Spring 2017

The influences of musical training and spectral centroid on perceptual interactions of pitch and timbre

Heather Daly
James Madison University

Recommended Citation

Daly, Heather, "The influences of musical training and spectral centroid on perceptual interactions of pitch and timbre" (2017). *Masters Theses*. 516.
https://commons.lib.jmu.edu/master201019/516

The Influences of Musical Training and Spectral Centroid on Perceptual Interactions of Pitch and Timbre

Heather Daly

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

May 2017

FACULTY COMMITTEE:

Committee Chair: Michael D. Hall, Ph.D.

Committee Members/ Readers:

Jeff Dyche, Ph.D.

Allison Ames, Ph.D

Acknowledgments

I would like to thank my advisor, Dr. Michael Hall, for his endless support and patience throughout the past two years and for teaching me how to be a better scientist. I would also like to thank my committee members, Dr. Jeff Dyche and Dr. Allison Ames for their help and encouragement throughout the writing process. Finally, I would like to thank Cameron Meyers for staying up with me through the long nights, listening to my ramblings about acoustics and statistics, and for always believing in my ability to complete this project.

Table of Contents

List of Tables

List of Figures

Abstract

Perceptual interactions of pitch and timbre have frequently been observed, and the nature of these interactions differs between musicians and nonmusicians. Yet, few researchers have investigated which aspects of timbre or musical training contribute to such interactions. Recently, Becker and Hall (2014) demonstrated that the spectral centroid contributed to pitch-timbre interactions in missing-F0 experiments, particularly for nonmusicians. The present experiment investigated whether the centroid also accounted for previously observed interactions between pitch and timbre (see Pitt, 1994) in a Garner speeded classification task designed to evaluate the perceptual independence of dimensions. There were two sets of synthetic stimuli involving orthogonal combinations of A4 and D#4 tones derived from violin and trumpet. Timbres in one set varied with respect to spectral envelope, amplitude envelope, and spectral centroid, whereas the other equated spectral centroids through slight manipulations of spectral slope. Tones with the same centroid were expected to reduce the magnitude of observed interference and redundancy gain effects. Contrary to hypotheses, such an effect was not observed, suggesting that the spectral centroid was not the aspect of timbre driving perceptual interactions in the current investigation.

While musical training has been proposed to enhance the ability to distinguish pitch from timbre changes, the aspect of training that contributes to such enhancement remains unclear. This is complicated by most studies only considering total years of experience as a means of categorizing musicians versus nonmusicians, which directly impacts conclusions regarding statistical significance. The current investigation addressed these issues by introducing a musical training survey that measures a more

diverse range of musical experiences (e.g., ensemble performance, recency/frequency of practice, level of coursework) in a more continuous manner (i.e., without fixed response options). This permitted statistically appropriate (regression) analyses of the relationship between years of training and perceptual independence of timbre and pitch, which was intended to identify relevant experiences for observed interactions. Increased amounts of musical training in general were associated with smaller interference effects with the adjusted stimulus set. Although specific experiences were not identified in the current investigation due to high correlations between musical predictor variables, such correlations raise the possibility that a single factor may be underlying the musical training items examined in this study. Collectively, these findings demonstrate that pitch and timbre perceptually interact regardless of level of musical training, although training can reduce the size of the interference effect in certain conditions. Additionally, rather than the spectral centroid being the attribute of timbre responsible for perceptual interactions, current results suggest that the spectral envelope may have a stronger influence.

The Influences of Musical Training and Spectral Centroid on Perceptual Interactions of Pitch and Timbre

Pitch is typically defined as the subjective experience of fundamental frequency (F_0), although other spectral characteristics may also contribute to the percept (Houtsma, 1997). In fact, pitch can even be perceived accurately just from the periodicity of the harmonics associated with F_0 (Seither-Preisler et al., 2007). The perception of pitch is ordered from low to high, tracking the low to high ordering of frequency. Timbre, on the other hand, traditionally has been defined by what it is *not,* rather than what it is. The American National Standards Institute (ANSI, 1973) has defined timbre as "that attribute of auditory sensation in terms of which a listener can judge that two sounds, similarly presented and having the same loudness and pitch, are different" (p. 56). Similar to this interpretation, some researchers have defined timbre as the aspect of sound that differs between sound sources that have the same pitch, loudness, and duration (e.g., see Krumhansl & Iverson, 1992). Other researchers have defined it more simply as tone quality, or the unique quality of a sound that contributes to its identification, such as the distinction between different musical instruments playing the same note (Pitt, 1994; Warrier & Zatorre, 2002).

Still other researchers have described timbre in terms of its acoustic correlates (McAdams, Winsberg, Donnadieu, De Soete, & Krimphoff, 1995; Seashore, 1938). Although pitch is typically perceived as unidimensional, timbre consists of multiple perceptual dimensions, each correlated with distinct physical characteristics. These physical dimensions can be grouped into three main categories: spectral, temporal, and spectro-temporal (e.g., see McAdams et al., 1995). Temporal attributes include dynamic characteristics collectively referred to as the amplitude envelope, such as attack, decay, and release times. A primary spectro-temporal attribute is spectral flux, which can be defined as variability of the spectral envelope over time. Finally, spectral attributes include spectral envelope shape and spectral centroid. The spectral centroid can be defined as the mean of all of the amplitude-weighted frequencies in a sound, including the fundamental frequency and corresponding harmonics (which are integer multiples of the fundamental frequency). Each sound has characteristic patterns of harmonic and inharmonic acoustic energy related to the size and shape of the body of the source that produced it. The patterns of harmonic energy are manifested in the harmonics of each sound, such that some groups of harmonics are attenuated more than others, and thus have lower amplitude. If most of the amplitude attenuation is in harmonics at higher frequencies, then the spectral centroid will be lower. Likewise, if harmonics at lower frequencies are attenuated, then the centroid will be higher. The spectral centroid is generally argued to contribute to the perceived "brightness" of a sound, given that sounds with higher centroids (such as trumpet) tend to be perceived as brighter than sounds with lower centroids (such as tuba). Thus, the dimension of brightness, like pitch, is ordered from low- to high-frequency dominant signals, such that increases in a single acoustic measure (centroid, rather than F_0) tend to produce corresponding nonlinear increases along the perceptual dimension.

Given that pitch and some aspects of timbre are both at least partially rooted in frequency, it seems natural that the two attributes would perceptually interact. Although some researchers initially suggested that pitch and timbre are perceptually independent (Demany & Semal, 1993; Semal & Demany, 1991), there is a considerably larger body of evidence suggesting that they interact. In fact, timbre variation has been shown to influence pitch judgments in a number of different paradigms (Krumhansl & Iverson, 1992; Pitt, 1994; Russo & Thompson, 2005; Warrier & Zatorre, 2002).

One demonstration of the perceptual interaction between timbre and pitch involves the phenomenon of the missing fundamental, which allows for dissociation of the fundamental frequency from the corresponding harmonics of a tone (Becker & Hall, 2014; Seither-Preisler et al., 2007). If the fundamental is removed from a tone, the harmonics still convey information about that frequency because of their shared periodicity (i.e., they are in phase at the rate of the fundamental), which allows pitch perception roughly corresponding to the fundamental. Removing the fundamental, which represents the lowest-frequency partial (and typically the most intense partial), from a complex tone also influences spectral properties, such as the spectral envelope and spectral centroid, which in turn should influence timbre perception.

Missing-F_0 stimuli have been used to demonstrate perceptual interactions of pitch and timbre (Becker & Hall, 2014; Seither-Preisler et al., 2007). Seither-Preisler et al. (2007) used stimuli where the direction of missing-F_0 change was incongruent with the direction of harmonic spectrum change. For example, in a tone pair with a descending missing-F_0, the first tone had the lower harmonics associated with a higher F_0 and the second tone had the upper harmonics associated with a lower F_0. Thus, a descending missing-F_0 was associated with a rising harmonic spectrum, and an ascending missing-F_0 was associated with a falling spectrum. The resulting effect was that each tone pair could be heard as either "ascending" or "descending," depending on whether one was listening to the direction of the missing-F_0 or to the direction of the harmonic spectrum.

Interactions between pitch and timbre were evidenced by listeners making pitch direction decisions based upon the direction of the harmonics rather than the missing-F_0. Becker and Hall (2014) further used missing-F_0 stimuli to demonstrate how spectral envelope and spectral centroid differentially influenced pitch perception. When the two tones being compared were based on the same fundamental, then spectral envelope changes influenced pitch judgments. When the tones were based on different fundamentals, then the spectral centroid was more influential as demonstrated by the lack of impact of removing the F_0 when centroid shifts were also eliminated via filtering. These results suggest that there is not a single universal effect of timbre on pitch judgments in all situations. Rather, the most influential aspect of timbre is determined by whether the two tones to be compared share the same spectral centroid.

Another demonstration of perceptual interactions between pitch and timbre can be found in the tritone paradox, which uses pairs of Shepard tones as stimuli (Deutsch, 1986; Repp, 1997). A single Shepard tone consists of a group of sinusoidal tones, each at a different octave of the same pitch class (chroma) such that within each Shepard tone, all octaves for a single chroma are present (i.e., all octaves for C). Additionally, each Shepard tone has a bell-shaped spectral envelope that amplifies harmonics in the middle of the spectrum and gradually attenuates harmonics at the low and high ends of the spectrum. These characteristics should result in a clear pitch class but ambiguous tone (i.e., pitch) height (Shepard, 1964). In the tritone paradox, pairs of Shepard tones separated by half of an octave (i.e., a tritone) are sequentially presented, so that the distance in both ascending and descending directions of pitch change across any pair of tones is the same. Perception of these pairs around the chroma circle is consistent within any given listener, but the point along the chroma circle where perception shifts from ascending to descending is highly inconsistent across listeners. Deutsch (1986) explained this result as differences in individual pitch-class templates, suggesting that listeners typically hear some pitch classes as "higher" than others. On the other hand, Repp (1994; 1997) attributed the observation primarily to differences in the spectral envelope. He found that perception of whether the tone pair was ascending or descending depended on placement of the spectral envelope. Listeners were more likely to perceive pitch as descending if the peak of the bell-shaped spectral envelope was centered on the second tone due to it having a lower spectral centroid, thus demonstrating yet another influence of timbre on pitch perception.

The current investigation focuses on a third type of task that has been used to demonstrate pitch-timbre interactions: Garner speeded classification (Garner, 1974; for timbre and pitch interactions in the task, see Melara & Marks, 1990; also see Pitt, 1994). The task is designed to evaluate the perceptual independence of dimensions by determining if sensitivity along one dimension is helped and/or hindered by variability along another dimension. The paradigm is typically limited to a 2 x 2 matrix of stimulus values across two perceptual dimensions, contributing to a total of four stimuli to be used in the task. Two focus conditions are employed to allow for observation of possible processing asymmetry between the two dimensions. In the case of pitch and timbre, pitch-focus and timbre-focus conditions are used. In pitch-focus conditions, participants classify based on pitch differences, and in timbre-focus conditions they classify based on timbre differences. The Garner speeded classification task uses three types of trials: baseline, correlated, and orthogonal. In baseline trials participants classify stimuli on the task-relevant dimension (i.e., low/high pitch) as fast as they can in the absence of any variation on the irrelevant dimension (i.e., same instrument timbre) in order to assess speed and accuracy of classification in a single dimension without any contribution from the other dimension.

In correlated trials, variation on the relevant dimension is accompanied by predictable variation in the irrelevant dimension. For example, if pitch (low/high) is the task-relevant dimension and timbre (violin/trumpet) is the irrelevant dimension, then in correlated trials the low-pitch violin would always be presented in the same block as the high-pitch trumpet, or vice versa. If the two dimensions are perceptually integral, then the values of both pitch and timbre provide information to indicate the correct response. This redundant information regarding stimulus identity could result in "redundancy gain," that is, greater accuracy and faster response times relative to baseline.

Finally, in orthogonal trials, variation in the relevant dimension is accompanied by unpredictable variation in the irrelevant dimension. Continuing with the pitch-timbre example, either the low or high pitch could be presented in either timbre. Thus, the dimensions vary orthogonally. If the two dimensions are perceptually integral, then the unpredictable variation in the irrelevant dimension is expected to interfere with classification of the relevant dimension, thus reducing classification accuracy and response time. If a redundancy gain in the correlated condition and an interference effect in the orthogonal condition are observed, the two dimensions are said to be integral. If there is neither redundancy gain nor an interference effect, then the dimensions are perceptually separable. It is also possible to obtain asymmetric effects indicating the influence of one dimension on another, but not the reverse effect. For example, there could be a large interference effect in the pitch-focus condition, but only a mild effect in the timbre-focus condition. This would indicate that timbral variation influences pitch judgments more than pitch variation influences timbre judgments.

