits and false alarms dictate a "correct" and "incorrect" response. Therefore, if listeners were using the $F_0$ for tones including lower frequency components low in amplitude, d' should have been reduced. However, sensitivity for these tones generally remained high (see Table 2). If d' was calculated according to $F_0$ in these cases, then what was originally considered a hit would become a false alarm and false alarms would become hits. Thus, d' would become negative, but the measured sensitivity (indicated by its distance from 0) would not change. Overall, sensitivity was best explained by listeners using the second component and increased amplitude of the $F_0$ was related to increased d'.

This argument about relative amplitude of the lowest component becomes more complicated since loudness is frequency dependent. Equal loudness contours indicate that listeners need more energy below 100 Hz and above 10,000 Hz to hear isolated frequencies at a resolvable level (Fletcher & Munson, 1933; Robinson & Dadson, 1956). With more complex stimuli, even more energy is necessary. Since intensity and frequency are integral properties (Grau & Nelson, 1988; Melara & Marks, 1990), this directly impacts how listeners judge pitch for the Shepard tones. Depending on the frequency value of the $F_0$, certain tones may be more easily resolved since the bell-shaped envelope limits the amplitude in the lower frequency region.

The relationship between frequency and intensity relates back to the idea of virtual pitch, that is, how listeners infer pitch from harmonics other than $F_0$ (see Terhardt, 1991). Malek (2018) expanded upon this notion with the idea of “lower frequency region prominence,” in which a listener’s ability to access the lowest frequency component determines how pitch judgments are made. While Malek’s idea was never subsequently tested directly, how listeners process Shepard tones supports that listeners systematically process lower frequency components starting at the $F_0$ until they can make a pitch judgment since, as Terhardt (1991) stated, listeners can infer pitch from harmonics. Contrary to Terhardt (1991) and Malek (2018), it is important to note that listeners are not necessarily reliant on components where the energy is greatest in the spectrum based on how listeners judge Shepard tones using lower frequency components rather than the centroid. That said, listeners typically try to listen for the $F_0$ which often corresponds with the greatest amount of energy within the spectrum, at least for naturally occurring tones.

## Centroid and Musicianship

Previous demonstrations attributed the individual variation seen in the tritone paradox to the spectral centroid while the current investigation shows amplitude of the $F_0$ may be more significant. It was hypothesized that for all tones, both centroid values would perceptually shift pitch higher ($F^\#_5$ centroid value more so than $C_5$), and the centroid match condition would impact sensitivity depending on the chroma. Specifically, the centroid mismatch condition would increase sensitivity for tone pairs with a starting chroma of C, $C^\#$, D, and $D^\#$ at the $C_5$ centroid and for tone pairs with a starting chroma of $F^\#$, G, $G^\#$, and A at the $F^\#_5$ centroid. Results showed that centroid value did influence sensitivity, varying at the $F^{\#}_5$ centroid value compared to the $C_5$ centroid value for the Shepard tones. In contrast, centroid value did not impact sensitivity for the slope and peak tones, possibly because d' was already so high. For all tones, the matched centroid condition had no impact.

It is possible these centroid effects were not as drastic as previous studies due to how only the two highest centroid values ($C_5$ and $F^{\#}_5$) from Deutsch's original stimuli were used in the current study. Deutsch (1986) included centroid values of $C_4$, $F^{\#}_4$, $C_5$ and $F^{\#}_5$, Repp (1997) tested tones centered around $A_4$ and $D^{\#}_5$ and another set around $G_4$ and $C^{\#}_5$, and Malek (2018) used $A_3$ and $C_6$ centroid values. These studies found significant pitch shifts and non-musicians were particularly dependent upon centroid as made evident in Repp (1997) and Malek (2018). It is important to note that Deutsch (1986) averaged across centroid and reported results without accounting for their effects. For the current study, including the two centroid values farthest apart in value would have maximized observable centroid effects. Still, any pitch shift from the centroid would have had to be more than six semitones to overcome both the amplitude effects and the chroma distance of the tritone pair. It would also be difficult to differentiate influences from centroid and amplitude at/near the fundamental since both would have a similar impact on sensitivity.

Centroid effects were also expected to become apparent when examining individual differences based upon musical training since people with less musical training are more susceptible to timbral influences. It was thought that listeners with less musical training would use the centroid to make their pitch judgments for Shepard tones since the bell-shaped envelope de-emphasized the lower frequency regions of the spectrum. Results showed that musical training predicted pitch judgment sensitivity though it was unclear the extent to which listeners were following centroid.

Historically, non-musicians performed worse on pitch judgment tasks compared to musicians (see Pitt, 1994 and Ryalls & Lieberman, 1982), but this could be due to their confusing timbral elements when judging pitch, saying pitch is higher when they are actually referring to timbre. Evidence has been obtained that non-musicians will reliably follow centroid changes to judge pitch (see Seither-Preisler et al., 2007). Repp (1997) posited that non-musicians were synthetic listeners, especially reliant on spectral characteristics, using centroid information in lieu of the $F_0$.

While it is widely disputed how non-musicians process pitch, non-musicians can be trained to perform just as well as people with great amounts of musical training (e.g. on categorical perception tasks, see Yashaswini & Maruthy, 2020 and perceiving masked speech, see Boebinger et al., 2015). Even something as simple as adjusting the language used in the instructions can make pitch judgment performance of non-musicians resemble that of musicians (Bigand & Poulin-Charronnat, 2006). This brings into question what kind of musical training allows “musicians” to make more accurate pitch judgments. Is pitch judgment accuracy based on ear training, exposure to different instruments (e.g. ensemble setting), years of training, or is it simply a matter of having the terminology to properly express what they are hearing? The data from the current study suggest that the number of years involved with music most strongly explains why musicians show greater sensitivity for pitch judgments. Exposure to music alone may be sufficient to increase sensitivity, but it is still unclear what aspect of that training is more important related to pitch judgment accuracy. Still, to accurately interpret data in pitch perception studies, researchers need to take precautions to ensure all respondents are given a fair chance to convey their pitch judgment ability despite musical training.

In light of the training evidence previously mentioned, it is likely non-musicians could be trained to separate the effects of $F_0$ and centroid for the peak and slope conditions in the current study. Based upon the conditions in the current study, non-musicians could hypothetically improve their pitch judgment accuracy when presented with different tone types with variable sloping. For example, Li and Pastore (1995) found that spectral slope was perceptually separable from $F_0$ for harmonic tones. Table 3 depicts the dB loss per octave necessary to achieve the proper centroid values for the peak and slope tones at each chroma value. In the current study, the peak tones were rendered with steeper spectral slopes at the higher pitch classes (e.g. F#, G, G#, A) to achieve the appropriate centroid value. This could make the peak tones more easily distinguished from slope tones that have more shallow spectral slopes, but additional research would be necessary to test this idea.

