Perceived Affect of Musical Instrument Sounds

by

Chelsea Douglas

Music Technology Area, Department of Music Research
Schulich School of Music
McGill University, Montreal
April 2015

A thesis submitted to McGill University in partial fulfillment of the requirements
of the degree of Master of Arts

Abstract

Listeners use a combination of global and local factors to assess emotional intent in music. As little as 250 milliseconds of a musical excerpt can hold enough information to perceive an emotional tone, which emphasizes the importance of examining how local acoustic factors, such as timbre, contribute to emotion perception. This thesis investigates the role of different aspects of orchestral timbre on perceived affect, preference, and familiarity. Specifically, register, attack, and playing technique, as well as spectral and temporal descriptors of sound are examined. The stimuli consisted of recorded instrumental samples with a constant duration, pitch class, and dynamic. Samples were chosen from the entire range of each instrument with different attacks (weak, normal or strong). Participants completed six ratings per sound on bipolar scales. Four scales measured perceived emotion (valence: negative/positive; valence: displeasure/pleasure; energy arousal: tired/awake; tension arousal: tense/relaxed) and two measured participants' preference for (dislike/like) and familiarity with (unfamiliar/familiar) the sounds. Sounds in higher registers were perceived as more awake. String and percussion sounds tended to have more positive valence and higher energy ratings than brass and woodwind sounds. Timbre is a multidimensional property comprising numerous acoustic features. With the purpose of studying which acoustic descriptors correlate with perceived affect, two toolboxes (MIRToolbox, Timbre Toolbox) were verified and compared, and 23 reliable measures were used in a principal components analysis. Five principal components (PC) explained 84.7 % of the variance in these descriptors. The first four PCs were highly correlated with spectral measures (PC1: centroid, decrease, rolloff, skewness), temporal descriptors (PC2), spectral flatness and crest (PC3), and spectral variation (PC4). PC5 was only moderately correlated with spectral flatness and crest. Perceived energy ratings were correlated with PC1 and PC3. Perceived valence ratings were moderately correlated with PC1-PC4. Perceived tension ratings were correlated with PC1, PC2, and PC4. Preference

ratings were only correlated with PC2. Despite the short duration of the stimuli, participants assigned ratings representative of a three-dimensional affect scale to the various timbres. Timbre should therefore be considered as a prominent vehicle of affective expression in music.

Résumé

Les auditeurs se servent d'une combinaison de facteurs globaux et locaux pour évaluer les intentions émotives dans la musique. Même 250 millisecondes d'un extrait musical possèdent assez d'informations pour percevoir un ton émotionnel, ce qui souligne l'importance étudier comment les facteurs locaux comme le timbre contribuent à la perception de l'émotion. Cette thèse étudie le rôle d'aspects divers du timbre orchestral sur l'affect perçu, la préférence et la familiarité. Plus spécifiquement nous étudions le registre, l'attaque et le mode de jeu, ainsi que les descripteurs spectraux et temporels des sons. Les stimuli étaient des échantillons enregistrés d'instruments de musique ayant une durée, classe de hauteur et dynamique constantes. Les échantillons ont été choisis à travers toute la gamme de chaque instrument avec des attaques différentes (douce, normale, mordante). Les participants ont effectué six évaluations par son sur des échelles bipolaires. Quatre échelles mesuraient l'émotion perçue (valence: négative/positive; valence: plaisant/déplaisant; éveil d'énergie: fatigué/éveillé; éveil de tension: tendu/détendu) et deux mesuraient la préférence pour (me plait/ne me plait pas) et familiarité avec (familier/pas familier) les sons. Les auditeurs ont jugé les timbres dans les registres aigus plus éveillés. Les timbres des cordes et des percussions avaient tendance vers une valence plus positive et une énergie accrue par rapport aux timbres des cuivres et des bois. Le timbre est une propriété multidimensionnelle qui comprend plusieurs caractéristiques acoustiques. Afin d'étudier quels descripteurs acoustiques sont corrélés avec l'affect perçu, deux boîtes à outils informatiques (MIRToolbox, Timbre Toolbox) ont été vérifiées et comparées. 23 mesures fiables ont été utilisées dans une analyse par composantes principales. Cinq composantes principales (CP) ont expliqué 84.7 % de la variance de ces descripteurs. Les quatre premiers CP étaient corrélées avec des mesures spectrales (CP1: centroïde, décroissance, rolloff, asymétrie), des descripteurs temporels (CP2), la platitude et la crête spectrales (CP3) et la variation spectrale (CP4). CP5 n'était que

modérément corrélée avec la platitude et la crête spectrales. L'énergie perçue étaient corrélée avec les CP1 et CP3. La valence perçue était modérément corrélée avec les CP1-CP4. La tension perçue était corrélée avec les CP1, CP2 et CP4. La préférence n'était corrélée qu'avec la CP2. Malgré la durée courte des stimuli, les évaluations des timbres par les participants se déployaient sur les trois dimensions de l'espace affectif. Le timbre devrait donc être considéré comme un véhicule proéminent de l'expression affective en musique.

Acknowledgments

First and foremost, I would like to thank my wonderful advisor, Stephen McAdams. He constantly proved to be an extraordinarily intelligent and truly kind person, and it was an honour to work with him. Additionally, I would like to thank all of the members of the Music Perception & Cognition Lab (MPCL) for consistently creating a fantastic and supportive environment. Especially Bennett Smith for always providing sound technical advice and an endless supply of charged batteries, Cecilia Taher for her immense willingness to help at all times and incredible (often chocolate-based) support, David Sears and Yinan Cao for our Linear Mixed Model (nerd) brainstorms, Kai Siedenburg for sharing Matlab tricks and teaching me the meaning of timbre through song/vocalizations, and all of the other past and present members, specifically: Jason Noble, Meghan Goodchild, Sven-Amin Lembke, Eddy Kazazis, Christopher Wood, and Song Hui Chon. I'd like to extend my gratitude to all of my best friends in Montreal and elsewhere, specifically Henry Gras, Philippe Dubost, Gabriel Cartier, Marine Hercouet, Gwen Decat-Beltrami, Hannah Robertson, Paula Georgieva and Austin Causey, for their unlimited love and laughter. Lastly, I'd like to thank my family, especially Patricia Rogers and Lindsay Douglas, for their infinite love and support.

Author contributions

As the author of this thesis, I was responsible for every step of designing and running the experiment, statistical analyses and interpretations of the results, and the writing and editing of the thesis. As the thesis advisor, Stephen McAdams, provided laboratory equipments and guidance regarding experimental design, data analysis and interpretation of the results. Additionally, I was financially supported by a grant from NSERC (RGPIN 312774-10) awarded to S. McAdams, as well as his Canada Research Chair.

Contents

List of Figures

List of Tables

List of Acronyms

ANOVA	Analysis of variance
ERB	Equivalent Rectangular Bandwidth
LMM	Linear mixed-effects model
ML	Maximum likelihood
PC	Principal component
PCA	Principal component analysis
SD	Standard deviation
SPL	Sound pressure level

Chapter 1

Introduction

The relationship between music and emotion has become a widely studied topic. The existence of such a relationship is undeniable and is supported by multiple studies that have revealed that for most people the predominant motivation for listening to and engaging in music is its emotional impact (Sloboda & O'Neill, 2001; Krumhansl, 2002; Juslin & Laukka, 2004). Although there is an increasing amount of research on music and emotion, it remains difficult to draw decisive conclusions about how musical factors contribute to emotion in a piece. In addition to global factors of a piece, it is likely that acoustic factors of a sound can relay affective information as well. This thesis aims to examine timbre, and the acoustic components that contribute to the timbre of a sound, in relation to perceived affect of musical instruments.

This chapter presents a brief overview of music and emotion research, comparing felt and perceived emotion research before honing in on perceived emotion research involving the musical attribute: timbre. An experiment aimed to further contribute to research involving timbre and perceived affect of sound is outlined in Chapter 2. A linear mixed model analysis is used to examine the role of instrumental family and pitch on participants' perceived affect judgments of short instrumental sounds. In Chapter 3, timbre is broken down into spectral, temporal, and

spectrotemporal components. A review of two Matlab toolboxes used to calculate acoustic components is presented prior to a principal components analysis used to further relate timbre to the participants' affect ratings.

1.1 Music and Emotion

Various dimensional models of affect have been applied to music and emotion research including a two-dimensional model and a three-dimensional model. The two-dimensional circumplex model represents affect as a function of two core dimensions: valence and arousal (Russell, 1980; Russell, Weiss & Mendelsohn, 1989). This model can be further labeled with pleasure, excitement, arousal, distress, displeasure, depression, sleepiness, and relaxation falling around the circle at intervals of 45 degrees (Russell, 1980). A three-dimensional model of affect measures emotion as a function of valence, tension arousal, and energy arousal (Schimmack & Grob, 2000). This model likely provides a more complete representation of affect than the two-dimensional model because tension arousal and energy arousal were shown to be two distinct measures of activation that should not be collapsed into a single measure of arousal (Schimmack & Reisenzein, 2002). Schubert (1999) completed a series of experiments applying a dimensional model of affect to music research and found the dimensional model to be a valid and reliable measure for research involving music and emotion. Furthermore, the three-dimensional model of affect was recently applied to multiple perceived emotion and music studies that will be discussed later in this chapter (Ilie & Thompson, 2006; Eerola, Ferrer & Alluri, 2012).

In a recent review of music and emotion studies, Eerola & Vuoskoski (2013) discuss numerous questions that fuel music and emotion research, emphasizing a core question: "how does music evoke emotion in listeners?" Breaking this question down, they reveal two important sub-questions: (1) which emotions can be induced by music and (2) how are emotions conveyed

by music? These two questions highlight the concept of the locus of emotion (Evans & Schubert, 2008), which distinguishes felt emotions (internal locus) from perceived emotions (external locus).

1.1.1 Felt versus Perceived Emotion

According to Sloboda (1991) "The ability to judge mood is logically and empirically separable from the ability to feel emotion in response to music. It is quite possible to judge a piece of music to represent extreme grief, yet be totally unmoved by it." Although the relationship between felt emotions and perceived emotions has yet to be clearly defined (see Gabrielsson, 2002 for suggested models), many studies have empirically demonstrated that listeners' felt emotions can differ from the emotions they perceive in a piece (Gatewood, 1927; Martindale & Moore, 1989; Schubert, 2007; Evans & Schubert, 2008; Zentner, Grandjean & Scherer, 2008). Factors such as musical training (Kawakami, Furukawa, Katahira, Kamiyama & Okanoya, 2013) and empathy (Egermann & McAdams, 2013) may contribute to the ability to distinguish felt emotions from perceived emotions; whereas increased musical training may promote differentiation of felt and perceived emotions, increased empathy may reduce that ability.

Furthermore, a distinction between felt and perceived emotion could be related to listeners' enjoyment of sad music. A few studies have examined listeners' felt emotions and preferences of negatively valenced music. Early research observed a relationship between pleasure and intensity of perceived emotion, where extremely happy or sad excerpts were preferred to neutrally valenced music (Gatewood, 1927). Schubert (1996) suggests that negative emotions in music can be pleasurable for listeners because they perceive a negative emotion, but know there is no risk of harm or danger. He further suggests an associative network model based on the theory that emotional arousal (or the induction of emotion) is generally pleasurable

(Berlyne, 1971). Schubert hypothesizes that when a listener perceives a negative emotion in music, a negative "emotion node" is activated, and this activation would typically activate the displeasure center. However, the listener's awareness that no harm or danger exists disables the displeasure center and instead enables the pleasure center, allowing the listener to experience a positive emotion. Additionally, Vuoskoski & Eerola (2011) suggest that listeners do feel sadness while listening to a negative valenced excerpt, but also simultaneously experience more positive emotions such as nostalgia, peacefulness, and wonder.

1.1.2 Felt-Emotion Responses

It is almost unanimously reported in questionnaires, experiments, and interviews that listeners, at some point, experience emotions while listening to music (reviewed in Juslin & Laukka, 2004; Juslin & Västfjäll, 2008). However, there are few empirical results that explain when, why, or how emotions are induced in listeners. Such experiments can be difficult to conduct because individual differences regarding felt musical emotions are immense. Juslin & Laukka (2004) found that participants reported feeling strong emotions while listening to music between 5% and 100% of the time. Although it is clear that nearly all listeners are capable of experiencing a strong emotional response, the aforementioned result emphasizes the difficulty in forming generalizable conclusions about the musical factors that induce a strong emotional response in listeners.

When investigating felt emotions, researchers typically use self-report questionnaires or physiological responses (Eerola & Vuoskoski, 2013). Sloboda (1991) used a combination of these methods by surveying 83 listeners on physical responses experienced as emotional reactions to music and concluded that most participants experienced a variety of physical responses when listening to self-selected music and some responses could be linked to the musical structure: for example, tears can be induced by melodic appoggiaturas. Limitations of

this study include the inability to further generalize these results beyond the experimental sample because 67 of the 83 participants were musicians and the majority of the participants chose to report on classical music.

In addition to examining felt emotion via physical responses, brain activation research has demonstrated that music listening can induce activation in regions of the brain that are associated with emotional response (Blood & Zatorre, 2001; Menon & Levitin, 2005; Koelsch, Fritz, Müller & Friederici, 2006).

In general, research regarding felt emotion in the literature is highly confirmatory but lacks conclusive results about which musical elements contribute to emotion reactions that can be generalized across individual listeners and music genres.

1.1.3 Perceived-Emotion Responses

Emotional perception refers to a listener recognizing an expressed emotion, but does not necessitate feeling an emotion (Juslin & Västfjäll, 2008). When examining expressed emotion in an entire piece, pitch combination and order were largely focused on, which led to the understanding that form, mode, melody, and harmony play an important role in emotion perception of music (reviewed in Gabrielsson & Lindström, 2010). Furthermore, tempo (Juslin, 1997; Gagnon & Peretz, 2003) and rests (Margulis, 2007) are additional structural elements that can influence emotion expression.

However, listeners' judgments of perceived emotion are not solely based on structural elements of music. By altering factors such as amplitude, pitch level, pitch contour, tempo, envelope, and filtering, in synthesized tone sequences, Scherer & Oshinsky (1977) concluded that over two-thirds of the variance in listener's perceived emotion ratings could be explained by the manipulation of the acoustic cues. Further research supports the notion that, in addition to structural musical features, finer acoustic features, such as articulation, dynamics, loudness,

spectrum, and attack are factors listeners consider when making emotion judgments (Juslin & Laukka, 2004; Gabrielsson & Lindström, 2010). The latter two factors are components that contribute to the timbre of a sound (McAdams, Winsberg, Donnadieu, De Soete & Krimphoff, 1995).

1.2 Timbre and Perceived Emotion

Timbre is a multidimensional acoustic attribute that is composed of spectral, temporal, and spectrotemporal dimensions (McAdams et al., 1995). The term timbre is often used as an explanation of why a listener can discriminate different sounds of the same pitch and loudness, and therefore contributes to source identity (McAdams, 1993). Additionally, the timbre of acoustic instruments varies with pitch register, i.e., a given instrument played in a low register can have a drastically different timbre when played in a high register (Risset & Wessel, 1999; Marozeau, de Cheveigné, McAdams & Winsberg, 2003). Although timbre has not yet been studied extensively in relation to emotion perception in music, based on related literature, it is likely a contributing factor.

