# Timbre in the Communication of Emotions Among Performers and Listeners from Western Art Music and Chinese Music Traditions


Lena Heng

Master of Arts

Schulich School of Music

McGill University

Montreal, Quebec


April 2018

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

## ACKNOWLEDGEMENTS

This thesis and my Masters studies would not have been possible without the many wonderful people around me. First, I would like to express my gratitude for my supervisors, Professor Stephen McAdams and Professor Kok Roe-Min, both of whom generously shared their time, expertise, and amazing wealth of knowledge, and inspired in me the desire to continue delving more deeply into music and research. Bennett Smith has been immensely helpful in programming the experimental interface, rushing to finish it before I left for Singapore and wonderfully patient as I constantly bugged him over technical issues. Eddy Kazazis, Etienne Thoret, and Vincent Perreault helped tremendously with MATLAB and answering my endless stream of questions. All the lovely human beings in the Music Perception and Cognition Lab: Tanor Bonin, Hannah Darroch, Manda Fischer, Erica Huynh, Behrad Madahi, Jason Noble, Kit Soden, and Moe Touizrar, for interesting discussions on issues of music and/or perception and cognition, sharing their valuable experiences, sometimes talking about weightier things like life and our goals, other times about lighter stuff like what good shows there are to watch, and once in a while, letting me tell stories about my cats.

## AUTHOR CONTRIBUTIONS

As the sole author, I was responsible for designing and running the experiments, performing the data analyses, interpreting the results, and writing this thesis. My supervisors Professor Stephen McAdams and Professor Kok Roe-Min guided and oversaw the study.

## ABSTRACT

Timbre has been identified by music perception scholars as an important aspect in the communication of emotions in music. While its function as a carrier of perceptually useful information about sound source mechanics has been established, whether and how it functions as a carrier of information for communicating intended emotions in music is still uncertain. If timbre functions as a carrier of intended emotion content in music, how it is used may vary across different musical traditions and has to be learned by musicians.

To investigate whether there is a difference in the use of timbre by musicians of different musical traditions, two parallel groups of performers (*erhu*, violin, *pipa*, guitar, *dizi*, flute) from the Chinese and Western art music traditions are recorded as they perform excerpts to express different intended emotions (happy, sad, angry, and neutral). Three groups of listeners (trained in Chinese and Western art music traditions, and Western nonmusicians; 30 per group) listened to the recorded excerpts and classified the intended emotions performed. Listeners trained in the two musical cultures showed a significant difference in identifying the intended musical intent. It appears likely that certain aspects of timbre provide pertinent information for musical communication.

Finally, sound profiles of the recorded excerpts were analyzed to determine acoustic aspects that are correlated with timbre characteristics. Analysis of the sound profiles also revealed consistent manipulations of timbre by performers in their expression of intended emotions in music. It appears therefore that timbre not only functions as a carrier of information for musical communication but is also implicated in the learning of musical meanings that differs across musical traditions.

## RÉSUMÉ

Le timbre a été identifié par les spécialistes de la perception musicale comme un aspect important de la communication des émotions dans la musique. Alors que sa fonction en tant que support d'informations perceptivement utiles sur la mécanique des sources sonores a été établie, il est encore incertain si et comment il fonctionne comme porteur d'informations pour communiquer les émotions dans la musique. Si le timbre fonctionne comme porteur du contenu émotif dans la musique, son utilisation peut différer selon les traditions musicales et doit être apprise par les musiciens.

Pour déterminer s'il existe une différence dans l'utilisation du timbre par des musiciens de différentes traditions musicales, deux groupes parallèles d'interprètes (*erhu*, violon, *pipa*, guitare, *dizi*, flûte) issus des traditions musicales chinoises et occidentales sont enregistrés alors qu'ils exécutent des extraits afin d'exprimer différentes émotions (heureuse, triste, colère et neutre). Trois groupes d'auditeurs (formés dans les traditions de musiques chinoises et occidentales, et non-musiciens occidentaux, 30 par groupe) ont écouté les extraits enregistrés et classifié les émotions voulues. Les auditeurs formés dans différentes cultures musicales ont montré une différence significative dans leurs bonnes réponses vis-à-vis de l'intention musicale de l'interprète. Il semble probable que certains aspects du timbre fournissent des informations pertinentes pour la communication musicale.

Enfin, les profils sonores des extraits enregistrés ont été analysés quant aux aspects corrélés avec des caractéristiques du timbre. L'analyse des profils sonores a également révélé des manipulations analogues du timbre par les interprètes à travers les cultures dans leur expression des émotions dans la musique. Il apparaît donc que le timbre servirait non seulement de support

d’information pour la communication musicale, mais qu’il serait également impliqué dans l’apprentissage des significations musicales, qui diffère d'une tradition musicale à l’autre.

TABLE OF CONTENTS

List of Tables

List of Figures

# CHAPTER 1
# Introduction

## 1.1 Motivations for this Study

Drawing from a large body of scholarly and anecdotal evidence, I observe timbre playing a prominent role in Chinese music (Fan, 2014; Zhao, 2017). The only records we have of musical performances before any methods of sound recording were invented are written descriptions. Even though these descriptions are unable to provide us with clear acoustic representations of the performance, they are valuable in a different way—they reflect aspects that these writers believed were important in a musical performance. It is from these written records that we can observe the numerous instances in which musical timbre has been mentioned, as well as the myriad ways in which it has been talked about and emphasized as an important musical parameter. In one poem from the *Tang* dynasty for instance, we see repeated references to how timbre is more important than the actual melody: “Turning the pegs, she strums the strings, even before the music starts, the sound touches our hearts” (轉軸撥絃三兩聲，未成曲調先有情), and further on, “The lower strings patter like torrential rain, the higher strings whisper sweet affections” (大絃嘈嘈如急雨，小絃切切如私語) (Chang Qing collection of Bai’s poems [白氏

长庆集]). Descriptions focused on how timbral qualities evoke emotional reactions from the listeners, more than the actual pitches that are being played.

Being first trained in Western art music before learning a Chinese instrument, I also observed how musicians in the Western art music tradition and those in the Chinese music tradition utilize timbre in different ways. Although there are discussions of how timbre can be manipulated within Western art music performances to suggest a certain musical intent, timbral manipulations are often treated with even more precise intentions in Chinese music performance. There are at times certain timbral manipulations that are allowed and others that are prohibited, as well as combinations of certain timbres that are consistently used in particular musical phrases, which are often definitively articulated as a teacher guides a student. The literature on Chinese music also often brings up timbre as an important element in performance and musical aesthetics.

These differences in the use of timbre led me to wonder how timbre functions in communicating musical intent. Which acoustic attributes within timbre carry pertinent information for musical meaning, and, if possible, what aspects of these attributes are learned in music?

## 1.2 Musical Communication

### 1.2.1 Musical meanings

Communication of musical meaning is an important aspect of music, not only in areas of hermeneutics and interpretation, but also in analysis, performance, and studies on music perception, to name a few. The discourse surrounding musical communication, however, has been complex and filled with differing perspectives. On the surface, music appears to share many features with language, and philosophers such as Susanne Langer (1948) have put forward theories of how music is the precursor to language. Linguists such as Charles Hockett (1960) have outlined characteristics of language that he called “design features” and have shown that most of these are shared by music. More recent research has argued that even though both language and music can be said to have “evolved in the human lineage” (Patel, 2010, p. 356), this does not imply that music is another form of “language” or that we should think of understanding music the way we do language.

Following this argument then, musical meaning has to be discussed in a different way from linguistic meaning. Peter Kivy (2002) believes that we cannot talk about meaning in music because it does not have the possibility of being meaningful in the linguistic sense. Jean-Jacques Nattiez (2003) also believes that musical meanings should not be simply equated to linguistic meanings, although his conception of meaning differs from that of Kivy who looks at meaning as signification. Meaning is not inherent in the object; rather, it is the act of signification that creates meaning. Nattiez while in concordance to the above, also believes that signification is simply one amongst the many discourses of meaning in music. The semiologist Jean Molino believes that not all the traces from poiesis left in the symbolic form can be perceived by the receiver, and that communication is no more than any one possible result of the symbolic process within the various cases of exchange (as cited in Nattiez, 1990, p. 17). Heavily influenced by Molino’s ideas, Nattiez proposed a tripartitional model of semiology in which poiesis by the encoder and esthesis by the decoder of the neutral trace or the signifier is constantly taking place at different levels. Meaning, or the signified, is produced through these different layers of signification. There are others who hold differing views, such as that the intended symbolic process has to occur from the communicator to the receiver. George Herbert Mead, for example, believes that there is a “triadic relationship” between a stimulus, the thing to which it refers, and the individual for whom the stimulus has meaning. Communication takes place only where the gesture made has the same meaning for the individual who makes it and the individual who responds to it (Mead, 1934). Patrik Juslin and Renee Timmers (2010) appear to sum up these differing views, although they unpack Molino’s idea of communication into two parts—expression and recognition. Expression is the process that arises from the individual (the encoder) who creates the stimulus. Recognition makes up the other part of communication and involves the interpretation of the stimulus by the individual (the decoder) who responds to it. The notion of expression thus does not require that there be a correspondence between the encoder and the decoder, and in the case of music, of what the listener perceives in a performance and what the performer intends to express. “Communication” on the other hand, following Mead’s conceptualization, requires there be “both a performer’s intention to express a specific emotion and recognition of this emotion by a listener” (p. 455). Communication may also occur at different levels, as suggested by Nattiez, whether it is from the poietic intent of the composer to the esthesic process by the reader, or the poiesis from the performer to the esthesic process of the listener.

### 1.2.2 Musical meanings as a socio-cultural practice

The question about musical meanings functioning within socio-cultural practice has been part of discourse on music for a long time. Theodor Adorno (1976) sees the autonomy of music as the ideal but bemoans the ubiquitous commodification of music that occurs because the extraartistic aspect of music is necessary to reach the autonomous musical structures within. More recent sociological, feminist views question this concept of musical autonomy, and David Clarke (2003) attempts to look at music in a more relativistic framework. Music is situated within a web of meanings and is dependent on the relationships it enacts within a culture. Benjamin Piekut (2014) talks about actor-network theory as a methodology that can be used to understand the many "entanglements [with extramusical elements] that [music] is necessarily caught up with" (p. 192).

Conventional musicological discourses take meaning as socially constructed or as grounded in the material rather than the social. Both positions are deterministic, somewhat Saussurean, and ultimately arbitrary. Musical meaning is neither entirely determined by extramusical elements, nor is the relationship between the signifier and the signified totally arbitrary. Musicologist Nicholas Cook (2001) criticized that musical analyses have been placing too much emphasis "on the interpretation and not the analysis that underlies it" (p. 171). He proposes understanding musical meanings through music's emergent properties and thinking of signs and structure as interacting through some kind of dialogical relationship. In a different field, anthropologist Clifford Geertz (1973) proposes the concept of culture to be "a semiotic one" (p. 5), being made up of "interworked systems of construable signs" (p. 14). Culture thus involves the process of signification, without which there will be no understanding. Musical signs then, can be seen as one of these systems of signification, deeply entangled within the context of a culture. In summary, the process of signification in music is grounded in understanding the semiology of musical structures, which in turn is predicated on their entanglements with socio-cultural networks.

### 1.2.3 Meyer's concept of style and music in musical communication

Leonard Meyer, an important scholar who contributed influential ideas and works on aspects of aesthetic theory, music analysis, and musicology, believes that "human communication is for the most part dependent upon learning" (1989, p. 349), and interpretations are subjective and dependent on the questions formed and hypotheses selected among all the available ones arising from the communicative act. The hypotheses selected are influenced by what the receiver has learned, and, if we were to extend it further, to the experiences and enculturation to which they have been exposed since birth. Within musical communication, musical style "constitutes the universe of discourse within which musical meanings arise" (Meyer, 1994, p. 7). The knowledge and understanding of particular musical styles is a learned response and is based on the "psychology of human mental processes—the ways in which the mind, operating within the context of culturally established norms, selects and organizes the stimuli that are presented to it" (p. 7). The structure governing a piece of music is created through rules of a musical practice and arises as certain musical elements are arranged in syntactical positions. Style is made up of patterns created by constraints of structure and the choices made within those constraints. According to Meyer, the two most important concepts governing style are thus constraints and choices. There is a hierarchy here—laws and rules, which are constraints, and strategies which are choices—and they delimit the different possibilities a composer can make.

Laws in Meyer's view are universals. They may be physical or psychological principles that govern perception and cognition of musical patterns and are therefore innate and similar across cultures. Rules on the other hand are intra-cultural. They have to be learned. Strategies are the choices that are made within the possibilities established by rules and are also influenced by extramusical parameters. Different degrees of these various constraints used in structuring a work make up the style, and style analysis attempts to discover and formulate the rules and strategies that function as the basis of characteristics replicated in some group of works and to explain in the light of these constraints how they are related to one another.

How music can be structured into these different levels of hierarchical principles requires the presence not just of semantic associations but also of syntax organized by different parameters of music. Meyer differentiates between primary and secondary parameters: primary ones have elements that can be "segmented into discrete, nonuniform relationships so that the similarities and differences between them are definable, constant, and proportional"; and secondary ones "cannot be segmented into perceptually proportional relationships [and have] no specific closural states" (Meyer, 1989, p. 14). Within the Western tonal art music tradition for instance, elements such as rhythm, melody, and harmony are usually considered to be primary musical parameters, while others like timbre or dynamics are usually considered to be secondary. Meyer, however, also acknowledges that what constitutes a primary or secondary musical parameter is fluid. He considered primary parameters to be those that are governed by syntactic constraints, while secondary ones to be governed more by statistical constraints. This means that primary parameters are those that have been learned as rules, whereas secondary parameters still function within the laws of human auditory perception or statistical learning and have not yet been constrained into rules and syntax. This is not to say that secondary parameters cannot be learned syntactically, but rather, within the musical culture in question, it *has not* functioned as such. If stylistic change creates certain recurring patterns of secondary parameters with enough consistency and for a long enough period of time, a parameter that was originally secondary might become primary.

### 1.2.4 Emotional meaning and emotional responses to music

"Most major topics in psychology and every major problem faced by humanity involves emotion." It is, however, an extremely complex field of study with "deep controversies on fundamental questions" (Russell, 2003, p. 145). Emotional meaning in music is also an extremely complex area in musical communication.

In the field of human emotions and affect, psychologists hold differing views on the types of models that can best describe it. Paul Ekman (1992) argues for the existence of "universal basic emotions." He believes that emotions have a biological basis, and from studies looking at universals in facial expression, believes that there are basic emotions that are common across cultures. Studies on activities of the autonomic nervous system further point to the existence of basic emotions in humans.