## Limitations and Future Directions

While the peak tone condition did not function as expected, it is assumed that sensitivity to the peak tones would show comparable results to the missing $F_0$ research (see Seither-Preisler et al., 2007) had the $F_0$ been rendered at an amplitude that was not able to be resolved. Another consideration is that this rendering limited the centroid values that could be achieved in the peak tone condition. As a result, the current study did not include all centroid conditions that were originally included in the tritone paradox and was constrained to rely on only the two highest centroid values from Deutsch's stimuli (see Deutsch, 1986). Therefore, any observable centroid effects could have been minimized or reduced relative to the typical range of conditions that have been used to evaluate the tritone paradox. That being said, this necessary limitation did not completely negate all effects of centroid, as there was significant variation in sensitivity obtained as a function of centroid for the Shepard tone condition.

Another consideration is that adapting the overall loudness level for each spectral envelope shape may be necessary to make the Shepard tones and harmonic tones more perceptually similar (before average RMS processing is completed). Certain participants reported overall loudness level differences between the sawtooth-like (slope/peak) and Shepard tones so matching the loudness levels would help lessen possible confounding effects of loudness differences. Unfortunately, there is no way to manipulate loudness which depends upon frequency (e.g. see equal loudness contours, Fletcher & Munson, 1933; Robinson & Dadson, 1956) without impacting spectral envelope shape. Since pitch and loudness have been demonstrated to be integral properties, there is no way to truly match loudness (Grau & Nelson, 1988; Melara & Marks, 1990). The anecdotal reports of loudness differences between conditions, when combined with corresponding differences in pitch performance, can be argued to represent further support for perceptually integral dimensions. Without having a simple and direct method to completely control for loudness without impacting spectral envelope shape, future studies could assess loudness differences along with pitch for the same Shepard tone conditions to determine if there was a predictable relationship between observed pitch patterns and loudness changes across stimuli.

Despite these limitations, the current investigation shows that Shepard tones are not truly ambiguous with respect to pitch. Rather, listeners make systematic pitch judgments for Shepard tones that appear to reflect the amplitude of the $F_0$. This implies there is a shared pitch processing mechanism that all listeners use, regardless of tone type, including Shepard tones. Therefore, a main question is what makes Shepard tones different, if anything, from other tone types? The fact that findings were obtained from the current investigation that contradict traditional explanations that assume the perceptual ambiguity of Shepard tones should not be considered particularly surprising. After all, if the pitch of Shepard tones were truly ambiguous, then individual listeners should not exhibit such reliable pitch patterns in response to them. The current investigation further supports the idea of shared pitch processing across all tone types since it is unlikely listeners would change their method for judging pitch mid-trial simply because a Shepard tone was present.

In light of the current findings further study is necessary to explore the direct impact of amplitude at the $F_0$ on pitch perception across various types of tones. This would permit a more complete determination of the degree to which the amplitude of lower frequency components (at, or proximal to, the fundamental) influences pitch judgments. As previously discussed, this could be accomplished by manipulating/raising the amplitude within a missing $F_0$ stimulus, or by widening the spectral envelope shape of Shepard tones within a tritone paradox evaluation. For Shepard tones, one possible method to achieve the intended centroid values while preserving the amplitude of the $F_0$ would be to increase the width of the bell-shaped spectral envelope while maintaining centroid placement. If such a study were conducted, it would be expected that sensitivity level should correspond with amplitude. For instance, a C-$F^{\#}$ Shepard tone pair would be judged as ascending if the first component was able to be resolved. If the first component was not able to be resolved, and the listener judged pitch based on the second component, the same tone pair would be judged as descending. Currently, the amplitude(s) at which most listeners could resolve frequency components to make pitch judgments within a complex tone, including Shepard tones, is unknown.

How auditory illusions function are not supposed to be inexplicable, yet the reasoning for how listeners process the tritone paradox are largely accepted without consistent interpretation in part, due to the use of the Shepard tones and the attribution that they are ambiguous to pitch. All explanations surrounding the tritone paradox then becomes dependent upon observed individual variation. However, this does not account for the consistent pattern of results found in previous tritone paradox research. The current study suggests performance can be explained by amplitude and raises the idea of a shared pitch processing mechanism despite tone type. Future research must consider that centroid and chroma, combined with the other factors examined in the current study, all influence spectral envelope shape, a timbral characteristic well known to influence pitch judgments especially for listeners with less musical training. Considering how changes to spectral envelope shape influence pitch judgements should ultimately lead us to a more complete understanding on how listeners are generally able to use and understand pitch for both naturally occurring and synthesized tones.

## References

Ableton Live 11 Suite. (2021). *Ableton.* (Version 11.2) [Computer software]. Retrieved from https://www.ableton.com/

Adobe Audition 3. *Adobe.* [Computer software]. Adobe, Ltd. Retrieved from https://www.adobe.com/

Allen, E. J., & Oxenham, A. J. (2014). Symmetric interactions and interference between pitch and timbre. *The Journal of the Acoustical Society of America*, *135*, 1371-1379. https://www.doi.org/ 10.1121/1.4863269.

American National Standards Institute (1994). *American national standard acoustical terminology*. New York: American National Standards Institute.

Becker, C. J. & Hall, M. D. (2014). Effects of fundamental frequency removal and low-pass filtering on pitch comparisons. *Psychomusicology, 24(3)*, 195-206.

Bigand, E., & Poulin-Charronnat, B. (2006). Are we "experienced listeners"? A review of the musical capacities that do not depend on formal musical training. *Cognition*, *100*, 100-130. https://www.doi.org/10.1016/j.cognition.2005.11.007.

Boebinger, D., Evans, S., Rosen, S., Lima, C. F., Manly, T., & Scott, S. K. (2015). Musicians and non-musicians are equally adept at perceiving masked speech. *The Journal of the Acoustical Society of America*, *137*(1), 378-387.

Boersma, P. & Weenink, D. (2022). Praat: doing phonetics by computer (Version 6.2.23) [Computer Software]. Retrieved from https://www.praaat.org

Bonfils, P. , Bertrand, Y., & Uziel, A. (1988). Evoked otoacoustic emissions: Normative data and presbycusis. *Audiology*, *27(1)*, 27-35, DOI: 10.3109/00206098809081571.

Brahma, S. M., De Leon, P. L., & Kavasseri, R. G. (2009). Investigating the option of removing the antialiasing filter from digital relays. *IEEE Transactions on Power Delivery*, *24*(4), 1864-1868.