The notion that perceived emotion can be judged by non-structural acoustic features is supported by listeners' ability to make emotional judgments in an extremely short amount of time, and therefore, with limited acoustic information. In certain cases, as little as 250 milliseconds of a musical excerpt holds enough information to perceive an emotional tone (Peretz, Gagnon & Bouchard, 1998; Filipic, Tillmann & Bigand, 2010) and even a single note provides listeners with enough cues to form an emotional judgment (Bigand, Vieillard, Madurell, Marozeau & Dacquet, 2005). Furthermore, musical expertise did not have an impact on musical recognition and emotional judgments based on minimal acoustic information (Filipic et al., 2010). Peretz et al. (1998) completed experiments with 32 short classical excerpts, averaging

about 15 seconds in length. The excerpts were specifically chosen to convey a happy or sad tone. After participants completed both emotion classification tasks (happy or sad) and emotion judgments on a 10-point scale (sad to happy), Peretz et al. (1998) concluded that emotional judgments are not only immediate, but also highly consistent across listeners, including a participant with bilateral cerebral damage. The ability to recognize emotion in such a short stimulus emphasizes the importance of examining how individual acoustic factors, such as timbre, contribute to emotion perception in music.

Furthermore, acoustic factors such as loudness and timbre are focal points in the comparison of musical expression of emotion to vocal expression of emotion. Juslin & Laukka (2003) suggest musical expression and vocal expression use the same emotion-specific patterns of acoustic factors, and Patel & Peretz (1997) note that timbre, unlike tonality, is a factor of music that shares the same neural resources as speech. Ilie & Thompson (2006) examined acoustic cues in music and speech in relation to a three-dimensional model of affect and found that in both domains increased loudness led to higher pleasantness, energy, and tension ratings and increased speed resulted in higher energy ratings. Pitch height behaved oppositely across the domains: high pitch height in speech was rated as more pleasant, whereas high pitch height in music was rated as less pleasant. Relatedly, Patel, Gibson, Ratner, Besson & Holcomb (1998) examined pitch contour instead of pitch height and suggested neural processing similarities of pitch contour for both speech and music. Coutinho & Dibben (2013) had listeners rate perceived emotion in music and speech clips and found that the emotion ratings of both music and speech could be predicted by psychoacoustic factors, namely loudness, tempo, contour, and spectral descriptors.

Although listeners and performers have identified timbre as a musical factor contributing to emotion perception (Gabrielsson, 2001; Juslin, 2001; Holmes, 2011), little empirical research has been done to identify how timbre systematically expresses affect. Huron, Anderson &

Shanahan (2014) asked one group of participants to judge acoustic properties of 44 Western instruments and another group to judge those instruments' ability to express sadness. All judgments were made based on participants' familiarity and knowledge of the instruments. Using the acoustic property judgments, such as the darkness of timbre, from the first group of participants as predictors for the sadness judgments of the second group of participants. Huron et al. (2014) concluded that acoustic properties of the instruments, the ability to make small pitch movements, the ability to play low pitches, and the ability to play quietly predicted sadness judgments.

In addition to emotional perception from the listener's perspective, from a performance perspective, timbre is a critical device of emotional expression. A performer has greater control over acoustic factors compared to compositional or structural elements of a piece (Juslin & Laukka, 2003). This is an important element of emotion perception because it has been shown that listeners typically perceive a performer's intended emotional expression in a piece (Gabrielsson & Juslin, 1996). Holmes (2011) emphasized timbre as a fundamental element of affective communication between performers and listeners when examining a subjective account from a classical guitar performer. As a musical factor that contributes to both performers' expression of emotion and listeners' perception of emotion, timbre should be isolated and closely examined in regards to perceived affect in order to better understand how it contributes to the emotional identity of a piece.

A few studies have attempted to empirically examine timbre's contribution to perceived affect. Hailstone, Omar, Henley, Frost, Kenward & Warren (2009) studied timbre as a main factor contributing to emotion perception in music. Listeners were presented with melodies that possessed a strong emotional intent and labeled the melodies with an emotion through a forced-choice paradigm. There was a significant interaction between instrument and emotion judgment. However, the experiment only looked at four timbres and introduced the timbre

variable to participants via various novel melodies, which could possibly confound the emotional expression of the timbre alone.

Eerola et al. (2012) recently examined 110 instrument timbres in relation to a two-dimensional affect model. Affect ratings were consistent among participants and were related to spectral, temporal and spectrotemporal acoustic features. The results support the hypothesis that timbre systematically contributes to affect perception and will be further examined in the following chapter.

Chapter 2

Experiment

Many important vocabulary terms, including emotion and affect, are only loosely defined throughout the literature. Furthermore, these terms are often used interchangeably. Juslin & Västfjäll (2008) define affect as "an umbrella term that covers all evaluative—or valenced—states such as emotion, mood and preference" and emotions as "relatively intense affective responses that usually involve a number of sub-components—subjective feeling, physiological arousal, expression, action tendency, and regulations—which are more or less synchronized." Although much of the literature review used the term perceived emotion in accordance with the cited works, for the purposes of our experiment we will refer to perceived valence and arousal ratings with the term perceived affect rating and the preference rating as a felt affect rating.

In a review of literature regarding emotions in musical expression, Juslin & Laukka (2004) name timbre as a musical feature correlated with perceived, discrete emotions. They note that, in general, bright timbres are associated with happiness, dull timbres with sadness, sharp timbres with anger, and soft timbres with both fear and tenderness. These ideas have recently been empirically studied by isolating instrument timbres from a musical context and relating the timbres to a dimensional affect model.

Eerola et al. (2012) examined individual timbres in relation to a two-dimensional affect model of valence and energy-arousal. The stimuli consisted of 110 real instrument sound samples. Pitch and duration were kept constant at D#4 (311 Hz fundamental frequency) and one second, respectively, across all stimuli, and loudness was equalized. Participants listened to the individually presented stimuli and performed affect and preference ratings. The scales were labeled on their ends and included valence (pleasant/unpleasant), energy arousal (awake/tired), tension arousal (tense/relaxed), and preference (like/dislike). The three-dimensional model of affect (Schimmack & Grob, 2000) used to collect ratings was reduced to a two-dimensional model for analysis purposes due to a highly collinear relationship between the energy-arousal and tension-arousal dimensions. The tension-arousal dimension was eliminated from further analysis in their study.

Furthermore, the valence and preference ratings in their study had a nearly perfect correlation, $r(108) = .97, p < .001$. It is important to note here that the bipolar emotional valence scale was labeled from unpleasant to pleasant. This provides a clear methodological difference from Bigand et al. (2005) where emotional valence and pleasantness were rated on two separate scales. Bigand et al. (2005) stimuli were extremely short in duration (1 s), sometimes consisting of one single tone, and are comparable to the stimuli of Eerola et al. (2012), although not identical. The difference in definition of the scales may have contributed to a key difference in the findings because Bigand's group found that the emotional valence dimension was not correlated with pleasantness judgments, suggesting happy music is not necessarily identified with pleasant emotions or sad music with unpleasant emotions. These findings influence the prediction that perceived valence will not be completely correlated with preference ratings in the current experiment.

This experiment aimed to further contribute to and clarify research regarding the role of timbre in affect perception in music by showing that participants' judgments regarding perceived

affect vary systematically with timbral qualities of short instrument sounds. First we examined affect ratings in relation to broader variables such as pitch register and instrument family with a linear mixed model analysis. We expected to find different patterns in the tension-arousal ratings compared to the energy-arousal ratings, supporting a three-dimensional model of affect, instead of a two-dimensional model. Additionally, we expected to find a difference in the perceived valence ratings compared to the preference ratings, highlighting a difference between perceived measures and felt measures. Finally, we expected a significant interaction between pitch register and instrument family for each of the perceived affect ratings, showing perceived emotion ratings may not be the same for all instruments across pitch registers.

2.1 Method

The experiment examined perceived affect of instrument sounds. The experimental design isolated timbre as an independent variable, similar to Experiment 1 presented in Eerola et al. (2012), and allowed us to examine register, attack, and playing technique as factors contributing to timbre. Modifications, such as an added valence measure and increased range of pitch register, facilitated a comparison between emotional valence and preference scales as well as examining how changes in register contribute to changes in timbral components that influence arousal ratings.

2.1.1 Participants

Forty participants, 24 females, were between 18 and 35 years of age ($M = 23$, $SD = 4.4$). Twenty participants reported formal musical training ranging from 7 to 25 years of practice ($M = 16$, $SD = 5.3$), and 14 reported formal training with multiple instruments. The remaining 20 participants reported no musical training at a collegiate level and no more than one year of

formal music training during childhood.

2.1.2 Stimuli

One hundred and thirty seven recorded instrument sounds were chosen from the Vienna Symphonic Library (Vienna Symphonic Library GmbH., 2011). The sounds consisted of timbres of orchestral instruments from four instrument families: brass, woodwinds, strings, and percussion. Audio signals were sampled at 44.1 kHz with 16-bit amplitude resolution. The stimuli were edited to have a consistent duration of 500 ms with a raised-cosine ramp applied to fade them out over the final 50 ms. The attack of each sound was unaltered. The brass stimuli varied by an attack parameter, having a weak, normal or strong attack, as labelled by the VSL. The percussion stimuli varied by mallet material, using a felt, wood or metal mallet. Pitch was kept consistent at pitch class D#, and the dynamic level was forte, as labelled by the VSL. Samples were chosen from the entire range of the instruments, thus the stimuli ranged from D#1 to D#7 (A4 has a fundamental frequency of 440 Hz). Most instruments cannot successfully play from D#1 to D#7, so stimuli were only taken from appropriate and playable registers for each instrument. Although some instruments can play outside of that range, there were not enough samples to create useful, balanced groups. Furthermore, various techniques, such as flutter-tonguing for brass and woodwinds and vibrato and pizzicato for strings, were also included. A detailed list of the stimuli is provided in Appendix A.

2.1.3 Procedure

All participants passed a pure-tone audiometric test using a MAICO MA 39 (MAICO Diagnostic GmbH, Berlin, Germany) audiometer at octave-spaced frequencies from 125 Hz to 8 kHz (ISO, 2004) and were required to have thresholds at or below 20 dB HL in order to proceed to the experiment.

The interface was created in TouchOSC (Hexler.net, 2011) and consisted of a play button, six clearly labeled 9-point, analogical-categorical scales (Weber, 1991), and a next button. The next button was not activated until all six ratings were completed; pressing this button would reset the display to the original position and play the next sound. All 137 stimuli were presented in a randomized order for each participant and each sound could be played multiple times. Participants completed six ratings per sound on the 9-point, bipolar analogical-categorical scales. The first four ratings measured perceived emotion and reflected affect dimensions from the three-dimensional model of affect (Schimmack & Grob, 2000) with an additional measure of an overall negative to positive valence. The scales were labeled at the left and right ends with the following pairs: negative and positive, displeasure and pleasure, tired and awake, and tense and relaxed. The participants were also reminded that a rating of 5 would equate to a neutral rating. A full copy of the instructions is provided in Appendix B. These four scales were labeled in blue on the iPad interface. The last two ratings measured participants' preference and familiarity for each sound. These scales were labeled with the pairs: dislike - like and unfamiliar - familiar. These two scales provided a felt rating of personal preference and familiarity and were labeled in purple to differentiate them from the perceived affect ratings. An example of the interface is displayed in Fig. 2.1. The protocol was certified by the McGill Review Ethics Board (Certificate 67-0905) and all participants gave written informed consent prior to the experiment. Each participant completed the task within an hour and was compensated $10 CAD.

2.1.4 Apparatus

Participants completed the experiment individually inside an isolated, soundproofed room (Industrial Acoustics model 1203). The sound samples were played from a Macintosh G5 computer and amplified with a Grace Design m904 monitor system and heard over circumaural Senneiser HD280 Pro headphones (Sennheiser Electronic GmbH, Wedemark, Germany) set at 65

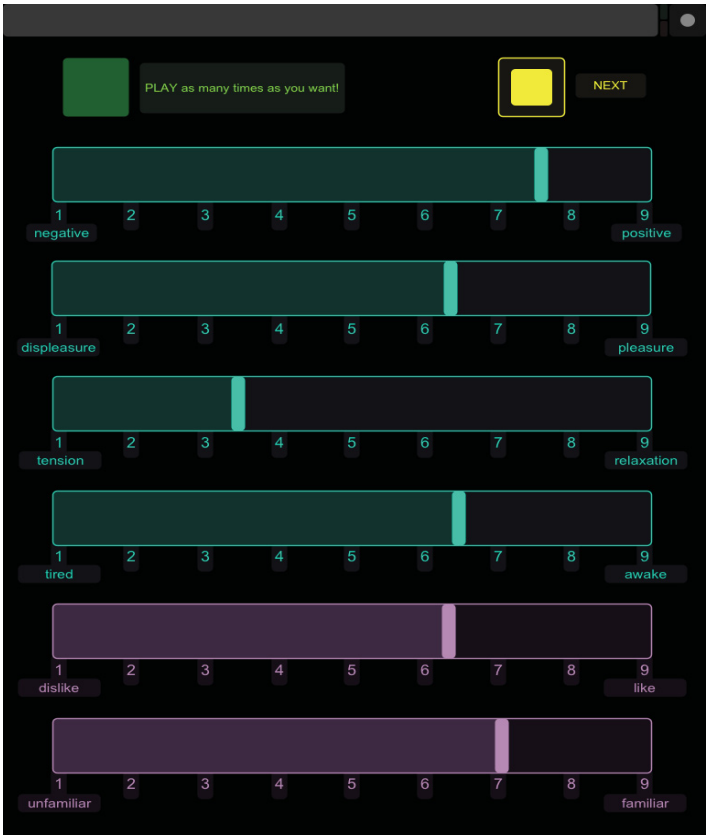

Fig. 2.1 Screenshot of Experimental Interface on iPad.

dB. The participants were not allowed to adjust the volume. Sound levels were measured with a Brüel and Kjær Type 2205 sound-level meter (A-weighting) connected to a Type 4152 artificial ear (Brüel and Kjær, Nærum, Denmark) to which the headphones were coupled. Stimuli ranged between 59.8 and 77.5 dB SPL ($M = 65.3, SD = 5.4$). The participants completed the experiment on an iPad interface (Apple Computer, Inc., Cupertino, CA). The iPad communicated via OpenSoundControl (Center for New Music and Audio Technologies, Berkley, CA) messages over a wifi network with a Max/MSP version 5.1.9 (Cycling '74, San Francisco, CA) patch run on a Macintosh G5 computer. The Max/MSP patch was designed to randomize and play the stimuli as well as record and output the ratings.

2.2 Control Experiment

A control experiment was completed after the original experiment to validate the original interface. The main purpose was to confirm, with a correlation analysis between the control ratings and the original ratings, that no bias resulted from the order and orientation of the rating scales, which remained in a fixed position for every trial and every participant in the original experiment.

2.2.1 Participants

Twenty participants, 12 females, were between 18 and 42 years of age ($M = 25$, $SD = 6.5$). Ten participants reported formal musical training ranging from 13 to 19 years of practice ($M = 16$, $SD = 2.5$), and seven of those reported formal training with multiple instruments. The remaining ten participants reported no musical training at a collegiate level and no more than a year of formal music training during childhood.

2.2.2 Stimuli

Forty stimuli were selected from the original 137 samples to create a group that was representative of the entire set. Therefore, the control stimuli were brass, woodwind, string, and percussion samples ranging from D#1 to D#7 with weak, normal, and strong attacks. The exact samples are marked with an asterisk in Appendix A.