Melara and Marks (1990) were the first to use the Garner speeded classification paradigm to investigate potential interactions between pitch and timbre. Two levels of timbre, defined by duty cycle values (.1878, "twangy" vs .3128, "hollow"), were crossed with two levels of pitch (900 Hz vs 920 Hz F_0) to create four stimuli. Significant Garner interference effects and redundancy gains were observed in both pitch- and timbre-focus conditions. Taken together, these results suggest that pitch and timbre are perceptually integral dimensions.

Pitt (1994) used the Garner speeded classification task to evaluate whether the integrality of pitch and timbre was dependent upon musical training experience. Nonmusicians exhibited significant interference effects within both focus conditions, with stronger effects in the pitch-focus condition indicating that timbral variation influenced pitch judgments more strongly than the reverse. For musicians, accuracy was high and did not significantly vary across conditions, but response times were slower in the orthogonal condition than in the baseline condition, indicating perceptual integrality. Unlike nonmusicians, musicians showed evidence of a redundancy gain effect within the timbre-focus condition, suggesting that musicians were better able to capitalize on predictable variation in pitch when it was the irrelevant dimension. Additionally, response times for musicians were significantly shorter than those for nonmusicians, indicating that across conditions, musicians were able to process the relevant dimension more quickly than nonmusicians. Pitt concluded that regardless of level of musical training, pitch and timbre are integral dimensions. However, the processing asymmetry observed in the nonmusician group indicates that timbre is a perceptually more salient dimension than pitch for musically-untrained listeners.

Pitt (1994) demonstrated that although pitch and timbre were integral dimensions for both musicians and nonmusicians, the musicians were more efficient at correctly classifying stimuli in the orthogonal condition. Many researchers have found complementary results; some have reported that pitch and timbre perceptually interact regardless of musicianship (Russo & Thompson, 2005; Singh & Hirsh, 1992; Vurma, Raju, & Kuuda, 2010). For example, Russo and Thompson (2005) found that for musicians, the pitch-timbre interactions were only observed in descending intervals, which are typically encountered less frequently in musical training than ascending intervals. This suggests that musically-trained individuals are not universally superior at perceptually separating pitch and timbre, but that they only have an advantage in familiar musical situations. Others, however, have found comparable interactions in musicians and nonmusicians in situations that could be musically-relevant, such as determining the direction of a pitch change or judging whether a comparison tone was in tune to that of the standard tone (Singh & Hirsh, 1992; Vurma et al., 2010).

Far more researchers have reported that musical experience enhances the ability to cognitively distinguish between the dimensions since for nonmusicians, pitch discrimination abilities diminish once harmonics are added to the fundamental frequency (Beal, 1985; Becker & Hall, 2014; Fine & Moore, 1993; Pitt & Crowder, 1992; Platt & Racine, 1985; Micheyl, Delhommeau, Perrot, & Oxenham, 2006; Preisler, 1993; Zarate, Ritson, & Poeppel, 2013). Some researchers have reported that musicians were better at detecting pitch differences than nonmusicians when timbre varied, but that both groups exhibited similar performance when timbre remained the same (Beal, 1985; Fine & Moore, 1985; Pitt & Crowder, 1992). Similarly, other researchers have reported that musicians are more accurate at tuning comparison tones to the pitch of standard tones when their timbres differ (Platt & Racine, 1985; Preisler, 1993). Musicians also exhibit lower pitch and interval discrimination thresholds for complex tones than nonmusicians (Micheyl et al., 2006; Zarate et al., 2013).

While there clearly are discrepancies concerning the influence of musical training on pitch-timbre interactions, the cause of the inconsistencies is less clear. One possibility is the fact that musically-trained individuals simply tend to process pitch more accurately than individuals without such training (Fine & Moore, 1993; Itoh, Okumiya-Kanke, Nakayama, Kwee, & Nakada, 2012; Schön, Magne, & Besson, 2004). Some researchers have reported superior performance on pitch discrimination tasks by musicians (Bidelman, Gandour, & Krishnan, 2011; Bidelman, Krishnan, & Gandour, 2011; Strait, Kraus, Parbery-Clark, & Ashley, 2010; Wayland, Herrera, & Kaan, 2010). Similarly, others have reported lower pitch discrimination thresholds for musicians (Demany & Semal, 1993; Zarate et al., 2013). Still others have reported that musicians are better at detecting pitch violations and mistunings in both tones and language (Habibi, Wirantana, & Starr, 2013; Marques, Moreno, Castro, & Besson, 2007; Schellenberg & Moreno, 2009). All of these results suggest that musically-trained individuals are able to detect smaller deviations in pitch than nonmusicians. If musical training enhances pitch perception, then musicians would naturally have an advantage in distinguishing pitch from other perceptual characteristics, including timbre. It is important to note that although superior processing in one dimension does not guarantee perceptual independence across dimensions, it could make integrality more difficult to observe.

There is evidence that musicians exhibit superior performance on timbre discrimination tasks as well (Chartrand & Belin, 2006; Crummer, Walton, Wayman, Hantz, & Frisina, 1994). Although some suggest that musically-trained individuals have generalizable perceptual advantages for timbre, Pantev, Roberts, Schulz, Engelien, and Ross (2001) demonstrated that musicians tend to have enhanced neural activity in response to the timbres of familiar instruments. The researchers used magnetoencephalography to demonstrate that violinists and trumpeters had greater neural responses averaged across hemispheres for the timbre of their primary instrument than for sounds presented in other timbres, in addition to larger responses for instrumental timbres in general than for sine tones. These conflicting results suggest that while musicians may have perceptual advantages for timbre, such advantages may not generalize to non-musical contexts. Together, the existing literature suggests that while pitch and timbre perceptually interact in all individuals, musical training may enhance the ability to cognitively distinguish pitch changes from timbre changes. However, the aspect of training that contributes to such an enhancement is presently unclear.

Previously-Developed Scales to Measure Musicianship

Perhaps the most likely cause for the observed discrepancies in the relationship between musicianship and pitch-timbre interactions is the immense variability in how researchers have defined and measured musicianship. A recent survey of 38 published studies investigating pitch perception differences between musicians and nonmusicians revealed this lack of consistency (Daly & Hall, 2016; see Table 1). Although there are several validated scales available to assess auditory and musical skills, such scales are not always feasible to implement in addition to a full experiment, and others may not collect all of the information of interest. As a result, it is common practice for each laboratory to create their own surveys to collect information regarding musical experience, and there is no formal consensus regarding the types of items to be included in these surveys.

Musicianship is typically classified based upon aptitude, skill level, or experience. Aptitude tests, such as the Seashore Tests of Musical Talents (Seashore, 1919; 1960), are most appropriately used to measure natural ability and predict the potential level of success an individual may have in musical training, which can make it difficult to separate the effects of aptitude from the effects of experience and training. If a researcher wants to make the argument that musical training itself is correlated with some sort of performance measure, then aptitude is not the most appropriate measure to use, since it is thought to be unrelated to skills gained through training. Rather, a measure of specific musical training experiences *controlling* for aptitude would provide a clearer indication of the direct influence of musical training. Additionally, aptitude tests tend to take a great deal of time to administer, and thus are not ideal to use in conjunction with other experimental tasks. For example, one of the more popular musical aptitude tests, the Gordon Musical Aptitude Profile, consists of three major divisions (Tonal Imagery, Rhythm Imagery, Musical Sensitivity), and each division takes approximately 50 minutes to administer (Gordon, 1965). Even the alternative Measures of Music Audiation, which are intended to be brief tests of musical aptitude, take approximately 20 minutes for researchers or teachers to administer (Gordon, 1982).

Another popular method for classifying participants as musicians or nonmusicians involves the assessment of skill level. Such measures typically assess specific knowledge and skills (that are generally developed in musical training), which is appealing in experimental research wanting to connect specific musical skills to task performance. However, these assessments also take a long time to administer, and many are not clearly different from musical aptitude tests. For example, validity evidence for the Profile of Music Perception Skills (PROMS, Law & Zentner, 2012) test was established by examining intercorrelations with musical aptitude tests such as the Gordon Musical Aptitude Profile, even though aptitude and ability represent different constructs. Aptitude is proposed to remain stable over time, and is an indicator of potential for skill growth (Boyle & Radocy, 1987), whereas ability is subject to change based upon experience and training. It can be useful to think of musical aptitude as a trait that facilitates the acquisition of specific musical abilities. Thus, aptitude and ability are related, but not identical.

A newer method for distinguishing musicians from nonmusicians is to categorize them based on their level of musical sophistication, a general term that subsumes performance and aural skills, involvement with music, ability to appreciate and evaluate music, and commitment to improving musical abilities (Ollen, 2006). Two psychometrically analyzed scales to measure this construct are the Goldsmiths Musical Sophistication Index (Gold-MSI) and the Ollen Musical Sophistication Index (OMSI) (Müllensiefen, Gingras, Stewart, & Musil, 2014; Ollen, 2006). Although these scales include a variety of items to assess musical background and experience, they also include items assessing musical preference and enjoyment, which introduces affect into the measurement.

In addition to conceptual issues regarding the construct being measured, each scale is plagued by fundamental psychometric concerns. For example, determination of which items would be retained in the final version of the OMSI was based solely upon which items best predicted an expert's ratings of participants' musical sophistication in a logistic regression model. Expert raters were professors in the music school, and each professor rated all of the students in their class. Inter-rater reliability could not be calculated since each student was assessed only by one rater, nor was the lack of independence between ratings within each classroom taken into consideration. Additionally, because backward elimination was used to determine which predictors would remain in the final model, there is no guarantee that the items retained in the OMSI are actually good predictors of musical sophistication.

The primary issue with the Gold-MSI is that it includes some poorly-worded items according to some guidelines proposed by Bandalos (2017), which elicit uncertainty from participants that can in turn introduce measurement error (which may influence statistical significance by reducing power). Some items are vague, such as, "I am able to identify what is special about a given musical piece." Different individuals likely have different conceptualizations of what the word, "special," means, and would likely interpret and respond to the item in different ways. For example, one respondent may interpret "special" as a quality that is unique to the piece in terms of music theory, whereas another respondent may interpret it as what makes the piece special to them personally. Other items on the scale include more than one complete thought, such as, "I don't like to sing in public because I'm afraid that I would sing wrong notes." Participants may not like to sing in public for reasons other than fear of singing the wrong notes, but are only permitted to agree/disagree with the statement in its entirety.

Statistical Considerations

There also are a number of statistical issues with the manner in which musicians and nonmusicians have been defined. Although a variety of indicators can be used to estimate amount of musical experience, by far the most popular one is total years of formal musical training, which is typically operationally defined as time spent learning to play an instrument or sing via formal music lessons or music classes (see Table 1). However, despite the popularity of this method, there is a distinct lack of psychometrically evaluated scales to measure such experience. Measured properly, total years of training should ideally be a continuous variable with good variability between participants. Unfortunately, it is common practice to collect such information, then split participants into musician/nonmusician groups based upon number of years. Median splits are a common method for creating groups, but occasionally researchers choose a cut-off that will yield two groups of approximately the same size.

Different studies use different points of dichotomization of years of experience and the end result is a large number of studies that cannot appropriately be compared to one another, and quite possibly different conclusions regarding the nature of pitch-timbre interactions in musicians and nonmusicians. For example, in one study musicians may be defined as anyone with three or more years of musical training, and nonmusicians would be defined as anyone with fewer than three years of training. Another study may classify musicians as anyone with 10 or more years of training, and nonmusicians as anyone with fewer than five years of training. Both studies may find significant differences between musicians and nonmusicians, but because their groups are defined in such a discrepant manner, comparing the results of the two studies makes little sense. Alternatively, only one study may find a significant difference between groups, leading to different conclusions about the relationship between musicianship and pitch-timbre interactions. As a result, this practice of dichotomization has been argued to be statistically inappropriate in nearly every situation (e.g., see MacCallum, Zhang, Preacher, & Rucker, 2002).

One of the chief concerns with categorizing continuous variables is the loss of information regarding individual differences (Humphreys, 1978, MacCallum et al., 2002). Conceptually, this should be alarming to researchers incorporating musical experience as an independent variable in their studies, since such studies typically hypothesize individual differences as a result of the total amount of training. Collapsing individual differences into discrete categories (musician v. nonmusicians) is also detrimental to ecological validity. For example, if two individuals who differ in musical training by only one year are on either side of the point of dichotomization, then one of those individuals would be classified as a musician, and the other as a nonmusician. Furthermore, if the range of musical training in the musician group was 10 years, then a person with three years of experience would be treated as having the same amount of experience as a person with 13 years. Making any conclusions regarding the influence of more years of musical training is clearly inappropriate in such a situation.