Other researchers such as James Russell (1980) talk about a circumplex model of affect. Based on evidence from a lay conceptualization and multivariate analyses, affective states can be best represented on a circle in a two-dimensional bipolar space in this model, along the dimensions of pleasure-displeasure and degree of arousal. Ulrich Schimmack and Alexander Grob (2000) describe a three-dimensional model of pleasure-arousal-tension of core affect after comparing three-dimensional versus two-dimensional spaces for affect representation. They believe that at least three dimensions are necessary for explaining core affects. Two-dimensional models neglect variance from the arousal and tension dimensions by assuming them to accrue in only a single dimension. These three dimensions, however, are not likely to be orthogonal to one another. Although there are many compelling arguments for both the circumplex model and the three-dimensional model of affect, this study's focus is not on which models better explain human emotions; rather, it seeks to foreground emotions as one of the components that can be communicated via music. I will therefore be adopting Ekman's model of basic emotions as it offers a simple set of emotion descriptors that can be utilized for musical communication.

There is also the perennial dichotomy about induced and perceived emotions in music. Klaus Scherer (2004) believes that there are different mechanisms through which music can elicit emotions and also that the emotions induced by music may be very different from emotions experienced in real life. Vladimir Konečni (2008) on the other hand argues that the body of research supporting induced emotions in music is not convincing. "Basic emotions … may be induced … but probably only through the mediation of personal associations to emotionally compelling events" (p. 127). Patrik Juslin (2013) attempts to integrate these different approaches by proposing a framework with several mechanisms through which music might arouse emotions. In addition to these mechanisms, Juslin adds another dimension, aesthetic judgement; he believes that all these work in conjunction to form preference judgements. Taken together, the framework for emotions aroused by music, aesthetic judgement, and preference, can be used to account for induced and perceived emotions in music.

Music as a symbol can definitely hold many different types of meanings and offer different processes of signification. Emotional meanings and responses are but one, and as Nick Zangwill (2011) argues, "description of music in terms of emotion … is the metaphorical description of aesthetic properties of music" (p. 14). Regardless of whether music actually *possesses* emotions, the fact that it is possible to *represent* emotions in music, and also that it is an aspect that performers and listeners of music are able to describe, regardless of their level of musical expertise. Emotions appear to be a suitable dimension for use in the study of musical communication. It is also an area on which a great deal of research has focused. Meyer (1956), who also wrote at length about emotions in music, believes that because emotional behaviour may be subjected to different kinds of social conditioning, and because there is no “theory as to the relation of musical stimuli to affective responses, observed behaviour can provide little information as to either the nature of the stimulus, the significance of the response, or the relation between them” (pp. 12-13). Researchers on emotions in music have noted this, and Patrik Juslin and Daniel Västfjäll (2008) concluded in their framework for studying emotional responses to music that “amassing data on listeners’ emotional reactions to music is not fruitful, unless one is able to *interpret* these data in the light of an explanatory theory” (p. 575). In this study, my purpose is to understand the function of timbre in contributing to listeners’ understanding of emotions. I decided therefore that it is appropriate to use listeners’ explicit judgement on the emotional intent they think the performer is trying to express as I am attempting to acquire information concerning listeners’ perception of emotion intent in music.

#### 1.2.5 Timbre and communication of emotions in music

“Timbre” has been defined in many ways (Sandell, n.d.) and comprises a complex set of auditory attributes. For the purposes of this study, I use the recent definition by McAdams and Goodchild (2017): “a set of auditory attributes...that both carry musical qualities, and collectively contribute to sound source recognition and identification” (p. 129). I will focus on timbre’s function in carrying “musical qualities.”

Music perception scholars have also identified timbre as an important aspect in the communication of emotions in music. Herbert Spencer (1854/2015) proposed that vocal music, and by association instrumental music, is intimately related to vocal expression of emotions, and extending upon that, Patrik Juslin and Petri Laukka (2003) conducted a meta-analysis on vocal expression of emotions and musical expression looking at how vocal expression and musical expression are related. In contrast to Spencer's belief, they found that while vocal expression shares certain attributes with musical expression, it cannot account for all of music's expressiveness. There are certain acoustic cues that are specific to vocal expression, and others to music performances, that contribute to the communication of emotions. Tuomas Eerola, Rafael Ferrer, and Vinoo Alluri (2012) looked at the role of timbre in the perception of broad affect dimensions in music and found that brief isolated sounds were sufficient for affect ratings and that these affect ratings were moderately well explained using a small set of acoustic features such as attack slope, spectral characteristics, and spectral flux.

**1.3 Previous Studies**

Although Pierre Schaeffer did not specifically discuss timbre in the communication of emotions in music, his treatise on musical objects (1966) pioneered other writings in the phenomenology of sound. It opens up possibilities for looking at aspects of timbre and discussions about timbral descriptions and paved the way for timbre studies. Before establishing if timbre is integral in the communication of musical meanings, we have to ensure whether and what dimensions of timbre are perceptually salient. Multidimensional scaling is a class of techniques that uses listeners' ratings of how similar two objects or events are in order to create a spatial representation of their distances (Kruskal & Wish, 1978). John Grey (1977) was one of the first researchers to use multidimensional scaling on the perception of musical timbres and coined the concept of a “timbre space” that could be utilized in studying its perceptually salient dimensions. Three dimensions: temporal pattern of attack and decay, spectral pattern, and spectrotemporal pattern were found to be perceptually salient in his set of sounds derived from wind and string instruments. Carol Krumhansl (1989) also found three dimensions that accounted for listeners’ perception of timbre dissimilarities in a set of synthesized sounds: temporal envelope, spectral flux, and spectral envelope. Stephen McAdams, Suzanne Winsberg, Sophie Donnadieu, Geert De Soete, and Jochen Krimphoff (1995) also used multidimensional scaling on a subset of complex musical sounds from the Krumhansl study and found that a three-dimensional model, highly correlated with Krumhansl’s model, with additional “specificities” for individual timbres adequately accounted for listeners’ responses to timbre dissimilarities. These multidimensional scaling models suggest that a few dimensions relating to temporal, spectral, and spectrotemporal dimensions are perceptually salient and consistently used in timbre perception, although the dimensions found depend on the set of sounds employed.

Given its perceptual salience, another aspect that is important for timbre’s function in carrying musical meaning is whether listeners are able to utilize it in learning. David Ehresman and David Wessel (1978) used synthesized timbres and found that listeners were able consistently to perceive timbre analogies in a two-dimensional timbre space. Stephen McAdams and Jean-Christophe Cunibile (1992) tested this on a three-dimensional space and found a vector model to be fairly successful at prediction although it appears that varying amounts of musical training lead listeners to evaluate dimensions of timbre differently. Barbara Tillmann and Stephen McAdams (2004) looked at implicit learning of musical timbre sequences and found that listeners were able to learn statistical regularities with complex multidimensional auditory material.

As mentioned earlier, one problem in timbre studies is that it comprises numerous complex acoustic attributes. However, with evidence that a few dimensions are regularly and consistently utilized by listeners, the possibility of creating a "timbre space," the capability to perceive timbre analogies, and the ability to learn timbre sequences, timbre appears to have the potential to be segmented into discrete salient units so that they can form relationships that are definable, constant, and proportional. There is, therefore, the potential of timbre functioning as a primary musical parameter as defined by Meyer. If timbre is able to function as a primary musical parameter, it follows that patterns of how it is used to communicate musical meanings, including emotion intents, can be learned as rules within a particular musical tradition.

There has been a great deal of previous research on timbre and emotions. Some of these studies talk about timbre as derived from discrete sound events, while others look at it as having a relational aspect. Emmanuel Bigand, Sandrine Vieillard, François Madurell, Jeremy Marozeau, and Alice Dacquet (2005) used multidimensional scaling on dissimilarity ratings of emotional responses to musical excerpts. In addition to thirty-second excerpts, they also used short one-second excerpts and found that even one second was sufficient for listeners to perceive emotional cues. Eerola, Ferrer, and Alluri (2012) presented brief, isolated instrument samples and had listeners rate them on affect dimensions of valence, energy-arousal, tension-arousal, and preference. They found that participants were able to do so with consistency. Anders Askenfelt (1986) measured bow motion and force in violin playing over an entire excerpt as different moods are expressed. This study focused on the performer's, instead of the listener's, use of timbre. Performers on the violin were found to consistently utilize different bowing gestures in different moods. Alf Gabrielsson and Patrik Juslin (1996) and Juslin (1997) designed experiments in the standard content paradigm in which the encoder (performer) is asked to express different musical intents and the decoder (listener) listens to the recorded stimuli and judges the musical intent. The assumption of the standard content paradigm is that because the musical materials remain the same and only the intent is modified, the effects that appear in listeners' judgements and acoustic measures will be the result of the expressive intent. In their studies, they found overall accuracy and consistency in judgements of performers' intent by listeners over a range of instruments and expressive intentions. However, some instruments appeared less suitable for expressing certain emotions: there were tendencies to confuse certain emotions such as tenderness and sadness. These studies also looked at the amplitude envelopes and frequency spectra of the sounds but discussed differences mostly in qualitative terms. Sergio Canazza, Giovanni De Poli, Stefano Rinaldin, and Alvise Vidolin (1997) performed sonological analyses on clarinet performances and synthesized various dimensions of timbre to attempt to create expressive intentions (Canazza, De Poli, Rodà, & Vidolin, 1997). They found consistent patterns in the performances in terms of temporal variances, articulations, and spectral content in expressive clarinet performance. Mathieu Barthet, Philippe Depalle, Richard Kronland-Martinet, and Sølvi Ystad (2010) looked at a range of acoustic descriptors in expressive clarinet performance and found significant differences in several timbre descriptors in an expressive performance as compared to a mechanical (inexpressive) performance. However, in the study, they simply defined the performance to be "expressive" or "mechanical" for the performer and there were no specific musical intentions the performer was asked to express. Therefore, although we might say that timbre varies considerably in expressive performances, it is uncertain how and to what extent each timbre dimension may vary for a particular musical intent.

## 1.4 Current Study

### 1.4.1 Objectives

Meyer's musicological writings proposed a way of considering musical meanings as "psychological processes ingrained as habits in the perceptions, disposition, and responses of those who have learned through practice and experience to understand a particular style" and that the constants between styles reveal to us how "the mind, operating within the context of culturally established norms, selects and organizes the stimuli that are presented to it" (Meyer, 1994, p.7). Effective musical communication therefore requires a shared lexicon of learned psychological processes such that the process of signification from the signifier to the signified possesses certain similarities between the encoder and the decoder.

From the works cited above, communication in music can therefore be understood as a process in which both the encoder (which may include composers, performers, sound engineers, etc.), and the decoder (which may include performers, listeners, etc.) share a socially transmitted understanding of a common set of laws and rules about certain musical parameters that they use regularly and consistently such that the process of signification or meaning making is similar for both.

Previous studies all look at how performers and listeners utilize timbre in musical communication. There is however a lack of a set of studies specifically focusing on timbral manipulation in the communication of emotional intentions in music, or sonological analyses looking at the specific dimensions of timbre that are used in communicating emotion intents. At the same time, many current theories of music perception and cognition "posit that individuals use a process of 'statistical learning' to generate rule structures" (Morrison & Demorest, 2009, p. 68). Musical enculturation, or the "process by which individuals acquire culture-specific knowledge about the structure of the music they are exposed to through everyday experiences" (Hannon & Trainor, 2007, p. 466) therefore requires a sufficient amount of "statistical learning" to have taken place such that rule structures governing the understanding of music can be generated. There is, however, a lack of studies that examine whether timbre manipulations are specifically learned and whether they vary across different musical traditions.

### 1.4.2 Research questions and hypotheses

If expressive timbre rules are learned, both through enculturation and specific musical training, performers trained in different musical traditions will likely manipulate timbre differently in their performances to express different emotional intentions. Similarly, if these rules are learned, listeners enculturated in different musical traditions will likely have different understandings of the expressive intent of various timbral manipulations in performance. Although there will be differences in timbre utilization on different instrument types, I hypothesize that performers trained in different musical traditions (Western art music and Chinese music in the context of this study) will use timbre differently in expressing musical emotions. For instance, they might collectively use a brighter sound for a certain type of emotion or a sharper attack for another. They might also use different combinations of these timbre manipulations to varying degrees. Listeners enculturated in different musical will also have different understanding of the use of timbre in emotion expression such as attributing a brighter sound to a particular emotion. They might also differ in how they weight different aspects of timbre—listeners from one culture might emphasize the function of the attack in certain emotions or downplay the contributions of spectral flux for instance. My study aims to explore these two questions:

1. Do performers trained in different musical traditions, specifically Western art music and Chinese music, diverge in their use of timbre when expressing musical emotions on similar instrument types from the two musical traditions?
2. When listeners are enculturated in different musical traditions, will their understanding of how timbre is deployed differ?

Through a series of experiments, I hope to understand in quantitative terms, a) what information timbre provides in the communication of emotion intent in music, and b) whether aspects of timbre manipulation and recognition function in rule-based learning by musicians and listeners enculturated in different musical traditions. As musical communication involves the agency of the performer as well as the listener, both groups are included in my study.

I hypothesize that:

a) although the inherent timbre characteristics of instruments with different modes of sound production will be different, performers trained in the same musical tradition will converge on certain aspects regarding timbre manipulation for expressing certain emotions;

b) the various acoustic parameters that have previously been found to be salient for timbre perception will likely be systematically manipulated by performers to differing degrees—for instance, an angry emotion intent is likely to have a shorter and sharper attack, while a sad emotion intent is likely to be the opposite;

c) listeners trained in the Chinese musical tradition will be more successful in judging the emotion intent of excerpts played by musicians trained in the Chinese musical tradition; listeners trained in the Western musical tradition will be more successful in judging the emotion intent of excerpts played by musicians trained in the Western musical tradition; and nonmusicians will perform the least accurately but will be better at excerpts played by musicians trained in the Western musical tradition as this group of listeners will have had almost no prior exposure to Chinese music.

# CHAPTER 2
# Method

## 2.1 Pre-Experiment

### 2.1.1 Participants

Twenty listeners participated in this experiment (ten musicians and ten nonmusicians). Musicians were defined as having more than 5 years of formal music training (mean = 12.3, *SD* = 6.57), whereas nonmusicians were those with less than a year of formal music training (mean = .2, *SD* = .42). All the participants signed a written consent and met the required hearing threshold of 20 dB HL on a pure-tone audiometric test with octave-spaced frequencies from 125 Hz to 8 kHz (ISO 389-8, 2004; Martin & Champlin, 2000). They were compensated $10 for their participation. This study was certified for ethical compliance by McGill University's Research Ethics Board II.

### 2.1.2 Stimuli

In preparing stimuli for use in this study, thirty-two excerpts of film music (thirty as experimental stimuli and two as practice stimuli), selected from Eerola and Vuoskoski's study (2011) and Charlie Chaplin's films were used (Appendix 1). Eerola and Vuoskoski's study systematically compares perceived emotions in music, providing a set of excerpts that are appropriate in carrying various emotional intentions. Charlie Chaplin's film music from the silent film era also provides a set of excerpts with strong connotations of various stock emotions.