Daly, H. R., & Hall, M. D. (2018). Not all musicians are created equal: Statistical concerns regarding the categorization of participants. *Psychomusicology: Music, Mind, and Brain*, *28*(2), 117-126.

Deutsch, D. (1986). A musical paradox. *Music Perception, 3*, 275-280.

Deutsch, D. (1991). The tritone paradox: An influence of language on music perception. *Music Perception, 8,* 335-347. https://www.doi.org/ 98.244.112.229.

Deutsch, D. (1995). *Diana Deutsch Musical Illusions and Paradoxes* [Album]. Philomel Records.

Deutsch, D., Dooley, K., & Henthorn, T. (2008). Pitch circularity from tones comprising full harmonic series. *The Journal of the Acoustical Society of America*, *124*, 589-597. https://www.doi.org/ 10.1121/1.2931957.

Eggermont, J. (2015). Periodicity pitch. In *Auditory Temporal Processing and its Disorders* (pp. 144-164). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198719090.003.0008.

Fletcher, H., & Munson, W. A. (1933). Loudness, its definition, measurement and calculation. *Journal of the Acoustical Society of America*, *5*, 82
108. https://doi.org/10.1121/1.1915637.

Fugiel, B. (2011). Waveform circularity from added sawtooth and square wave acoustical signals. *Music Perception*, *28*, 415-424. https://www.doi.org/10.1525/mp.2011.28.4.415.

Garner, W. R. & Felfoldy, G. L. (1970). Integrality of stimulus dimensions in various types of information processing. *Cognitive Psychology*, *1*, 225-241.

Giordano, B. L. && McAdams, S. Material identification of real impact sounds: Effects of size variation in steel, glass, word, and plexiglass plates. *The Journal of the Acoustical Society of America, 119(2),* 1171-1181.

Grau, J. W., & Nelson, D. K. (1988). The distinction between integral and separable dimensions: evidence for the integrality of pitch and loudness. *Journal of Experimental Psychology: General*, *117*(4), 347.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and Psychophysics.* John Wiley.

Grey, J. M. (1977). Multidimensional perceptual scaling of musical timbres. *The Journal of the Acoustical Society of America*, 61(5):1270–1277. https://doi.org/10.1121/1.381428

Grey, J.M. & Gordon, J.W. (1978). Perceptual effects of spectral modifications on musical timbres. *The Journal of the Acoustical Society of America*, 63(5):1493–1500. https://doi.org/10.1121/1.381843.

Hall, M. D. & Beauchamp, J. W. (2009). Clarifying the spectral and temporal dimensions of musical instrument timbre. *Canadian Acoustics, 3(37),* 3-22).

Hall, M. D. & Becker, C. (2010). A basis for perceptual interactions between timbre and pitch. [poster presentation].

Hall, M. D. & Redpath, T. (2016). *FormAnt* (Version V1.012720) [Computer software]. Retrieved from https://psyc.jmu.edu/undergraduate/auditoryperception/formant.html

Jarvis, B. G. (2018a). DirectRT (Version 2018.1.117) [Computer Software]. New York, NY: Empirisoft Corporation. Retrieved from https://www.empirisoft.com/

Jarvis, B. G. (2018b). MediaLab (Version 2018.1.109) [Computer Software]. New York, NY: Empirisoft Corporation. Retrieved from https://www.empirisoft.com/

Krumhansl, C. L. (1989). Why is musical timbre so hard to understand? In S. Nielsen and O. Olsson (Eds.), *Structure and Perception of Electroacoustic Sound and Music* (pg. 43-53). Amsterdam: Elsevier.

Li, X., & Pastore, R. E. (1995). Perceptual constancy of a global spectral property: Spectral slope discrimination. *The Journal of the Acoustical Society of America*, *98*(4), 1956-1968.

Macmillan, N. A., & Creelman, C. D. (2004). *Detection theory: A user's guide*. Psychology press.

Malek, S. (2018). Pitch class and envelope effects in the Tritone ParadoxTritone paradox are mediated by differently pronounced frequency preference regions. *Frontiers in Psychology*, *9*, 1-18. https://www.doi.org/10.3389/fpsyg.2018.01590.

McAdams, S. & Siedenburg, K. (2019). Perception and cognition of musical timbre. In P. J. Rentfrow & D. J. Levitin (Eds.), *Foundations of music psychology: Theory and research* (pp. 71-120). MIT Press.

McAdams, S. (1999). Perspectives on the contribution of timbre to musical structure. *Computer music journal*, *23*(3), 85-102.

McAdams, S. (2019). The Perceptual Representation of Timbre. In K. Siedenburg et al. (eds.), *Timbre: Acoustics, Perception, and Cognition, Springer Handbook of Auditory Research*, 69. Springer Nature Switzerland AG. https://doi.org/10.1007/978-3-030-14832-4_2.

McAdams, S., Winsberg, S., Donnadieu, S., De Soete, G., & Krimphoff, J. (1995). Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes. *Psychological Research, 58*(3), 177-192.

Melara, R. D., & Marks, L. E. (1990). Interaction among auditory dimensions: Timbre, pitch, and loudness. *Perception & psychophysics*, *48*(2), 169-178.

Moore, B. C. J. (2013). *An Introduction to the Psychology of Hearing Sixth Edition.* Emerald Group Publishing Limited: Bingley, UK.

Peeters, G. (2004). A large set of audio features for sound description (similarity and classification) in the CUIDADO project. *CUIDADO Ist Project Report*, *54*(0), 1-25.

Pitt, M. A. (1994). Perception of pitch and timbre by musically trained and untrained listeners. *Journal of experimental psychology: human perception and performance*, *20*, 976. https://www.doi.org/10.1037//0096-1523.20.5.976.

Repp, B. H. (1997). Spectral envelope and context effects in the tritone paradoxtritone paradox. *Perception*, *26*, 645-665. https://www.doi.org/10.1068/p260645.

Robinson, D. W., & Dadson, R. S. (1956). A re-determination of the equal-loudness relations for pure tones. *British Journal of Applied Physics*, *7*(5), 166.

Ryalls, J. H. & Lieberman, P. (1982). Fundamental frequency and vowel perception. *The Journal of the Acoustical Society of America*, *72*, 1631-1634. https://www.doi.org/10.1121/1.388499.

Schön, D., Magne, C., & Besson, M. (2004). The music of speech: Music training facilitates pitch processing in both music and language. *Psychophysiology*, *41*, 341-349. https://www.doi.org/10.1111/1469-8986.00172.x.