The loss of data regarding individual differences also raises several other statistical issues including loss of power, attenuated correlations and effect sizes, and potentially spurious statistical significance (e.g., see MacCallum et al., 2002). For example, Cohen (1983) demonstrated that dichotomization of a single variable at the mean can result in a loss of power equivalent to discarding data from approximately 38% of participants in large samples (*N* = 80), or up to 60% of participants in smaller samples (*N* = 25). Because sample sizes tend to be fairly small in perceptual studies evaluating the influence of musicianship, it is likely that the loss of power due to dichotomization is equivalent to discarding data from approximately half of the participants. Note that dichotomization of a single variable at the mean is a best-case scenario, and only if the sample is perfectly normal. This loss of power increases the further the point of dichotomization moves from the mean, or as additional variables are dichotomized (due to the further loss of information about individual differences within those variables).

Categorizing continuous variables can also influence the magnitude of correlations between variables, which in turn influences effect size and increases the chance of committing Type I error (Cohen, 1983). Even if a single independent variable is dichotomized at its mean, the resulting population correlation will still be attenuated by approximately 20% (e.g., see MacCallum et al., 2002). However, it is possible to obtain larger correlations and effect sizes due simply to sampling error, especially for small sample sizes and small population correlations (size of the actual effect in the population is small). Such a favorable result does not mean that dichotomization was beneficial. Rather, it is indicative that the sample was not representative of the population, and reflected a correlation larger than that of the true population correlation.

Maxwell and Delaney (1993) also demonstrated that in certain situations, spurious significance can result from dichotomization. This risk is particularly high when two independent variables are correlated with one another, but only one of those variables is correlated with the outcome. When both predictor variables are dichotomized and submitted to an ANOVA, significant main effects for *both* can emerge, despite the fact that only one was substantially related to the outcome variable. Additionally, if either predictor variable is nonlinear, ANOVA also may reveal a spuriously significant interaction between the two variables. Regression models, on the other hand, can easily incorporate nonlinear terms in addition to properly modeling interactions between continuous variables (see Aiken & West, 1991).

All of the issues described above contribute to a general difficulty of comparing results across studies, especially for meta-analyses. Because population correlations are distorted by dichotomization, aggregating such results across studies is statistically and conceptually inappropriate. Hunter and Schmidt (1990) presented a potential solution to this issue in the form of complicated statistical corrections for attenuated effect sizes. The corrections involve a weighting system in which correlations involving continuous variables are given the largest weights, those involving variables with near-median splits are given moderate weights, and those involving variables with extreme splits are given the smallest weights. However, the authors ultimately concluded that the ideal solution is for researchers to report correlations among the original continuous variables in addition to the dichotomized versions of those variables.

Many studies that have divided participants into “musician” and “nonmusician” groups explicitly report high variability with respect to the dependent measure within each of those groups (Baumann, Meyer, & Jäncke, 2008; Beal, 1985; Singh & Hirsh, 1992; Spiegel & Watson, 1984; Vurma et al., 2010). For example, Wayland et al. (2010)

attributed variability within the musician group to differences in the number of instruments played and the amount of practice/study time with each of those instruments. Variability within the nonmusician group was attributed to some participants having a small amount of musical experience. Micheyl et al. (2006) posited that different types of musical backgrounds (e.g. classical vs. contemporary) likely contributed to different auditory performance enhancements. The researchers thus used more stringent selection criteria for their musician group (classical musicians with at least 10 years of experience), but still found substantial intra-group variability in pitch discrimination based upon instrument played such that pianists performed more poorly at a pitch discrimination task relative to other instrumentalists (winds and strings). Different instruments require different performance demands, further supporting the notions that musicians are not a homogenous group (Carey et al., 2015). For example, it has been suggested that string, wind, and brass players typically need to pay closer attention to intonation while playing, whereas percussionists and pianists can focus more on timing and precision (Ehrlé & Samson, 2005). Because of the substantial variability in perceptual and motor skills required by different types of musicians, it does not seem appropriate to put them all into one group and treat them as though they were equivalent.

Similarly, it is probably not appropriate to include participants in a “nonmusician” group if they have several years of musical training. There is no clear divide regarding how many years of training are required to be considered a “musician,” so it makes little sense to treat a participant with several years of musical training as possessing the same skills and experiences as a participant with no musical training whatsoever. Thus, it is most appropriate to model years of training as a continuous variable in addition to including other variables to capture the different facets of musicianship.

A final reason that musical training should be modeled as a continuous variable concerns the shape of the relationship between the amount of training and skill acquisition. It has been commonly reported that the rate of learning or improvement declines the longer a certain skill is practiced or studied (Karni et al., 1998; Mazur & Hastie, 1978). Improvements occur rapidly and easily during the early stages of learning, but eventually taper off later in training. Thus, the same increment of improvement that might be made early in the learning process is likely to take an exponentially larger amount of effort later in the learning process. Although this possibility has yet to be investigated with regards to musical training, it seems likely that skill acquisition would follow the same exponential pattern. Because the relationship between amount of training and observed skills could be curvilinear, it is important to use a statistical model that can incorporate such nonlinearity, which requires continuous variables. As mentioned earlier, failure to model a nonlinear variable continuously could lead to a spuriously significant result when that variable is categorized and submitted to an ANOVA (Maxwell & Delaney, 1993).

The Current Investigation

The current investigation attempts to address these issues by evaluating musicianship in a more continuous fashion while exploring a potential explanation of pitch-timbre interactions. The primary goal of this study was to determine whether the spectral centroid, which has previously been suggested as an explanation of pitch-timbre interactions in missing-F_0 experiments (Becker & Hall, 2014), also accounts for interactions observed in Garner speeded classification. Additionally, the current study evaluated whether this centroid-based explanation accounts for musical training-based performance differences on the Garner task.

Two sets of the Garner speeded classification task were used to evaluate perceptual interactions of pitch and timbre in individuals with different levels of musical training. The first set used stimuli that vary in spectral envelope, amplitude envelope, and spectral centroid, and served as a control condition. Given that timbre has been demonstrated to be perceptually more salient to nonmusicians than to musicians, performance in this set was expected to replicate the results of Pitt (1994). In the pitch-focus condition, individuals with less musical training were expected make more errors in the orthogonal trials relative to fixed trials. Individuals with more musical training were expected to make fewer errors, but would show an interference effect in their response times.

The second set used stimuli that had been equated to have the same spectral centroid. Thus, spectral envelope and amplitude envelope were the only available timbre cues. If the centroid is the basis of timbre interference, then equating tones to have the same centroid should remove interference effects. If interference effects were removed, then performance differences between musicians and nonmusicians were also expected to vanish. Thus, performance of all participants would be similar, regardless of level of musical training. All participants were expected have similar error rates and response times for all conditions of the experiment.

A secondary goal of the current investigation was to evaluate whether training-based performance differences on the Garner task changed when training was examined in a statistically more appropriate way. To address measurement and statistical problems, a new musical training survey with an emphasis on including continuous items and capturing a more diverse range of musical experiences was developed. This was intended to maximize the amount of information available for statistical analyses, which should have in turn helped pinpoint specific musical learning experiences that contribute to reduced perceptual interactions between pitch and timbre. As of right now, musical experience is typically assessed using categorical analyses such as ANOVA, which do not allow for assessment of individual differences. By creating a survey that collects information on a continuous scale of measurement, regression analyses that highlight individual differences can more easily be employed.

Method

Participants

Fifty-five participants were recruited through the JMU Department of Psychology participant pool, which allows undergraduate students in introductory psychology courses to satisfy a course requirement by participating in research. Data from seven participants were excluded from analyses due to a failure to follow instructions or due to a failure to reach the minimum average of 70% accuracy in the Baseline condition. As a result, statistical analyses were restricted to data from a total of 48 participants. All participants were between 18 and 40 years old so that they could provide informed consent and to reduce potential impacts of presbycusis. Additionally, participants were required to understand written and spoken English, as all instructions were provided in English. Participants were asked to self-report any known hearing deficits. Reported deficits were not be used to exclude anyone from participating, but were used to exclude data from statistical analyses. Listeners were not selected on the basis of musical training, but they were surveyed on a host of training-related behaviors (years of training, type of training, age of training onset, etc.). This nonspecific selection yielded a slightly positively skewed distribution (more listeners in the lower range of experience) along the various musical training variables collected by the musical training survey (e.g., total years of experience, years in a musical ensemble, hours of practice per week).

Materials

Three surveys were administered to all participants, the first of which was a questionnaire developed as part of the current investigation that was designed to capture a variety of aspects of musical training. The other two surveys were administered to collect validity evidence for the new scale. They are measures of musical sophistication, rather than musical training. However, they are frequently used by researchers to collect information regarding musical training, and are thus the most similar measures available to provide convergent validity evidence.

Musical Training and Experience Survey (MUTE). The complete version of the survey of musical training which includes items assessing type and duration of auditory/musical training experiences can be found in the Appendix. The items were designed to capture different facets of experience that may contribute to enhanced auditory perception such as attendance in formal music classes, performance experience, practice habits, composition experience, and musical style most frequently-played. Whereas traditional surveys of musical training typically employ categorical items, this new survey includes open-response items whenever possible. This allows for the items to be measured in a more continuous manner, which in turn enables enhanced measurement of individual variation. Rather than have participants select a category in which their level of training falls, respondents were asked to generate their own values to indicate amount of training. For example, one item from the survey reads:

> If you are currently involved in musical activities, about how many hours do you spend playing music per week, including rehearsal and individual
>
> (1) practice? _____ hours/week

This survey represents an exploratory attempt to identify different components of musical training (which aren't typically measured) that could influence perception of pitch and timbre. Thus, there are not distinct categories that are being separately measured. Often, a categorical item will be followed by a continuous item. This allows those without certain experiences to advance more quickly through the survey, while respondents with more experience should give more detailed answers regarding their experiences. For example:

> Are you currently or have you ever received private music lessons? (if answering
>
> *no*, skip to question #11)
> (2)
>
> ☐ yes (currently receiving)
> ☐ yes (received in the past)
> ☐ no (no history of private lessons)
>
> Over what approximate dates and for how long has this private instruction taken
>
> place?
> (3)

Ollen Musical Sophistication Index (OMSI). Scores on the OMSI were calculated and correlated with items on the MUTE to assess convergent validity. The OMSI is a 10-item scale designed to provide a single indicator of musical sophistication

(Ollen, 2006). A total score is created by multiplying selected response options by the corresponding regression coefficients that were determined during the validation of the scale, with higher scores reflecting higher levels of musical sophistication. The OMSI primarily includes objective items assessing musical background and experience, such as years of study, amount of college-level musical coursework completed, and amount of composition experience. However, it also includes items that seem more tangential to musical training, such as live concert attendance and self-categorization as a musician or nonmusician. Four of the items use the continuous scale of measurement, five are ordinal, and one is nominal. Here is a sample ordinal item with its corresponding response options:

> Which category comes nearest to the amount of time you currently spend practicing an instrument (or voice)? Count individual practice time only; not group rehearsals.
>
> (4)
>
> - ☐ I rarely or never practice singing or playing an instrument
> - ☐ About 1 hour per month
> - ☐ About 1 hour per week
> - ☐ About 15 minutes per day
> - ☐ About 1 hour per day
> - ☐ More than 2 hours per day

Goldsmiths Musical Sophistication Index (Gold-MSI) Musical Training subscale. The Gold-MSI is a multi-part assessment of musical sophistication that was administered to collect additional convergent validity evidence for the MUTE (Müllensiefen, Gingras, Stewart, & Musil, 2011). It has five subscales to measure different aspects of sophistication: Active Engagement, Perceptual Abilities, Singing Abilities, Emotions, and Musical Training. The Musical Training subscale was designed to assess musical background and experience, but only includes seven items, two of which are affective items unrelated to actual training experiences. This subscale has been psychometrically validated to be administered on its own. The two affective items on the Musical Training subscale are measured with a 7-point Likert scale ranging from 1 (*completely disagree*) to 7 (*completely agree*). For example, one item from the subscale reads:

I have never been complimented for my talents as a musical performer.

(5)

- ☐ 1 Completely Disagree
- ☐ 2 Strongly Disagree
- ☐ 3 Disagree
- ☐ 4 Neither Agree nor Disagree
- ☐ 5 Agree
- ☐ 6 Strongly Agree
- ☐ 7 Completely Agree

The other five items on the subscale are more objective and measure specific aspects of musical background and experience. Although the questions reference continuous variables, all five items are asked in a multiple-choice, ordinal format:

At the peak of my interest, I practiced **0 / 0.5 / 1 / 1.5 / 2 / 3-4 / 5 or more** hours per day on my primary instrument.