2.2.3 Procedure

The instructions and procedure were identical to the original experiment, however participants were randomly given one of four different interfaces. The interfaces included the same “play” and “next” buttons as the original but the order of the six scales was changed as well

as the orientation (i.e., the end labels) of some of the scales. However, the blue "perceived" scales and the purple felt scales were always grouped together to avoid confusion between perceived and felt ratings. An example of one of the control interfaces is presented in Fig. 2.2.

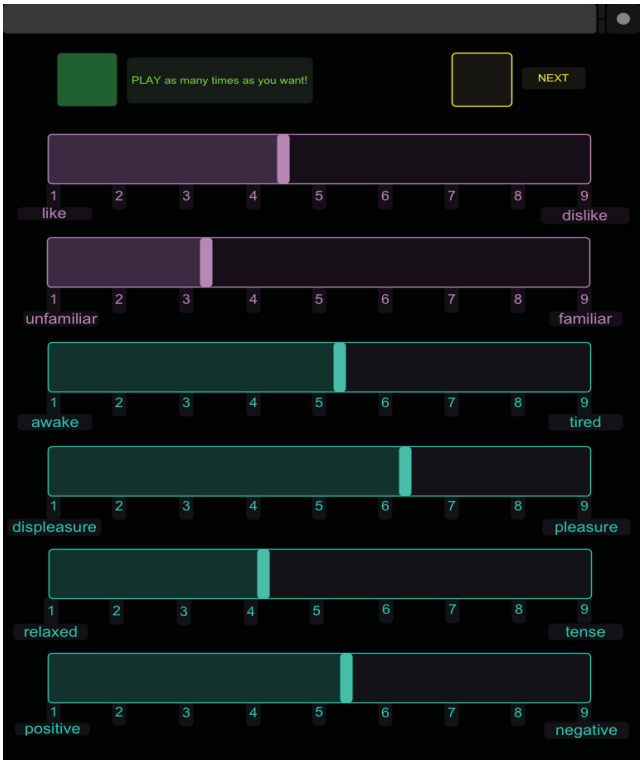

Fig. 2.2 Screenshot of Control Interface on iPad.

2.3 Data Analysis

Initial reliability measures and correlations of the scales were completed in SPSS v22.0 (IBM Corp., Armonk, NY). Further statistical analyses utilized a linear mixed model method (West, Welch & Galecki, 2006), which performs a regression-like analysis while controlling for random variance caused by differences in factors such as participant and stimuli. Because each participant rated all of the stimuli, the model included crossed random effects for participant and

item (Baayen, Davidson & Bates, 2008). Specifically, a maximal random effects structure was implemented due to the confirmatory hypothesis nature of the analyses and to reduce Type I errors, i.e., false positives (Barr, Levy, Scheepers & Tily, 2013). The basic equation (eq. 2.1), and a brief description of the terms as discussed in Baayen et al. (2008), are presented below.

$$y_{ij} = X_{ij}\beta + S_i s_i + W_j w_j + \varepsilon_{ij} \tag{2.1}$$

The term y_{ij} is a vector of the responses of subject i for item j. X_{ij}, an experimental design matrix, includes columns representing the factor contrasts. β is a vector of population coefficients. S_i and W_j represent the random effects structure for subject and item, respectively, and are copies of the X_{ij} matrix. The terms s_i and w_j are vectors composed of adjustments for each subject (i) and each item (j), respectively. Finally, the vector ε_{ij} is composed of error values for each subject-item pair. Analyses were completed with the software R v3.0.2 (www.r-project.org) using the lmer function from the lme4 package (Bates, Maechler, Bolker & Walker, 2014) and the Anova function from the Companion to Applied Regression (car) package (Fox & Weisberg, 2011).

2.4 Results

All scales had good internal consistency (Cronbach's α for 40 participants = .93 for positive/negative, .91 for pleasure/displeasure, .92 for relaxed/tense, .90 for awake/tired, .97 for like/dislike, and .99 for familiar/unfamiliar).

Ratings were averaged across participants as an initial analysis, so each of the 137 timbres had a measure for each of the six rating scales. This allowed for examination of correlations between the scales and also comparison to previous literature. When comparing the averaged

ratings, the valence scale, labeled negative and positive, and the valence scale, labeled displeasure and pleasure, had a Pearson's correlation of $r(135) = .97, p < .001$. This was expected as the latter scale was the measure used in Schimmack & Grob (2000), but both scales were included in the experiment to verify that displeasure to pleasure was an appropriate descriptor for musical stimuli. In the following analyses, the valence measure will only refer to the negative - positive scale.

The tension-arousal and energy-arousal ratings were significantly, but only weakly, correlated, Pearson's $r(135) = .30, p < .001$. Interestingly, there was a strong, positive correlation between preference and valence ratings, $r(135) = .72, p < .001$, and a strong negative correlation between preference and tension ratings, $r(135) = -.75, p < .001$. However, there was only a very weak correlation between preference and energy ratings, $r(135) = .18, p = .029$. Although valence and preference ratings and tension and preference ratings are linearly related, energy and preference ratings do not exhibit a similar linear relationship. Furthermore, there was a moderate, negative correlation between the valence ratings and the tension arousal ratings, $r(135) = -.45, p < .001$, and a strong positive correlation between the valence ratings and energy-arousal ratings, $r(135) = .67, p < .001$.

Table 2.1 displays the within-subject correlations of both the main and control experiments as well as the between-subjects correlations of the main experiment with the control experiment for the 40 timbres common to both studies. Although the main experiment included 137 timbres, the following correlations reflect only the 40 timbres from the main experiment that were also used in the control experiment. As all scales in the experiment were very strongly correlated with the designated control, $r(38) = .89 - .92$, $p < .001$, the original interface was confirmed to be valid and reliable. Further analysis is completed on data from the main experiment only.

A linear mixed model analysis was completed for each of the three perceived affect

Table 2.1 Pearson's Correlation Coefficients Among Ratings of Perceived Valence, Tension Arousal, Energy Arousal, Preference, and Familiarity for Both the Main and Control Experiments

	Main Valence	Main Tension Arousal	Main Energy Arousal	Main Preference	Main Familiarity	Control Valence	Control Tension Arousal	Control Energy Arousal	Control Preference
Main Tension Arousal	-.51**	1							
Main Energy Arousal	.67**	.24	1						
Main Preference	.70**	-.81**	.11	1					
Main Familiarity	.54**	-.40*	.27	.64**	1				
Control Valence	.89**	-.65**	.44**	.81**	.60**	1			
Control Tension Arousal	-.28	.89**	.47**	-.67**	-.31	-.51**	1		
Control Energy Arousal	.56**	.37*	.92**	.01	.18	.33*	.57**	1	
Control Preference	.57**	-.72**	.02	.89**	.67**	.80**	-.69**	-.06	1
Control Familiarity	.46**	-.37*	0.2	.57**	.90**	.56**	-0.31	.14	.65**

Notes. $N = 40$. $^{*}p < .05$ $^{**}p < .01$ $^{***}p < .001$.

ratings as well as the preference and familiarity ratings. Fixed factors examined in these models include instrument family and pitch register of the timbres and musical training of the participants. Attack and playing technique parameters were not included in order to simplify this model. Furthermore, those factors are not necessarily comparable across instrument family; for example, the brass technique of flutter-tonguing is not comparable to the string technique of vibrato. Therefore those factors were only included in the individual instrument family models, not the full models. Because all participants rated all timbres, a crossed random effects design was implemented and therefore, the maximal random effects structure included random intercepts for participant with random slopes for family and register, and random intercepts for the stimuli (timbres) with random slopes for training. An example of the model equation in R syntax is: Valence $\sim 1 +$ training $*$ family $*$ register $+ (1 +$ family $+$ register $|$ participant) + $(1 +$ training $|$ timbre).

Valence $\sim$ represents the dependent variable (valence ratings) that is predicted by the factors in the following equation. An interaction between fixed factors (i.e., training, family, and register) is notated with $*$, and followed by the random effects structure. The random effects structure is represented by $(1 +$ random slopes $|$ random effect). In this equation, there were two random effects: participants and timbres.

Type III F test results from the five models are displayed in Table 2.4. Musical training alone was only a significant predictor of familiarity ratings, although the interaction of training and register was a significant predictor for valence, tension-arousal and preference ratings. Family was a significant predictor for all ratings except energy-arousal ratings. Register alone, and the interaction between register and family both significantly predicted all of the perceived affect ratings but not the preference and familiarity ratings. Register was especially influential for energy-arousal ratings. The three-way interaction between training, family and register was significant for tension-arousal, energy-arousal and preference ratings.

Table 2.2 Linear Mixed Effects Model Type III Wald F Tests for Perceived Ratings of Valence, Tension Arousal, Energy Arousal, Preference, and Familiarity.

	df	F	p-value	*df*	F	p-value
	Valence (R^2 = .53)			Tension Arousal (R^2 = .43)		
Intercept	1, 124.4	145.63	< 0.001***	1, 121.70	121.96	< 0.001***
Training	1, 136.88	0.43	0.512	1, 132.04	1.56	0.214
Family	3, 120.66	7.3	< 0.001***	3, 121.11	4.04	0.009**
Register	6, 122.01	7.98	< 0.001***	6, 120.75	4.15	< 0.001***
Training * Family	3, 104.62	0.24	0.871	3, 95.27	1.38	0.253
Training * Register	6, 69.38	3.71	0.003**	6, 56.34	2.27	0.050*
Family * Register	16, 111.0	2.41	0.004**	16, 111.0	2.41	0.004**
Training * Family * Register	16, 111.0	1.04	0.417	16, 111.0	2.13	0.011*
	Energy Arousal (R^2 = .50)			Preference (R^2 = .51)		
Intercept	1, 124.02	381.45	< 0.001***	1, 135.08	120.15	< 0.001***
Training	1, 137.48	0.05	0.819	1, 96.02	1.97	0.164
Family	3, 118.78	1.67	0.178	3, 125.81	10.19	< 0.001***
Register	6, 112.83	13.27	< 0.001***	6, 120.9	1.65	0.139
Training * Family	3, 91.45	1.44	0.239	3, 94.77	1.1	0.353
Training * Register	6, 53.83	2.1	0.068	6, 58.39	3.89	0.002**
Family * Register	16, 111.0	3.09	< 0.001***	16, 111.0	1.44	0.135
Training * Family * Register	16, 111.0	2.3	0.006**	16, 111.0	1.99	0.02*
	Familiarity (R^2 = .59)					
Intercept	1, 149.78	112.47	< 0.001***			
Training	1, 70.46	5.89	0.018*			
Family	3, 131.8	6.33	< 0.001***			
Register	5, 120.36	0.82	0.54			
Training * Family	3, 81.11	2.09	0.108			
Training * Register	5, 69.51	0.58	0.716			
Family * Register	16, 111.0	1.85	0.039*			
Training * Family * Register	16, 111.0	1.62	0.084			

Notes. N = 5480. All predictors are sum-coded factor variables. The following random effects were included: a) random intercepts for Participant and Timbres, b) random slopes for Family and Register (within Participants) and Training (within Timbres).

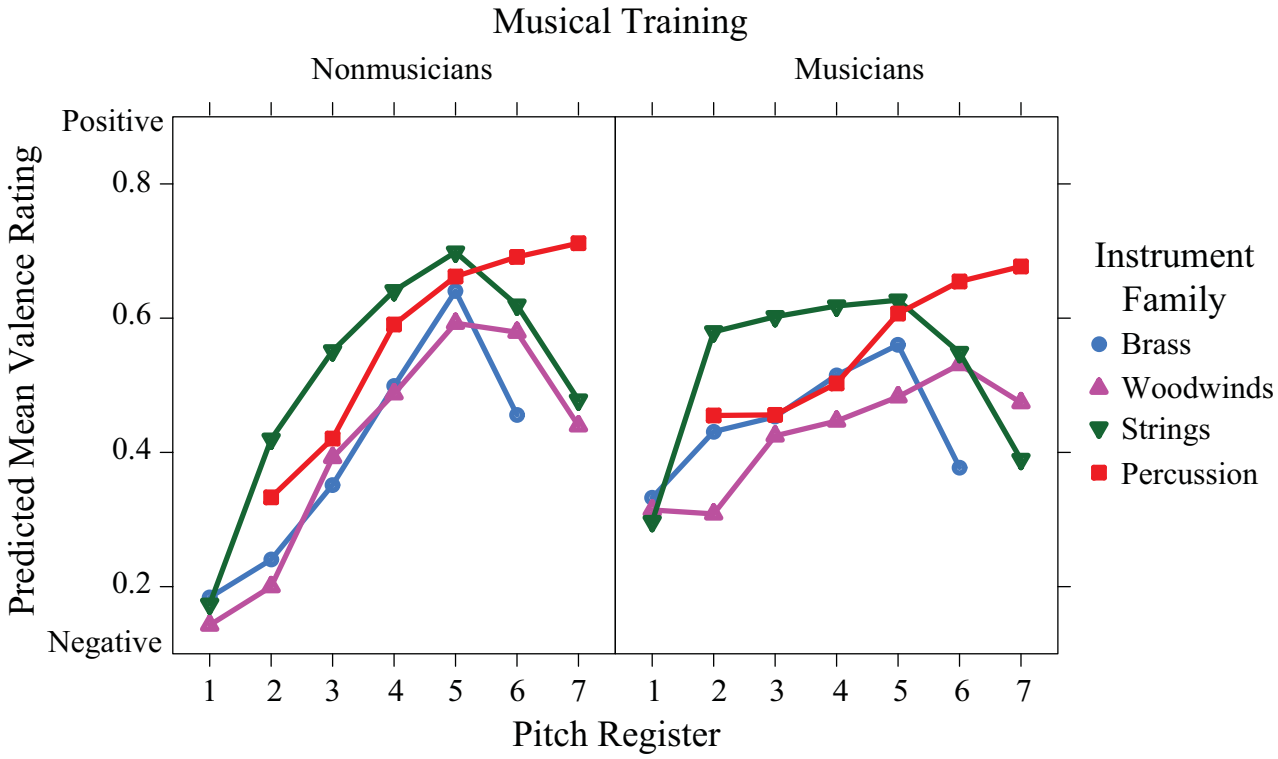

Fig. 2.3 Predicted Means of Perceived Valence Ratings Across Pitch Register for Each Instrument Family and Each Musical Training Group.

Figures 2.3-2.7 display the graphs for predicted means of each rating scale across register for each family and training group. For Valence ratings (Fig. 2.3), register was highly significant and presents an inverted U-shaped form with a peak around octave 5 or 6. The exception to this pattern is the percussion family for which valence continues to increase at higher registers, leading to a significant interaction between register and family. The effect of register is also affected by training as indicated by their significant interaction: although the inverted U-shape for strings, brasses and woodwinds is present for both groups, there is a greater effect of register overall for musicians whose peak is more pronounced. There is a general tendency for the families to be ordered in terms of decreasing valence: strings - percussion - brass - woodwinds, with the exceptions mentioned for percussion in the higher registers.