Seither-Preisler, A., Johnson, L., Krumbholz, K., Nobbe, A., Patterson, R., Seither, S., & Lütkenhöner, B. (2007). Tone sequences with conflicting fundamental pitch and timbre changes are heard differently by musicians and nonmusicians. *Journal of Experimental*

*Psychology: Human Perception and Performance*, *33*, 743. https://www.doi.org/10.1037/0096-1523.33.3.743

Shamma, S. A. (2004). Topographic organization is essential for pitch perception. *Proceedings of the National Academy of Sciences*, *101*(5), 1114-1115. https://www.doi.org/10.1073/pnas.0307334101.

Shepard, R. N. (1964). Circularity in judgments of relative pitch. *The Journal of the Acoustical Society of America journal of the acoustical society of America*, *36*, 2346-2353. https://www.doi.org/10.1121/1.1919362.

Siedenburg, K., Jacobsen, S., & Reuter, C. (2021). Spectral envelope position and shape in sustained musical instrument sounds. *The Journal of the Acoustical Society of America*, *149*, 3715-3726. https://www.doi.org/10.1121/10.0005088.

Sona Systems (n.d.). *Sona Systems: Cloud-based Participant Management Software* [Computer software]. Sona Systems, Ltd. https://www.sona-systems.com/

Yashaswini, L., & Maruthy, S. (2020). Effect of Music Training on Categorical Perception of Speech and Music. *Journal of Audiology & Otology*, *24*(3), 140.

**Table 1**

*Example of stimuli combinations for one chroma pair*

| Pitch Pair | $C_3$-$F^{\#}_3$ | | | | $F^{\#}_3$- $C_3$ | | | |
|---|---|---|---|---|---|---|---|---|
| Ordering | $C_3$ | $F^{\#}_3$ | $C_3$ | $F^{\#}_3$ | $F^{\#}_3$ | $C_3$ | $F^{\#}_3$ | $C_3$ |
| | **std** | comp | comp | **std** | **std** | comp | comp | **std** |
| Matched Centroid | $C_5$ | $C_5$ | $C_5$ | $C_5$ | $C_5$ | $C_5$ | $C_5$ | $C_5$ |
| | $F^{\#}_5$ | $F^{\#}_5$ | $F^{\#}_5$ | $F^{\#}_5$ | $F^{\#}_5$ | $F^{\#}_5$ | $F^{\#}_5$ | $F^{\#}_5$ |
| Mismatched Centroid | $C_5$ | $F^{\#}_5$ | $C_5$ | $F^{\#}_5$ | $C_5$ | $F^{\#}_5$ | $C_5$ | $F^{\#}_5$ |
| | $F^{\#}_5$ | $C_5$ | $F^{\#}_5$ | $C_5$ | $F^{\#}_5$ | $C_5$ | $F^{\#}_5$ | $C_5$ |

*std = standard; comp = comparison

**Table 2**

*Amplitude height of $F_0$ for each Shepard tone according to centroid and chroma value with corresponding sensitivity*

| Centroid Value | Chroma | Frequency (Hz) | $F_0$ Amplitude (relative dB) | d’ | SE |
|---|---|---|---|---|---|
| $C_5$ | $C_3$ | 130.81 | 50.5 | 1.73 | 0.35 |
| | $C^{\#}_3$ | 138.59 | 51.7 | 1.77 | 0.33 |
| | $D_3$ | 146.83 | (18.4) 52.9* | 1.95 | 0.34 |
| | $D^{\#}_3$ | 155.56 | (26.6) 53.9* | 2.40 | 0.33 |
| | $F^{\#}_3$ | 185 | 38.7 | 2.49 | 0.28 |
| | $G_3$ | 196 | 41.4 | 2.02 | 0.31 |
| | $G^{\#}_3$ | 207.65 | 44.2 | 2.02 | 0.31 |
| | $A_3$ | 220 | 46 | 1.79 | 0.33 |
| $F^{\#}_5$ | $C_3$ | 130.81 | 39 | -0.17 | 0.34 |
| | $C^{\#}_3$ | 138.59 | 41.7 | 0.55 | 0.34 |
| | $D_3$ | 146.83 | 43.8 | 0.94 | 0.36 |
| | $D^{\#}_3$ | 155.56 | 45.7 | 1.56 | 0.35 |
| | $F^{\#}_3$ | 185 | 50.3 | 3.22 | 0.22 |
| | $G_3$ | 196 | 51.7 | 3.23 | 0.23 |
| | $G^{\#}_3$ | 207.65 | (19.9) 52.9* | 2.95 | 0.25 |
| | $A_3$ | 220 | (26.8) 53.8* | 3.11 | 0.25 |

Note: The $F_0$ values in parentheses represent the relative dB for the first component in the spectrum while the values followed by an asterisk (*) symbol represent the relative dB for the second component. All values were obtained using the spectral slice function in Praat (Boersma & Weenink, 2022). This provides a dB measure relative to the maximum possible signal amplitude/voltage. However, the resulting values are expressed in an usual way, where larger positive values are closer to the maximum amplitude.

**Table 3**

*Spectral sloping depicted as a loss of dB per octave necessary to achieve the centroid value for the slope and peak tones at each chroma value*

| Centroid Value | Chroma | Frequency (Hz) | slope | peak |
|---|---|---|---|---|
| $C_5$ | $C_3$ | 130.81 | -10.3334 | -15.8915 |
| | $C^{\#}_3$ | 138.59 | -10.584 | -16.7262 |
| | $D_3$ | 146.83 | -10.8395 | -17.687 |
| | $D^{\#}_3$ | 155.56 | -11.101 | -18.8242 |
| | $F^{\#}_3$ | 185 | -11.9325 | -24.511 |
| | $G_3$ | 196 | -12.2297 | -28.361 |
| | $G^{\#}_3$ | 207.65 | -12.5399 | -35.625 |
| | $A_3$ | 220 | -12.8654 | -60.68 |
| $F^{\#}_5$ | $C_3$ | 130.81 | -8.8947 | -12.2702 |
| | $C^{\#}_3$ | 138.59 | -9.13 | -12.7718 |
| | $D_3$ | 146.83 | -9.3663 | -13.3028 |
| | $D^{\#}_3$ | 155.56 | -9.60412 | -13.8701 |
| | $F^{\#}_3$ | 185 | -10.3335 | -15.8918 |
| | $G_3$ | 196 | -10.584 | -16.7263 |
| | $G^{\#}_3$ | 207.65 | -10.8394 | -17.687 |
| | $A_3$ | 220 | -11.10105 | -18.8246 |

**Figure 1**

*Example of tone chroma pairs used in the current investigation across the chroma circle*