As all participants for Experiment 1 and participants trained in Chinese music (some of these participants were also trained in Western art music) for Experiment 2 were from Singapore, whereas participants trained in Western art music (none were trained in Chinese music) and nonmusicians for Experiment 2 were from Montreal, choosing either Western art music or Chinese music would therefore bias one or the other group of participants. Although excerpts from other musical traditions might have been chosen, I decided against this because they might not be easily performed on either the Chinese or Western instruments or both. Non-idiosyncratic performances might create extraneous factors that might play a part in biasing the communication of emotions. Additionally, excerpts from other musical traditions were not used because unfamiliar scales or tonalities might also introduce other confounding variables. Popular music, which is likely to be familiar to both groups, was also not the most ideal as lyrics in the songs might create unwanted confounds. Therefore, non-diegetic film music, used to heighten the emotional effect within the film, was chosen as the most suitable type of music for carrying emotions that could be easily performed on all the instruments. It was presumed to be of equal familiarity to participants from both Singapore and Montreal and would therefore create the least bias to any group of participants within this study. Although it is not possible to ascertain the exact amount of exposure and expertise each participant had with to music in Hollywood films, the current generation of Singaporeans have the same availability to such movies produced between 1930s to the present, as the population in Montreal. Therefore, it was decided that even though this music is familiar, there is little bias in terms of the level of familiarity with any of the groups of participants.

### 2.1.3 Procedure

Participants were asked to rate the excerpts presented on a Likert scale of one to seven, with one as "not at all" and seven "very much so," on each of the five emotion descriptors. They were instructed by the experimenter not to let the rating for one emotion affect that of another, i.e., they could rate an excerpt as 7 on both happy and sad if they felt that it expressed both. The thirty-two excerpts were presented in random order through OpenSesame (Mathôt, Schreij, & Theeuwes, 2012), a Python-based computer environment. Sounds stored on a MacBook Air laptop running on OSX (Apple Computer, Inc., Cupertino, CA) were presented over Sennheiser HD280 Pro earphones (Sennheiser Electronic GmbH, Wedemark, Germany) in the Recital Hall of the Nanyang Academy of Fine Arts in Singapore. In order to select excerpts that can express different emotion intents easily, five out of the thirty excerpts that had the lowest variance (Appendix 2) in ratings across the different emotion descriptors were chosen as stimuli for Experiment 1 (Appendix 3).

## 2.2 Experiment 1: Performance of Intended Emotions

### 2.2.1 Participants

Eighteen professional musicians from Singapore, three for each instrument, were recruited for this experiment. Two groups of musicians with different instrument cultures were employed. The first group included practitioners in the Chinese music tradition (hereafter termed "Chinese performers"), and the second group of musicians was from the Western art music tradition, (hereafter "Western performers"). The instruments played by these two groups were matched based on the mode of sound production for the instruments: aerophones, bowed chordophones, and plucked chordophones (hereafter "instrument category"). The Chinese instruments were the *dizi, erhu,* and *pipa,* respectively, and the corresponding Western instruments were the flute, violin, and guitar, respectively. All but one of the participants had more than 10 years of formal musical training on their particular instrument (that participant had 7 years of formal training on the flute), with the Chinese group having slightly more years of musical training (mean years = 18.1, SD = 6.29) than the Western group (mean years = 14.7, SD = 4.21), although the difference was not statistically significant ($t(16) = 1.36$, $p = .19$). Each recording session took from 1 to 1.5 hours, and all the participants signed a written consent form and were compensated $15 for their time. This study was certified for ethical compliance by McGill University's Research Ethics Board II.

### 2.2.2 Stimuli

From the film-music recordings of the five excerpts selected in the pre-experiment, the main melody line was transcribed for use in the rest of the study (Appendix 1). To avoid having an experiment that was too long, only three of the five emotion descriptors from the pre-experiment were used in Experiment 1—angry, happy, and sad—as studies have shown that these emotions are most reliably identified in musical performances (Gabrielsson & Juslin, 1996; Terwogt & van Grinsven, 1991). These will hereafter be termed "emotion intent."

### 2.2.3 Procedure

Four recordings were made of each of the five excerpts for each musician, expressing angry, happy, sad, and a neutral emotion intent. A total of 360 recorded excerpts were thus obtained. Each performer was given the scores of the five excerpts at least a week prior to the recording day. All the recording sessions were conducted in the Recital Hall of the Nanyang Academy of Fine Arts in Singapore, a room with soundproof insulation appropriate for sound recording. Excerpts were recorded with a Zoom H4n Handy Recorder (Zoom Corporation, Tokyo, Japan). Performers were instructed to express the music according to their understanding of the emotions angry, happy, and sad, as well as to render a fourth one in a neutral way. Participants were told that they could record a given excerpt as many times as they wished and then decide with which performance for each emotion category they were most satisfied. That recording was used in subsequent parts of the study. Interpretation of what each emotion descriptor meant was left open to the performer, and most did not ask what was implied by each emotion. Only one of the performers asked what kind of sadness was required, and he was told to just express whatever he thought sadness meant to him. For the neutral emotion intent, performers were asked to play what was indicated on the score without any additional interpretation of the music on their part. This was to be used as a control.

## 2.3 Experiment 2: Perception of Intended Emotion

### 2.3.1 Participants

Ninety participants were recruited. Thirty were musicians trained in the Chinese music tradition, thirty were musicians trained in the Western music tradition, and thirty were nonmusicians. The Chinese musician listeners were all from Singapore while the Western musician listeners and the nonmusician listeners were from Montréal with the majority having a Western background. Only one of the participants is an international student from Korea. These participant groups will hereafter be termed “Chinese musician listeners,” “Western musician listeners,” and “nonmusician listeners.” Similar to the group of participants in Experiment 1, musicians were those who had more than five years of formal musical training in either the Chinese tradition (mean = 12.4, SD = 4.52) or the Western tradition (mean = 13.2, SD = 5.53).

Nonmusician listeners (mean = .17, SD = .38) were those with less than a year of formal training in any form of music. There was no significant difference between the number of years of musical training between the Chinese and Western musician listeners ($t(58) = -.59$, $p = .56$). All the participants signed a written consent form and met the required hearing threshold of 20 dB HL on a pure-tone audiometric test with octave-spaced frequencies from 125 Hz to 8 kHz (ISO 389-8, 2004; Martin & Champlin, 2000). They were compensated $15 for their participation. This study was certified for ethical compliance by McGill University's Research Ethics Board II.

#### 2.3.2 Stimuli

If recordings from all the performers were used, the experiment would be too long for listeners, with subsequent problems of listening fatigue and lack of attention (Paus et al., 1997). It was therefore decided to use only recordings from a single musician for each instrument. Recordings were chosen from the performer with the highest number of years of musical training. The "neutral" excerpts were not included in this listening experiment as they had only been intended for use as a control in the acoustic analyses. With six different instruments, five different excerpts, and three different emotion intents, there were a total of 90 stimuli for this experiment. A further three were selected from outside these 90 excerpts to be used as practice stimuli. Prior to running the experiment, three volunteers participated in a loudness-matching experiment in which the loudness of all the different instruments playing the same emotion intent and excerpt were matched. This is crucial in the experiment as the loudness of a stimulus might influence the perception of emotion intent (Kamenetsky, Hill, & Trehub, 1997). Median loudness values were calculated across the three volunteers and were used to adjust the stimulus levels so that the loudness across the different instruments were equalized.

### 2.3.3 Experimental design

The experiment was a mixed-measure design with one between-subjects factor—three listener groups (Chinese musician listeners, Western musician listeners, and nonmusician listeners)—and four within-subjects factors—five different musical excerpts, three emotion intents (angry, happy, and sad), two instrument cultures (Chinese and Western), and three instrument categories (aerophones, bowed chordophones, and plucked chordophones).

### 2.3.4 Procedure

The experimental session was run with the PsiExp computer environment (Smith, 1995). For the part of the experiment conducted with participants in Singapore, sounds stored on a MacBook Air running on OSX (Apple Computer, Inc., Cupertino, CA) were presented over Sennheiser HD280 Pro earphones (Sennheiser Electronic GmbH, Wedemark, Germany). Participants were seated within the same sound-proof space used for recording in Experiment 1. For the other part of the experiment conducted with participants in Montréal, sounds were stored on a Mac Pro computer running on OSX (Apple Computer, Inc., Cupertino, CA), amplified through a Grace Design m904 monitor (Grace Digital Audio, San Diego, CA), and presented over Sennheiser HD280 Pro earphones. Participants were seated in an IAC model 120act-3 double-walled audiometric booth (IAC Acoustics, Bronx, NY). Sound levels were measured with a Bruel & Kjær Type 4153 artificial ear to which the headphones were coupled (Bruel & Kjær, Nærum, Denmark) to ensure similar levels of presentation for both the experiments conducted in Singapore and in Montréal.

Participants were briefly instructed on the procedure and then exposed to three practice trials, during which they could clarify any questions about the experimental procedure before the start of the actual experiment. The instructions for the participants specifically asked that they judge the emotion they thought the performer of the recorded sound was trying to express, instead of any actual emotions they were feeling. As seen in Fig. 2-1, the participants were presented with three emotion descriptors from which they had to select one after the presentation of each stimulus. Each stimulus was presented only once, and within the experiment, the order of presentation of the stimuli was randomized. The positions of the three boxes for the emotion intents were also randomized across listeners. When participants clicked on the emotion they thought best described the performer's intention, the experiment automatically proceeded to the next stimulus. The entire experiment was divided into three experimental blocks and participants were told they could take a break in between the blocks if they so wished.

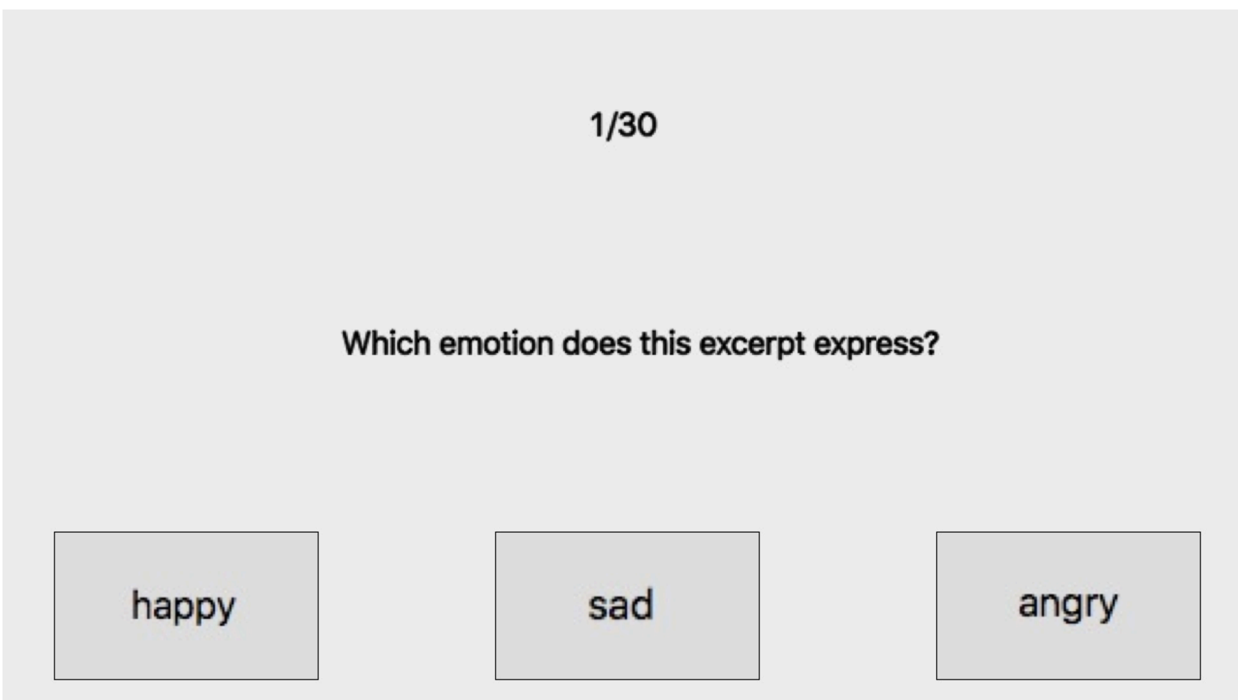


*Figure 2-1*. Experiment 2 interface.

## 2.4 Acoustic Analysis

### 2.4.1 Procedure

Individual notes from the set of sound recordings used for Experiment 2 were extracted by looking for the note onsets in Audacity® (2016). Each note comprises the duration from its onset to the next note's onset (Fig. 2-2). These extracted notes were then analyzed for the energy envelope, spectral, and spectrotemporal descriptors of the sound using the Timbre Toolbox in MATLAB (http://www.cirmmt.org/l/research-tools/timbretoolbox/).

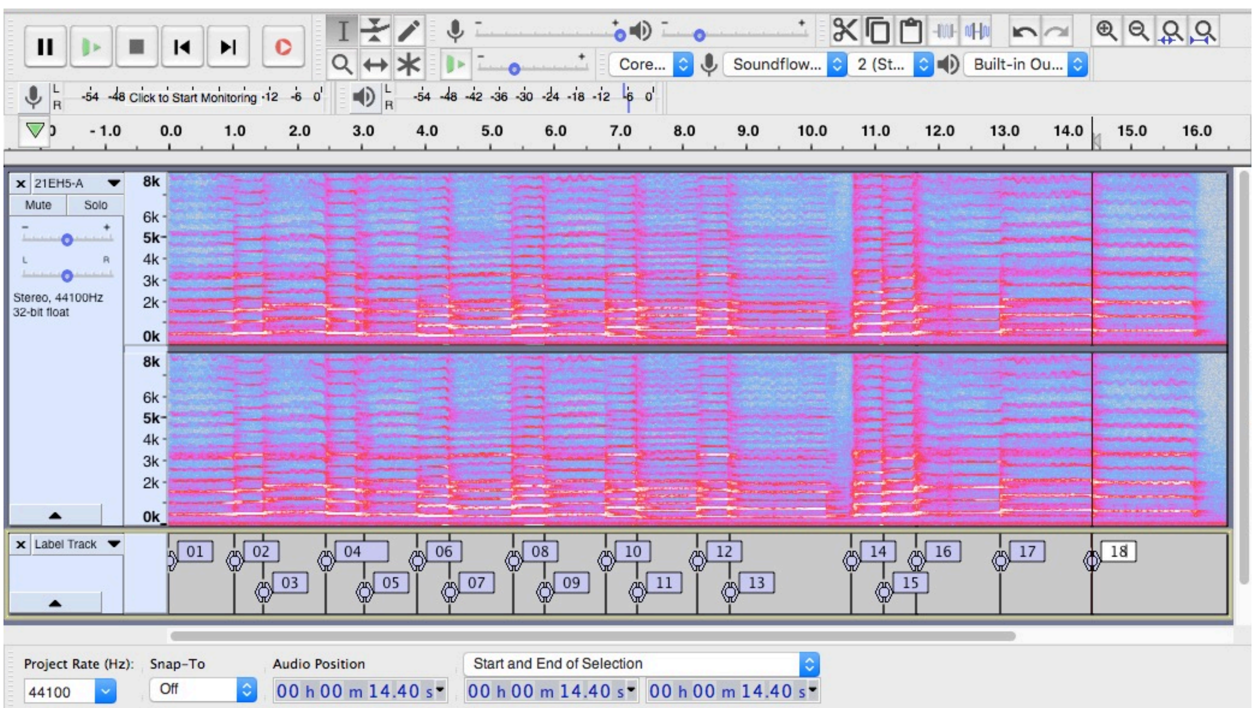


*Figure 2-2*. Note onset positions in Audacity® from a spectrogram of an excerpt.