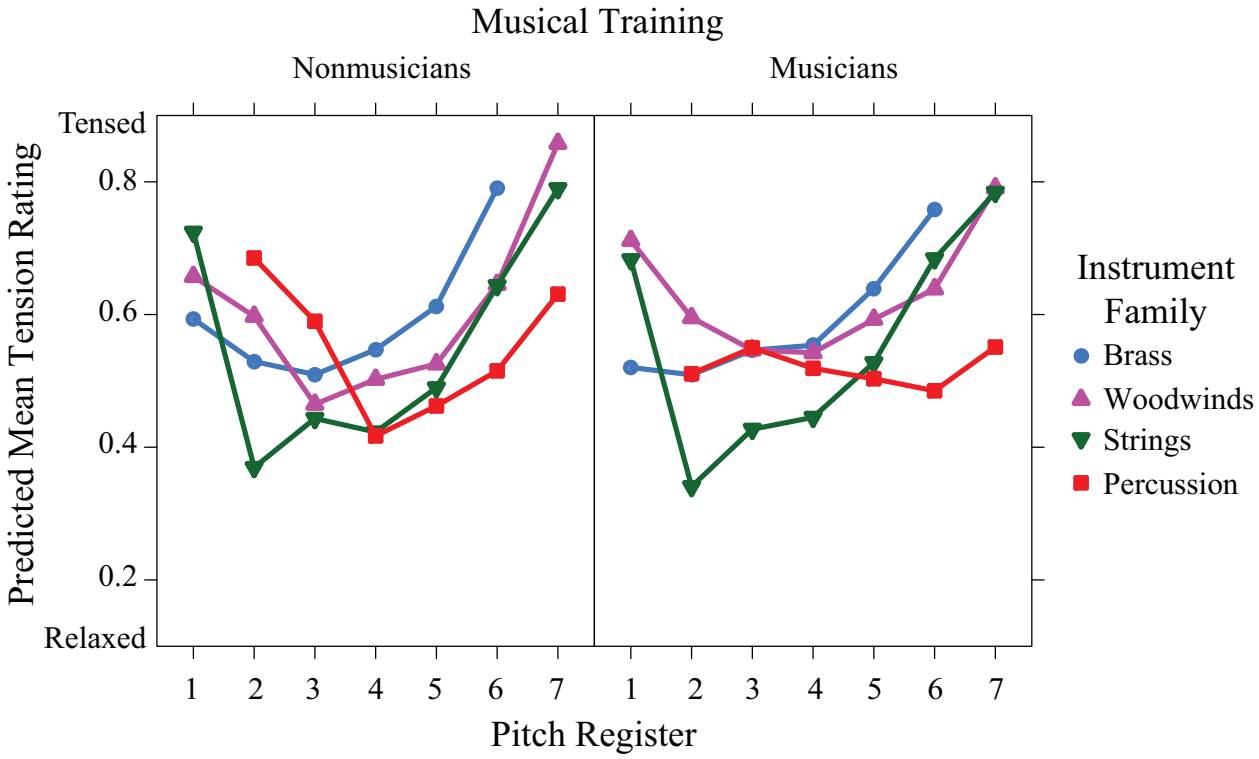

Fig. 2.4 Predicted Means of Perceived Tension-Arousal Ratings Across Pitch Register for Each Instrument Family and Training Group.

Tension-arousal ratings (Fig. 2.4) were highly significant for register and followed a U-shaped form, with most families peaking at the lowest and highest octaves. This trend was apparent for all families in the non-musician training group, but only in the woodwind and string families in the musician group. This result relates to the significant three-way interaction between training, family, and register. In the musician group, tension ratings for brass instruments remained neutral in the low and mid register and then increased in the fifth register and peaked at the sixth register, like the non-musician group's ratings. Furthermore, the musician group's tension ratings for the percussion family remained relatively neutral across all the registers.

Register was a highly significant predictor for energy-arousal ratings (Fig. 2.5), and a linear trend is visible across registers with lower registers perceived as more tired and high

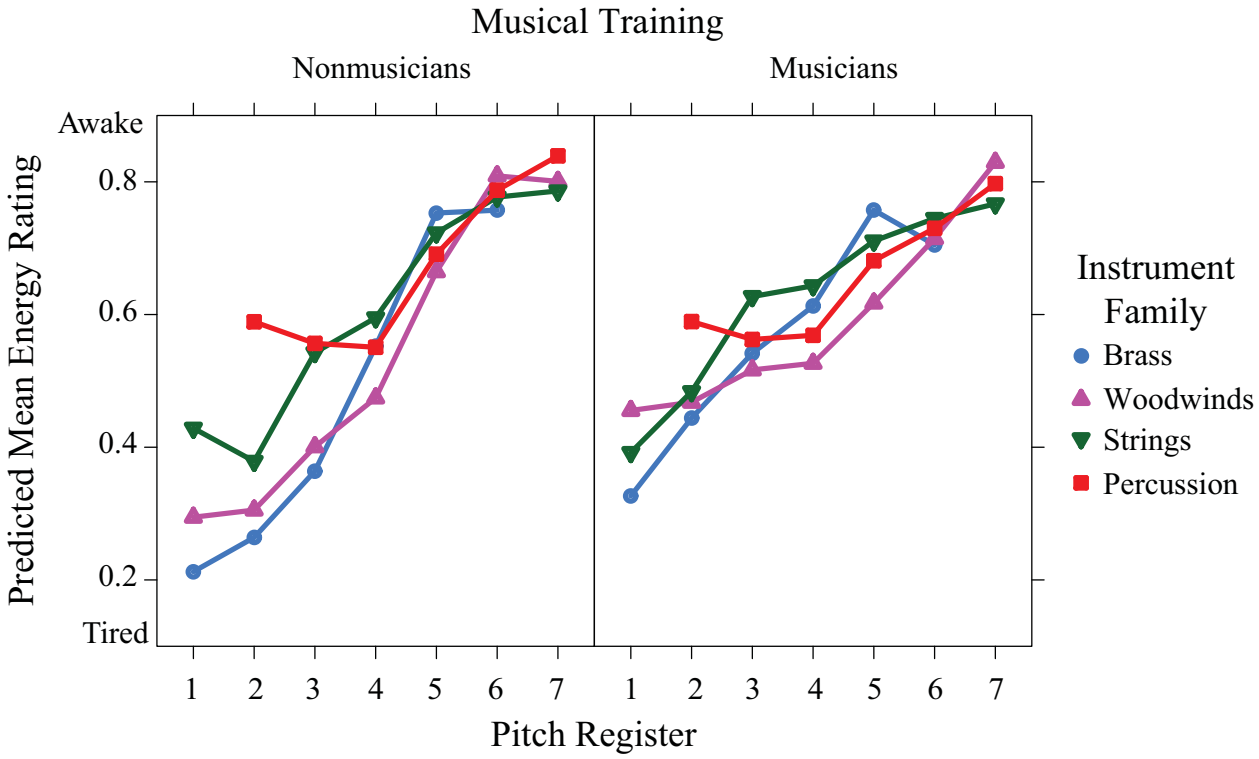

Fig. 2.5 Predicted Means of Perceived Energy-Arousal Ratings Across Pitch Register for Each Instrument Family and Training Group.

registers perceived as more awake. Additionally, the interaction between register and family was significant. This can be seen, specifically in the percussion family ratings in the second register (the lowest register for percussion timbres in this experiment), which were higher than the ratings of the other families in this register. Furthermore, the significant three-way interaction between training, family and register can be seen when comparing the variance of the energy ratings of the different families in the low registers: in the non-musician group, the families are more spread out in the first three registers, whereas the ratings of the musician group are more similar across families, even in the low registers.

Unified trends were slightly less apparent in the graphs of preference and familiarity ratings compared to the perceived affect ratings. Instrument family was a significant predictor of

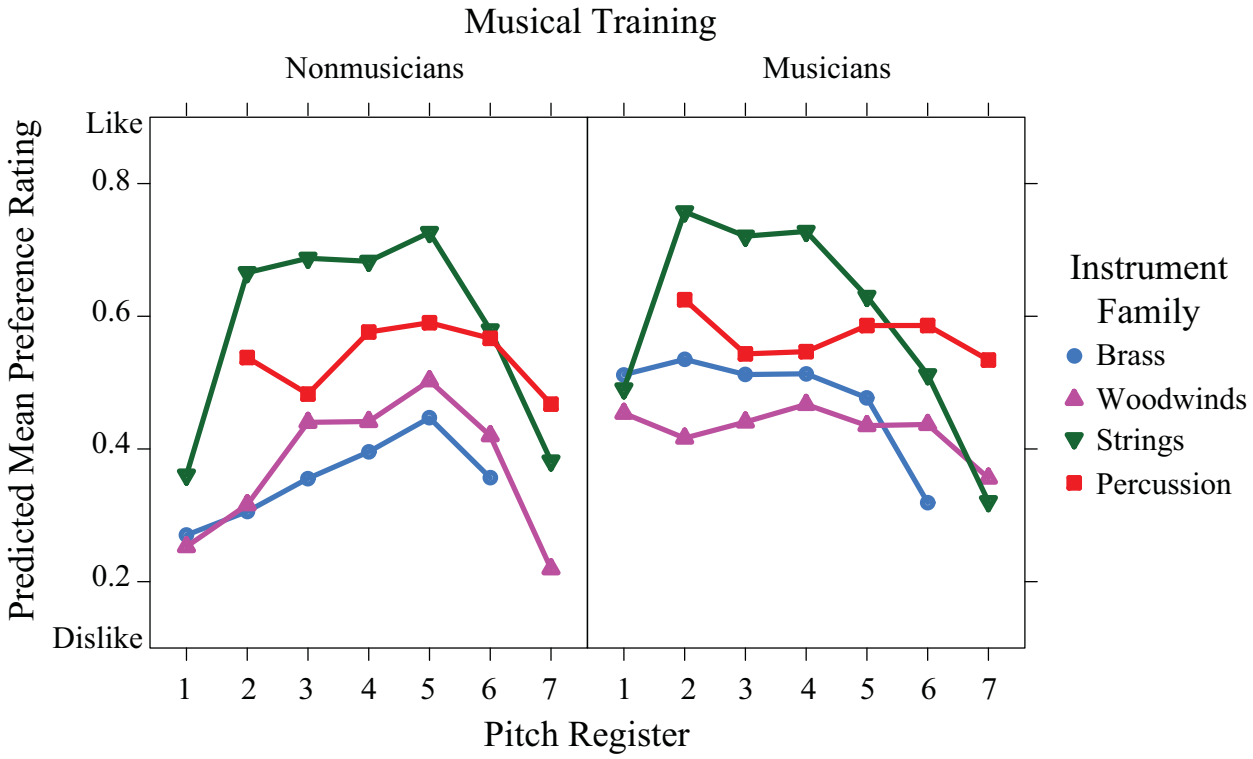

Fig. 2.6 Predicted Means of Preference Ratings Across Pitch Register for Each Instrument Family and Training Group.

preference ratings (Fig. 2.6), and mid-register string timbres were the most preferred by both musicians and non-musicians. In line with a significant three-way interaction between training, family, and register, the musicians' preference ratings for brass and woodwind timbres were relatively neutral for lower and mid-register timbres, then decreased for high-register timbres, whereas non-musicians preference ratings for brass and woodwind timbres were low for low-register and high-register timbres, but increased to a neutral rating. This pattern in the non-musicians' ratings is similar to that found in the perceived valence ratings.

Training and family were both individually significant predictors of familiarity ratings (Fig. 2.7). The familiarity ratings were generally higher for the musician group than the non-musician group. There was also a significant interaction between family and register,

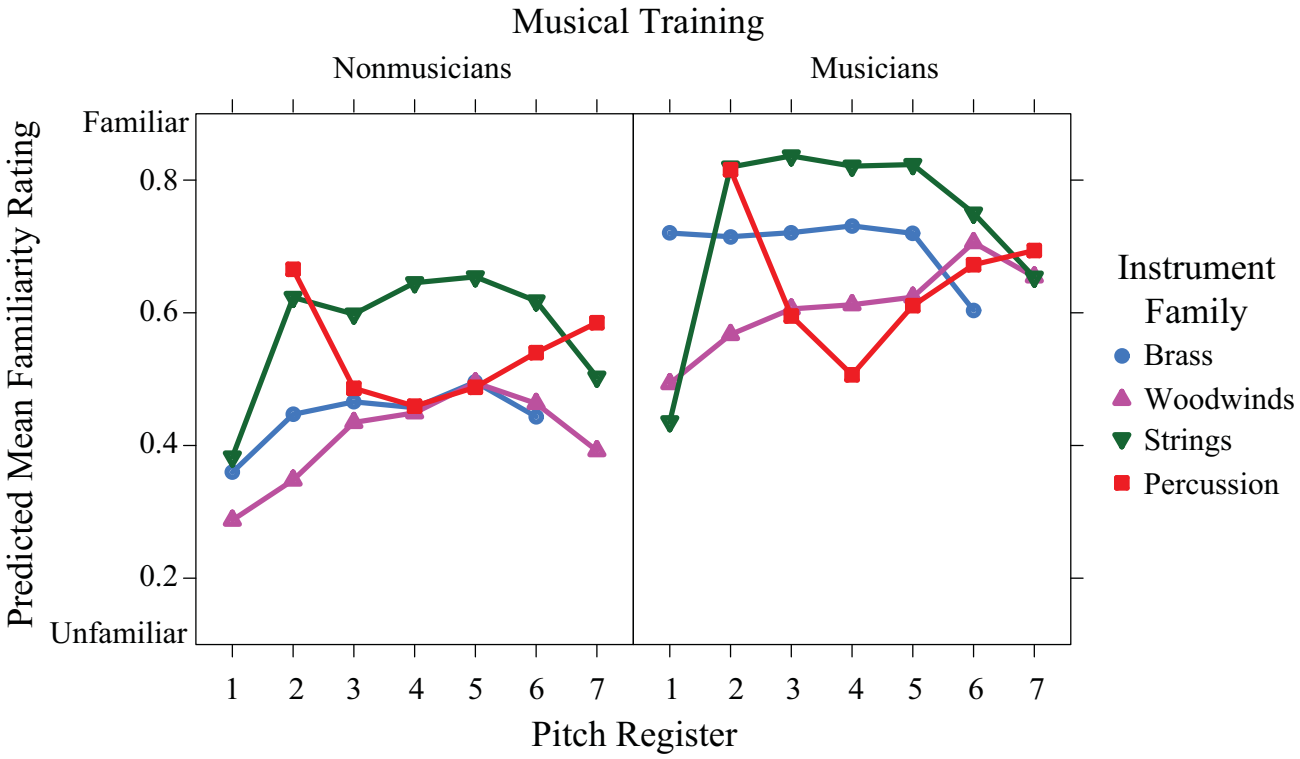

Fig. 2.7 Predicted Means of Familiarity Ratings Across Pitch Register for Each Instrument Family and Training Group.

depicted by the higher ratings of string timbres in octaves two through six compared to the string timbres in the first and seventh octaves. This trend appears to be the opposite for percussion timbres, where the highest familiarity ratings occurred in the lowest and highest octaves (second and seventh, respectively).

In order to examine attack strength and technique as possible factors contributing to affect ratings, the dataset was separated by instrument family and 20 linear mixed models were created: one for each of the five ratings for each of the four instrument families. All of the models included fixed factors of training and register. Attack strength was included in the brass and percussion models, and technique was included in the brass, woodwind and string models. This separation was necessary because of the lack of different attack strengths in the woodwind and

string samples and the lack of different playing techniques for the percussion samples. As with the full models, a maximal random effects structure was specified for each instrument family model. The brass familiarity, string preference and familiarity, and percussion familiarity models did not converge. Reducing the random effects structure by removing register as a random slope allowed the models to converge, although this removal increases likelihood of Type I errors. That being said, neither attack strength nor technique was a significant predictor for any individual instrument family model.