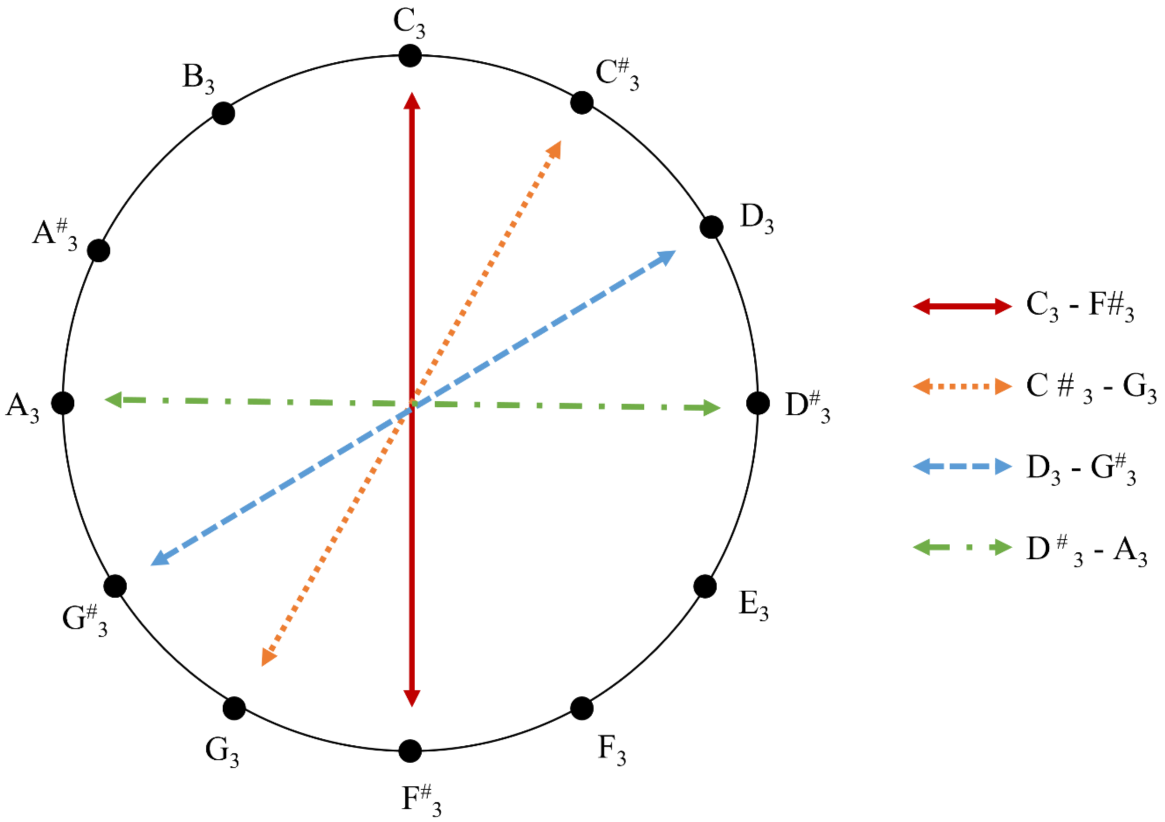


**Figure 2**

*Average d' values by centroid and starting chroma as a function of spectral envelope shape*

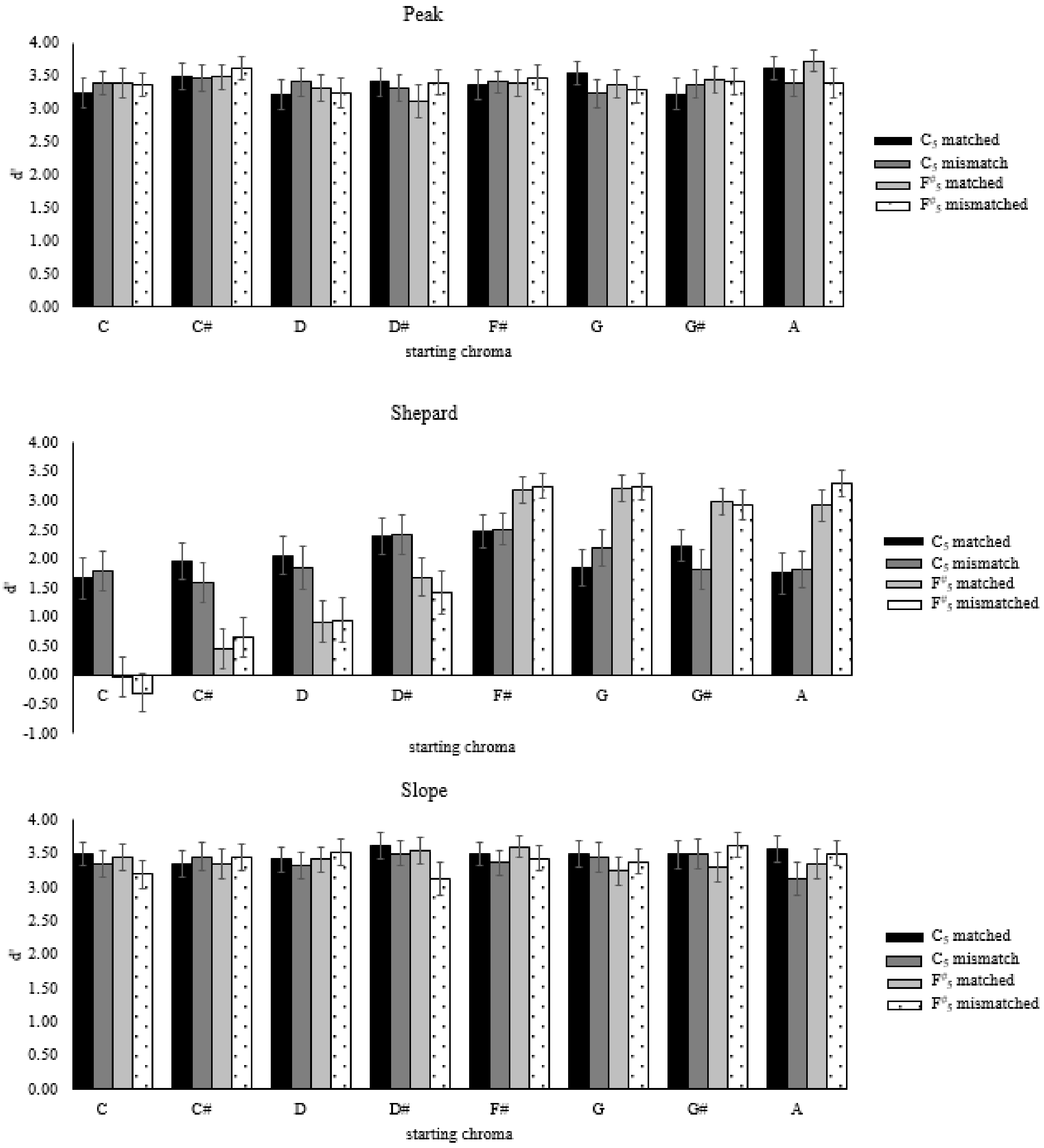


**Figure 3**

*Difference scores for spectral envelope shape derived from d' for slope, peak, and Shepard tones for matched low and high centroid values*