(6)

The Musical Training subscale of the Gold-MSI is scored by simply summing all of the responses. The two affective items are reverse-scored, and the objective items receive points corresponding to the response category in ascending order. For example, in example item 6 above, the first category, “0,” would receive a score of 0, and the second category, “0.5,” would receive a score of 1. Higher scores on this subscale represent a higher level of musical training.

Stimuli

Stimuli consisted of two sets of four synthesized instrumental tones: two different instruments (derived from samples performed on violin or trumpet) at two different pitches (A_4, $D^{\#}_4$). Violin and trumpet were chosen on the basis of timbral dissimilarity: the two tones come from different musical instrument families and involve different methods of sound production. These were the same instruments used by Pitt (1994), and have distinct spectral envelope shapes, spectral centroids, and rise times, as can be seen by the spectral profiles and corresponding centroid measurements displayed in Figure 1. As can be seen in the figure, violin tones have higher spectral centroids than the trumpet tones, in addition to having more peaks in their spectral profile. Thus, the two timbres should have been distinguishable even when their spectral centroids were equated. All tones had a duration of 1*s* and had linear attack and release amplitude ramps. Both timbres had 20*ms* release ramps created using Adobe Audition CS6 v.5.0 (2012), but the duration of the attack ramps differed between timbres: 60*ms* for trumpet, and 400*ms* for violin. These differing attack times were expected to be representative of the two timbres, since a bowed violin typically has a more gradual attack than a trumpet.

All stimuli were presented at a peak amplitude of 80 dB[A] and rendered with a 44.1 kHz sampling rate (16-bit depth resolution). At the time of presentation, stimuli were submitted to an anti-aliasing, low-pass filter (Butterworth) with a -24dB/octave slope and a cut-off frequency of 11 kHz. All stimuli were equated for average RMS amplitude using Adobe Audition to roughly equate them for loudness without compromising spectral manipulations. Stimuli were presented to participants via Sennheiser HD 25-SP II Headphones while in a single-walled sound-attenuated chamber.

An original set of unedited synthesized tones was obtained from instrument samples within Ableton Live's Orchestral Instruments Collection, and were only used to obtain harmonic amplitude measurements. The samples were obtained from natural recordings of a solo legato violin and a solo legato trumpet. These samples were played back within Ableton Live's 9.6.2 (2015) Sampler virtual studio technology at the two specified fundamental frequencies – one at A_4 (440 Hz) and one at $D^{\#}_4$ (311.13 Hz).

An initial set of tones was a simplified set of synthesized tones, created by modeling a static spectral envelope via the harmonic profile of the original sounds. A static spectral envelope was expected to produce a constant effect of filtering when adjusting the centroid, thus making such adjustments more precise. To create the simplified tones, the relative amplitude of each harmonic of each original tone was measured using Camel Audio's *Alchemy* v1.50.1, a sample-based virtual instrument and VST plug-in. Amplitude values for the first 50 harmonics were then uploaded into FormAnt v.1.010117, a formant synthesizer plug-in device designed in the Max for Live 7.2.4 (2015) programming environment (Hall & Redpath, 2016). Tones were synthesized with static spectral envelopes and 1*ms* releases. To create representative attack times, linear attack ramps were added to the tones within the FormAnt device: 60*ms* for trumpet, and 400*ms* for violin.

A second set of tones consisted of centroid-adjusted tones, where all four tones had the same spectral centroid. This set of tones was used to evaluate the contribution of the centroid to pitch-timbre interactions. Specifically, if the centroid were responsible for such interactions, then interference effects should have been minimized if two tones with different timbres were equated to have the same centroid. Spectral centroids of the simplified tones were calculated using the following formula:

$$Centroid = \frac{\sum(HarmHz * dB)}{\sum nHarm} \qquad (1)$$

Centroids were equated by adjusting all 50 harmonic amplitudes of simplified tones (except for F0) to reflect the centroid shift. This is conceptually similar to applying a low-pass filter with a very shallow slope such that a small amount of energy is removed from each harmonic. By removing energy from each harmonic rather than applying a standard low-pass filter, slope could be changed by as little as a fraction of a dB, which allowed for more precise centroid adjustment. Such a shallow slope should have also helped to maintain the general spectral envelope shape, which would help maintain distinguishability of the two instrumental timbres. The centroids of both violin tones and the $D^{\#}_4$ trumpet were equated to match that of the A_4 trumpet, the tone with the lowest spectral centroid (1481 Hz).

Procedure

After giving informed consent, participants first completed the three surveys prior to beginning the experiment. Order of the surveys was counterbalanced across participants. Stimuli were presented, timing was controlled, and responses were collected and stored using E-Prime v.2.0 (SP1; Psychology Software Tools, Inc., 2012) experiment generation software.

The experiment began with a familiarization task in which participants listened to examples of the simplified stimuli five times each in order to encourage recognition of the intended instrument. Stimuli were presented in a fixed order: trumpet A_4, violin A_4, trumpet D#$_4$, violin D#$_4$. This presentation order was designed to encourage discrimination between the two different timbres, but no responses were recorded during the familiarization. The familiarization task was followed by a brief categorization task in which participants were asked to categorize each stimulus based on timbre (violin or trumpet). The categorization task consisted of 40 stimuli: 10 of each tone. This categorization task was intended to further familiarize participants with the simplified timbres in addition to ensuring that they could reliably distinguish between the two timbres.

The main experimental procedure consisted of the Garner speeded classification task, where listeners were asked to classify stimuli as rapidly as possible while maintaining accuracy. Table 2 presents a summary of the different blocks of trials that were used for a single set of stimuli in this experiment. In total, 20 blocks of 48 trials each will were presented: 10 blocks of trials with the simplified stimuli, and 10 blocks of trials with the centroid-adjusted stimuli. For each stimulus set, five blocks of trials were timbre-focused blocks, in which participants were asked to classify stimuli on the basis of musical instrument timbre (violin or trumpet). The other five blocks of trials were pitch-focused blocks, in which participants were asked to classify stimuli on the basis of pitch (low or high).

Within each focus condition, two of the blocks of trials reflected fixed (Baseline) conditions, two were Correlated, and one was Orthogonal. In the Baseline blocks of the pitch-focus condition, participants were asked to classify tones presented in a single timbre as having low (D#$_4$) or high (A$_4$) pitch. In Baseline blocks of the timbre-focus condition, participants classified tones presented at a single F$_0$ as having trumpet or violin timbre. In one of the Correlated blocks of the pitch-focus condition, participants classified tones as having low or high pitch, but the low pitch was always presented in the violin timbre and the high pitch was presented in the trumpet timbre. In the other Correlated block of the pitch-focus condition, the low pitch was always presented in the trumpet timbre and the high pitch was always presented in the violin timbre. This distribution of stimuli was the same in the timbre-focus condition, but participants were asked to classify tones as having trumpet or violin timbre rather than classifying based on pitch. In the Orthogonal blocks of both focus conditions, any combination of the two dimensions was permissible, and participants were asked to classify stimuli on the relevant dimension. Ordering of these blocks was counterbalanced across participants, with additional counterbalancing of the two Baseline and Correlated blocks. Optional rest breaks were provided after each block of trials, and brief rest breaks of approximately five minutes each were mandated after completion of every five blocks (after each focus condition).

On each trial, participants heard a single tone. Depending upon the condition, they were then asked to report whether the tone was low/high pitch or violin/trumpet timbre. Because this was a speeded classification task, participants were asked to make their response as soon as they were able to categorize the tone, regardless of whether the tone had finished playing. There was a 500ms intertrial interval following each response. All responses were made using a DirectIN High Speed Button-Box v2012 from Empirisoft to ensure measurement of response times within millisecond timing accuracy. The first/left-most button on the response box were used to advance through instructions slides and familiarization examples. The second and third buttons were used in the pitch-focus conditions, and were labeled “L” and “H” for low and high timbre, respectively. The seventh and eight buttons were used in the timbre-focus condition, and were labeled “V” and “T” for violin and trumpet timbre, respectively.

Within each experimental block of trials, each stimulus was repeated 24 times. The Baseline and Correlated blocks consisted of 48 trials since only two stimuli were presented in those blocks. The Orthogonal block consisted of 96 trials due to the fact that all four stimuli were presented within that block. The entire experiment took no longer than 90 minutes to complete.

Results

Speeded Classification Performance: Overall Effects

For both stimulus sets, accuracy was calculated for each participant as percentage of correct responses to each stimulus. These correct responses were then averaged across each block of trials to obtain accuracy for each classification condition. Because Baseline blocks of trials involved classifying tones without any variation in the irrelevant dimension, a failure to reach 70% accuracy would represent either noncompliance with task instructions or a true inability to discriminate between the tones. As a result, if a participant had less than 70% accuracy in any of the averaged data from blocks of trials within the Baseline conditions, then their data were excluded from analyses.

Median response times were also calculated for correct responses within each block of trials for each participant. All response times shorter than or equal to 150*ms* were omitted because it has been empirically demonstrated that it takes a minimum of 150*ms* to choose a response option and press the correct button (Luce, 1986). Thus, any times shorter than 150*ms* would not represent a decision-making process, and should not be included in the measurement of perceptual differences.

IBM SPSS Statistics v.22 was used for all statistical analyses. For both sets of stimuli, accuracy and median response times were submitted to separate 2 x 3 repeated measures analyses of variance (ANOVAs) with focus condition (Pitch, Timbre) and classification condition (Baseline, Correlated, Orthogonal) as factors. Bonferroni-corrected pairwise comparisons were computed for all significant interactions. Greenhouse-Geisser corrections were used for situations in which the assumption of sphericity was violated, and partial eta squared values were calculated as a measure of effect size. Partial eta squared effect sizes are traditionally considered small if $\eta_p^2 = .10$, medium if $\eta_p^2 = .30$, and large if $\eta_p^2 = .50$ (Cohen, 1988).

Figure 2 presents accuracy for each classification condition within each focus condition for both sets of stimuli. As can be seen in the figure, overall accuracy in pitch-focus and timbre-focus conditions and was similar, as indicated by a non-significant main effect of focus condition, $F(1, 47) = 2.920$, $p = .094$, $\eta_p^2 = .058$ (Table 3). Across both stimulus sets, accuracy significantly varied between classification conditions, simplified: $F(1.353, 63.605) = 48.610$, $p < .001$, $\eta_p^2 = .508$ (Table 3), adjusted: $F(1.183, 55.596) = 45.501$, $p < .001$, $\eta_p^2 = .492$ (Table 4). Corrected degrees of freedom are reported here because Mauchly's test indicated that the assumption of sphericity had been violated, and using corrected degrees of freedom help ensure that the standard errors and corresponding statistical inferences remain unbiased, $\chi^2(2) = 10.578$, $p < .05$. Subsequent pairwise comparisons of means (collapsed across focus conditions) revealed that accuracy was significantly lower in the Orthogonal condition relative to the Baseline condition for both stimulus sets, indicative of an interference effect, $p < .001$. Contrary to hypotheses, there were no significant differences between Correlated and Baseline conditions, $p > .05$, suggesting the lack of a redundancy gain effect.

Figure 3 presents the median response times for each classification condition within each focus condition and for each stimulus set. As can be seen in the figure, the overall median response times were faster for the pitch-focus condition than for the timbre-focus condition, suggesting that participants were able to classify tones on the basis of pitch faster than they could classify on the basis of timbre, simplified: $F(1, 47) = 22.957$, $p < .001$, $\eta_p^2 = .328$ (Table 5), adjusted: $F(1, 47) = 34.188$, $p < .001$, $\eta_p^2 = .421$ (Table 6). Response times also varied between classification conditions for each stimulus set, simplified: $F(1.659, 77.981) = 128.632$, $p < .001$, $\eta_p^2 = .732$ (Table 5), adjusted: $F(1.642, 77.153) = 223.200$, $p < .001$, $\eta_p^2 = .832$ (Table 6). As can be seen in Figure 3, pairwise comparisons for both stimulus sets revealed that response times were significantly longer in the Orthogonal condition relative to the Baseline condition, consistent with an interference effect, $p < .001$.

Similar to accuracy data for the simplified stimulus set, response times failed to show evidence of redundancy gain, as indicated by the lack of significant differences between Correlated and Baseline conditions, $p > .05$. Unlike the simplified stimulus set, pairwise comparisons for the adjusted stimulus set revealed evidence of redundancy gain, as indicated by significantly shorter response times in the Correlated condition relative to the Baseline condition, $p = .005$. However, pairwise comparisons following a small but significant interaction between focus condition and classification condition, $F(1.438, 67.567) = 6.867$, $p = .005$, $\eta_p^2 = .127$ (Table 6), revealed that the redundancy gain effect was only significant for the timbre-focus condition, $p < .001$, but not for the pitch-focus condition, $p = .551$.