### 2.4.2 The Timbre Toolbox

The Timbre Toolbox "provides a comprehensive set of [audio] descriptors that can be useful in perceptual research" (Peeters, Giordano, Susini, Misdariis, & McAdams, 2011, p.

2902). Log attack time is estimated by taking the logarithm of the duration from the start to the end of the attack segment while attack slope is the average temporal slope of the energy envelope during this attack segment. The decrease slope measures the rate of decrease of the signal's energy envelope, the temporal centroid is the centre of gravity of the energy envelope, and the effective duration measures the the duration of the sound above a perceptually salient threshold and may differ from the actual notated durations—staccatos will have a much shorter effective duration than legatos even when their notated values are identical. Decrease slope, temporal centroid, and effective duration can also be used to distinguish percussive from sustained sounds. Energy modulation measures the amplitude and frequency of modulation in the energy envelope, providing information about possible tremolos and vibratos. Log attack time, attack slope, decrease slope, temporal centroid, effective duration and energy modulations are all audio descriptors that characterize the temporal envelope of the whole sound event. Spectral centroid, spectral spread, and spectral flux are time-varying descriptors of the shape of the frequency spectrum that are computed using a short-time Fourier transform with a window function. Spectral centroid is the centre of gravity of the spectrum and the spectral spread represents the spread of the spectrum around its mean value. Spectral flux represents the amount of variation of the spectrum over time. As the different notes extracted all had different frequency content due to the different pitches in the melodies, the normalized spectral centroid was calculated so that a meaningful comparison could be made across pitches, because spectral centroid varies systematically with fundamental frequency.

# CHAPTER 3

# Results

Results from Experiment 2 will be presented first before those of the acoustic analyses on the recorded sounds from Experiment 1. This is because part of the acoustic analyses will depend on the listeners' responses from Experiment 2.

## 3.1 Experiment 2: Listening Results

### 3.1.1 Main effects

#### *3.1.1.1 Organization of data and analyses for main effects*

Listeners' responses provided a set of categorical data which were first coded in terms of their correspondence with the performers' emotion intent. The number of correct responses were grouped according to the five independent variables: three listener groups (Chinese musician listeners, Western musician listeners, and nonmusician listeners), five musical excerpts, three emotion intents (angry, happy, and sad), two instrument cultures (Chinese and Western musical traditions), and three instrument categories (aerophones, bowed chordophones, and plucked chordophones). The Kruskal-Wallis test on ranks used for between-subjects measures is a nonparametric method of performing an analysis of variance to compare whether the data in to-be-compared conditions come from the same distribution, thus allowing us to find out if there is a significant effect from each of the variables. The Friedman test, another nonparametric method of performing an analysis of variance, was used for the within-subject measures. Post-hoc tests to further compare between two specific groups were also performed, using the Mann-Whitney test for between-subjects measures and the Wilcoxon Signed Ranks test for within-subject measures. The Bonferroni-Holm method was used to adjust the alpha-level for the multiple pair-wise comparisons.

#### *3.1.1.2 Between listener groups*

As seen in Fig. 3-1, the Chinese musician listeners had a higher proportion of correct responses followed by the Western musician listeners. Nonmusician listeners had the lowest proportion of correct responses in judging the emotion intent of the performer. The differences among listener groups were significant, $\chi^2(2, N = 90) = 35.57$, $p < .05$. Further pairwise comparisons also revealed a significant difference between each of the pairs of groups: the comparison between Chinese musician listeners and Western musician listeners ($U = 168$, $p < .025$) had a large effect size ($r = .54$), that between Chinese musician listeners and nonmusician listeners ($U = 81.5$, $p < .0167$) also had a large effect size ($r = .70$), and the one between Western musician listeners and nonmusician listeners ($U = 278.5$, $p < .05$) had a medium effect size ($r = .33$).

The significant difference in the overall number of correct responses by the Chinese musician listeners as compared to the Western musician listeners suggests a combination of factors related to the different musical traditions, as well as other adjunct factors that may be a result of socio-cultural influences. The significant difference between Western musician listeners as compared to nonmusician listeners, however, can likely be attributed to differences in musical training and exposure, as both groups were from Montréal. The Western musician group and the nonmusician group both reported listening to a variety of popular and jazz music. Western musician listeners, however, almost all reported listening to Western classical music, whereas only two nonmusician listeners reported listening to Western classical music. This implies effects of timbral manipulation that are learned in musical training, practice, and passive exposure as Western musician listeners perform more accurately in the identification of emotion intent than do nonmusician listeners.

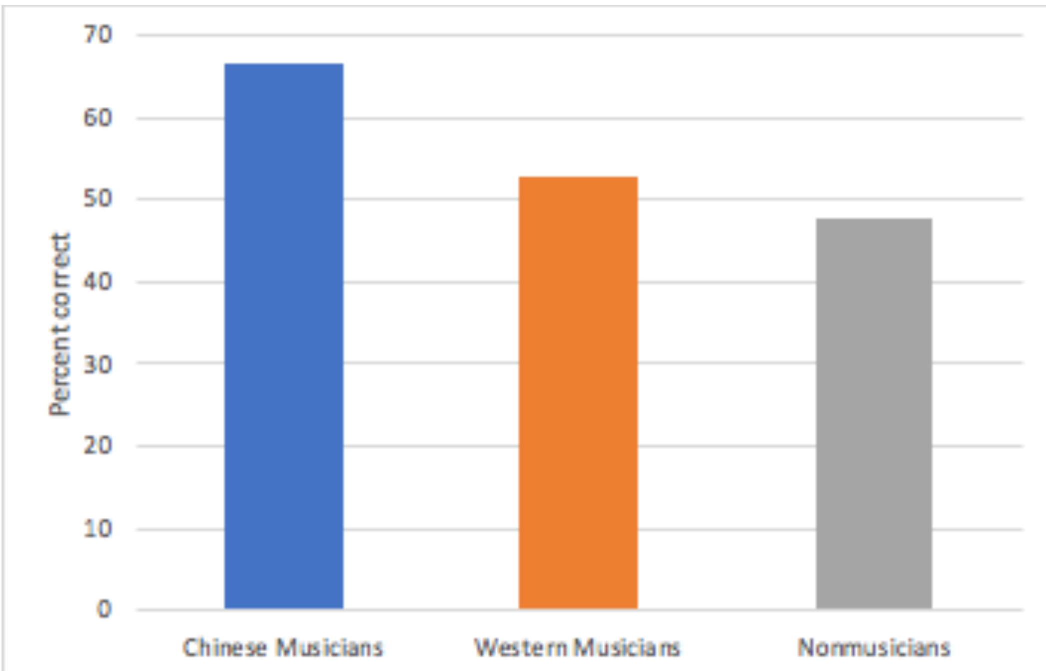


*Figure 3-1*. Percentage of correct responses by listeners.

#### *3.1.1.3 Between performer's instrument cultures*

There was a significant difference between listeners' overall correct responses for the Chinese instruments and Western instruments ($Z = -3.73$, $p < .05$) with excerpts played by Western instruments more often correctly identified than those played by Chinese instruments, although the effect size was small ($r = .28$). Within the Chinese musicians, there was no significant difference between correct responses for the excerpts played by Chinese or Western instruments ($Z = -1.44$, $p = .150$) (see blue bars in Fig. 3-2). Both the Western musicians (orange bars in Fig. 3-2) and the nonmusicians (grey bars) had a larger proportion of correct responses in the performances on a Western than on a Chinese musical instrument. Significant differences were found between the excerpts played by the different instrument cultures: Western musicians ($Z = -2.38$, $p < .05$) had a medium effect size ($r = .31$), and nonmusicians ($Z = -2.40$, $p = .016$) also had a medium effect size ($r = .31$).

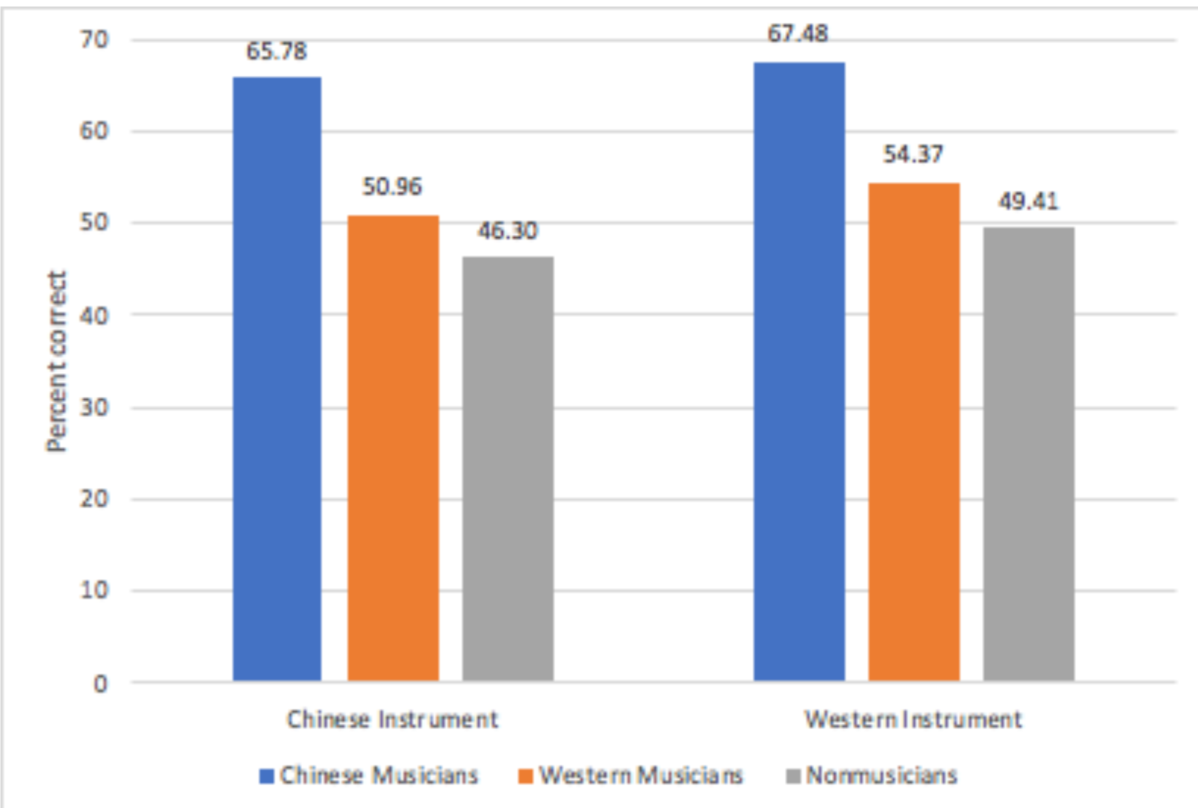


*Figure 3-2*. Percentage of correct responses by listener groups according to performer's instrument culture.

As all Chinese musician listeners had varying degrees of exposure to Western art music, the comparable accuracy in excerpts played by instruments from the two musical traditions might reflect their musical enculturation in both. The presence of significant differences in accuracy for judgement of emotion intent in both the Western musician listeners and the nonmusician listeners, with both groups demonstrating greater accuracy in the excerpts played by Western instruments as compared to those played by Chinese instruments, suggests that musical training contributes significantly to the accuracy of decoding a performer's emotion intent. For the excerpts performed on Chinese instruments, Western music listeners also had a significantly greater number of overall correct responses than nonmusician listeners. This implies that there are commonalities between these two musical traditions' use of timbre in expressing and decoding musical intents, commonalities that are learned through musical training.

#### *3.1.1.4 Between performer's emotion intents*

Different emotion intents elicited significant differences in their proportion of accurate identification, $\chi^2(2, N = 90) = 75.98$, $p < .05$. As shown in Fig. 3-3, performances having a "Sad" emotion intent were most accurately identified across all the participants while those having a "Happy" emotion intent were the least accurate. Pairwise comparisons showed significant differences between all the emotion intents: Angry and Happy ($Z = –5.58$, $p < .05$, $r = .42$), Angry and Sad ($Z = –5.95$, $p < .025$, $r = .44$), and Happy and Sad ($Z = –7.81$, $p < .0167$, $r = .58$). Each excerpt was performed with all the emotion intents to control for the notated rhythmic and melodic content. Expressive tuning and timing, however, could be manipulated by the individual performers. The higher accuracy for the Sad emotion intent may mean that the factors that performers are able to control here such as timbral manipulations, expressive tuning, and tempo, are utilized more in the communication of a Sad emotion intent. The other factors that the performer is unable to manipulate in this study, such as the melody, may play a more important role in helping listeners differentiate between the expression of Happy or Angry emotion intent.

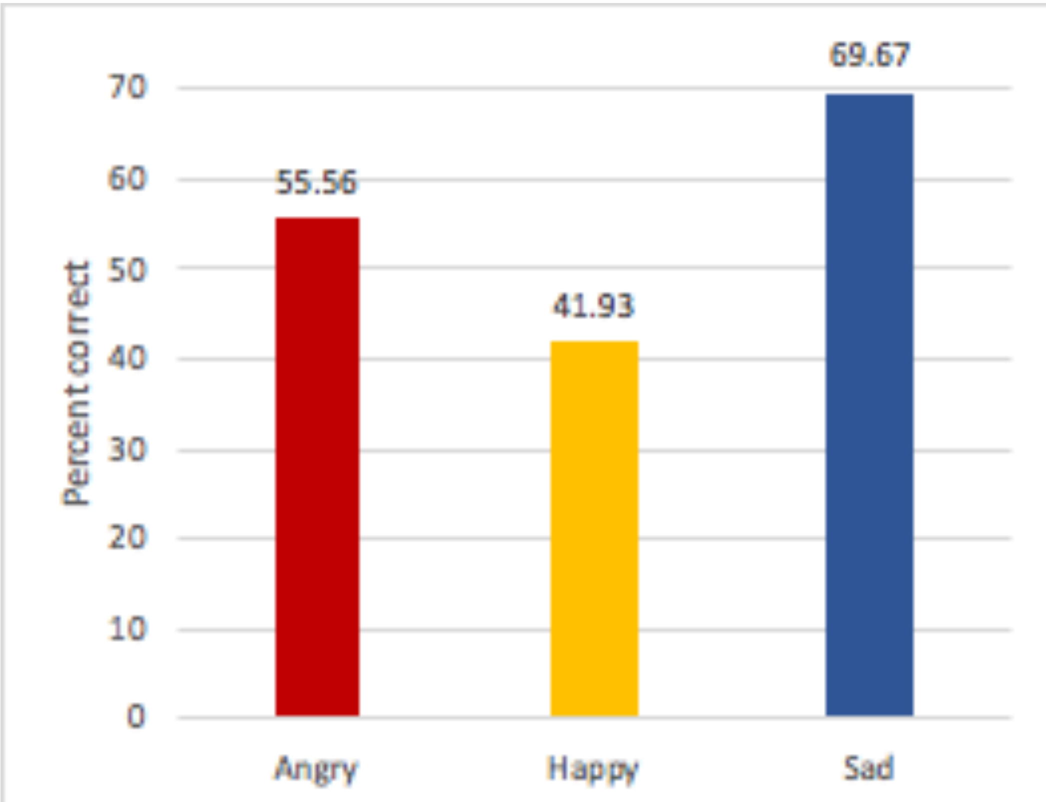


*Figure 3-3*. Percentage of correct responses by performer's emotion intent.