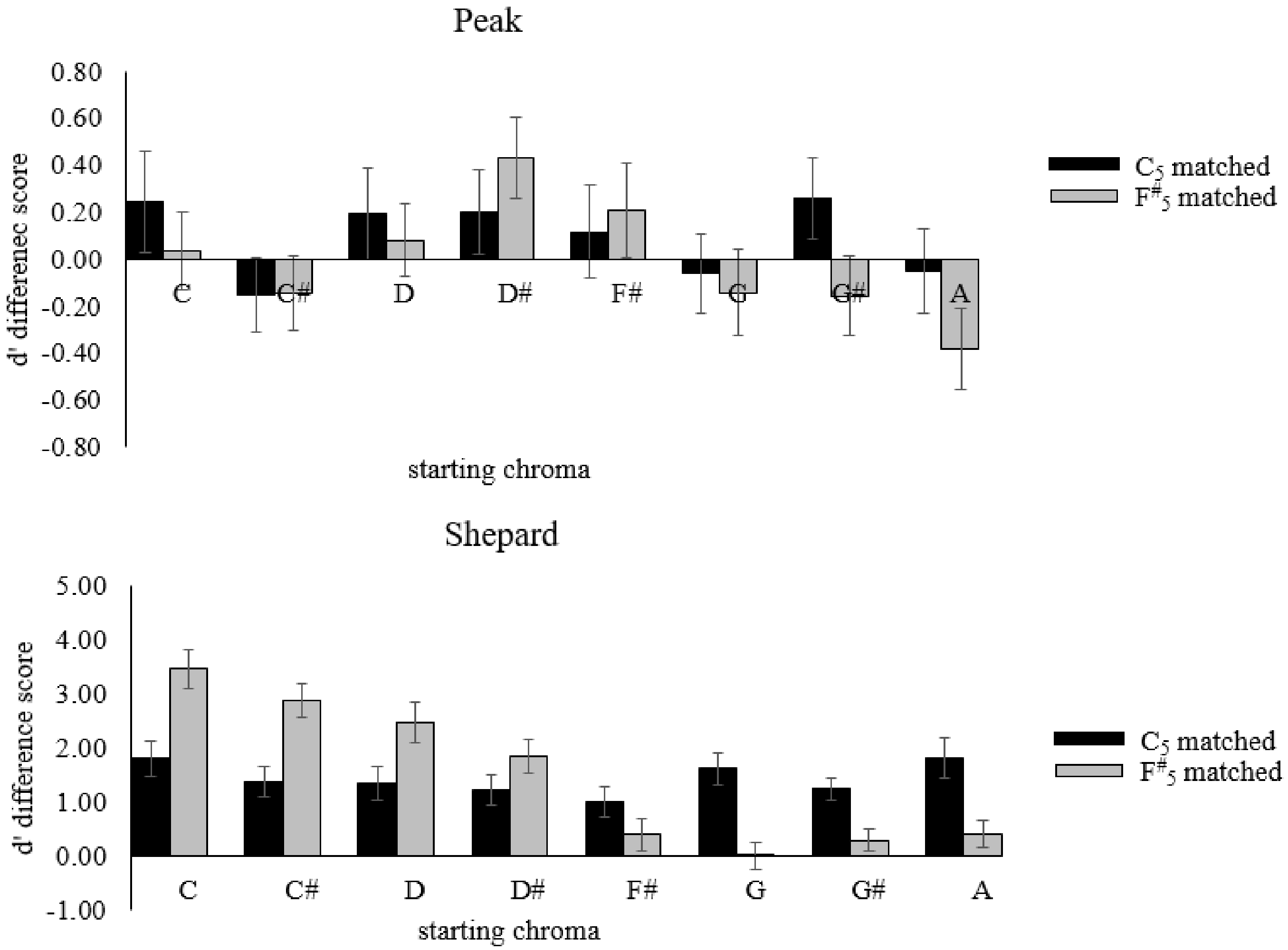


**Figure 4**

*D' per spectral envelope shape correlated with musical training*

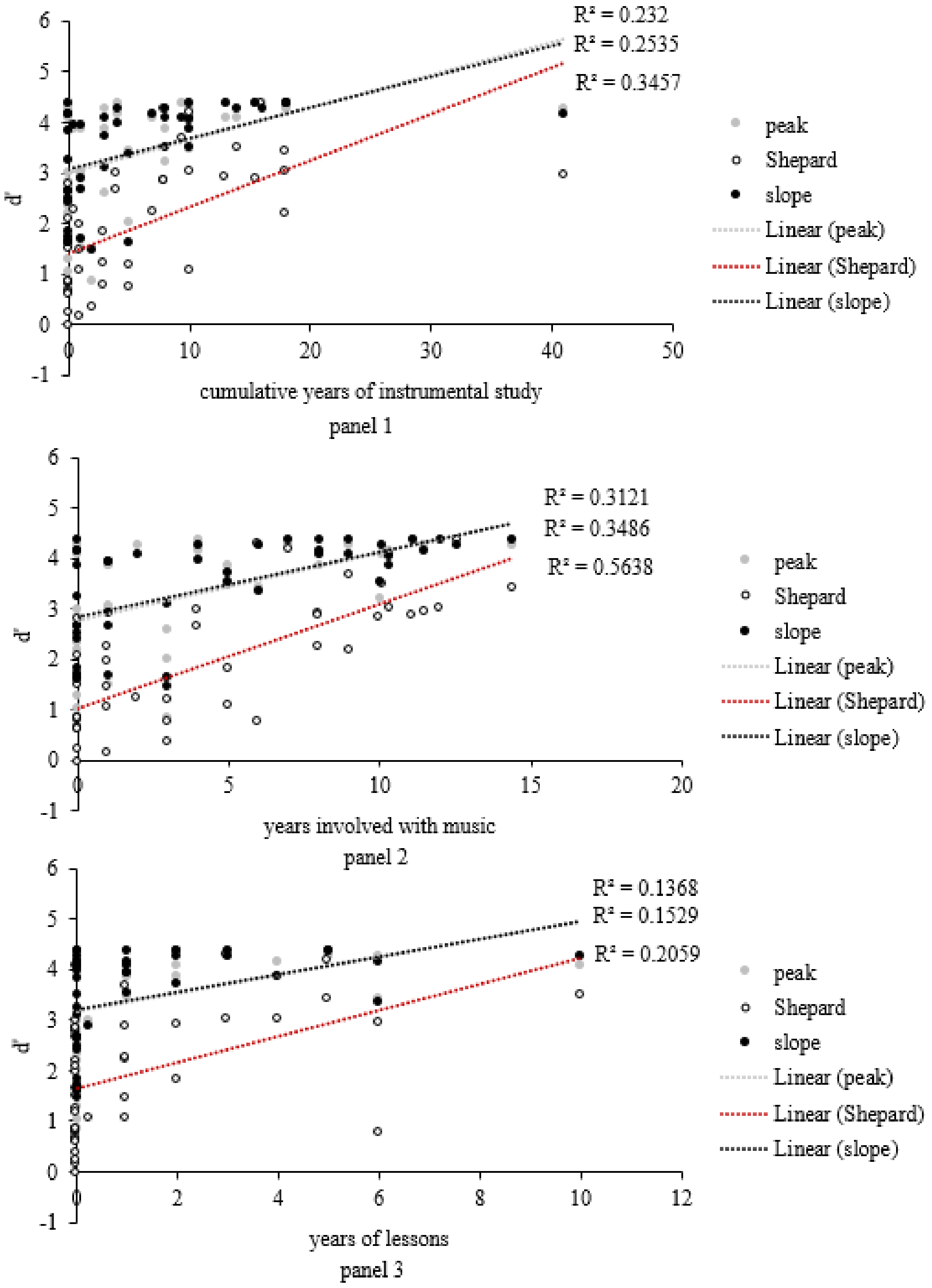