2.5 Discussion

There were a few key differences in these results, compared to the results from Experiment 1 in Eerola et al. (2012). First, the tension-arousal and energy-arousal ratings were significantly, but only weakly, correlated. As the energy-arousal dimension was highly predicted by register, listeners likely used spectral cues when making energy ratings and incorporated additional acoustic information when making tension ratings. Furthermore, valence and preference ratings in this experiment were moderately correlated. Listeners did not necessarily prefer timbres with the highest perceived valence. It is important to differentiate these measures, because valence is a perceived affect measure, whereas preference is a felt rating.

Musical training alone did not influence perceived affect ratings. However, the interaction between training and register influenced perceived valence and tension-arousal ratings. Regarding the valence ratings, musicians tended to perceive low-register timbres as less negative than nonmusicians. Musicians are more familiar with timbres in extreme registers than nonmusicians, and this familiarity could potentially play a role in the perceived affect ratings.

The register of the timbre influenced all perceived affect ratings. Listeners rated mid-register (i.e., registers four and five) timbres as more positive and more relaxed.

Energy-arousal ratings regularly increased as the register increased. Additionally, both groups preferred mid-register timbres to those in very low or high registers. This mid-register preference emphasizes the strong correlations found between preference ratings and both perceived valence, on the one hand, and tension-arousal ratings, on the other hand. These results can be compared with the weak correlation between energy-arousal ratings and preference ratings.

All instrument families were rated similarly on the energy-arousal scale. However, they were rated significantly differently on the valence and tension-arousal scales. String timbres were perceived as more positive than the other instrument families in the low and mid registers. Additionally, unlike the other instrument families which exhibited inverse U-shaped ratings on the valence scale, peaking around register five, valence ratings for percussion timbres were increasingly positive as register increased. Interestingly, this trend is not present in the preference ratings for percussion timbres.

As expected, familiarity ratings were higher in the musician group and varied across instrument family.

In this analysis, instrument family and pitch register were the only timbre-related factors take into account (register affecting timbre in the sense that spectral centroid, for example, varies with register, but not directly as a function of fundamental frequency). However, timbre is a multidimensional entity, and it is important to consider spectral, temporal, and spectrotemporal descriptors when examining timbre. The following chapter will further analyze the perceived affect, preference and familiarity ratings in relation to several timbre descriptors.

Chapter 3

Audio Descriptors: Tools and Analysis

3.1 Toolbox Verification

Due to its multidimensionality, it is necessary to account for multiple acoustic features when examining timbre (McAdams et al., 1995). There are numerous acoustic features that can typically be categorized as spectral, temporal, or spectrotemporal aspects of a sound. The goal of this chapter is to investigate the relationship between those descriptors and the perceived affect, preference and familiarity ratings.

Recently, researchers have packaged signal processing measures into useful tools that can be applied to psychoacoustic research. Timbre Toolbox (Peeters et al., 2011), MIRtoolbox (Lartillot, 2013), and PsySound 3 (Cabrera, Ferguson & Schubert, 2007) are run in or with Matlab (The MathWorks Inc., Natick, MA) to calculate values of acoustic descriptors from audio files. These toolboxes are often used in the music perception literature, but to our knowledge have not been validated against one another or with synthesized sounds having known values of the descriptors. We therefore felt it was necessary to compare and examine these toolboxes for accuracy so that only accurate descriptors are used in the experimental analysis with the affect,

preference and familiarity ratings.

Like the other toolboxes, PsySound 3 runs with Matlab, but unlike the Timbre Toolbox and MIRtoolbox, PsySound 3 is run on a separate interface so users with little to no Matlab experience can use it. PsySound 3 provides musical and rhythmic information about a sound file in addition to values for acoustic descriptors. However, it employs the MIRtoolbox for many of the acoustic descriptor calculations and thus is not further included in the comparisons below.

3.1.1 Toolbox Descriptions

The Timbre Toolbox calculates temporal descriptors, such as attack time, spectral descriptors, such as spectral centroid, and spectrotemporal descriptors, such as spectral variation over time. To calculate descriptors, there are three stages of computation. First, the input representations of the signal are computed. These input representations include the temporal energy envelope, Short-Term Fourier Transform (STFT), Equivalent Rectangular Bandwidth (ERB), and Harmonic measure, a representation of the signal in terms of a sum of harmonically related sinusoids. In order to calculate the temporal energy envelope of a given audio signal, the amplitude of the analytic signal, i.e., the signal with no negative-frequency components (Smith, 2007), is given by the Hilbert transform of the audio signal. The amplitude of the analytic signal is then low-pass filtered by a third-order Butterworth filter with a cutoff frequency of 5 Hz, resulting in the temporal energy envelope input representation. For the STFT input representation, STFT magnitude and STFT power measures are calculated. The Short-Term Fourier Transform of the signal is calculated and the resulting amplitude spectrum is used linearly (STFT magnitude calculations) or squared (STFT power calculations) for the spectral descriptor calculations. For the ERB input representation, ERB gammatone and ERB fft measures are calculated. The ERB gammatone and ERB fft measures utilize an Equivalent Rectangular Bandwidth model of peripheral auditory processing, which aims to represent the

auditory response of the basilar membrane with a series of bandpass filters with widths determined by psychoacoustic properties (Moore & Glasberg, 1983). The ERB gammatone measure uses gammatone filters, whereas the ERB fft measure utilizes a Fast Fourier Transform, but the frequency dimension is transformed to a physiological scale related to the distribution of frequencies along the basilar membrane in the inner ear as modeled by ERB-rate. Finally, the Harmonic measure is calculated via representing the audio signal as a sum of sinusoidal components with varying frequency and amplitude and then applying a frame analysis (Peeters et al., 2011).

In the second stage of computation, scalar and time-series descriptors are extracted from different input representations. To estimate the attack portion of the signal, the "weakest-effort method" (Peeters, 2004) is applied so that thresholds to detect the start and end time of the attack are not fixed but determined as a proportion to the maximum of the signal's energy envelope. Attack time, log-attack time, attack slope, temporal centroid, and Root Mean Square of the energy envelope are calculated from the temporal energy envelope input representation. Attack time refers to the duration (in seconds) of the attack portion of the signal, log-attack time is the log_{10} of the attack time, and attack slope is the averaged temporal slope of the energy during the attack portion of the signal (Peeters et al., 2011). Additionally, the temporal centroid is a measure of the center of gravity of the energy envelope of the signal (Peeters et al., 2011).

As described by Peeters et al. (2011), spectral centroid is a measure of the center of mass of the spectrum and is perceptually related to the "brightness" of the sound. Spectral spread refers to the standard deviation of the spectrum around the spectral mean value and spectral skewness refers to the degree of asymmetry of the spectrum around the mean. Spectral kurtosis examines the flatness of the distribution around the mean value of the spectrum and can indicate a flat, normal or peaky distribution. Spectral slope is a linear regression over the spectral amplitude values. Spectral decrease is the average of the set of spectral slopes between the fundamental

frequency and the frequency of the k^{th} harmonic. Spectral rolloff refers to the frequency below which 95% of the signal energy is contained. Spectrotemporal variation is a measure of the variation of the spectrum over time.

Spectral centroid, spectral spread, spectral skewness, spectral kurtosis, spectral slope, spectral decrease, spectral rolloff, and spectrotemporal variation are calculated from each of the following input representation: STFT magnitude, STFT power, ERB gammatone, ERB fft, and Harmonic. Spectral crest and spectral flatness are calculated from the STFT magnitude, STFT power, ERB gammatone and ERB fft input representations. Fundamental frequency, harmonic energy, noisiness, inharmonicity, tristimulus, harmonic spectral deviation and odd-to-even harmonic ratios are calculated with just the harmonic input representation (Peeters et al., 2011).

Finally, the third stage of computation considers the median and interquartile range (IQR) values of time-series descriptors to represent both central tendency and variability, respectively. Time-series descriptors include spectral centroid, spectral spread, spectral skewness, spectral kurtosis, spectral slope, spectral decrease, spectral rolloff, spectrotemporal variation, frame energy, spectral flatness, and spectral crest. A median and IQR value is calculated for each input representation used to calculate the given descriptor (Peeters et al., 2011).

The MIRtoolbox (version 1.5) calculates dynamic, rhythmic, timbre, pitch, and tonal descriptors. Although this toolbox is capable of analyzing longer sound files with multiple successive notes and rhythmic properties, for the purpose of this verification, we will only examine functions specific to timbre descriptors, applied to single-note sound files.

Like the Timbre Toolbox, input representations of the signal are calculated as an initial step, followed by the calculation of acoustic descriptors based on the input representation. To calculate the energy envelope input representation, the MIRtoolbox function, mirenvelope, first applies a Hilbert transform to the audio signal to retrieve the amplitude of the analytic signal, and then applies an infinite impulse response low-pass filter to the amplitude of the analytic signal.

The MIRtoolbox function, mironsets, uses mirenvelope to calculate the amplitude envelope and produces an attack onset detection curve. The mironset function is subsequently used in the MIRtoolbox functions to calculate temporal descriptors such as attack time and attack slope (Lartillot, 2013).

The default input representation for spectral descriptor calculations in the MIRtoolbox is an ERB filterbank similar to the ERBfft representation in the Timbre Toolbox (Lartillot, 2013). This process is carried out with the function mirfilterbank. Statistical descriptions, such as centroid, spread, skewness, kurtosis, flatness, and entropy, of the spectral distribution can then be calculated from this representation.

In order to compare the Timbre Toolbox and MIRtoolbox, we examined measures that could be calculated by both toolboxes. These included the temporal descriptors: attack time, log-attack time, attack slope and root mean square (of the temporal energy envelope), and the spectral descriptors: fundamental frequency, spectral centroid, spectral spread, spectral skewness, and spectral kurtosis.

3.1.2 Verification with Novel Sounds

Two stages of validation were completed with the toolbox results. The first stage utilized three novel sounds, labeled Note 1, Note 2 and Note 3, which were created in Pure Data (Puckette, 1996) using a patch created by Bennett Smith, similar to the one used in Caclin, McAdams, Smith & Winsberg (2005). The three novel sounds were all 0.5 s each and had a linear onset followed by a steady state time then a 0.05 s exponential decay to -60 dB. They varied by fundamental frequency (220 Hz, 440 Hz, 440 Hz), number of harmonics (40, 40, 0), linear attack time (0.05 s, 0.1 s, 0.2 s), spectral slope (-10 dB/octave, -5 dB/octave, N/A), and attenuation of even harmonics (-20 dB, -30 dB, 0 dB). Note 3 was the simplest sound and modeled a single sine wave.

The timbre descriptors that could be computed by both toolboxes were calculated for each of the three novel sounds, and then the calculated values were compared to the results computed by the Timbre Toolbox and MIRtoolbox. All calculated values for each of these measures are displayed in Table 3.1. Attack time, log-attack time and attack slope were calculated based on the designated linear attack time used to create each sound. The spectral measures: centroid, spread, skewness and kurtosis, were calculated based on Eq. 3.1, Eq. 3.2, Eq. 3.3, and Eq. 3.4 respectively, where K is the number of harmonics, f_k is the frequency of the k^{th} harmonic and a_k is the amplitude of the k^{th} harmonic.

$$\mu_1 = \frac{\sum_{k=1}^{K} a_k \cdot f_k}{\sum_{k=1}^{K} a_k} \tag{3.1}$$

$$\mu_2 = \sqrt{\frac{\sum_{k=1}^{K} a_k (f_k - \mu_1)^2}{\sum_{k=1}^{K} a_k}} \tag{3.2}$$

$$\mu_3 = \frac{\sum_{k=1}^{K} a_k (f_k - \mu_1)^3}{\mu_2^3} \tag{3.3}$$

$$\mu_4 = \frac{\sum_{k=1}^{K} a_k (f_k - \mu_1)^4}{\mu_2^4} \tag{3.4}$$

Overall, ten temporal and spectral descriptors were compared. The values for each measure calculated by hand, the MIRtoolbox and the Timbre Toolbox are displayed in Table 3.1. The Timbre Toolbox spectral descriptors that are displayed in Table 3.1 were calculated with the

harmonic input representation because the results are given in Hz, and thus in the same unit as the MIRtoolbox measures.

In this first stage of validation we are looking for absolute accuracy in both toolboxes. The attack measures from the Timbre Toolbox were closer to the true value than the attack measures from the MIRtoolbox. Both toolboxes measured the fundamental frequency of all the notes accurately. For the four spectral measures, the MIRtoolbox results were typically more accurate. Neither toolbox produced very accurate results for the spread, skewness and kurtosis for Note 3, which interestingly was the simplest note, a sine wave at 440 Hz. For additional comparison, and to specifically examine the remaining STFT- and ERB-based Timbre Toolbox results against the MIRtoolbox results, we completed a second method of validation, focusing on a proportional, instead of direct, comparison.

Table 3.1: Comparison of the MIRtoolboxand Timbre Toolbox for Ten Audio Descriptors for Three novel Sounds.

Measure	Calculated Measure	MIRtoolbox	Timbre Toolbox	Comments
Attack	Linear Onset Time (S)	Attack Time (S)	Attack Time Attack (S)	mironsets generates a curve with an apparent
Note 1	0.050	0.256	0.053	proper length but
Note 2	0.100	0.270	0.065	the attack prediction
Note 3	0.200	0.313	0.080	is incorrect.
Log-Attack Time	Log-Attack Time (log(S))	Log-Attack Time (log(S))	Log-Attack Time (log(S))	MIRtoolbox Log-Attack time is log base 10 of
Note 1	-1.301	-0.592	-1.020	the (incorrect) linear
Note 2	-1.000	-0.569	-0.940	attack time. For the
Note 3	-0.699	-0.504	-0.746	Timbre Toolbox it is not not exactly Log base 10 of the linear attack measure, however the result is more accurate.
Attack Slope	Attack Slope (dB/S)	Attack Slope weighted by Gaussian Curve	Temporal Energy Envelope Attack Slope (dB/S)	The MIRtoolbox attack slope calculation does not seem to be accurate or
Note 1	7.250	0.785	9.614	proportionally correct
Note 2	6.020	1.285	7.993	
Note 3	1.289	1.767	4.922	
continued on next page				

Measure	Calculated Measure	MIRtoolbox	Timbre Toolbox	Comments
Root Mean Square Note 1 Note 2 Note 3	RMS: Calculated with Matlab Function 0.157 0.13 0.139	RMS Energy 0.157 0.13 0.139	Harmonic RMS Energy Envelope Median 0.686 0.326 0.805	MIRtoolbox RMS is the same as the Matlab RMS function. The Timbre Toolbox does not seem to be accurate or proportionally related
Fundamental Frequency Note 1 Note 2 Note 3	Fundamental Frequency (Hz) 220 440 440	Pitch (Hz) 219.978 440.197 439.964	Harmonic F0 Median (Hz) 220.027 440.045 440.737	Both calculate the fundamental frequency accurately.
Spectral Centroid (Hz) Note 1 Note 2 Note 3	Spectral Centroid (Hz) 700.923 4461.133 440	Spectral Centroid (Hz) 714.115 4455.629 463.482	Spectral Centroid Median (Hz) 539.767 2453.443 455.751	The Timbre Toolbox calculated a significantly lower spectral centroid for Notes 1 & 2
Spectral Spread (Hz) Note 1 Note 2 Note 3	Spectral Spread (Hz) 1193.2 4733.8 4.40E-48	Spectral Spread (Hz) 1265.267 4737.049 564.395	Harmonic Spectral Spread Median (Hz) 705.385 2365.028 241.055	The spectral spread of Note 3 was not calculated correctly by either toolbox
Spectral Skewness Note 1 Note 2 Note 3	Spectral Skewness 5.418 3.721 1.00E+50	Spectral Skewness 4.515 1.145 28.37	Harmonic Spectral Skewness Median 2.928 1.059 22.451	The spectral spread of Note 3 was not correctly estimated by either toolbox, although the MIRtoolbox was more accurate for Notes 1 & 2.
Spectral Kurtosis Note 1 Note 2 Note 3	Spectral Kurtosis 26.93 10.306 1.00E+100	Spectral Kurtosis 33.785 3.172 866.2209	Harmonic Spectral Kurtosis Median 12.008 2.949 594.798	The MIRtoolbox spectral kurtosis measure was slightly more accurate than the Timbre Toolbox.