***3.1.1.5 Between performer's instrument categories***

There were significant differences in correct responses for the different instrument categories, $\chi^2(2, N = 90) = 55.64$, $p < .05$. Pairwise comparisons revealed significant differences between all the instrument category pairs: Aerophones and Bowed chordophones ($Z = –4.55$, $p < .025$, $r = .34$), Aerophones and Plucked chordophones ($Z = –3.72$, $p < .05$, $r = .28$), and

Bowed chordophones and Plucked chordophones ($Z = -7.00$, $p < .0167$, $r = .52$). The bowed chordophones were the most accurate in eliciting the correct responses for emotion intent followed by the aerophones and then the plucked chordophones as shown in Fig. 3-4.

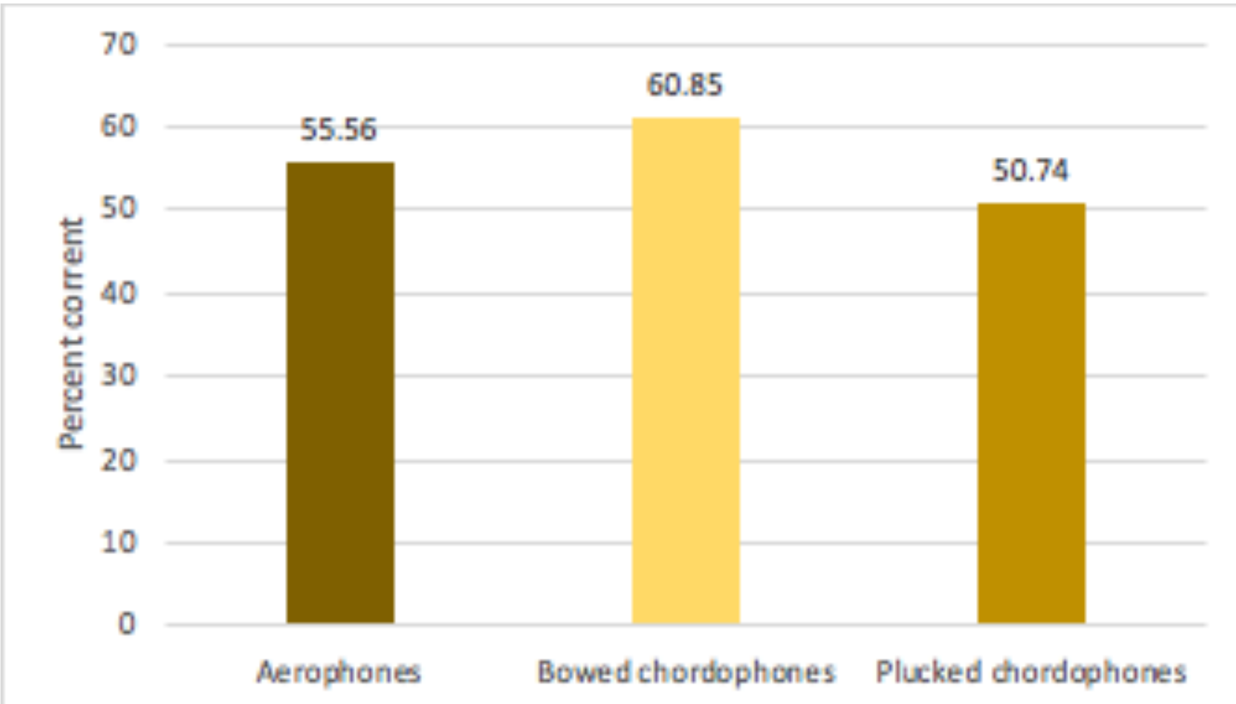


*Figure 3-4*. Percentage of correct responses by instrument category.

### 3.1.2 Analysis for interaction effects

The Aligned Rank Transform procedure by Wobbrock, Findlater, Gergle, and Higgins (2011) was created to circumvent the problem of Type I errors (false acceptance of a significant difference) for interaction effects in the Rank Transform method by Conover and Iman (1981). This is a method of stripping estimates of effects of other factors away to reveal the possible effect of a single factor, after which the new values of the dependent variable are ranked and then the effect of this single factor can be examined using an analysis of variance. The process is repeated for each of the factors of interest (Wobbrock et al., 2011).

### 3.1.3 Two-way interaction effects

There was no significant effect of the interaction between listener groups and instrument cultures, $F(2, 1479) = .058$, $p = .94$, nor of the interaction between listener groups and instrument categories, $F(2, 1479) = 1.02$, $p = .39$. The lack of significant interactions suggests that these factors contribute independently to the accuracy in judgement of emotion intents.

#### *3.1.3.1 Interaction between instrument cultures and instrument categories*

A significant interaction was found between instrument culture and instrument category, $F(2, 1479) = 3.51$, $p < .05$ (Fig. 3-5). Bowed chordophones appeared most effective in communicating the emotion intent of the performer, with the Western bowed chordophone (violin) having higher recognition performance than the Chinese bowed chordophone (*erhu*). The Western plucked chordophone (guitar) on the other hand was the least effective in communicating the emotion intent. Pairwise comparisons showed a significant difference between the violin and *erhu* ($Z = 4.47$, $p < .0056$, $r = .47$), but no significant difference between the Chinese and Western aerophones or plucked chordophones.

This interaction effect between instrument cultures and instrument categories implies that the accuracy in judgement of emotion intent for different instrument categories for a particular instrument culture does not vary uniformly with that of the other instrument culture. Bowed chordophones in the Western music tradition have the highest accuracy in judgements of emotion intent, whereas Western plucked chordophones have the lowest. It is not simply the instrument culture that contributes to a higher accuracy in judgement of emotion intent, but a combination of the mode of sound production and the instrument culture. The expression of emotion intents on a particular instrument involves not only learned timbral manipulations specific to a musical tradition, but also very specific kinds of techniques in performance unique to an instrument.

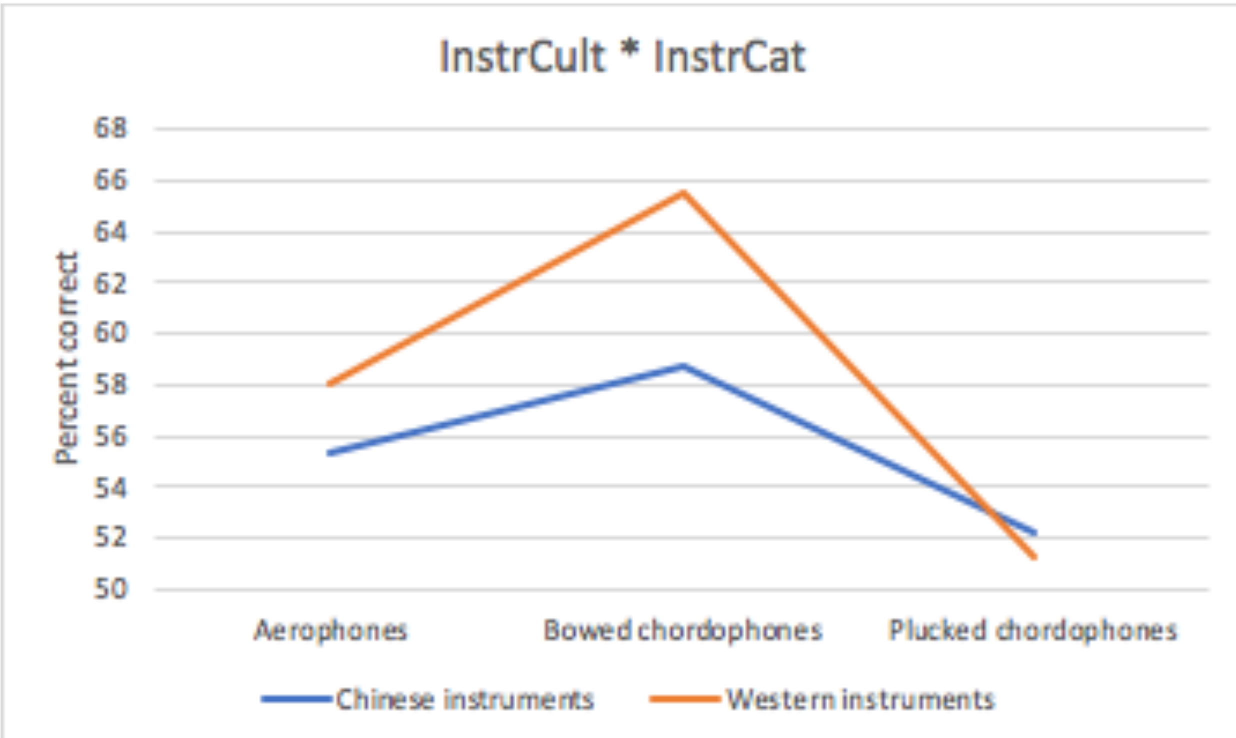


*Figure 3-5*. Interaction effects between performer's instrument culture and instrument category.

#### *3.1.3.2 Interaction between listener groups and emotion intents*

There was significant interaction between listener groups and emotion intents of performers, $F(4, 1479) = 7.33$, $p < .05$. Although Chinese musicians performed consistently better than the other two groups, the effect did not vary uniformly across the different emotion intents as seen in Fig. 3-6. Post hoc pairwise comparisons showed that in the Angry emotion intent, Chinese musicians performed significantly more accurately than did Western musicians ($U = 143.5$, $p < .006$, $r = .59$) or nonmusicians ($U = 60.0$, $p < .006$, $r = .75$). There were also medium effect sizes in the Happy emotion intent between the Chinese and Western musicians ($U = 280.0$, $p < .013$, $r = .33$) and the Chinese musicians and nonmusicians ($U = 257.5$, $p < .008$, $r = .37$). Similarly, there were also significant differences in the Sad emotion intent between the Chinese musicians compared to the Western musicians ($U = 263.5$, $p < .01$, $r = .36$) and to the nonmusicians ($U = 176.5$, $p < .006$, $r = .52$).

Western musicians were significantly more accurate than nonmusicians ($U = 285.0$, $p < .017$, $r = .32$) only in the Angry emotion intent. For both the Happy and Sad emotion intents, however, there were no significant differences between the Western musicians and nonmusicians: Happy ($U = 424.0$, $p = .70$), Sad ($U = 342.0$, $p = .11$).

The difference in effect size between Chinese musicians and the other groups for the Angry emotion intent was greater than that for Happy or Sad. Chinese musicians therefore appear to be much better at judgements of intended Angry emotion as compared to the other groups. Western musicians were also better than nonmusicians at recognizing the Angry emotion intent, although the effect size was smaller than that between Chinese musicians and both Western musicians and nonmusicians. Musical training therefore appears to contribute in some way to listeners' recognition of emotion intent.

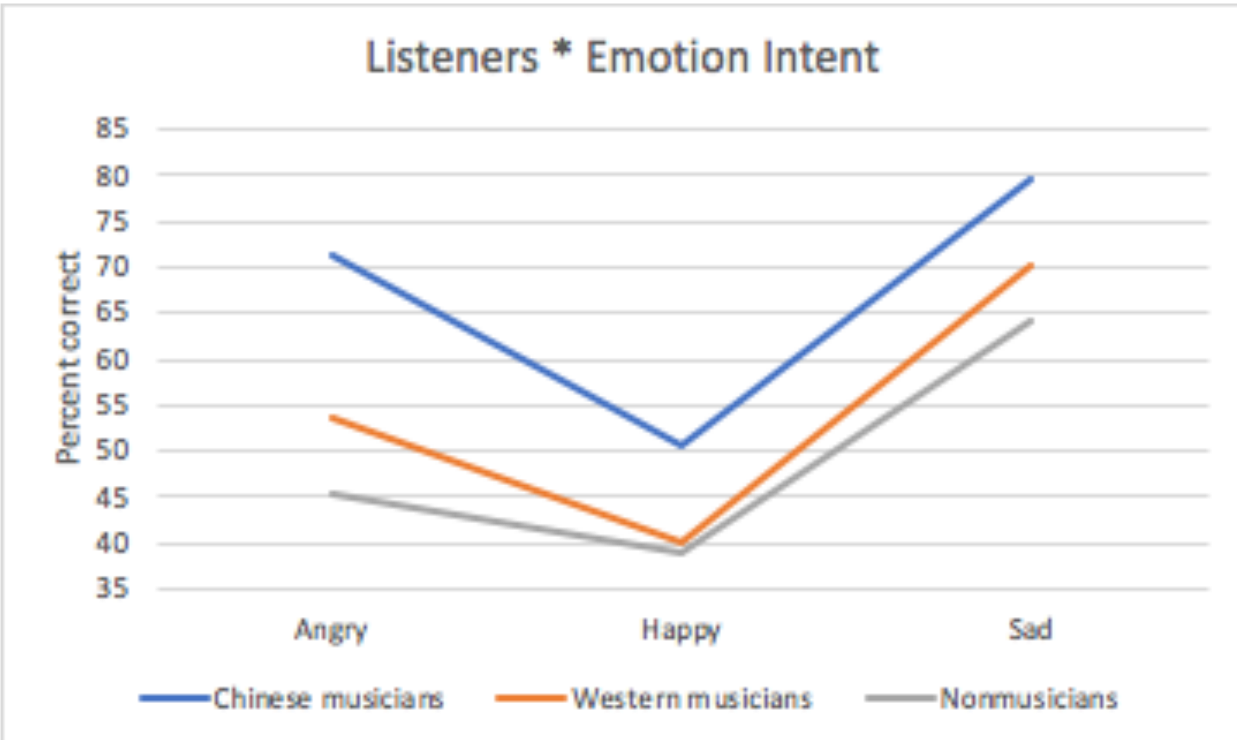


*Figure 3-6*. Interaction effects between listener groups and performer's emotion intent.