**Figure 5**

*Spectral slices for each Shepard tones according to $F_0$ at the $C_5$ centroid value*

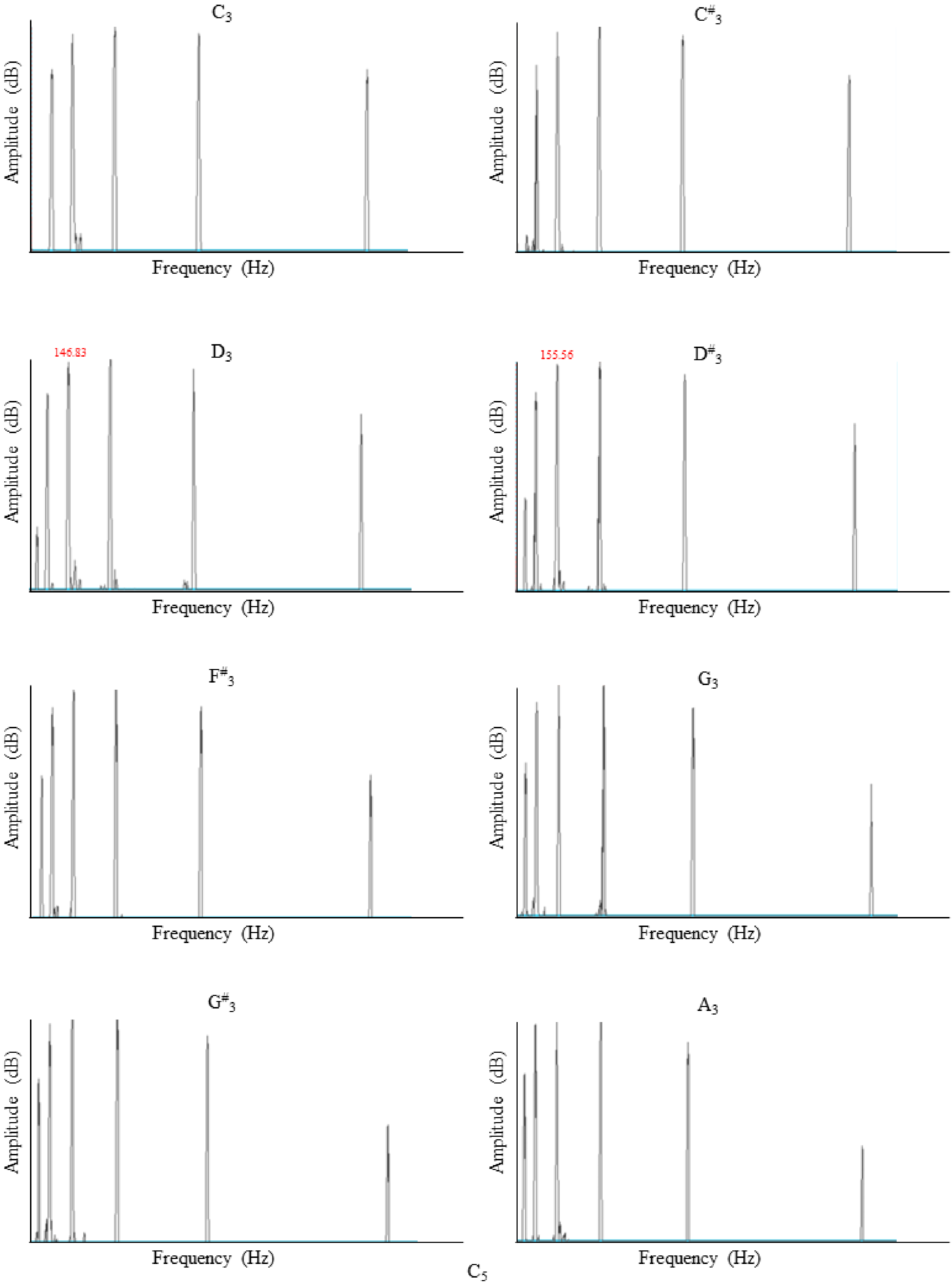
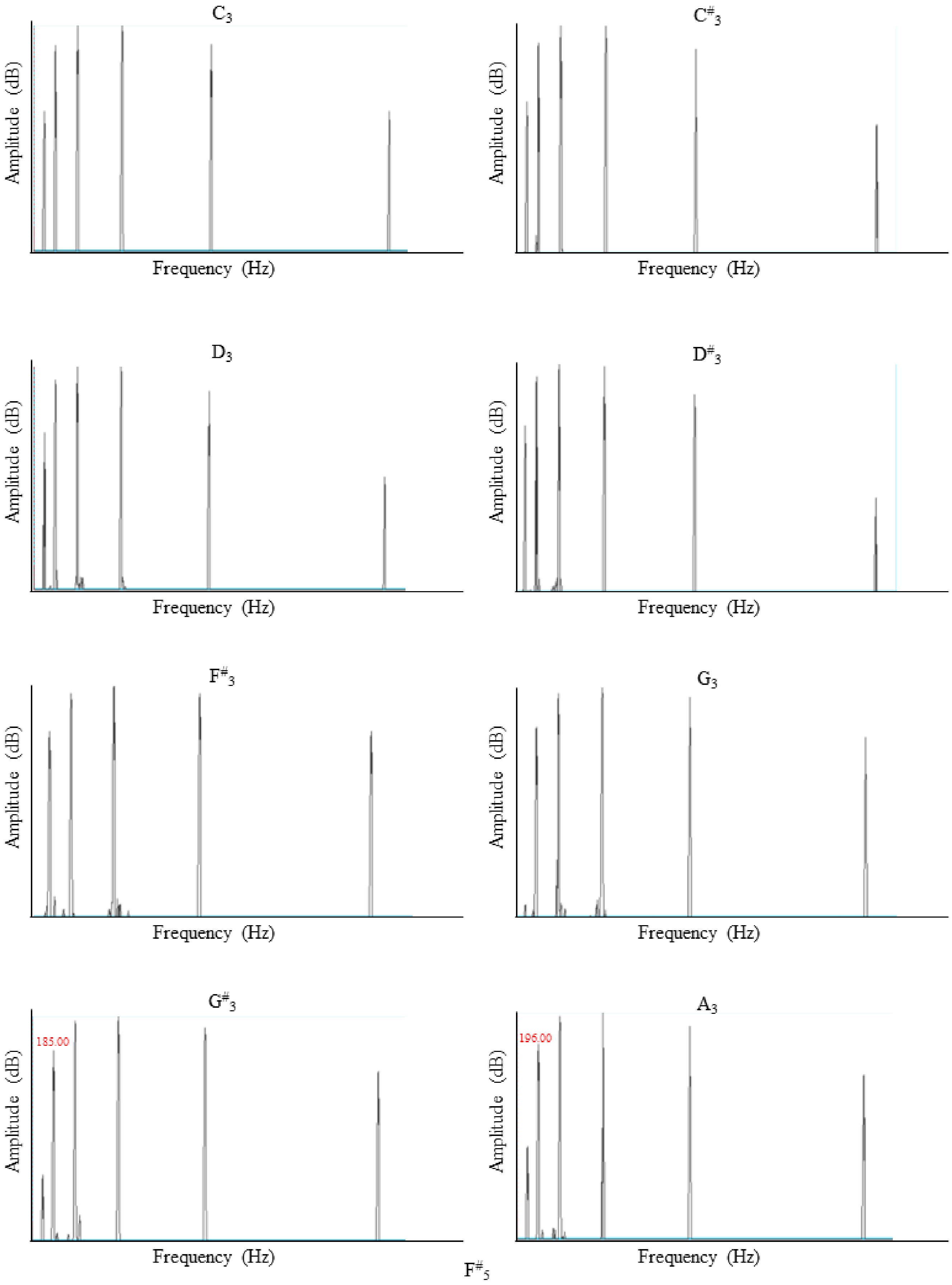