Musical Training and Interference Effects

Because most studies investigating the influence of musical training tend to use musician and nonmusician groups, the relationship between musical training and interference effect magnitude was first assessed using a set of independent-samples t-tests with musicianship (musician/nonmusician) as a dichotomous independent variable. These analyses were conducted for the sake of comparison with other studies using a musician/nonmusician dichotomy. Participants were classified into “nonmusician” and “musician” groups based upon their total years of musical training obtained from the MUTE. A cut-off of three years was selected to ensure that each group was roughly the same size. Participants with three or fewer years of experience were classified as nonmusicians ($N = 25$), and participants with more than four years of experience were classified as musicians ($N = 23$).

Remaining analyses will focus on interference effects as the dependent variables because they were the only effects consistently observed across stimulus sets in previous analyses. To calculate the effect for each focus condition within each stimulus set, the average median response times from both Baseline blocks of trials were averaged, then that combined Baseline average was subtracted from the average median response time from the Orthogonal block of trials. Thus, four interference effects were calculated for both accuracy and median response times, one for each combination of focus condition and stimulus set: simplified pitch-focus, simplified timbre-focus, adjusted pitch-focus, and adjusted timbre-focus. Only the response times for the adjusted pitch-focus condition showed significantly smaller interference effects for musicians (*M* = 125.54*ms*) relative to nonmusicians (*M* = 197.51*ms*), *t*(46) = 2.281, *p* = .027, η^2 = .102. A corresponding marginal tendency toward a reduced difference in accuracy for musicians (.03 v. .06 for nonmusicians) also was obtained within the adjusted pitch-focus condition, *t*(46) = 1.749, *p* = .087, η^2 = .062.

To assess the influence of musical training in a more continuous manner, simple regression analyses were computed with total years of musical training as the independent variable and the size of interference effects as dependent variables. Of the four interference effects, only one was significantly related to years of musical training: adjusted pitch-focus. Within that condition, the magnitude of the interference effects were significantly reduced for musicians for both dependent measures, accuracy: *F*(1, 46) = 5.286, *p* = .026, R^2 = .103, response times: *F*(1, 46) = 5.187, *p* = .027, R^2 = .101.

To further probe the relationship between musical training and the magnitude of observed interference effects, a multiple regression analysis was conducted with total years of training, years of private lessons, years of musical ensemble experience, and age at which training began as predictor variables. The equation for the planned multiple regression analysis is shown below:

$$Y' = b_0 + b_1(\text{TotalYrs}) + b_2(\text{Lessons}) + b_3(\text{Ensemble}) + b_4(\text{StartAge}) \qquad (2)$$

These variables were selected because they are among the most commonly used indicators of musical training in the literature. Prior to conducting the multiple regression, bivariate correlations between the independent variables were examined to assess the degree of multicollinearity. If predictor variables are highly correlated, then it is likely that they would each explain the same variance in the outcome, and thus should not all be retained in the final model. Table 7 presents the correlation matrix for all the predictors and the interference effects. As can be seen in Table 7, there were strong relationships between each of the predictors, particularly between Total Years and all other variables. This is not surprising, considering that each of the other variables is actually a function of total years of musical training. Given the strong intercorrelations, it is likely that each variable would explain the same variance in the interference effect, and the multiple regression would not be informative. However, this regression analysis was conducted anyway given that there was still a slight chance that one of the variables would be able to explain a significant portion of variance in the outcome above and beyond what was shared by the other variables.

As predicted, the overall analysis with the interference effect for response time as the dependent variable was not significant, $F(4, 43) = 2.375$, $p = .067$, $R^2 = .181$. Additionally, none of the individual predictors explained a significant portion of unique variance in the outcome, $p > .05$. A similar lack of effect was observed when the interference effect for accuracy was used as the dependent variable, $F(4, 43) = 1.747$, $p = .157$, $R^2 = .140$.

To explore whether any of the remaining musical training variables (i.e., those that have not traditionally been considered in research and that were left out of the general analysis summarized above) could help explain additional variance in the outcome(s), exploratory multiple regression analyses also were conducted. Given that the typical variables were so highly correlated, only Start Age was included as a predictor in the exploratory analyses since it was the most strongly related to the magnitudes of interference effects for both response time and accuracy measures ($r = .367$, $r = .355$, respectively). Correlations between other items from the MUTE and the interference effects were examined to select other variables that could have explanatory power in subsequent analyses. Due to the large number of variables examined, Table 8 presents only those that were significantly related to the interference effects: years elapsed since cessation of musical training, number of musical courses taken at any level, and number of instruments played throughout the course of training. The equation for the exploratory multiple regression analysis is shown below:

$$Y' = b_0 + b_1(\text{StartAge}) + b_2(\text{Recency}) + b_3(\text{Courses}) + b_4(\text{Instruments}) \qquad (3)$$

Although each variable was strongly related to the interference effects, they were even more strongly related to each other, suggesting that they would likely explain the same variance in magnitude of the interference effects. Nevertheless, all four predictors were submitted to multiple regression models with each of the interference effects as separate dependent variables. Again, neither the model for response times, $F(4, 43) = 2.494$, $p = .057$, $R^2 = .188$, nor for accuracy, $F(4, 43) = 1.748$, $p = .157$, $R^2 = .140$, reached statistical levels of significance. These nonsignificant results suggest that none of th2e musical training variables included in the analyses were capable of explaining unique variance in the size of the interference effects, nor could they explain a significant amount of variance as a group.

Comparisons of Assessment Instruments

To ensure that the MUTE was measuring musical training the way it was intended to, it was compared against two similar established instruments: the OMSI and the Gold-MSI. Prior to comparing data from items within the MUTE against those from the other instruments, the reliability of data obtained from the OMSI and Gold-MSI was reassessed for the current sample and compared to the reported reliability for those instruments. Although an estimate of reliability was not provided solely for the 10 items retained in the final version of the OMSI, those items were retained from a larger set of 29 items with a coefficient alpha value of .74. When one of the variables, “age at commencement of musical activity,” was temporarily removed, coefficient alpha increased to .78. However, that is not necessarily a reflection of the variable’s lack of relationship to the other items in the set, but rather of the way the item was phrased. Participants with no musical experience were instructed to respond “0,” to the item, and then their ages were used in place of the “0” to obtain regression weights for calculating the total score. When the raw response of “0” was included in reliability analyses instead of the replacement value, it was essentially implying that participants without musical experience actually began musical activities at birth, rather than never, which, unsurprisingly, was at odds with other items in the scale.

Coefficient alpha for all ten items with the current sample was .65, and alpha with the “age at commencement of musical activity” item removed was .68. There are a couple of reasons that internal consistency could have been lower in the present sample than in the original sample, the most likely being number of items. Coefficient alpha is known to increase with more items, as long as those additional items are related to the other items in the scale. Because the final version of the OMSI had 19 fewer items than the original sample, it is natural to expect reliability to be a bit lower. Additionally, the proportion of musically-inexperienced participants was lower in the original sample (65/633, 10.3%) than in the current sample (9/48, 18.8%). It is possible that the scale is more reliable for respondents with larger amounts of musical experience.

Reliability of the Musical Training subscale of the Gold-MSI for the original sample was quite high, as evidenced by a coefficient alpha of .90. That sample was composed of 147,633 respondents who completed a web-based version of the scale. Prior to computing coefficient alpha for the sample from the current investigation, both negatively-phrased items on the Musical Training subscale were reverse-scored (as is standard in Gold-MSI scoring) such that higher scores indicate higher levels of musical training. Coefficient alpha for the current sample was .91, indicating similar reliability across samples.

To briefly assess the validity of the established instruments on the current sample, a subset of the sample was selected to compare scores for "nonmusicians" with 0-.5 years of musical experience (*N*=11) to those of "musicians" with nine or more years of experience (*N*=12). As expected, average scores on each of the scales were higher for musicians (Gold-MSI: 26.54, OMSI: 24.25) than for nonmusicians (Gold-MSI: 2.45, OMSI: 9.95). For the Gold-MSI, this suggests that the scale did a good job of distinguishing between participants with vastly different levels of musical training. However, the same cannot necessarily be concluded for the OMSI. Although average scores were different for the two groups, it is important to remember that the total score on the OMSI is a probability, and a score of 24.25 is interpreted as a 24.25% chance of being classified as "more musically sophisticated." Considering that the musician group in this select subset had a minimum of nine years of musical training, it seems that the OMSI may have less accurately classified participants in the current sample.

To evaluate how the established instruments to measure musical training relate to performance on the Garner speeded classification task, bivariate correlations between measurements from the OMSI and the Gold-MSI were examined. Table 9 presents a correlation matrix of the relationships between the probability score from the OMSI, the total score from the musical training subscale of the Gold-MSI, and the interference effects for the adjusted pitch-focus condition. Analyses were restricted to that condition because it was the only one to show a significant difference due to duration of musical training in the aforementioned initial simple regression analyses. Because total years of musical training was used in the original analyses exploring the relationship between training and interference effects, it also was included in the current correlation matrix as a comparison. Out of the three musical training indicators, the total score from the Gold-MSI had the strongest relationship with the interference effect for response times (*r* = -.332). This negative correlation indicates that higher total scores on the Gold-MSI were associated with less Garner interference. However, total years of training had the strongest relationship with the interference effect for accuracy (*r* = -.321), indicating that longer durations of musical training were associated with less Garner interference. Both the OMSI and Gold-MSI were strongly related to total years of training (*r* = .468, *r* = .859, respectively). The OMSI was not strongly related to either indicator of the interference effect. Although the OMSI and Gold-MSI are both measures of musical sophistication, and were strongly correlated (*r* = .569), the OMSI was not strongly related to either the interference effect for response time or for accuracy. These results suggest that the OMSI was not capable of predicting performance on the Garner speeded classification task, and possibly that musical sophistication is unrelated to performance on pitch-timbre interaction tasks. Although the Gold-MSI is a measure of musical sophistication, it is important to note that only the Musical Training subscale was used in the current investigation, thus making it a less valid measure of musical sophistication as a construct.

Discussion

The primary goal of the current investigation was to test whether the spectral centroid is the dimension of timbre responsible for perceptual interactions between pitch and timbre. To isolate the centroid, two sets of stimuli were generated: one with centroid differences (simplified), and one where all four tones shared the same centroid (adjusted). There was strong evidence of an interference effect for both stimulus sets. Accuracy was significantly lower and response times were significantly slower in the Orthogonal condition relative to the Baseline condition. There was also slight evidence of a redundancy gain effect for the adjusted stimulus set. When participants were focused on classifying each tone based on timbre (violin/trumpet), they responded significantly faster in the Correlated condition relative to the Baseline condition. These interference and redundancy gain effects suggest that pitch and timbre are perceptually integral dimensions, even when centroid differences have been removed. The fact that integrality was demonstrated under both types of stimulus dimensions suggests that the spectral centroid is not the attribute of timbre driving the perceptual interactions.

Similar to previous studies demonstrating integrality, significant interference effects were observed in both pitch- and timbre-focus conditions (Melara & Marks, 1990). However, unlike the results from that study, redundancy gains in the current investigation were only observed in the timbre-focus condition (for the adjusted stimulus set), and only for response times. The most probable explanation for this discrepancy is the fact that the fundamental frequencies used in that earlier study (900 Hz vs. 920 Hz) were much closer to each other than were the fundamental frequencies used in the current study (313 Hz vs. 440 Hz). As a result, it is likely that pitch was more easily discriminable. In fact, as can be seen in Figure 2, accuracy for all Baseline conditions was near ceiling, with the highest accuracy in the pitch-focus condition for both simplified and adjusted stimulus sets. With performance already at ceiling levels in the Baseline condition, no further improvements could be achieved in the Correlated condition, thus leading to a lack of redundancy gain for accuracy in both pitch- and timbre-focus conditions. As can be seen in Figure 3, response times for the Baseline and Correlated conditions were almost identical for both pitch-focus conditions, which would make any sort of redundancy gain effect impossible.

In contrast, performance in the Correlated conditions for each of the timbre-focus conditions was slightly faster relative to the Baseline conditions, suggesting a redundancy gain effect. Although the average response time differences associated with redundancy gain were similar for both the simplified and adjusted stimulus sets, the effect was only statistically significant for the adjusted set due to the reduced variability for that set. Nevertheless, there was a corresponding non-significant trend for the simplified stimulus set, so it seems that there was always at least a trend toward redundancy gain, consistent with the findings of Melara & Marks (1990).