Interestingly, both the Happy and Sad emotion intents only showed a significant difference between the Chinese musicians and the other two groups, but not between the Western musicians and nonmusicians. Within each listener group, the Sad emotion intent on the whole is much more accurately identified than the Happy or Angry emotion intents.

#### *3.1.3.3 Interaction between instrument cultures and emotion intents*

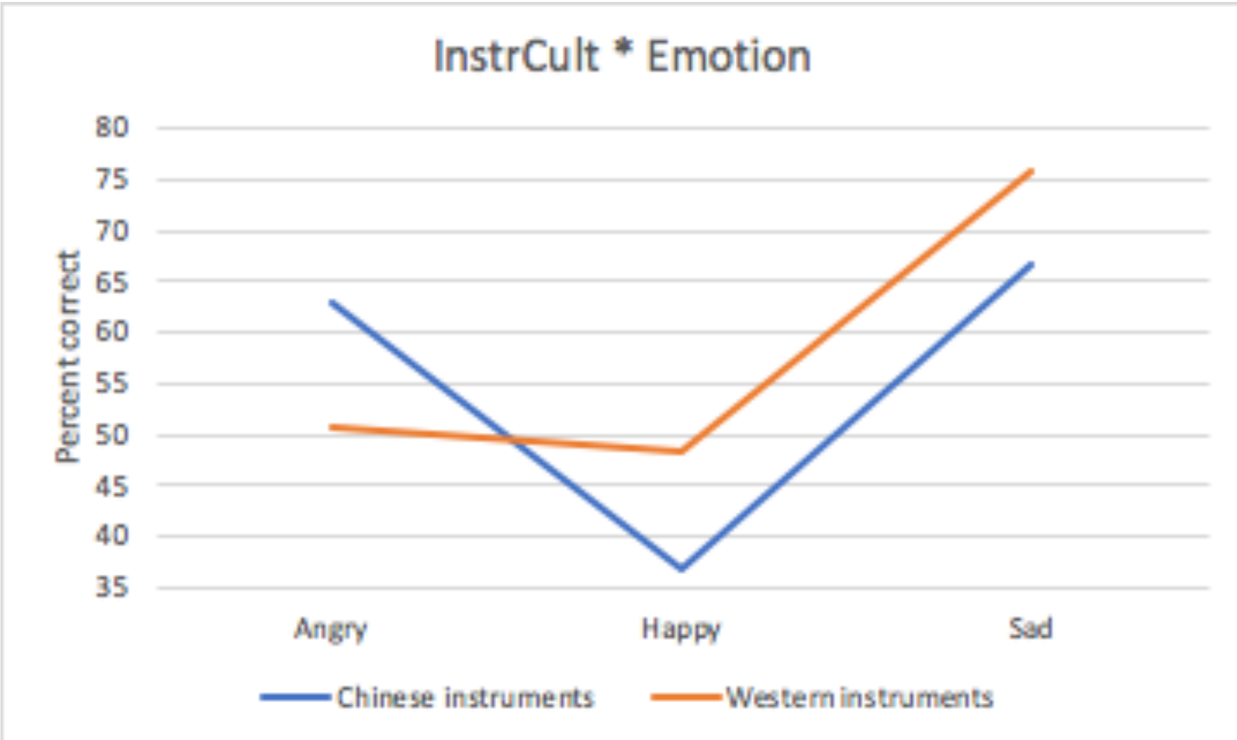


*Figure 3-7*. Interaction effects between performer's instrument culture and performer's emotion intent.

A significant interaction was found between instrument cultures and emotion intents of performers, $F(2, 1479) = 57.45$, $p < .05$ (Fig. 3-7). Chinese instruments were more effective at communicating an Angry emotion intent than were Western instruments, whereas Western instruments were more effective in the Happy and Sad emotion intents. Pairwise comparisons showed a large effect size in the differences between the instrument culture for all the emotion intents: Angry ($Z = –5.89$, $p < .003$, $r = .62$), Happy ($Z = –5.49$, $p < .003$, $r = .58$), and Sad ($Z = –5.25$, $p < .003$, $r = .55$).

### *3.1.3.4 Interaction between instrument categories and emotion intents*

Different instrument categories appeared to have significantly different effectiveness in communicating different emotion intents, $F(4, 1479) = 30.86$, $p < .05$ (Fig. 3-8). Bowed chordophones were better than the other two instrument categories in the Angry and Sad emotion intents but worst in the Happy emotion intent.

In the Angry emotion intent, pairwise comparisons revealed a highly significant difference between bowed chordophones and aerophones ($Z = –5.56$, $p < .0014$, $r = .59$), a significant difference between plucked chordophones and aerophones ($Z = –3.32$, $p < .013$, $r = .35$), and a highly significant difference between plucked and bowed chordophones ($Z = –7.05$, $p < .0014$, $r = .74$). Plucked chordophones performed the worst in the Angry emotion intent.

In the Happy emotion intent, bowed chordophones performed the worst. Pairwise comparisons showed the difference between bowed chordophones and aerophones to have a large effect size ($Z = –4.79$, $p < .0014$, $r = .51$), and the difference between plucked and bowed chordophones to have a medium effect size ($Z = –3.03$, $p < .017$, $r = .32$), whereas the difference between plucked chordophones and aerophones ($Z = –1.60$, $p = .11$) was not statistically significant.

In the Sad emotion intent, the difference between bowed chordophones and aerophones ($Z = –6.02$, $p < .0014$) had a large effect size ($r = .63$), the difference between plucked and bowed chordophones ($Z = –6.18$, $p < .0014$) also had a large effect size ($r = .65$), whereas the difference between the plucked chordophones and aerophones was not statistically significant ($Z = -.84$, $p = .40$).

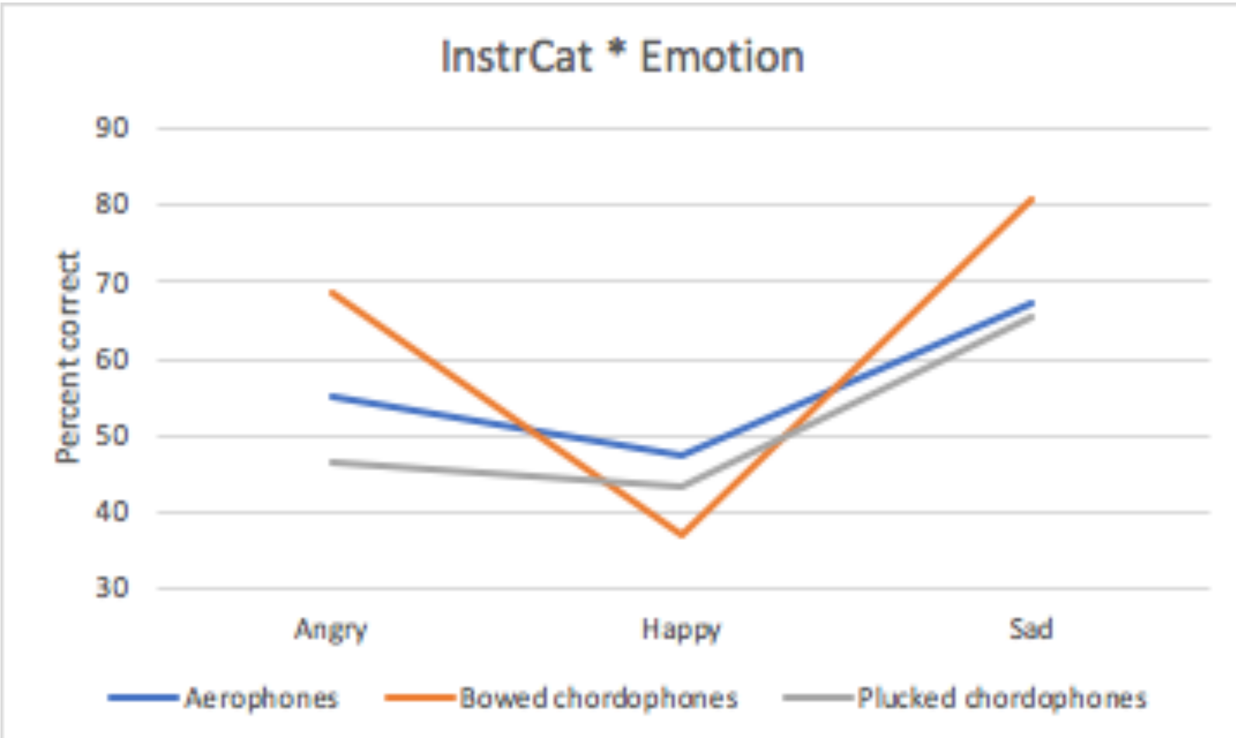


*Figure 3-8*. Interaction effects between performer's instrument category and performer's emotion intent.

These interaction effects between the instrument categories and emotion intents also bring up an interesting observation. Although bowed chordophones are the most accurate overall in communicating emotion intent, they fare the worst in communicating a Happy emotion intent. Instruments with different modes of sound production therefore appeared to vary with respect to the emotion intent they are more successful at expressing.

### 3.1.4 Three-way interaction effects

#### *3.1.4.1 Interaction between listener groups, instrument cultures, and instrument categories*

There was a significant interaction effect for listener groups x instrument cultures x instrument categories, $F(4, 1479) = 4.66$, $p < .05$. As seen in Fig. 3-9, the effectiveness of different instrument categories and instrument cultures varies across listener groups.

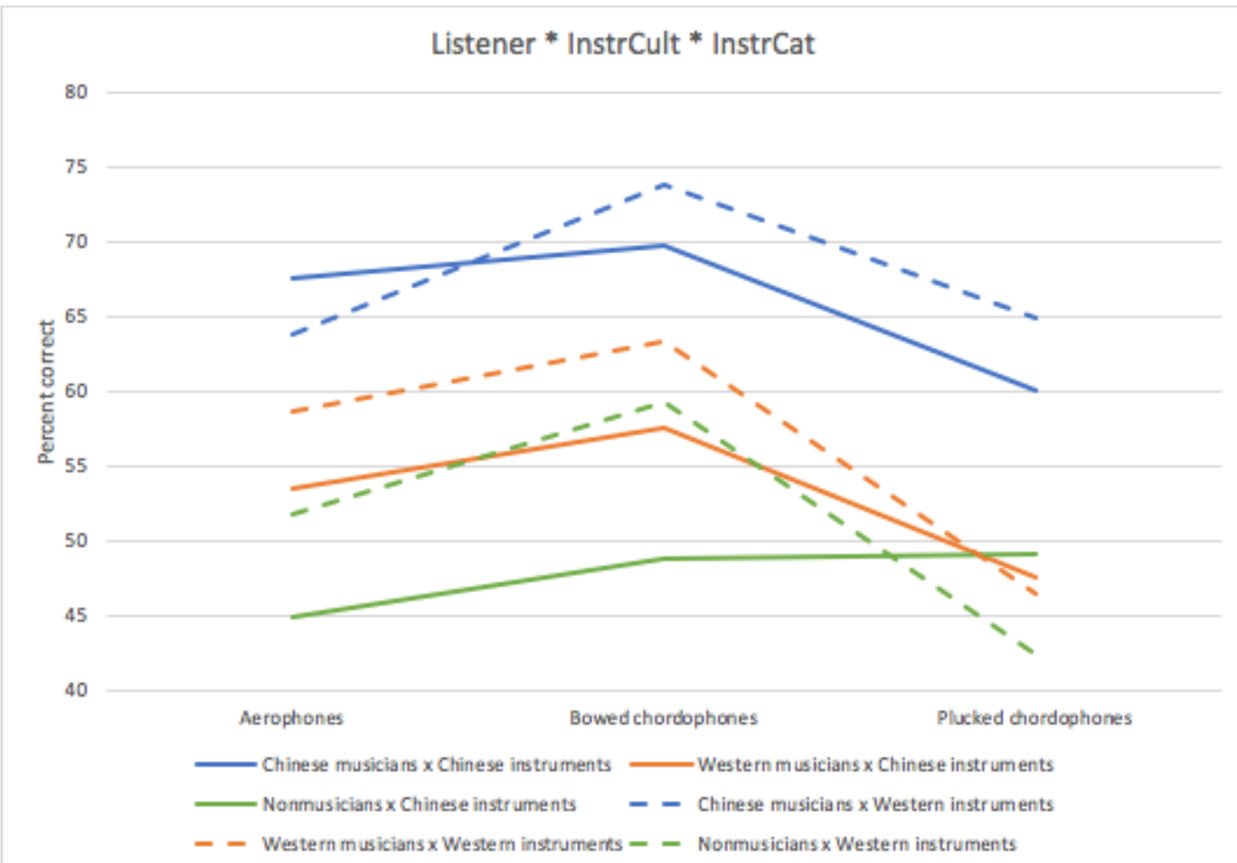


*Figure 3-9*. Three-way interaction effects between listener groups (different colours), performer's instrument culture (line texture), and instrument category.

Chinese musicians were generally more accurate than the other two groups of listeners (see blue lines in Fig. 3-9). They were better at the stimuli performed on Chinese aerophones than Western aerophones and better at stimuli performed on Western plucked chordophones than the corresponding Chinese ones. Interestingly, this effect was reversed in the other two groups of listeners—both the Western musicians and nonmusicians did better on stimuli performed on Western aerophones and plucked chordophones than on their Chinese counterparts. All the three listener groups were more accurate in judging emotion intent on Western than on Chinese bowed chordophones.

Both the Western musicians and nonmusicians perform worse on Western as compared to Chinese plucked chordophones, but interestingly, the Chinese musicians performed better on the Western plucked chordophone rather than its Chinese counterpart.

***3.1.4.2 Interaction between listener groups, instrument cultures, and emotion intents***

A significant interaction effect was found for listener groups x instrument cultures x emotion intents, $F(4, 1479) = 5.60$, $p < .05$. As seen in Fig. 3-10, all the three groups of listeners are more accurate at identification of the Angry emotion intent played by Chinese instruments as compared to its rendering on Western instruments. However, the Happy and Sad emotion intents were better with the Western instruments. Interestingly however, their performance for the Happy emotion intent is the lowest on Chinese instruments, regardless of the listener groups.