**Figure 6**

*Spectral slices for each Shepard tones according to $F_0$ at the $F^{\#}_5$ centroid value*



# Appendix A

## *Musical Training and Experience (MUTE) survey*

Musical Training Survey Participant #: ____

1. What is your birthdate? ___/___/______ (MM/DD/YYYY)

2. Have you ever taken a formal music class? (include elementary/grade school music classes, theory, choir, band, etc).

• yes • no

If answering yes, please select the types of classes that best match your experience (check all that apply):

• elementary class • music appreciation • music theory • ear training • band/choir

• music history • music composition • conducting • piano

• other (please describe) _____________________________

3. How many college-level music courses have you completed? _____ courses

Please list all courses and indicate if you are currently enrolled in any of them:

4. Have you ever played a musical instrument or studied singing? (if no, please skip to question #16)

• yes • no

5. What style of music do you play most often (select one):

• Classical • Pop • Jazz • Folk • Rock • Country

• Other ____________________

6. What instrument(s) have you played (including voice)?

How many months/years have you studied/played each instrument (or voice)? Please indicate both duration and the corresponding instrument:

7. At what age did you begin playing/studying music? _____ years old

8. Approximately how many hours per week did you spend practicing music during your first year of study? _____ hours/week

9. Are you currently involved in any musical activities? If not, at what age did you stop playing music?

• yes (currently involved) • no (not involved) ____ age that you stopped, if applicable

10. If you are currently involved in musical activities, about how many hours do you spend playing music per week, including rehearsal and individual practice? (if not currently practicing, skip to next question) _____ hours/week

11. Are you currently or have you ever received private music lessons?

• yes (currently receiving) • yes (received in the past) • no (no history of private lessons)

12. How many years/months of experience do you have taking private music lessons?

____years/____months

13. Are you currently or have you ever participated in a musical ensemble? (e.g. band/choir class, honor bands/choirs, informal musical ensembles, church music group, community ensembles, any situation in which you create music with others):

• yes (currently participating) • yes (participated in the past) • no (no history of participation)

14. How many years/months of experience do you have participating in a musical ensemble?

____years/____months

Please describe all ensembles and how many years you participated in each:

15. How many years/months of improvisation experience do you have? (playing music spontaneously, not following written musical notation)

____years/____months

16. How many years/months of experience do you have composing/writing music?

____years/____months

17. How many years/months of experience do you have creating or manipulating music using a computer? (DJ, electronic music, etc.):

____years/____months

18. How many year/months of experience do you have participating in musical theatre? How many musicals have you participated in?

____years/____months _____ musicals

Please describe your role in these musical theatre productions (performer, stage crew, orchestra):

19. How many years/months of dance experience do you have? (ballet, jazz, tap, color guard, etc.)

____years/____months

Please describe all dance styles and how many years you participated in each:

20. How many years/months of experience do you have playing musical video games? (Guitar Hero, Rock Band, etc.)

____years/____months

# Appendix B

## Debriefing Questions

1. Were some tone pairs easier or harder to make the pitch judgment?
    a. If so, how would you describe what you were listening for?
2. Did you get a sense of what you were listening for?
    a. What did you think the experiment was about?
3. Did you notice if any trials had problems with playback? (e.g. no sound playing, too much time between trials, only hearing silence, etc.)
4. Did you feel fatigue at any point or were you alert the whole session?
5. Any additional comments?