3.1.3 Comparison with Instrument Timbres

After the first stage of validation looked at absolute accuracy, a second stage of validation was run on the same 137 instrument samples from the experiment outlined in Chapter 2. This stage of validation served two main purposes. First, it allowed us to further examine results that were not absolutely accurate to see if there was a proportional relationship. A large sample size of timbres was useful for this comparison because we were looking for trends and proportional similarities by graphing the MIRtoolbox descriptors against the different Timbre Toolbox descriptors.

Additionally, this validation provided a method to proportionally compare the spectral descriptor results from the STFT and ERB input representation results from the Timbre Toolbox to the MIRtoolbox results. These could not be compared for absolute accuracy because the STFT and ERB input representations yielded results in different units. The MIRtoolbox applies an ERB-modeled filterbank to the signal (Lartillot, 2013). Therefore we expect the ERB measures from the Timbre Toolbox to be highly correlated with the MIRtoolbox spectral measures.

Table 3.2 Pearson's Correlation Results for Spectral Descriptors from the MIRtoolbox Compared to Those Derived from the Various Input Representations from the Timbre Toolbox

Timbre Toolbox Input Representation	Spectral Centroid	Spectral Spread	Spectral Skewness	Spectral Kurtosis
STFT magnitude	.92***	.15	.39***	.19*
STFT power	.86***	.62***	.76***	.73***
ERB gammatone	.90***	.73***	.93***	.92***
ERB fft	.94***	.79***	.94***	.93***
Harmonic	-.03	.18*	.32***	.23**

Notes. df=135. $*p < .05$ $**p < .01$ $***p < .001$.

Table 3.2 displays the correlations of the MIRtoolbox results with each method of calculation for the Timbre Toolbox for the spectral descriptors. The ERB fft Timbre Toolbox

measure was consistently the most correlated with the MIRtoolbox measure. As predicted, the ERB gammatone measure was strongly correlated as well. The Harmonic and STFT Magnitude measures were weakly or not at all correlated with the MIRtoolbox measures. The code for these measures should be further examined for discrepancies in both toolboxes.

Fig. 3.1 presents scatterplots comparing the spectral centroid results for each Timbre Toolbox measure plotted against the MIRtoolbox result. The MIRtoolbox spectral centroid value is plotted along the x-axis and the Timbre Toolbox value is plotted along the y-axis. The lack of correlation between the Timbre Toolbox Harmonic measure and the MIRtoolbox measure is clearly depicted in Fig. 3.1 e. Furthermore, the ranges of this plot should have been similar because the Harmonic measure is calculated in Hz like the MIRtoolbox but the range of the MIRtoolbox was about four times as large. Another problem present in this plot is the zero result calculated for several timbres in the Timbre Toolbox. All of the timbres in this plot should have a nonzero spectral centroid value. This measure should be further assessed in the Timbre Toolbox to correct for the zero calculations and incorrect range. For the remaining spectral plots the ranges cannot be directly compared because they are calculated in different units, however, the plots show reasonable correlation across all the timbres for spectral centroid.

Fig. 3.2 depicts a comparison between the MIRtoolbox and Timbre Toolbox for (a) linear attack time (b) log attack time and (c) attack slope. The Pearson's correlations for those measures were $r(135) = .75, p < .001$, $r(135) = .76, p < .001$ and $r(135) = .79, p < .001$, respectively. Spearmans rank correlation gave higher values of $\rho = .83$ and $\rho = .82$ for linear and log attack time, respectively, due to the nonlinear relationship between the variables. Although the correlation is relatively strong, the ranges displayed in the linear attack graph are drastically different. The Timbre Toolbox (y-axis) ranges from 0 to 0.14 seconds whereas the MIRtoolbox (x-axis) ranges from 0 to 0.45 seconds. This result supports the findings from the first validation, and the MIRtoolbox should be further examined for an error with attack calculations.

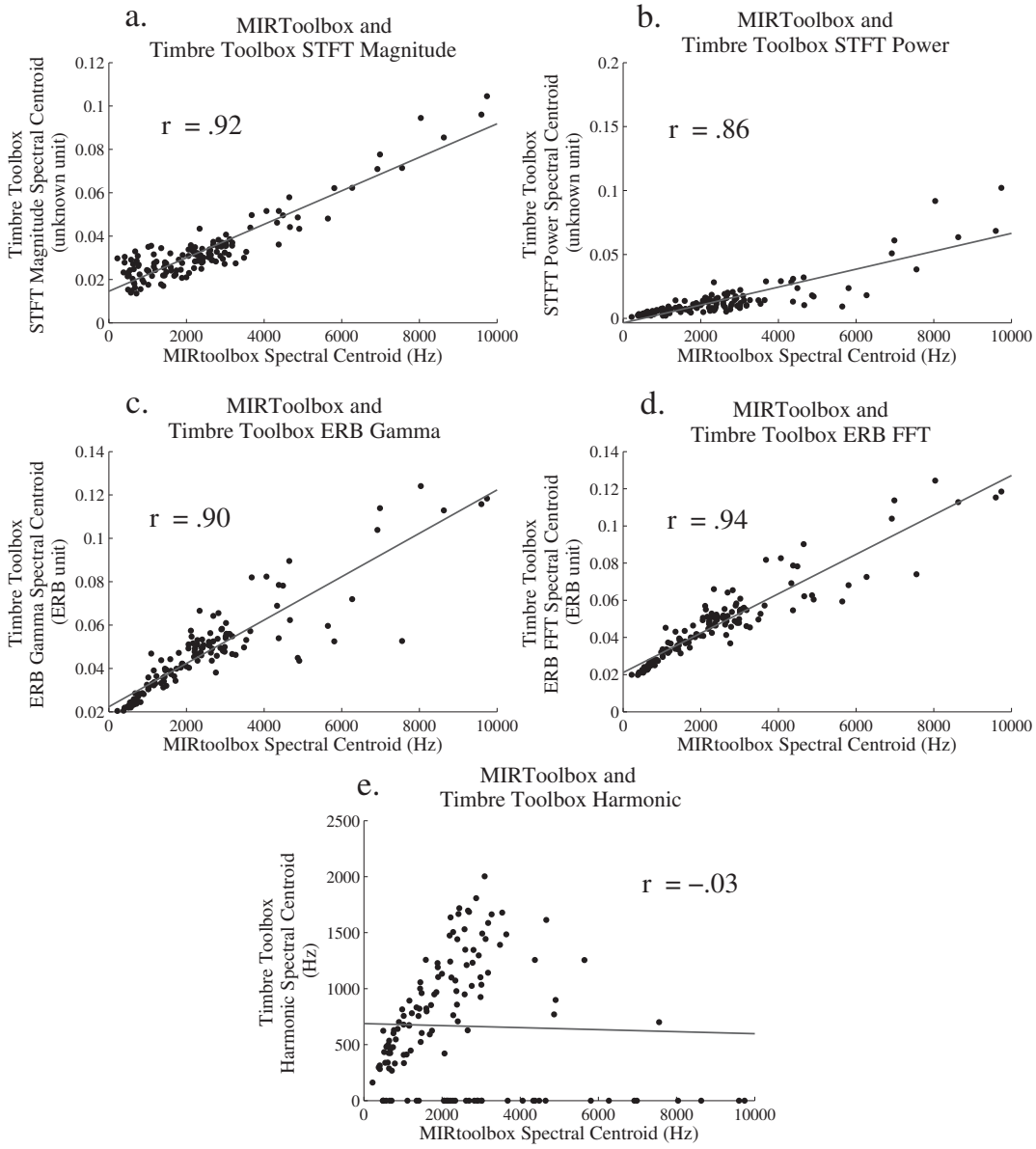

Fig. 3.1 MIRtoolbox Spectral Centroid Compared to Timbre Toolbox Spectral Centroid Calculated with STFT magnitude, STFT power, ERB gamma, ERB fft, and Harmonic Input Representations.

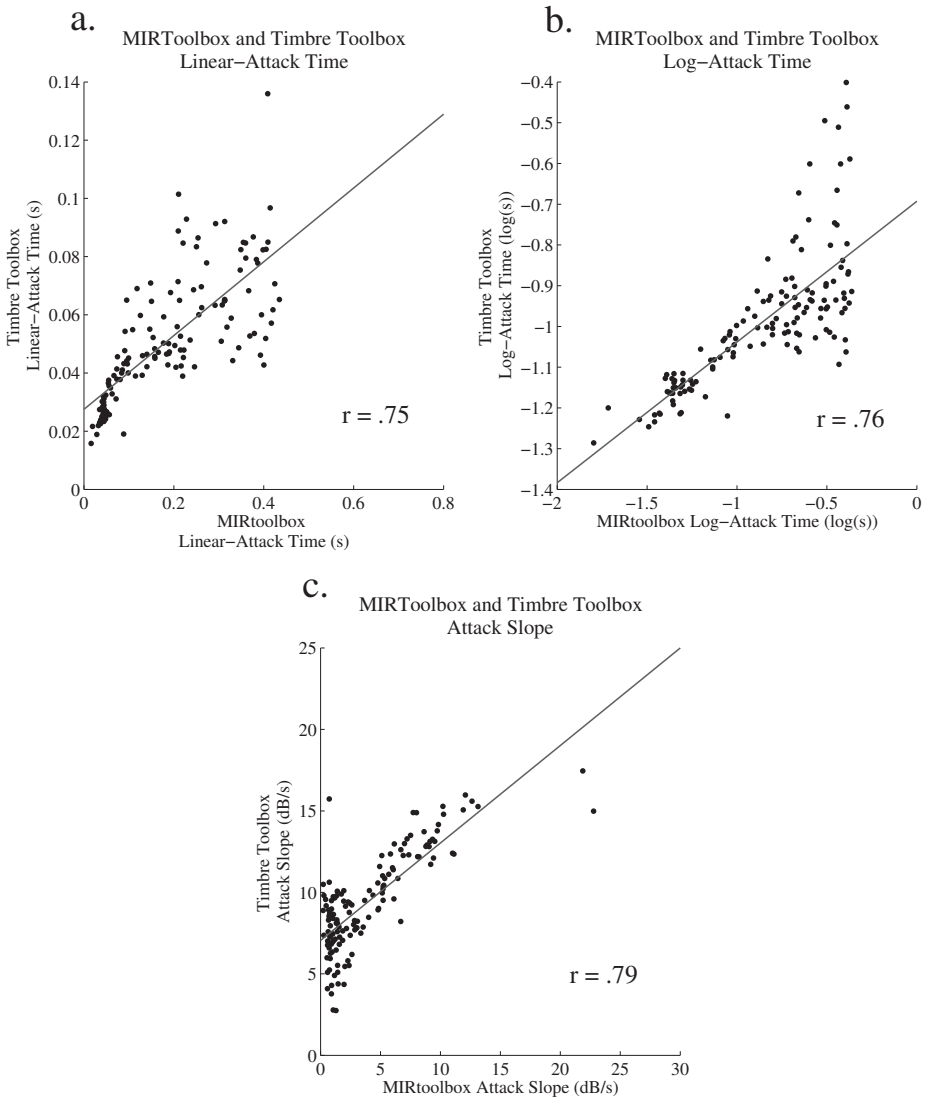

Fig. 3.2 MIRtoolbox Attack Measures Compared to Timbre Toolbox Attack Measures.

By comparing results from both toolboxes, we found a few errors present in each. The main problem in the MIRtoolbox seemed to be related to a function used to recognize the attack, mironsets, and in turn to calculate attack time and slope. This function was designed to detect multiple note onsets in a passage of notes, so it may not function as well when applied to a single

note. On the other hand, the main problem in the Timbre Toolbox was in the calculation of spectral measures, particularly in relation to the harmonic and STFT magnitude measures. Overall, the Timbre Toolbox temporal measures and the spectral ERB fft measures appear to be accurate and were used for further analysis of audio descriptors in relation to the perceived affect ratings.

3.2 Acoustic Descriptors: Analysis

To examine the relation of the different audio descriptors to the set of affect ratings, we completed a principal components analysis (Hubert, Rousseeuw & Vanden Branden, 2005). PCA is a statistical analysis that aims to explain covariance by combining correlated predictor variables into a component. This method of data reduction is useful when examining a multidimensional subject such as timbre.

Based on the results from the first validation we decided to use the Timbre Toolbox temporal descriptors in the acoustic descriptor analysis. Although the MIRtoolbox spectral measures displayed better absolute accuracy, the Timbre Toolbox spectral measures calculated with the ERB FFT input representation were proportionally accurate when compared to the MIRtoolbox measures. The Timbre Toolbox calculates both the median value and the interquartile range for each spectral descriptor, representing the central tendency and variability of each descriptor for a given sound. We decided to use the Timbre Toolbox ERB FFT spectral descriptors in order to examine both the median and IQR value of the spectral descriptors in relation to the affect ratings. Using the median value provides spectral information, and using the IQR measures adds spectrotemporal information because the IQR value of a spectral descriptor is representative of spectral variability over time. The PCA and a subsequent Pearson's correlation analysis were both completed in Matlab (The MathWorks Inc., Natick, MA).

Table 3.3 Definition of Acoustic Descriptors used in PCA as Defined by Peeters et al. (2011)

Acoustic Descriptor	Definition
Spectral Centroid	Center of gravity of the spectrum
Spectral Spread	Standard deviation of spectrum around the mean value
Spectral Skewness	Asymmetry of the spectrum around the mean value
Spectral Kurtosis	Flatness of spectrum around the mean value
Spectral Slope	Linear regression over the spectral amplitude values
Spectral Decrease	Average of the set of slopes between the fundamental frequency and frequency of the k^{th} harmonic
Spectral Rolloff	Frequency below which 95% of the signal energy is contained
Spectral Variation	Variation of the spectrum over time
Spectral Flatness	Ratio of the geometrical spectral mean to the arithmetical spectral mean
Spectral Crest	Ratio of maximum spectral value to the arithmetical spectral mean
Attack Time	Duration of the attack portion of a sound.
Attack Slope	Change of energy over time in the attack portion of a sound
Temporal Centroid	Center of gravity of the energy envelope

The principal components analysis was completed on a group of 23 measures, listed in Table 3.4, from 13 acoustic descriptors listed and defined in Table 3.3. The acoustic descriptors consisted of spectral, temporal, and spectrotemporal features and the median and interquartile range measures were used for the spectral and spectrotemporal features. The analysis generated a solution in which five components had eigenvalues of 1 or greater. This solution explained 84.7% of the variance of the original group of descriptors. Although spectral centroid is related to fundamental frequency, the relationship varies depending on instrument, attack, and technique, and thus, spectral centroid was included as a spectral measure in the PCA. The relationship between the fundamental frequency and spectral centroid for each of the 137 timbres is depicted in Fig. 3.3, which highlights the complexity of the relationship and the variation of spectral centroid for different instruments at the same fundamental frequency. Additionally, the spectral centroid for a given instrument at a given pitch can further vary based on the strength of attack, type of mallet (for percussion sounds) or playing technique.