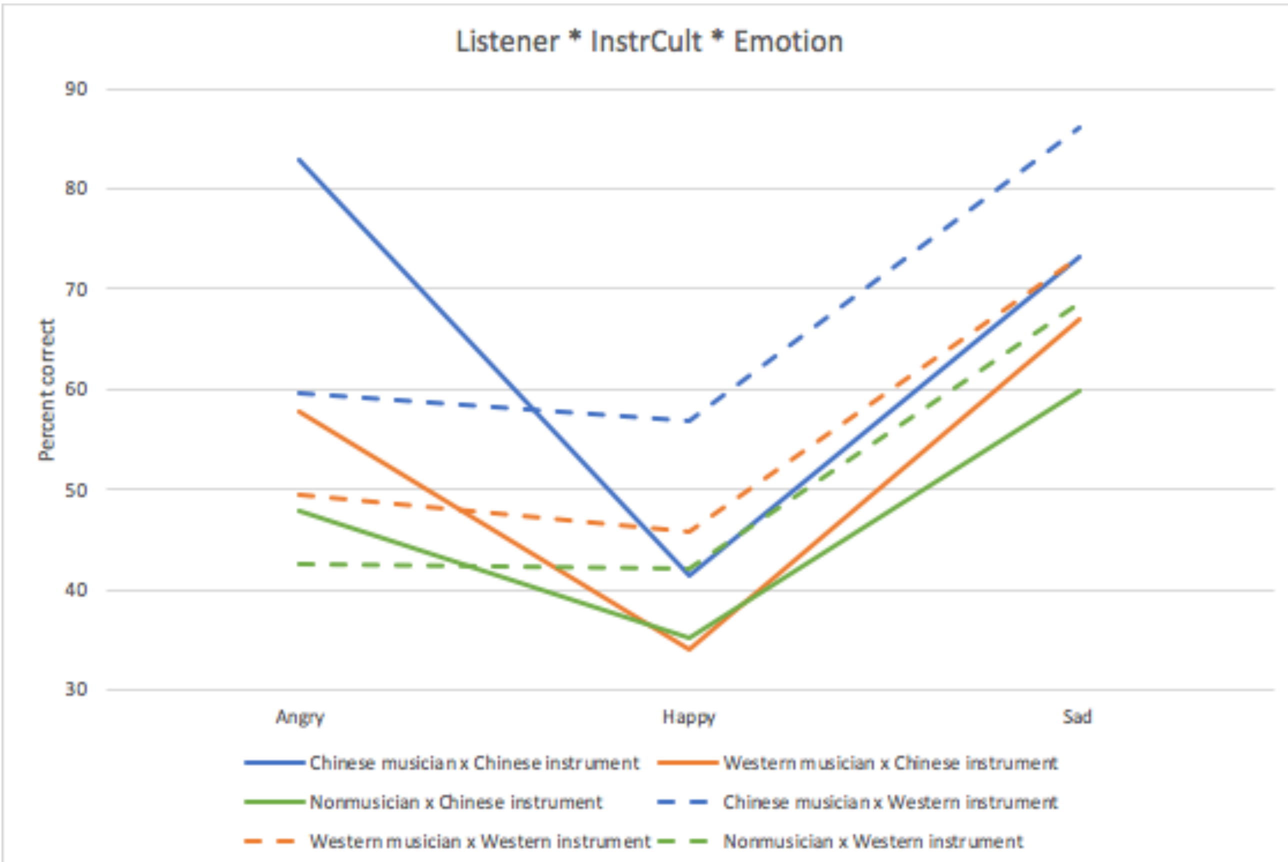


*Figure 3-10*. Three-way interaction between listener groups (different colours), performer's instrument culture (line texture), and performer's emotion intent.

*3.1.4.3 Interaction between instrument cultures, instrument categories, and emotion intents*

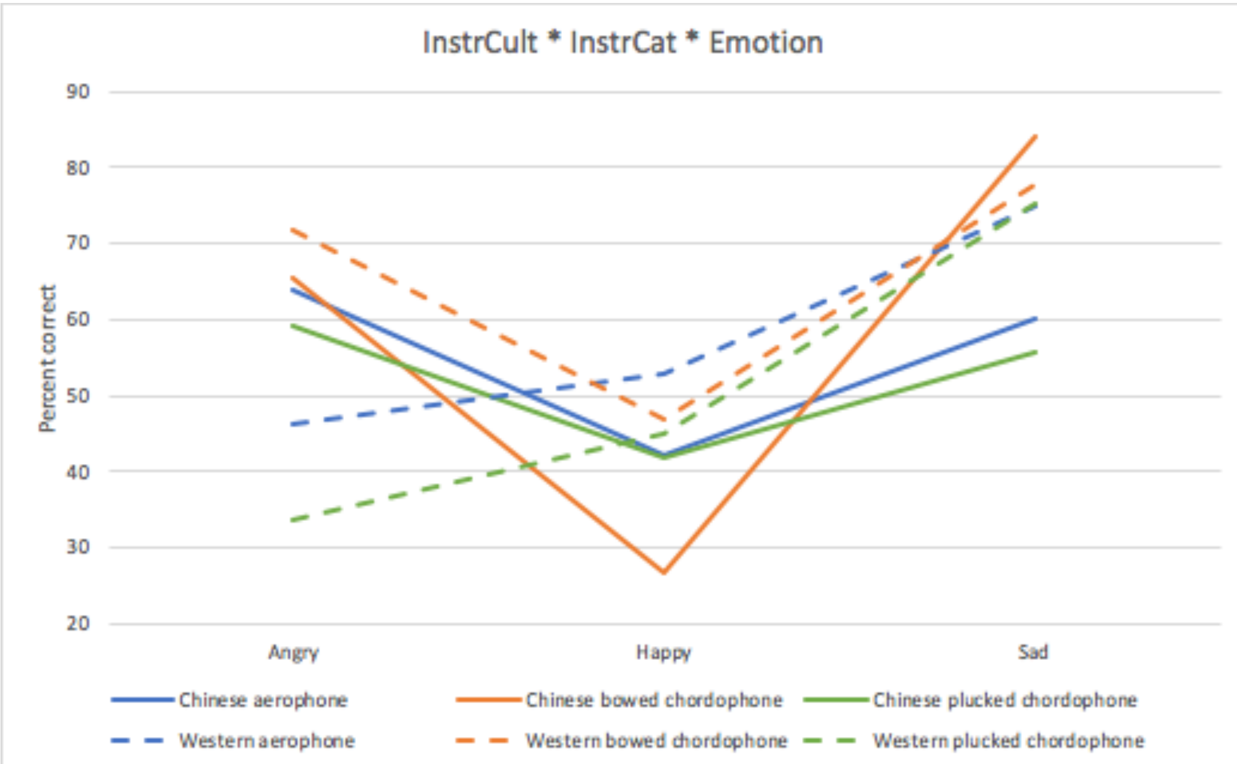


*Figure 3-11*. Three-way interaction between instrument culture (line texture), instrument category (different colours), and performer's emotion intent.

There was a significant interaction effect for instrument cultures x instrument categories x emotion intents, $F(4, 1479) = 23.45$, $p < .05$ (Fig. 3-11). As seen in the two-way interaction effects, the different categories of instruments are different between the two musical traditions in their overall accuracy at expressing emotion intents. In addition to these differences, it appears that the different emotion intents also vary in their accuracy for the different instrument categories across the two instrument cultures. On the whole, the Chinese bowed chordophone performed the worst for a Happy emotion intent but the best for a Sad emotion intent. The Western plucked chordophone is less successful at communicating an Angry emotion intent but relatively good at a Sad one.

### 3.1.5 Other factors

The five excerpts used have been selected for their comparable similarities in potential to carry different emotion intents. In the preceding analyses, the results of the listeners' responses were averaged across the five different excerpts. Any contributing factors that each individual excerpt has in carrying a certain emotion intent were thus minimized. However, an exploration of the proportion of correct responses of different excerpts by the different groups of listeners revealed some interesting observations.

For the Chinese musicians (blue bars in Fig. 3-12), pairwise comparisons showed significant differences between excerpts 1 and 2, excerpts 1 and 5, excerpts 2 and 3, excerpts 3 and 5, and excerpts 4 and 5, all with medium effect sizes. Compared to the Chinese musician listener group, there were more pairs that were significantly different within the Western musician listener group (orange bars in Fig. 3-12). Pairwise comparison within this group showed differences between: excerpts 1 and 2, excerpts 1 and 4, excerpts 2 and 3, excerpts 2 and 5, and excerpts 3 and 4 with a medium effect sizes; and excerpts 1 and 5, and excerpts 4 and 5 with a large effect sizes.

Finally, compared with the other two groups, pairwise comparisons within the nonmusician listener group showed the least number of pairs that were significantly different (grey bars in Fig. 3-12). Only excerpts 1 and 5, and excerpts 4 and 5 were significantly different, both having medium effect sizes.

The difference in the number of pairs of excerpts within each listener group is interesting as it may suggest that musical training influences the way listeners utilize different musical parameters in making judgements of emotion intents. Although notated rhythm and pitch were kept constant by having performers express all the different emotion intents on each excerpt, the differences in the number of responses for each emotion intent between the different excerpts suggests that these factors contribute to the expression of emotion intents. The characteristics of each of the notated excerpts might also influence how performers express the different emotion intents. In addition to timbre, expressive timing and intonation might be manipulated differently because of the nature of the melody and rhythm of each excerpts. From the results above, Western musician listeners seem to be influenced by the melodic and rhythmic (and perhaps implied harmonic) aspects of these excerpts much more than were the Chinese musician listeners. This may perhaps be related to how training in Chinese music often very explicitly emphasizes the different ways of utilizing timbre in performance. Even though melodic and rhythmic aspects contribute to influencing judgements of emotion intent, timbre's greater salience in this group of musicians may be the reason the differences between the excerpts are less marked than those for the Western musician listeners. Nonmusicians, interestingly, seem the least influenced by the different excerpts, suggesting perhaps that the use of timbre and other nonmelodic or rhythmic aspects, such as expressive tuning or tempo differences, are the cues that provide more information for this group of listeners.

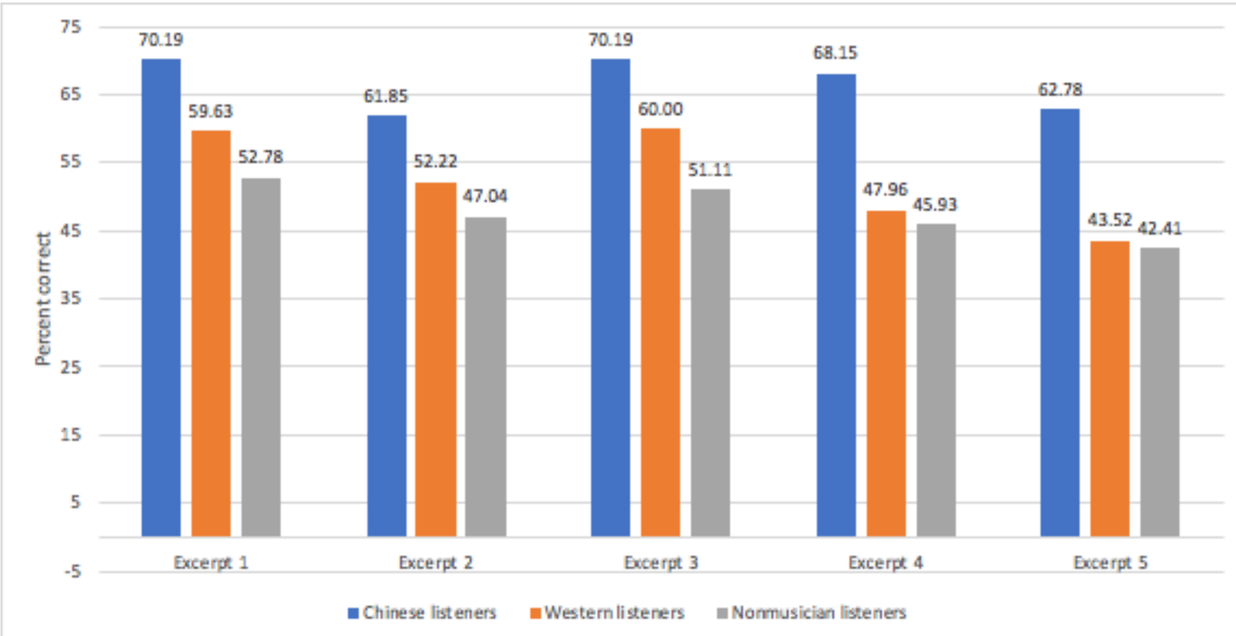


*Figure 3-12*. Percentage of correct responses by excerpts according to different listener groups.

As seen in Table 1, excerpts 1 and 3 appear to be better at carrying the various emotion intents, resulting in a higher proportion of correct responses across all the different emotion intents while excerpt 5 seems to fare the worst. Listeners' responses for excerpt 5 were the most unevenly distributed, with a large majority judging it to be a Sad emotion intent, regardless of performer's intended expression. Excerpt 2 interestingly also appears to fare the worst in terms of accuracy in judgement of emotion intent by Chinese musician listeners, whereas this is not the case for the other groups of listeners. It may be that melodic and rhythmic characteristics are musical parameters that are learned within particular musical traditions. A melody may have very different emotional connotations to the listeners of different musical traditions. Even though an attempt has been made to obtain excerpts that have the least variance across different emotion intents, it appears that there are still differences amongst the five excerpts and that melodic and rhythmic aspects are certainly important parameters listeners use when judging an emotion intent.

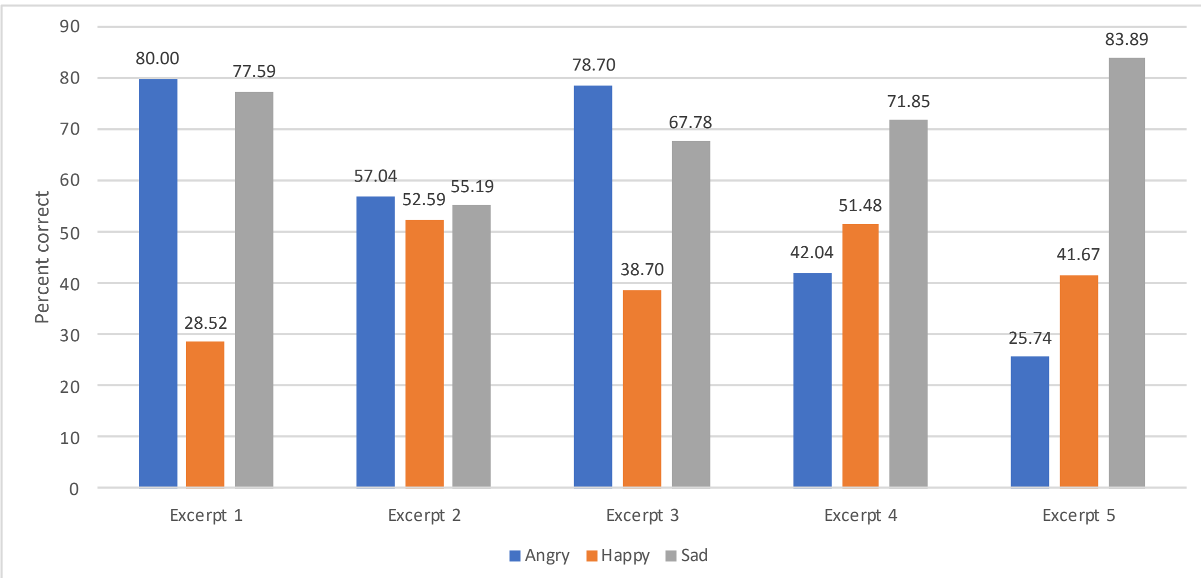


*Fig 3-13.* Percentage of correct responses by excerpts according to different emotion intents.

| | **Chinese musician listeners** | **Western musician listeners** | **Nonmusician listeners** |
|---|---|---|---|
| Most | 1 and 3 | 3 | 1 |
| | | 1 | 3 |
| | 4 | 2 | 2 |
| Least | 5 | 4 | 4 |
| | 2 | 5 | 5 |

*Table 1*. Excerpts ranked according to proportion of correct responses.