It is worth noting that performance in timbre-focus Baseline conditions was slightly slower than performance in pitch-focus Baseline conditions, which could explain why the redundancy gain effect emerged only for timbre-focus trials. One limitation of the current investigation concerns potential difficulty discriminating between the two simplified timbres. Although accuracy did not significantly differ between timbre-focus and pitch-focus conditions, response times were significantly longer on timbre-focus trials for both the simplified and the adjusted stimulus sets. This suggests that discriminating between timbres may have been more challenging than discriminating between pitches. Future work could investigate timbre discrimination thresholds in order to better equate discrimination difficulty between dimensions prior to collecting additional speeded classification data.

Since equating centroids neither reduced nor eliminated interference (contrary to hypotheses), that attribute cannot be responsible for the pitch-timbre interactions that were previously observed in Garner speeded classification. Fortunately, the simplified nature of the stimuli used in the present investigation only employed three attributes of timbre: spectral centroid, spectral envelope, and rise time. The fact that strong interference effects were observed once the centroid was removed suggests that either the spectral envelope or rise time must be driving those interactions. There are indications from the current data set that rise time was not responsible for pitch-timbre interactions. If participants were responding based on rise time differences, then there should have been systematically longer response times in the Baseline blocks of trials presenting just violin timbre relative those presenting just trumpet timbre (because violin had a 340*ms* longer onset than trumpet). If response times to the trumpet Baseline were substantially shorter than those to the violin Baseline, then the difference between the trumpet Baseline and the Orthogonal blocks of trials would have been larger than the difference between the violin Baseline and the Orthogonal blocks. No substantial difference was observed, suggesting that participants were most likely listening to the spectral envelope to distinguish between timbres. Because the spectral envelope is also partially rooted in frequency (like pitch), it is logical to predict that it is the attribute of timbre driving the perceptual interactions.

The spectral envelope and spectral centroid were treated as somewhat distinct dimensions of timbre in the current investigation, but it is relevant to note that any change in spectral envelope also usually impacts the centroid by changing the distribution of amplitudes for the harmonics. Thus, the spectral centroid can be characterized as a descriptor of the spectral envelope (Lembke & McAdams, 2015). The current results suggest that spectral envelope shape is responsible for the observed interactions between pitch and timbre. This explanation is at odds with initial hypotheses that were based upon previous suggestions about the role of the spectral centroid in pitch-timbre interactions proposed by Becker and Hall (2014).

Although this may appear to be a critical discrepancy, there may be a way to account for both sets of results. If the spectral envelopes between two different tones are not distinct, then the centroid may necessarily make a larger contribution to timbre perception. Even though Becker and Hall (2014) concluded that the spectral centroid influenced pitch judgments more than the spectral envelope for one of their experimental conditions, the two tones being compared were both based upon a violin spectral envelope shape. Therefore it is possible that the spectral envelope may always be responsible for pitch-timbre interactions, but that the centroid can still be observed to produce a large effect whenever it is a primary attribute that differs across the stimulus set. Future investigations could test this hypothesis for speeded classification by using a timbre discrimination task where the spectral envelope is held constant, but the centroid is manipulated by adjusting the spectral slope, similar to the manner in which Li and Pastore (1995) altered the slope.

Relationships with Musical Training

A secondary goal of the current investigation was to evaluate how musical training influenced pitch-timbre interactions. Although it was hypothesized that individuals with higher levels of musical training would have reduced interference effects for all conditions, it appears that the benefits were restricted to the pitch-focus condition for the adjusted stimulus set. There were significant negative correlations between several musical training-related variables and the magnitudes of the interference effects for both accuracy and response times, but because those training variables were strongly related to each other, the remainder of this discussion will focus on total years of musical training in general. Higher levels of musical training were associated with reduced interference effects, as evidenced by smaller accuracy and response time differences between the Orthogonal and Baseline conditions for the adjusted stimulus set. This finding starkly contrasts with the original hypothesis that the performance of nonmusicians would be the most improved with the centroid-adjusted stimuli. The presence of an interference effect even for musically-trained participants adds to the literature suggesting that pitch-timbre interactions occur regardless of level of musical training (Russo & Thompson, 2005; Vurma et al., 2010).

Similar to Pitt (1994), individuals from the current investigation who had less musical training provided indications of processing asymmetry. In Pitt (1994) nonmusicians exhibited stronger interference effects in the pitch-focus condition relative to the timbre-focus condition, suggesting that task-irrelevant timbral variation influenced pitch judgments more strongly than irrelevant pitch variation influenced timbre judgments. In the current study, individuals with less musical training only exhibited significantly larger interference effects in the pitch-focus condition with adjusted stimuli. This suggests that performance differences due to variation in musical training only emerged once stimuli were equated to have the same spectral centroid. Such a finding is contrary to expected outcomes, since the Garner task with the simplified stimulus set was expected to replicate the findings of Pitt (1994).

Although the stimuli used in both studies were similar in timbre, there were some slight differences in pitch and large differences in duration. The two tones used by Pitt (1994) had fundamental frequencies of 294Hz and 417 Hz, whereas the tones used in the current study had fundamental frequencies of 313 Hz and 440 Hz. These different fundamentals were chosen in part because 440 Hz is a stable tuning note commonly used for many instruments, including violin and trumpet. Even though the frequencies were not identical, the distance between them in Hertz was roughly similar, so the impact of pitch should have likewise been similar. The tones in Pitt (1994) had a duration of $250ms$, whereas the tones in the current investigation had a duration of $1000ms$. It is possible that the longer stimulus exposure aided the performance of individuals with less musical training since the additional time could compensate for perceptual differences, thus leading to similar performance across participants with the simplified stimuli. However, once centroid differences were removed, individuals with more musical training may have been able to take advantage of the removal of that source of variation, leading to faster response times for the pitch-focus condition. This advantage may not have been observed in the timbre-focus condition because the additional timbral variation may have actually been helpful in that condition, since the goal was to classify stimuli on the basis of timbre differences and additional variation would further distinguish the two timbres.

Some researchers have suggested that musical training can influence listening style such that highly-trained professionals tend to be "analytic" listeners, and nonmusicians tend to be "synthetic" listeners (e.g., Seither-Preisler et al., 2007). Seither-Preisler et al. (2007) demonstrated that analytic listeners were able to classify pitch change based on a missing F0, suggesting an ability to separate the different components of a tone (e.g. fundamental vs. harmonics). Synthetic listeners relied on the overtone spectra, suggesting a more holistic listening style. Assuming that musical training is positively correlated with a more precise listening style (analytic), it is possible that individuals with more musical training in the current study needed more time to process the simplified stimuli than those with less musical training. The presence of centroid differences in the simplified stimulus set contributed to increased complexity of the acoustic signal, which likely increased processing time necessary to classify pitch since participants with more musical training would have to filter out an additional timbral attribute in order to accurately process pitch. However, once centroid differences were removed, there was less variation to filter out of the signal, perhaps leading to faster pitch classification for musicians.

Measuring Musical Training

Another goal of the current investigation was to develop a new measure of musical training with an emphasis on including continuous items and capturing a more diverse range of musical experiences. To establish the potential necessity of such a scale, two psychometrically-validated musical sophistication scales (OMSI and Gold-MSI) were administered in addition to an exploratory version of the new scale (MUTE). Even though measures of musical sophistication are frequently used by researchers as a fast and easy way to distinguish musicians from nonmusicians, the total score (i.e., the probability of being classified as "more musically sophisticated") from the OMSI was not strongly related to any of the interference effects observed in the current study. Nevertheless, musical training was shown to be related to performance in the pitch-focus condition for the adjusted stimulus set. Although it could be argued that this could reflect a true construct difference between musical training and musical sophistication, it is more likely that the regression weights used to calculate the total score of the OMSI can create misleading results. Self-classification as a musician was weighted much more heavily than actual experience. In fact, Years of Private Lessons, Years of Regular Practice, and Current Practice Amount actually had negative regression coefficients, indicating that participation in such activities contributes to a lower probability of being classified as "more musically sophisticated." The participant with the second-lowest OMSI score in the current sample actually had 10 years of musical training.

To further demonstrate how the OMSI might misclassify participants, two artificial response profiles were created: one with a large amount of experience, and one with a small amount of experience. The "musical" response profile was a 22-year-old with 10 years of private lessons and regular practice who at the time of data collection practiced more than two hours per day, had taken more than three courses for music nonmajors, had a composition performed for a regional audience, had attended more than 12 musical performances in the past year, and self-identified as a semiprofessional musician. By all accounts, this profile reflects musical sophistication. However, it would only have a 2% chance of being classified as "more musically sophisticated" according to the OMSI. The "nonmusical" response profile was an 18-year-old with no private lessons and six years of regular practice who was not actively practicing, had taken one music appreciation course, had no composition experience, had attended one musical performance in the past year, and self-identified as a serious amateur. Even though this profile reflects disengagement with musical activities, it would have a 65% chance of being classified as "more musically sophisticated" according to the OMSI. Even if the self-classification changed from serious amateur to nonmusician, it would still have a 27% chance of being classified as "more musically sophisticated," which is substantially higher than the "musical" response profile.

Due to the potential of the OMSI giving misleading results, it does not seem to be an adequate proxy for musical training in its current form. Because backwards elimination logistic regression was used to select items to be retained in the final scale and assign regression weights, the OMSI is likely only a good measure of musical sophistication within the original sample used to develop the scale. The backwards elimination procedure capitalizes on chance variation within the sample. At the beginning of the procedure, all potential predictor variables are entered into the regression equation. For each variable, the program calculates how much additional variance (R^2) would be explained if that variable had been added to the equation last. If increment is statistically significant, the variable is retained for the final model. If the increment is not significant, then the variable is removed and the contributions of all remaining variables are recalculated. Often these decisions are based on very small numerical differences that could result from simple sampling error. Models developed using this method do not easily generalize to new samples, and it is likely that the items retained in the final version of the scale are not actually good predictors of musical sophistication. The OMSI has been used in a variety of studies where differences due to musical experience might be expected, but some have found it incapable of accurately discriminating between participants (e.g., see Dean, Bailes, & Schubert, 2011; Ladinig & Schellenberg, 2012).

Out of the three primary musical training predictor variables examined (OMSI, Gold-MSI, Total Years of Training), the total score on the Musical Training subscale of the Gold-MSI had the strongest relationship with the response time interference effect for the pitch-focus condition using the adjusted stimulus set. It surprisingly was not significantly related to the accuracy interference effect for that same condition, although the relationship was still negative. Although it is difficult to determine the specific cause of this discrepancy, it could be related to response patterns on the scale itself. As expected, the negatively-phrased items (e.g. “I would not consider myself a musician”) were confusing to some participants. It can be difficult to disagree with such statements, and at least two individuals who were actively involved in music performance responded “agree” to that survey item. If a participant failed to interpret the negatively-phrased statement properly, then they would most likely select the opposite response, thereby contributing to a misleading total score.

Additionally, the size of each response category and intervals between categories were not equal for the Gold-MSI. Some categories consisted of half a year of experience, and others consisted of three years of experience. One participant in the current sample circled the space between two categories to indicate number of years of formal training. Regardless of these slight issues, the Gold-MSI did significantly relate to performance on the speeded classification task, is easy to score, and has an intuitive, easily-interpretable total score where higher values indicate higher levels of musical training.

A major drawback of the Gold-MSI is that it does not provide any indication of which kinds of activities that typically occur during musical training may contribute to task performance differences. This limitation motivated the development of a new instrument that can capture different facets of training such as private instruction, group instruction, formal coursework, etc. A number of items from the MUTE were significantly related to performance on the speeded classification task: Total Years of Training, Years of Private Lessons, Years in Musical Ensembles, Start Age, Recency of Training, Number of Formal Music Classes, and Number of Instruments. However, all of these variables were also strongly related to each other, and thus were not capable of explaining unique variance in the size of the interference effect.

It is important to note that the multiple regression analyses to assess the contribution of individual independent variables were underpowered in the current investigation. For a medium effect size with four predictor variables, it is recommended to have approximately 110 participants for an adequately-powered analysis. The analyses in the current study only had 48 participants, and may have been unable to detect even strong effects. In fact, post hoc power analyses conducted using G*Power v.3.1.9.2 revealed that the power was approximately .60, indicating that the model only had a 60% chance of detecting a significant effect if one truly existed in the sample. Additionally, the relationship between musical training and size of the interference effect was fairly weak in the current investigation, so different facets of musical training may not have been as influential in this particular study. In order to determine whether all musical training variables explain the same variance, or whether some describe unique variance, future investigations should employ the MUTE with other tasks that have been known to show differential performance between musicians and nonmusicians.