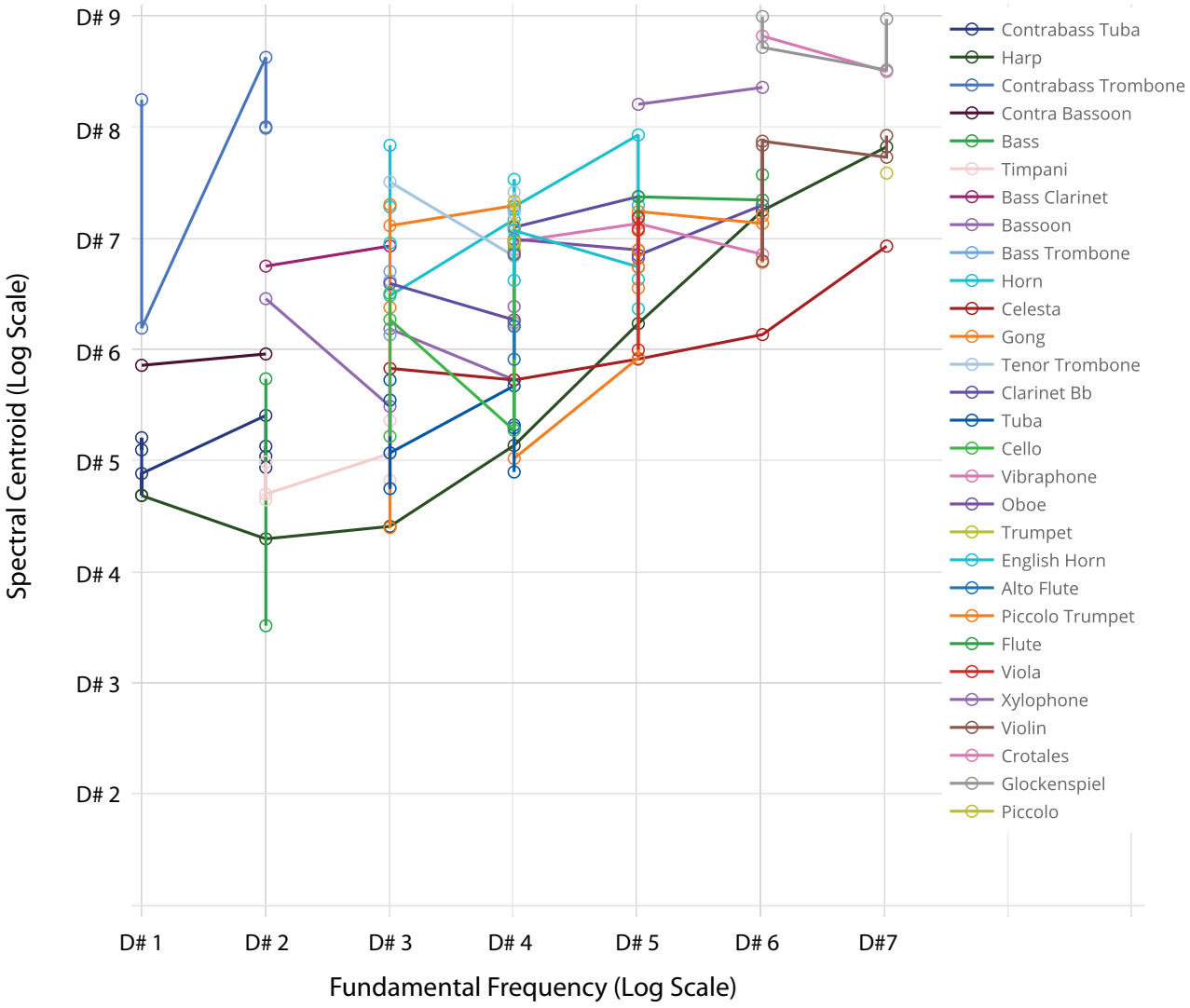

Fig. 3.3 Spectral Centroid and Fundamental Frequency for Each Sound Grouped by Instrument.

Table 3.4 displays the correlations between each principal component (PC) and each audio descriptor. PC1 was strongly correlated with the median values of spectral components: log

Table 3.4 Correlation Results for the Five Principal Components Correlated with the Acoustic Measures

Acoustic Descriptor	PC 1	PC 2	PC 3	PC 4	PC 5
Explained Variance	37.21%	22.04%	12.73%	6.70%	6.02%
Log Spectral Centroid Median	*.97***	*.14*	*.07*	*.08*	*.01*
Spectral Centroid IQR	.62***	-.56***	.41***	-.09	-.02
Spectral Spread Median	*.94***	*.01*	*.08*	*.17**	*.07*
Spectral Spread IQR	*.64***	*-.49****	*.45****	*-.05*	*-.08*
Spectral Skewness Median	*-.89****	*-0.2*	*-.13*	*.03*	*.13*
Spectral Skewness IQR	-.37***	-.73***	.26**	-.28***	.15
Spectral Kurtosis Median	-.86***	-.22*	-.14	.04	.15
Spectral Kurtosis IQR	-.50***	-.69***	.26**	-.22*	.13
Spectral Slope Median	.95***	.03	.04	.17	.08
Spectral Decrease Median	-.93***	-.23**	.06	.00	-.04
Spectral Decrease IQR	-.47***	-.58***	.36***	-.12	.05
Spectral Rolloff Median	.92***	.02	.15	.18*	.02
Spectral Rolloff IQR	.46***	-.49***	.55***	.02	-.11
Spectral Variation Median	-.46***	-.26**	.23**	.75***	.13
Spectral Variation IQR	-.39***	-.16	.31***	.78***	.22*
Spectral Flatness Median	-.29***	.08	.62***	-.08	-.53***
Spectral Flatness IQR	.24**	-.55***	.28***	-.08	.26**
Spectral Crest Median	.38***	-.12	-.59***	-.09	.59***
Spectral Crest IQR	.40***	-.61***	-.02	-.18*	.44***
Attack Time	-.16	.71***	.46***	-.09	.27**
Log-Attack Time	-.13	.63***	.56***	-.13	.40***
Attack Slope	.15	-.72***	-.54***	.15	-.28***
Temporal Centroid	-.29***	.76***	.40***	-.10	.14

Notes. df=135. $*p < .05$ $**p < .01$ $***p < .001$.

centroid, spread, skewness, kurtosis, slope, decrease, and rolloff. PC2 was correlated with the temporal components: temporal centroid, attack slope, and attack time, as well as the interquartile range measures of a number of spectral measures. PC3 was moderately correlated with the median measures of spectral flatness and spectral crest in addition to log attack time and attack slope. PC4 was strongly correlated with both the median and interquartile range measure of the spectrotemporal descriptor, spectral variation. Finally, PC5 was similar to PC3, correlating

more weakly with the spectral flatness and spectral crest median measures.

Table 3.5 Correlations of the Five Principal Components with Each of the Perceived Affect and Preference and Familiarity Ratings

	PC1	PC2	PC3	PC4	PC5
Valence	.38***	−.42***	−.40***	−.34***	.06
Tension Arousal	.38***	.32***	.14	.36***	.09
Energy Arousal	.67***	−.23**	−.34***	−.11	.15
Preference	−.17*	−.53***	−.19*	−.25**	.02
Familiarity	−.01	−.35***	.00	−.13	−.03

Notes. df=135. $*p < .05$ $**p < .01$ $***p < .001$.

After completing the PCA, we examined the correlation between the five resulting PCs and the affect, preference, and familiarity ratings collected in the main experiment described in Chapter 2. The summary of the Pearson's correlations are displayed in Table 3.5. Notably, the perceived energy-arousal ratings were moderately positively correlated with PC1, $r(135) = .67, p < .001$, and weakly negatively correlated with PC3, $r(135) = -.34, p < .001$. The perceived valence ratings were positively weakly correlated with PC1, $r(135) = .38, p < .001$). Additionally, the perceived valence ratings were weakly negatively correlated with PC2, $r(135) = -.42, p < .001, PC3$, $r(135) = -.40, p < .001$, and PC4, $r(135) = -.34, p < .001$. Perceived tension-arousal ratings were weakly positively correlated with PC1, $r(135) = .38, p < .001$, PC2, $r(135) = .32, p < .001$ and PC4. $r(135) = .36, p < .001$. Preference ratings were moderately negatively correlated with PC2, $r(135) = .53, p < .001$.

The results of these analyses will be discussed in the following chapter.

Chapter 4

Discussion

Although the perception of affect in timbre has not been studied extensively, many aspects of emotion and music research support the existence of a relationship. First, listeners can make perceived affect judgments almost instantaneously (Peretz et al., 1998; Bigand et al., 2005; Filipic et al., 2010). This hints that acoustic factors, such as timbre, that are present in very short sound samples influence affect ratings. Performers and composers reportedly use timbre as a means of communicating intended emotion to listeners (Holmes, 2011). Additionally, parallels involving timbral dimensions can be drawn between perceived emotion in music and research on perceived emotion in speech sounds, as compared by Juslin & Laukka (2003).

By isolating timbre, we were able to confirm that listeners can perceive affect in individual timbres and examine which components of timbre contribute to the different dimensions of the three-dimensional affect model. In accordance with the results from Experiment 1 presented in Eerola et al. (2012), and the aforementioned studies utilizing extremely short musical samples (Peretz et al., 1998; Bigand et al., 2005; Filipic et al., 2010), the participants were able to rate perceived affect in 500 ms instrument samples with great consistency. Additionally, we examined musicianship as a factor influencing affect ratings, but

found musical training only significantly contributed to familiarity ratings. Therefore, all participants, regardless of musicianship, rated perceived affect in timbre in a similar systematic fashion.

We found that the affect ratings were representative of the full three-dimensional model of affect (Schimmack & Grob, 2000). The energy-arousal and tension-arousal dimensions were only weakly correlated with each other. This finding was a key difference between this work and that of Eerola et al. (2012) where tension-arousal and energy-arousal ratings were highly correlated and consequently collapsed into a single arousal dimension. In contrast to Eerola et al. (2012), pitch register was varied in our study so we could examine pitch register as a factor of timbre and emotion. It is likely that both arousal dimensions applied to our results because of the addition of the pitch register variable. The linear mixed model analysis showed that the energy-arousal ratings were strongly influenced by pitch register, and unlike the tension-arousal ratings, were not significantly influenced by instrument family. This finding is a significant contribution to affect and timbre research because it shows that the two arousal dimensions are distinctly perceivable in timbre and not interchangeable, as they are influenced by different factors.

Furthermore, this study compared perceived affect ratings to a measure of participants felt preference. A primary goal of the experimental design was to clearly differentiate the perceived affect measures from the felt measures of preference and familiarity. As stated in a review of music and emotion studies (Eerola & Vuoskoski, 2013), there is often a lack of distinction between perceived emotion and felt emotion measures in the music and emotion literature. In Eerola et al. (2012), the perceived valence measure and preference measure were highly correlated and consequently collapsed into one measure. Collapsing across perceived valence and preference ratings reduces a distinction between a perceived affect measure and a felt measure.

In our experiment, there was a moderately strong positive correlation between the perceived valence ratings and preference, but the negative correlation between perceived

tension-arousal ratings and preference was slightly stronger. Although participants typically preferred more positive, less tense timbres, this finding demonstrates that there is not a clear one-to-one relationship between positive valence or tension and listeners preference. Furthermore, pitch register significantly influenced both perceived valence and tension-arousal ratings so that mid-register timbres were rated as more positive and more relaxed than timbres of an extreme high or low register. This finding can be applied to the literature that discusses preference for sad music (Schubert, 1996; Vuoskoski & Eerola, 2011).

In order to analyze timbre as a multidimensional attribute, we examined 23 acoustic measures, spanning spectral, temporal, and spectrotemporal dimensions. A principal components analysis allowed us to reduce the number of descriptors into five primary components, still representative of spectral, temporal and spectrotemporal dimensions. Calculating the correlation between these principal components and the affect ratings allowed us to identify which components of timbre were systematically influencing the various affect ratings.

This analysis further supported the inclusion of both tension-arousal and energy-arousal affect dimensions. Both arousal ratings were positively correlated with PC 1, but energy arousal was negatively correlated with PC 3 (a combination of spectral and temporal descriptors), whereas tension arousal was positively correlated with PC 2 and PC 4 (temporal and spectrotemporal descriptors). Energy arousal is more related to spectral components of timbre, whereas tension arousal is more related to its spectrotemporal elements.

Additionally, we can further emphasize the importance of separating perceived valence measures from preference ratings. Preference ratings were most related to the second principal component, temporal and spectrotemporal measures, whereas perceived valence ratings were correlated with the first four components to a similar degree. It is likely that listeners account for spectral, temporal and spectrotemporal information while judging perceived valence.

The differences in correlations between the principal components and the affect ratings

support the notion that participants systematically used different dimensions of timbre to complete their affect ratings. Familiarity ratings were weakly correlated to PC 2 but exhibited a lack of correlation to the other principal components, hinting that participants were not completing familiarity ratings based on inherent properties of timbre.

This study examined timbre and the complex covariance between timbre and register, shown in Fig 3.3, as musical elements capable of conveying emotion information. Future work should apply these results to increasingly ecological studies to validate the relationship between timbre and perceived affect in a greater context of music listening and examine how that relationship interacts with additional relationships between perceived affect and other musical variables such as pitch, dynamics, or tempo.

Appendix A

Experimental Stimuli

Table A.1 Description of Experimental Stimuli.