Despite the lack of power for the multiple regression analyses, the present investigation revealed that many of the musical training variables were highly correlated both with each other and with the interference effects. These intercorrelations suggest that a single factor may underlie many aspects of musical training. Future research with a larger sample size would allow for an exploratory factor analysis to investigate the factor structure underlying the MUTE. Factor analysis would then permit a total score to be computed from the items on the MUTE, thus maximizing explanatory power.

Although significant relationships were observed between total years of musical training and the magnitudes of some interference effects, those relationships were fairly weak, and visual inspection of the residuals plots revealed evidence of heteroscedasticity. One of the assumptions underlying general linear model analyses using Ordinary Least Squares estimation is homogeneity of variance (homoscedasticity). In other words, the variance of the residuals around the predicted scores should be constant across all levels of predicted scores. Violations of this assumption can bias standard errors, which may in turn bias inferences made from statistical tests. A likely cause of this heteroscedasticity could be the fact that a normal distribution was used to model musical training as a continuous variable, when it may have ideally been modeled as a count variable using a Poisson distribution. Count variables are typically positively-skewed, with many observations at the low end of the scale and few observations at the high end of the scale. This is typical of musical training variables when participants are recruited without regard to musicianship. Many individuals tend to have no experience whatsoever, many have low-moderate levels of experience, and few have high levels of experience. Future work should consider modeling musical training using techniques more suitable for Poisson distributions rather than normal distributions.

Conclusion

For now, based upon the primary findings from the current investigation, it appears that pitch and timbre perceptually interact, regardless of level of musical training. Additionally, the spectral centroid is most likely not the attribute of timbre responsible for perceptual interactions of pitch and timbre, since interference effects remained even after all stimuli had been equated to have the same spectral centroid. Rather, results suggest that spectral envelope shape may be responsible for pitch-timbre interactions. However, when differences in spectral envelope shape are minimal, then centroid differences may play a more important role in perceptual interactions of pitch and timbre. Future research exploring these possibilities and further evaluating the MUTE could help delineate which aspects of timbre are most influential in pitch-timbre interactions in addition to more fully understanding how musical training might reduce such interactions.

Appendix

Musical Training and Experience 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

References

Ableton AG (2015). Ableton Live Suite 9.6.2 [Computer software].

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Thousand Oaks, CA: Sage.

American Standards Association. (1973). *American standard acoustical terminology.* New York, NY: American Standards Association.

Bandalos, D. L. (2017). *Measurement theory and applications for the social sciences*. New York, NY: Guilford Publications.

Baumann, S., Meyer, M., & Jäncke, L. (2008). Enhancement of auditory-evoked potentials in musicians reflects an influence of expertise but not selective attention. *Journal of Cognitive Neuroscience, 20*(12), 2238-2249.

Beal, A. L. (1985). The skill of recognizing musical structures. *Memory & Cognition, 13*(5), 405-412.

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

Bidelman, G. M., Krishnan, A., & Gandour, J. T. (2011). Enhanced brainstem encoding predicts musicians' perceptual advantages with pitch. *European Journal of Neuroscience, 33*(3), 530-538.

Bidelman, G. M., Gandour, J. T., & Krishnan, A. (2011). Musicians and tone-language speakers share enhanced encoding but not perceptual benefits for musical pitch. *Brain and Cognition, 77*(1), 1-10.

Boyle, J. D., & Radocy, R. E. (1987). *Measurement and evaluation of musical experiences*. New York: Schirmer Books.

Carey, D., Rosen, S., Krishnan, S., Pearce, M. T., Shepherd, A., Aydelott, J., & Dick, F. (2015). Generality and specificity in the effects of musical expertise on perception and cognition. *Cognition, 137,* 81-105.

Chartrand, J., & Belin, P. (2006). Superior voice timbre processing in musicians. *Neuroscience Letters, 405,* 164-167.

Cohen, J. (1983). The cost of dichotomization. *Applied Psychological Measurement, 7*(3), 249-253.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences.* New York, NY: Routledge Academic.

Crummer, G. C., Walton, J. P., Wayman, J. W., Hantz, E. C., & Frisina, R. D. (1994). Neural processing of musical timbre by musicians, nonmusicians, and musicians possessing absolute pitch. *Journal of the Acoustical Society of America, 95*(5), 2720-2727.

Daly, H. R., & Hall, M. D. (2016). *Statistical considerations for evaluating the influence of musical training.* Talk presented at the Auditory Perception, Cognition, and Action Meeting (APCAM), Boston, MA.

Dean, R. T., Bailes, F., & Schubert, E. (2011). Acoustic intensity causes perceived changes in arousal levels in music: An experimental investigation. *PLoS ONE, 6*(4): e18591. doi:10.1371/journal.pone.0018591

Demany, L., & Semal, C. (1993). Pitch versus brightness of timbre: Detecting combined shifts in fundamental and formant frequency. *Music Perception: An Interdisciplinary Journal, 11*(1), 1-13.

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

Ehrlé, N., & Samson, S. (2005). Auditory discrimination of anisochrony: Influence of the tempo and musical backgrounds of listeners. *Brain and Cognition, 58*, 133-147.

Fine, P. A., & Moore, B. C. J. (1993). Frequency analysis and musical ability. *Music Perception: An Interdisciplinary Journal, 11*(1), 39-53.

Garner, W. R. (1974). *The processing of information and structure.* Potomac, MD: Erlbaum.

Gordon, E. (1965). The musical aptitude profile: A new and unique musical aptitude test battery. *Bulletin of the Council for Research in Music Education, 6*, 12-16.

Gordon, E. E. (1982). *Intermediate measures of music audiation.* Chicago, IL: G.I.A. Publications.

Habibi, A., Wirantana, V., & Starr, A. (2013). Cortical activity during perception of musical pitch: Comparing musicians and nonmusicians. *Music Perception, 30*(5), 463-479.

Houtsma, A. J. M. (1997). Pitch and timbre: Definition, meaning and use. *Journal of New Music Research, 26*(2), 104-115.

Humphreys, L. G. (1978). Doing research the hard way: Substituting analysis of variance for a problem in correlational analysis. *Journal of Educational Psychology, 70*(6), 873-876.

Hunter, J. E., & Schmidt, F. L. (1990). Dichotomization of continuous variables: The implications for meta-analysis. *Journal of Applied Psychology, 75*(3), 334-349.

Itoh, K., Okumiya-Kanke, Y., Nakayama, Y., Kwee, I. L., & Nakada, T. (2012). Effects of musical training on the early auditory cortical representation of pitch transitions as indexed by change-N1. *European Journal of Neuroscience, 36*, 3580-3592.

Karni, A., Meyer, G., Rey-Hipolito, C., Jezzard, P., Adams, M. M., Turner, R., & Ungerleider, L. G. (1998). The acquisition of skilled motor performance: Fast and slow experience-driven changes in primary motor cortex. *Proceedings of the National Academy of Sciences, 95*, 861-868.

Krumhansl, C. L., & Iverson, P. (1992). Perceptual interactions between musical pitch and timbre. *Journal of Experimental Psychology: Human Perception and Performance*, *18*(3), 739-751.

Ladinig, O., & Schellenberg, E. G. (2012). Liking unfamiliar music: Effects of felt emotion and individual differences. *Psychology of Aesthetics, Creativity, and the Arts, 6*(2), 146-154.

Law, L. N. C., & Zentner, M. (2012). Assessing musical abilities objectively: Construction and validation of the Profile of Music Perception Skills. *PLoS ONE, 7*(12): e52508. doi:10.1371/journal.pone.0052508

Lembke, S., & McAdams, S. (2015). The role of spectral-envelope characteristics in perceptual blending of wind-instrument sounds. *Acta Acustica United with Acustica, 101*, 1039-1051.

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.

Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization.* Oxford: Oxford University Press.

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods, 7*(1), 19-40.

Marques, C., Moreno, S., Castro, S. L., & Besson, M. (2007). Musicians detect pitch violation in a foreign language better than nonmusicians: Behavioral and electrophysiological evidence. *Journal of Cognitive Neuroscience, 19*(9), 1453-1463.

Maxwell, S. E., & Delaney, H. D. (1993). Bivariate median splits and spurious statistical significance. *Psychological Bulletin, 113*(1), 181-190.

Mazur, J. E., & Hastie, R. (1978). Learning as accumulation: A reexamination of the learning curve. *Psychological Bulletin, 85*(6), 1256-1274.

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,* 177-192.

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

Micheyl, C., Delhommeau, K., Perrot, X., & Oxenham, A. J. (2006). Influence of musical and psychoacoustical training on pitch discrimination. *Hearing Research, 219,* 36-47.

Müllensiefen D., Gingras B., Stewart L., & Musil J. (2011). The Goldsmiths Musical Sophistication Index (Gold-MSI): Technical Report and Documentation v0.9 [Technical report]. London, UK: Goldsmiths, University of London.

Müllensiefen D., Gingras B., Musil, J., & Stewart, L. (2014). The musicality of non-musicians: An index for assessing musical sophistication in the general population. *PLoS ONE 9*(2): e89642. doi:10.1371/journal.pone.0089642.

Ollen, J. E. (2006). *A criterion-related validity test of selected indicators of musical sophistication using expert ratings* (Unpublished doctoral dissertation). The Ohio State University, Columbus, OH.

Pantev, C., Roberts, L. E., Schulz, M., Engelien, A., & Ross, B. (2001). Timbre-specific enhancement of auditory cortical representations in musicians. *NeuroReport, 12*(22), 1-6.

Pitt, M. A. (1994). Perception of pitch and timbre by musically trained and untrained listeners. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(5), 976-986.

Pitt, M. A., & Crowder, R. G. (1992). The role of spectral and dynamic cues in imagery for musical timbre. *Journal of Experimental Psychology: Human Perception and Performance, 18*(3), 728-738.

Platt, J. R., & Racine, R. J. (1985). Effect of frequency, timbre, experience, and feedback on musical tuning skills. *Perception and Psychophysics, 38*(6), 543-553.

Preisler, A. (1993). The influence of spectral composition of complex tones and of musical experience on the perceptibility of virtual pitch. *Perception and Psychophysics, 54*(5), 589-603.

Psychology Software Tools, Inc. [E-Prime 2.0] (2012). Retrieved from http://www.pstnet.com.

Repp, B. H. (1994). The tritone paradox and the pitch range of the speaking voice: A dubious connection. *Music Perception: An Interdisciplinary Journal, 12*(2), 227-255.

Repp, B. H. (1997). Spectral envelope and context effects in the tritone paradox. *Perception, 26,* 645-665.

Russo, F. A., & Thompson, W. F. (2005). An interval size illusion: The influence of timbre on the perceived size of melodic intervals. *Perception and Psychophysics, 67*(4), 559-568.

Schellenberg, E. G., & Moreno, S. (2009). Music lessons, pitch processing, and g. *Psychology of Music, 38*(2), 209-221.

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.

Seashore, C. E. (1919). *The psychology of musical talent*. Boston, MA: Silver, Burdett and Company.

Seashore, C. E. (1938). *Psychology of music.* New York, NY: McGraw-Hill.

Seashore, C. E., Saetveit, J. G., & Lewis, D. (1960). *The Seashore measures of musical talents.* New York, NY: Psychological Corporation.

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*(3), 743-751.

Semal, C., & Demany, L. (1991). Dissociation of pitch from timbre in auditory short-term memory. *Journal of the Acoustical Society of America, 89*(5), 2404-2410.

Shepard, R. N. (1964). Circularity in judgments of relative pitch. *Journal of the Acoustical Society of America, 36*(12), 2346-2353.

Singh, P. G., & Hirsh, I. J. (1992). Influence of spectral locus and F0 changes on the pitch and timbre of complex tones. *Journal of the Acoustical Society of America, 92*(5), 2650-2661.

Spiegel, M. F., & Watson, C. S. (1984). Performance on frequency-discrimination tasks by musicians and nonmusicians. *Journal of the Acoustical Society of America, 76*(6), 1690-1695.

Strait, D. L., Kraus, N., Parbery-Clark, A., & Ashley, R. (2010). Musical experience shapes top-down auditory mechanisms: Evidence from masking and auditory attention performance. *Hearing Research, 261*, 22-29.

Vurma, A., Raju, M., & Kuuda, A. (2010). Does timbre affect pitch?: Estimations by musicians and non-musicians. *Psychology of Music, 39*(3), 291-306.

Warrier, C. M., & Zatorre, R. J. (2002). Influence of tonal context and timbral variation on perception of pitch. *Perception and Psychophysics, 64*(2), 198-207.

Wayland, R., Herrera, E., & Kaan, E. (2010). Effects of musical experience and training on pitch contour perception. *Journal of Phonetics, 38*, 654-662.

Zarate, J. M., Ritson, C. R., & Poeppel, D. (2013). The effect of instrumental timbre on interval discrimination. *PLoS ONE, 8*(9): e75410. doi:10.1371/journal.pone.0075410

Table 1

Commonly-Used Indicators of Musicianship

	Years	Type	Occupation	Practice	Age	Involvement	Skills
# of studies (out of 38)	27	15	15	7	7	7	1
% of studies	71%	39%	39%	18%	18%	18%	3%

Note. Years = total years of musical training; Type = type of musical training (e.g. formal lessons, group training, etc.); Occupation = music-related profession or music student; Practice = amount of daily musical practice; Age = age at which musical training commenced; Involvement = actively involved in a musical activity; Skills = measured musical skills.

Table 2

Experimental Blocks of Garner Speeded Classification

Pitch-Focus: Classify Low/High Pitch				
Fixed		*Correlated*		*Orthogonal*
Low/High V	Low/High T	Low V/High T	High V/Low T	All 4 stimuli
Timbre-Focus: Classify Violin/Trumpet Timbre				
Fixed		*Correlated*		*Orthogonal*
Low V/T	High V/T	Low V/High T	High V/Low T	All 4 stimuli

Note. Low = D#$_4$; High = A$_4$; V = Violin; T = Trumpet. Two sets of these blocks will be presented in the experiment: one for the simplified stimuli, and one for the centroid-adjusted stimuli.

Table 3

Source Table for 2x4 Repeated-Measures ANOVA: Accuracy for the Simplified Stimuli

Source	*df*	*F*	*p*	η_p^2
Focus Condition	1	2.920	.094	.058
Classification Condition	1.353	46.610	<.001	.508
Focus*Classification	1.370	3.295	.061	.066

Table 4

Source Table for 2x4 Repeated-Measures ANOVA: Accuracy for the Adjusted Stimuli

Source	*df*	*F*	*p*	η_p^2
Focus Condition	1	2.251	.140	.046
Classification Condition	1.183	45.501	<.001	.492
Focus*Classification	1.230	1.356	.256	.028

Table 5

Source Table for 2x4 Repeated-Measures ANOVA: Response Times for the Simplified Stimuli

Source	*df*	*F*	*p*	η_p^2
Focus Condition	1	22.957	<.001	.328
Classification Condition	1.659	128.632	<.001	.732
Focus*Classification	1.718	3.095	.058	.062

Table 6

Source Table for 2x4 Repeated-Measures ANOVA: Response Times for the Adjusted Stimuli

Source	*df*	*F*	*p*	η_p^2
Focus Condition	1	34.188	<.001	.421
Classification Condition	1.642	233.200	<.001	.832
Focus*Classification	1.438	6.867	.005	.127

Table 7

Correlations Between Musical Training Predictor Variables for Planned Multiple Regression Analysis and the Interference Effects

	Interference (RT)	Interference (Accuracy)	Total Years	Lessons	Ensemble	Start Age
Interference (RT)	-	.342*	-.318*	-.269	-.352*	.367*
Interference (Accuracy)		-	-.321*	-.298*	-.204	.355*
Total Years			-	.697**	.777**	-.855**
Lessons				-	.403**	-.589**
Ensemble					-	-.652**

Note. **p* < .05. ***p* < .01.

Table 8

Correlations Between Items from the MUTE and Interference Effects

	Interference (RT)	Interference (Accuracy)	Start Age	Years Uninvolved	# of Courses	# of Instruments
Interference (RT)	-	.342*	.367*	.406**	-.300*	-.372**
Interference (Accuracy)		-	.355*	.340*	-.176	-.249
Start Age			-	.787**	-.403**	-.691**
Years Uninvolved				-	-.432**	-.754**
# of Courses					-	.630**

Note. *p < .05. **p < .01.

Table 9

Correlations Between Interference Effects and Musical Sophistication Scales

	Interference (RT)	Interference (Accuracy)	OMSI	Gold-MSI	Total Years
Interference (RT)	-	.342*	-.195	-.332*	-.318*
Interference (Accuracy)		-	-.051	-.181	-.321*
OMSI			-	.569**	.468**
Gold-MSI				-	.859**

Note. *$p < .05$. **$p < .01$.

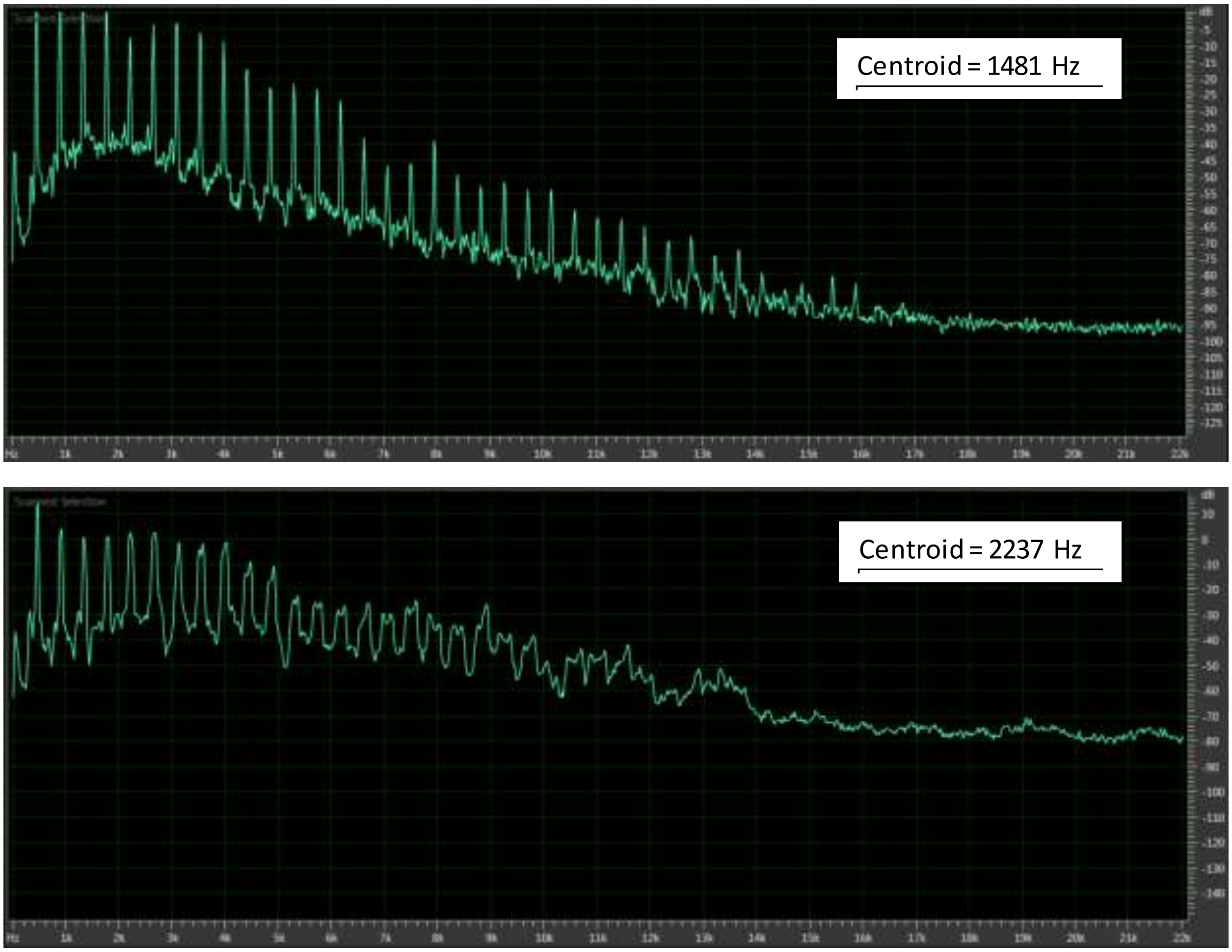

Figure 1. Spectral slices of the trumpet (top panel) and violin (bottom panel).

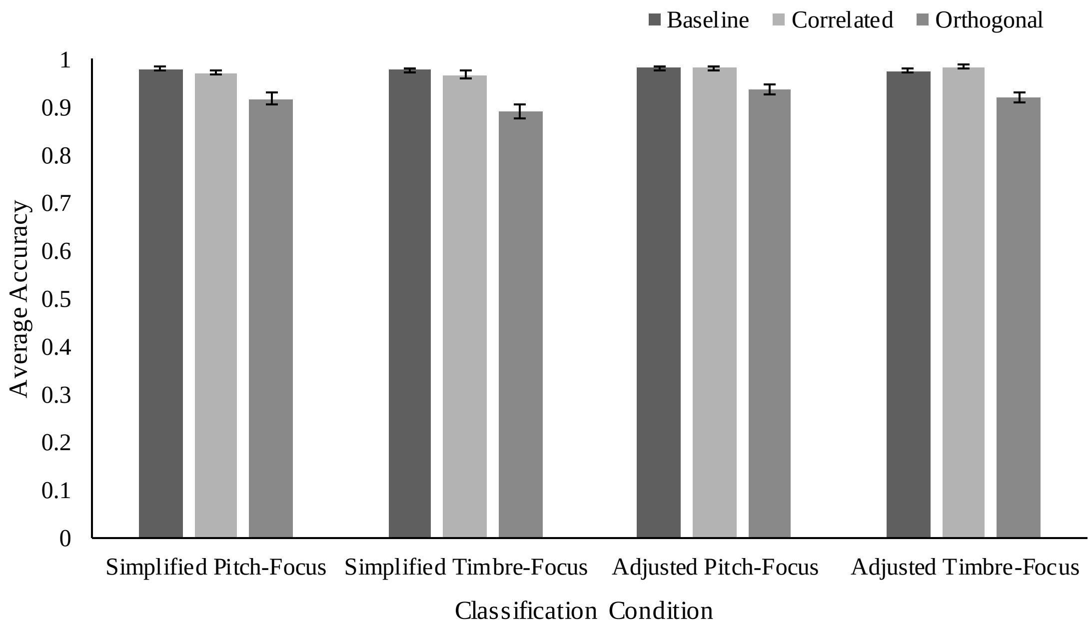

Figure 2. Average accuracy and corresponding standard errors for each classification condition in the pitch-focus and timbre-focus conditions for each stimulus set.

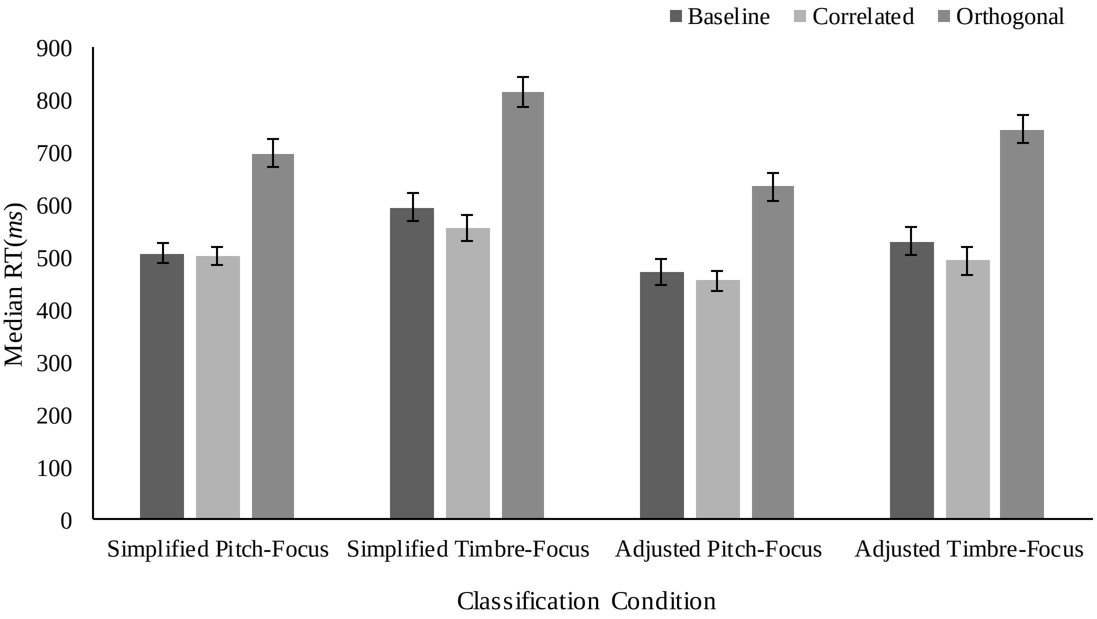

Figure 3. Average median response times and corresponding standard errors for each classification condition in the pitch-focus and timbre-focus conditions for the simplified and adjusted stimulus sets.