Instrument Family	Instrument	Pitch Registers**	Attack	Technique
Brass	Contrabass Tuba	1, 2	Strong	
	Contrabass Tuba	1, 2*	Normal	
	Contrabass Tuba	1, 2	Weak	
	Contrabass Tuba	1, 2	Normal	Flutter-Tonguing
	Tuba	3, 4	Strong	
	Tuba	3, 4	Normal	
	Tuba	3, 4	Weak	
	Tuba	3*, 4	Normal	Flutter-Tonguing
	Contrabass Trombone	1, 2	Strong	
	Contrabass Trombone	1, 2	Normal	
	Contrabass Trombone	2	Normal	Flutter-Tonguing
	Bass Trombone	3	Strong	
	Bass Trombone	3	Normal	None
	Bass Trombone	3	Normal	Flutter-Tonguing
	Tenor Trombone	4	Strong	
	Tenor Trombone	4	Normal	
	Tenor Trombone	3*, 4	Weak	
	Tenor Trombone	4	Normal	Flutter-Tonguing

Instrument Family	Instrument	Pitch Registers**	Attack	Technique
Brass	Horn	3, 4*	Strong	
	Horn	5	Strong	
	Horn	3, 4	Normal	
	Horn	4, 5*	Normal	
	Horn	3, 4, 5	Weak	
	Horn	3, 4	Normal	Flutter-Tonguing
	Trumpet	4*	Strong	
	Trumpet	4	Normal	
	Trumpet	4	Weak	
	Trumpet	4	Normal	Flutter-Tonguing
	Piccolo Trumpet	5*, 6	Strong	
	Piccolo Trumpet	5, 6	Normal	
	Piccolo Trumpet	5, 6*	Weak	
	Piccolo Trumpet	5*	Normal	Flutter-Tonguing
Woodwinds	Contra Bassoon	1, 2	Normal	
	Bassoon	3, 4*	Normal	
	Bassoon	2, 3, 4*	Normal	Flutter-Tonguing
	Bass Clarinet	2, 3	Normal	
	Clarinet Bb	4, 5*, 6*	Normal	
	Clarinet Bb	3, 4, 5*	Normal	Flutter-Tonguing
	Oboe	4, 5	Normal	
	Oboe	4, 5	Normal	Flutter-Tonguing
	English Horn	4*, 5	Normal	
	English Horn	4, 5	Normal	Flutter-Tonguing
	Alto Flute	4*	Normal	
	Alto Flute	4*	Normal	Flutter-Tonguing
	Flute	5, 6	Normal	
	Flute	5, 6*	Normal	Flutter-Tonguing
	Piccolo	7*	Normal	
Strings	Harp	1, 2*, 3, 4, 5, 6*, 7*	Plucked	
	Bass	2*	Bowed	
	Bass	2	Plucked	Vibrato
	Cello	3, 4	Bowed	
	Cello	3*, 4*	Bowed	Vibrato
	Cello	3, 4	Plucked	Vibrato

Instrument Family	Instrument	Pitch Register**	Attack	Technique
Strings	Viola	5*	Bowed	
	Viola	5	Bowed	Vibrato
	Viola	5*	Plucked	Vibrato
	Violin	6, 7	Bowed	
	Violin	6*, 7	Bowed	Vibrato
	Violin	6*	Plucked	Vibrato
Percussion	Timpani	2, 3	Metal	
	Timpani	2, 3	Wood	
	Timpani	2*, 3	Felt	
	Gong***	3*, 4, 5	Felt	
	Gong***	3, 4*, 5	Wood	
	Gong***	3, 4*, 5	Metal	
	Gong***	3, 4	Bowed	
	Celesta	3, 4, 5*, 6, 7	Wood	
	Glockenspiel	6, 7*	Metal	
	Glockenspiel	6, 7	Wood	
	Xylophone	5, 6*	Wood	None
	Vibraphone	4, 5*, 6	Metal	
	Crotales	6, 7*	Metal	

Notes: *The sound was also used in the control experiment.
**Pitch register ranged from D#1 - D#7.
***The pitch of the gong as labeled by the VSL.

Appendix B

Experimental Instructions

- You will listen to short sounds played by different musical instruments and perform 6 ratings for each sound. Ratings will be performed on an iPad on 9-point scales.
- The first 4 ratings are on PERCIEVED affect. You will be rating on the degree to which the sound expresses a feeling (NOT how it makes you feel).
- The first rating is the overall valence of the sound. The range is labeled from negative to positive. Below is a list of feelings that could be considered to have a negative valence (left) and positive valence (right). . . .

Unpleasant——————————————————————————————————Pleasant
Depressed-Angry-Tense-Tired-Boring-Neutral-Calm-Awake-Relaxed-Excited-Happy

- You will then rate sounds on scales for displeasure/pleasure, tensed/relaxed, and tired/awake.
- The last two ratings are how you feel about the sound. First you will rate your preference (dislike/like) for the sound then your familiarity with it (unfamiliar/familiar).
- You can listen to the sound as many times as you would like by pressing the play button on the iPad (upper left corner).
- After completing ratings on the sliders, press record responses (bottom) then next (top right) to move on.
- If you have any questions, please ask the experimenter now. . . .

Bibliography

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). *lme4: Linear mixed-effects models using Eigen and S4. R Package version 1.1-7*.

Berlyne, D. E. (1971). *Aesthetics and psychobiology.* New York, NY: Appleton-Century-Crofts.

Bigand, E., Vieillard, S., Madurell, F., Marozeau, J., & Dacquet, A. (2005). Multidimensional scaling of emotional responses to music: The effect of musical expertise and of the duration of the excerpts. *Cognition & Emotion*, *19*(8), 1113–1139.

Blood, A. J. & Zatorre, R. J. (2001). Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. *Proceedings of the National Academy of Sciences*, *98*(20), 11818–11823.

Cabrera, D., Ferguson, S., & Schubert, E. (2007). Psysound3: Software for acoustical and psychoacoustical analysis of sound recordings. In *Proceedings of the 13th International Conference on Auditory Display*, (pp. 356–363)., Montreal, Canada. Schulich School of Music, McGill University.

Caclin, A., McAdams, S., Smith, B. K., & Winsberg, S. (2005). Acoustic correlates of timbre space dimensions: A confirmatory study using synthetic tones. *Journal of the Acoustical Society of America*, *118*(1), 471–482.

Coutinho, E. & Dibben, N. (2013). Psychoacoustic cues to emotion in speech prosody and music. *Cognition & Emotion*, *27*(4), 658–684.

Eerola, T., Ferrer, R., & Alluri, V. (2012). Timbre and affect dimensions: Evidence from affect and similarity ratings and acoustic correlates of isolated instrument sounds. *Music Perception*, *30*(1), 49–70.

Eerola, T. & Vuoskoski, J. K. (2013). A review of music and emotion studies: Approaches, emotion models, and stimuli. *Music Perception*, *30*(3), 307–340.

Egermann, H. & McAdams, S. (2013). Empathy and emotional contagion as a link between recognized and felt emotions in music listening. *Music Perception*, *31*(2), 139–156.

Evans, P. & Schubert, E. (2008). Relationships between expressed and felt emotions in music. *Musicae Scientiae*, *12*(1), 75–99.

Filipic, S., Tillmann, B., & Bigand, E. (2010). Judging familiarity and emotion from very brief musical excerpts. *Psychonomic Bulletin & Review*, *17*(3), 335–341.

Fox, J. & Weisberg, S. (2011). *An R Companion to Applied Regression* (Second ed.). Thousand Oaks, CA: Sage.

Gabrielsson, A. (2001). *Emotions in strong experiences with music.* Oxford, UK: Oxford University Press.

Gabrielsson, A. (2002). Emotion perceived and emotion felt: Same or different? *Musicae Scientiae*, *5*(1 suppl), 123–147.

Gabrielsson, A. & Juslin, P. N. (1996). Emotional expression in music performance: Between the performer's intention and the listener's experience. *Psychology of Music*, *24*(1), 68–91.

Gabrielsson, A. & Lindström, E. (2010). The influence of musical structure on emotional expression. In P. N. Juslin & J. A. Sloboda (Eds.), *Handbook of music and emotion: Theory, research, application* (pp. 367–400). New York, NY: Oxford University Press.

Gagnon, L. & Peretz, I. (2003). Mode and tempo relative contributions to "happy-sad" judgements in equitone melodies. *Cognition & Emotion*, *17*(1), 25–40.

Gatewood, E. L. (1927). An experimental study of the nature of musical enjoyment. In M. Schoen (Ed.), *The effects of music* (pp. 78–120). New York, NY: Harcourt, Bruce & Company.

Hailstone, J. C., Omar, R., Henley, S. M., Frost, C., Kenward, M. G., & Warren, J. D. (2009). It's not what you play, it's how you play it: Timbre affects perception of emotion in music. *Quarterly Journal of Experimental Psychology*, *62*(11), 2141–2155.

Hexler.net (2011). *TouchOSC: Modular OSC and MIDI control surface for iPhone, iPod Touch, and iPad*. Available from `http://hexler.net/`.

Holmes, P. A. (2011). An exploration of musical communication through expressive use of timbre: The performer's perspective. *Psychology of Music*, *40*(3), 301–323.

Hubert, M., Rousseeuw, P. J., & Vanden Branden, K. (2005). Robpca: A new approach to robust principal component analysis. *Technometrics*, *47*(1), 64–79.

Huron, D., Anderson, N., & Shanahan, D. (2014). You can't play a sad song on the banjo: Acoustic factors in the judgment of instrument capacity to convey sadness. *Empirical Musicology Review*, *9*(1), 29–41.

Ilie, G. & Thompson, W. F. (2006). A comparison of acoustic cues in music and speech for three dimensions of affect. *Music Perception*, *23*(4), 319–329.

ISO (2004). Acoustics – Reference zero for the calibration of audiometric equipment – Part 8: Reference equivalent threshold sound pressure levels for pure tones and circumaural earphones (ISO 389–8). Technical report, International Organization for Standardization, Geneva.

Juslin, P. N. (1997). Perceived emotional expression in synthesized performances of a short melody: Capturing the listener's judgment policy. *Musicae Scientiae*, *1*(2), 225–256.

Juslin, P. N. (2001). Communicating emotion in music performance: A review and a theoretical framework. In P. N. Juslin & J. A. Sloboda (Eds.), *Music and emotion: Theory and research* (pp. 309–337). Oxford, UK: Oxford University Press.

Juslin, P. N. & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*, *129*(5), 770.

Juslin, P. N. & Laukka, P. (2004). Expression, perception, and induction of musical emotions: A review and a questionnaire study of everyday listening. *Journal of New Music Research*, *33*(3), 217–238.

Juslin, P. N. & Västfjäll, D. (2008). Emotional responses to music: The need to consider underlying mechanisms. *Behavioral and Brain Sciences*, *31*(5), 559–575.

Kawakami, A., Furukawa, K., Katahira, K., Kamiyama, K., & Okanoya, K. (2013). Relations between musical structures and perceived and felt emotions. *Music Perception*, *30*(4), 407–417.

Koelsch, S., Fritz, T., Müller, K., & Friederici, A. D. (2006). Investigating emotion with music: An fmri study. *Human Brain Mapping*, *27*(3), 239–250.

Krumhansl, C. L. (2002). Music: A link between cognition and emotion. *Current Directions in Psychological Science*, *11*(2), 45–50.

Lartillot, O. (2013). Mirtoolbox user's manual. Technical report, Finnish Centre of Excelence in Interdisciplinary Music Research.

Margulis, E. H. (2007). Silences in music are musical not silent: An exploratory study of context effects on the experience of musical pauses. *Music Perception*, *24*(5), 485–506.

Marozeau, J., de Cheveigné, A., McAdams, S., & Winsberg, S. (2003). The dependency of timbre on fundamental frequency. *Journal of the Acoustical Society of America*, *114*(5), 2946–2957.

Martindale, C. & Moore, K. (1989). Relationship of musical preference to collative, ecological, and psychophysical variables. *Music Perception*, 431–445.

McAdams, S. (1993). Recognition of sound sources and events. In S. McAdams & E. Bigand (Eds.), *Thinking in sound: The cognitive psychology of human audition* (pp. 146–198). Oxford, UK: Oxford University Press.

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

Menon, V. & Levitin, D. J. (2005). The rewards of music listening: Response and physiological connectivity of the mesolimbic system. *Neuroimage*, *28*(1), 175–184.

Moore, B. C. & Glasberg, B. R. (1983). Suggested formulae for calculating auditory-filter bandwidths and excitation patterns. *Journal of the Acoustical Society of America*, *74*(3), 750–753.

Patel, A., Gibson, E., Ratner, J., Besson, M., & Holcomb, P. (1998). Processing syntactic relations in language and music: An event-related potential study. *Journal of Cognitive Neuroscience*, *10*(6), 717–733.

Patel, A. & Peretz, I. (1997). Is music autonomous from language? A neuropsychological appraisal. In I. Deliège & J. Sloboda (Eds.), *Perception and cognition of music.* (pp. 191–215). Hove, UK: Psychology Press.

Peeters, G. (2004). A large set of audio features for sound description (similarity and classification) in the CUIDADO project. Technical report, IRCAM, Paris, France.

Peeters, G., Giordano, B. L., Susini, P., Misdariis, N., & McAdams, S. (2011). The timbre toolbox: Extracting audio descriptors from musical signals. *Journal of the Acoustical Society of America*, *130*(5), 2902–2916.

Peretz, I., Gagnon, L., & Bouchard, B. (1998). Music and emotion: Perceptual determinants, immediacy, and isolation after brain damage. *Cognition*, *68*(2), 111–141.

Puckette, M. (1996). Pure data: Another integrated computer music environment. In *Proceedings of the Second Intercollege Computer Music Concerts*, (pp. 37–41)., Tachikawa, Japan.

Risset, J.-C. & Wessel, D. L. (1999). Exploration of timbre by analysis and synthesis. In D. Deutsch (Ed.), *The psychology of music* (pp. 113–169). San Diego, CA: Academic Press.

Russell, J., Weiss, A., & Mendelsohn, G. (1989). Affect grid: A single-item scale of pleasure and arousal. *Journal of Personality and Social Psychology*, *57*(3), 493–502.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*(6), 1161–1178.

Scherer, K. R. & Oshinsky, J. S. (1977). Cue utilization in emotion attribution from auditory stimuli. *Motivation and Emotion*, *1*(4), 331–346.

Schimmack, U. & Grob, A. (2000). Dimensional models of core affect: A quantitative comparison by means of structural equation modeling. *European Journal of Personality*, *14*(4), 325–345.

Schimmack, U. & Reisenzein, R. (2002). Experiencing activation: Energetic arousal and tense arousal are not mixtures of valence and activation. *Emotion*, *2*(4), 412–417.

Schubert, E. (1996). Enjoyment of negative emotions in music: An associative network explanation. *Psychology of Music*, *24*(1), 18–28.

Schubert, E. (1999). Measuring emotion continuously: Validity and reliability of the two-dimensional emotion-space. *Australian Journal of Psychology*, *51*(2), 154–165.

Schubert, E. (2007). The influence of emotion, locus of emotion and familiarity upon preference in music. *Psychology of Music*, *35*(3), 499–515.

Sloboda, J. A. (1991). Music structure and emotional response: Some empirical findings. *Psychology of Music*, *19*(2), 110–120.

Sloboda, J. A. & O'Neill, S. A. (2001). Emotions in everyday listening to music. In P. N. Juslin & J. A. Sloboda (Eds.), *Music and emotion: Theory and research* (pp. 415–430). Oxford, UK: Oxford University Press.

Smith, J. O. (2007). *Mathematics of the discrete Fourier transform (DFT): With audio applications*. Stanford, CA: W3K Publishing.

Vienna Symphonic Library GmbH. (2011). *Vienna Symphonic Library*. Available from `http://vsl.co.at`.

Vuoskoski, J. K. & Eerola, T. (2011). The role of mood and personality in the perception of emotions represented by music. *Cortex*, *47*(9), 1099–1106.

Weber, R. (1991). The continuous loudness judgement of temporally variable sounds with an "analog" category procedure. In Schick, A., Hellbruck, J., & Weber, R. (Eds.), *Fifth Oldenburg Symposium on Psychological Acoustics*, (pp. 267–294)., Oldenburg, Germany: **BIS**.

West, B. T., Welch, K. B., & Galecki, A. T. (2006). *Linear mixed models: A practical guide using statistical software*. Boca Raton, FL: Chapman & Hall/CRC Press.

Zentner, M., Grandjean, D., & Scherer, K. R. (2008). Emotions evoked by the sound of music: Characterization, classification, and measurement. *Emotion*, *8*(4), 494.