## 3.2 Experiment 1: Acoustic Analyses

### 3.2.1 Organization of data

Data from the audio descriptors were grouped according to the performers' emotion intent and analyzed to see if any trends were present in how performers manipulated timbre in the expression of each intent. Listeners' responses were also coded according to their judgements of emotion intent, regardless of whether it corresponded to the performer's intended emotion. These data were then analyzed to explore whether performers and listeners converged on the same set of timbre dimensions in the communication of emotion intents, i.e., which timbral parameters were being varied for a given emotion intent by performers and which parameters were being used to identify an emotion intent by listeners.

### 3.2.2 Performer's emotion intent

#### *3.2.2.1 Acoustic descriptors for different emotion intents*

Analyses of variance were conducted over the different acoustic descriptors across the performers' emotion intents and post hoc tests for pairwise comparisons were conducted with the Tukey Honestly Significant Difference (HSD) test. Various audio descriptors—log attack time, attack slope, decrease slope, temporal centroid, effective duration, energy modulations, spectral centroid, spectral spread, and spectral flux—were all used to varying degrees in the expression of the different emotion intents by the performers. As each of the excerpts was performed with all the different emotion intents, the notated melodic and rhythmic factors were thus kept constant. The average values for each of the acoustic descriptors over the entire excerpt were used for the analyses.

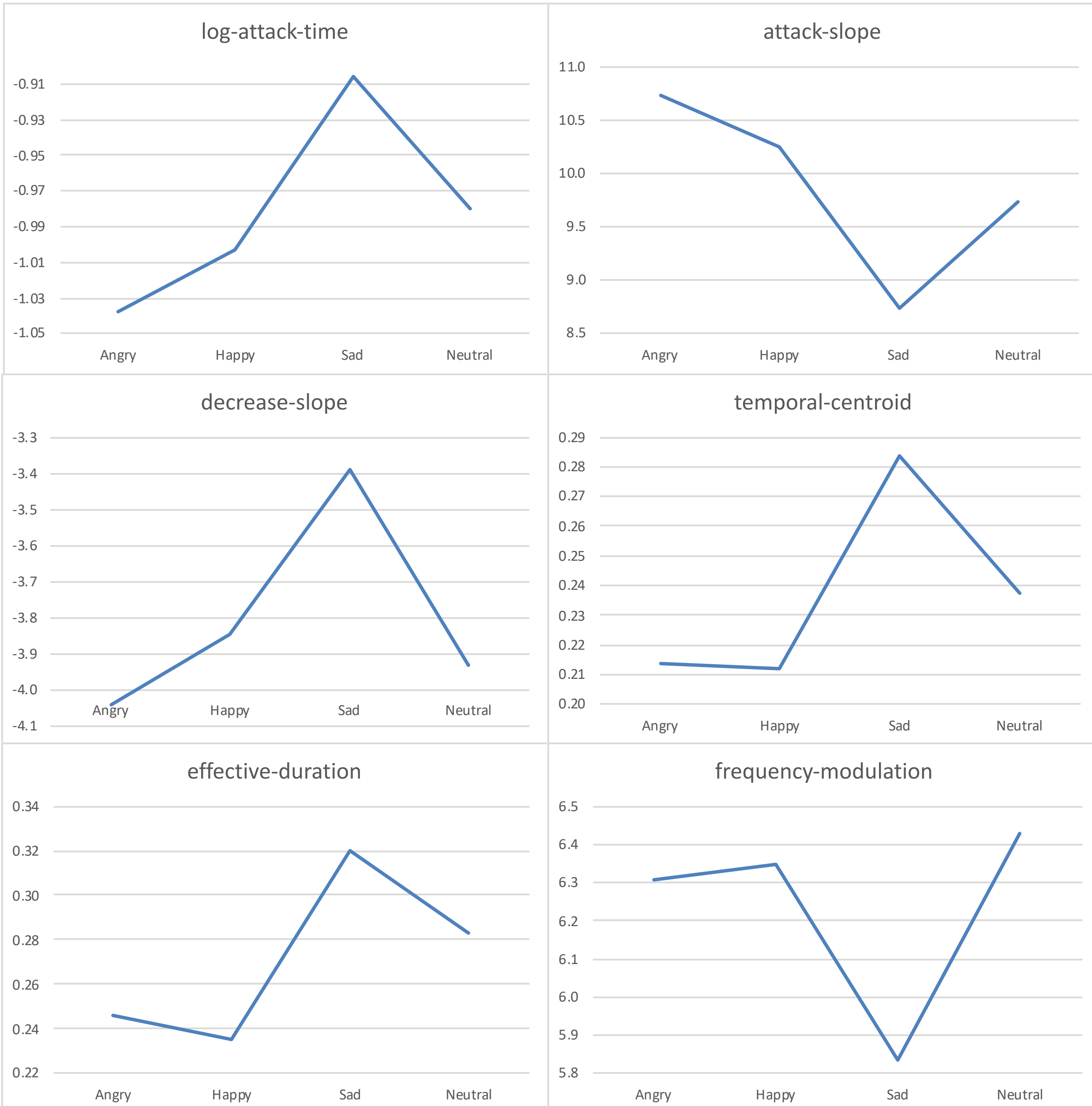


*Figure 3-14a.* Average audio descriptors for different emotion intents by performers.

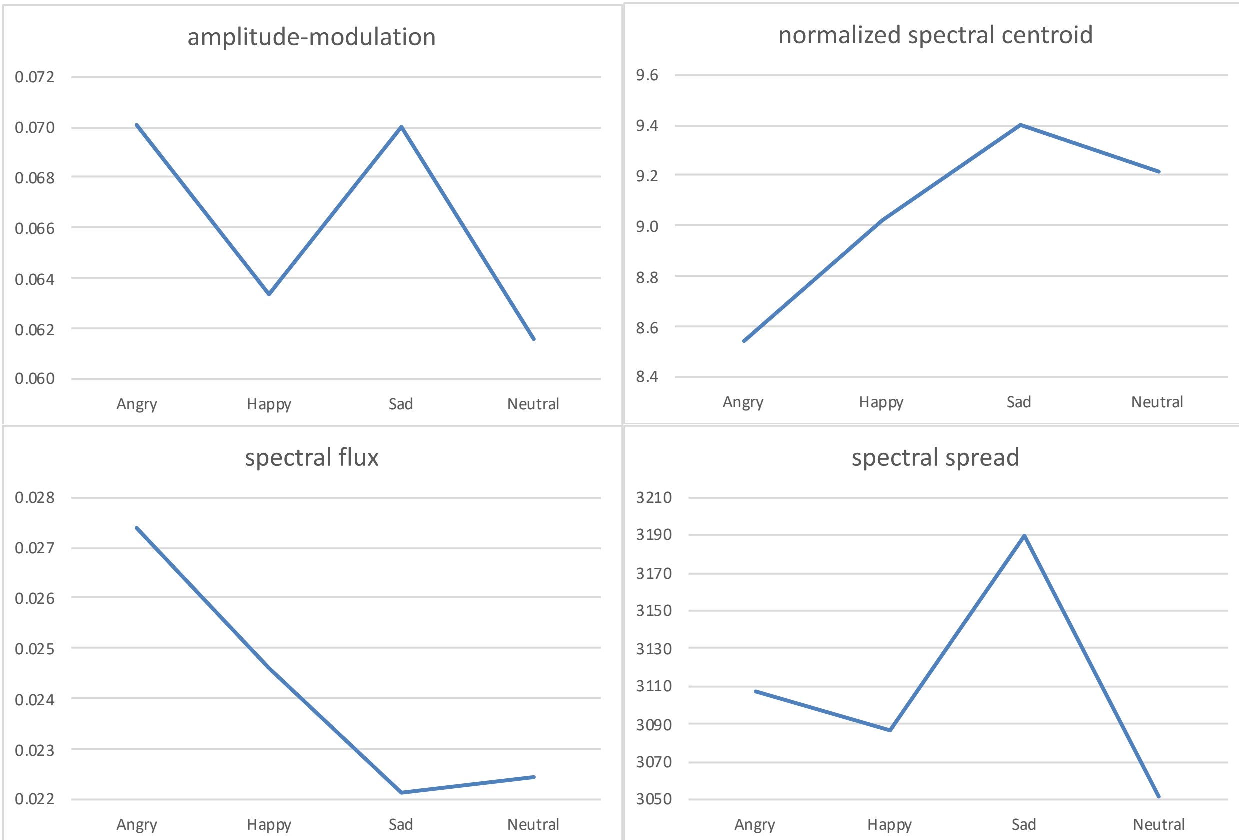


*Figure 3-14b.* Average audio descriptors for different emotion intents by performers.

#### *3.2.2.2 Angry emotion intent*

In the expression of an Angry emotion intent, performers consistently articulated with a more intense attack having a short attack time and steep attack slope (Figs. 3-14a). The relatively short effective durations of the notes suggest that they were likely to be played either at a faster tempo, or more detached, or a combination of both. A low temporal centroid also suggests that the sounds are somewhat more detached. Within the sustained part of the sounds, there was a high degree of frequency and amplitude changes, suggesting vibrato and possibly a higher rate of change in the amplitude of each sound. Low spectral centroid indicates a darker timbre, but there is a lot of variation in the spectral shape over time across each note as is reflected in the high spectral flux. The spectral spread, however, is low, meaning the spectra are centered closely around the mean value (Fig. 3-14b). Except for the spectral spread and amplitude modulation, all the other acoustic correlates of the Angry emotion intent contrast greatly with the Sad emotion intent. It seems likely, therefore, that these two emotion intents will be less confused with each other even when manipulations on other musical parameters are minimized.

#### *3.2.2.3 Happy emotion intent*

The values for many of the acoustic descriptors for the Happy emotion intent are close to those of the Angry intent. Only the log attack time, attack slope, and spectral flux are significantly different in the pairwise comparisons. This means that besides a less intense attack and a lower variation in the spectral shape over time, a performance with an Angry emotion intent is likely to sound quite similar to one with a Happy emotion intent (Fig. 3-14a,b).

#### *3.2.2.4 Sad emotion intent*

Acoustic correlates of the Sad emotion intent have the greatest difference compared to both the Angry and Happy emotion intents (Fig. 3-14a,b). Performers consistently use a gentler attack in expressing a Sad emotion intent. The high temporal centroid and long effective duration also suggest a more sustained treatment of the notes. Interestingly, the low frequency modulation seems to suggest less vibrato, but there seems to be greater change in the amplitude envelope of the notes. Furthermore, the high spectral centroid implies a brighter sounding timbre with a large spread across the mean, but there seems to be less variation of the spectra over time. This finding means that the acoustic correlates of the Sad emotion intent are quite distinct from those of the other two emotion intents.

### 3.2.3 Differences in acoustic correlates of emotion intents across performer's instrument culture

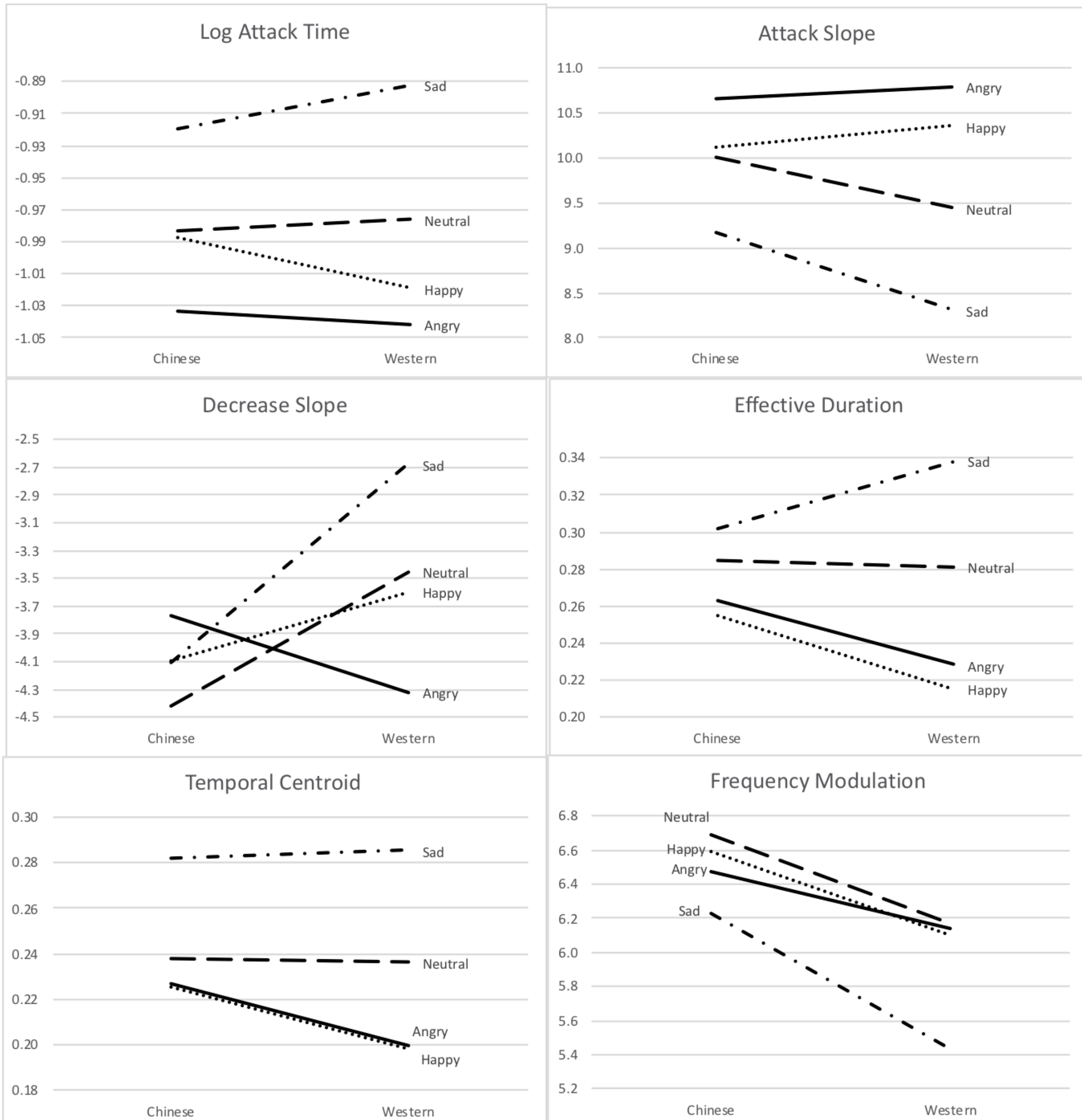


*Figure 3-15a*. Comparison between performer's instrument culture for different audio descriptors.

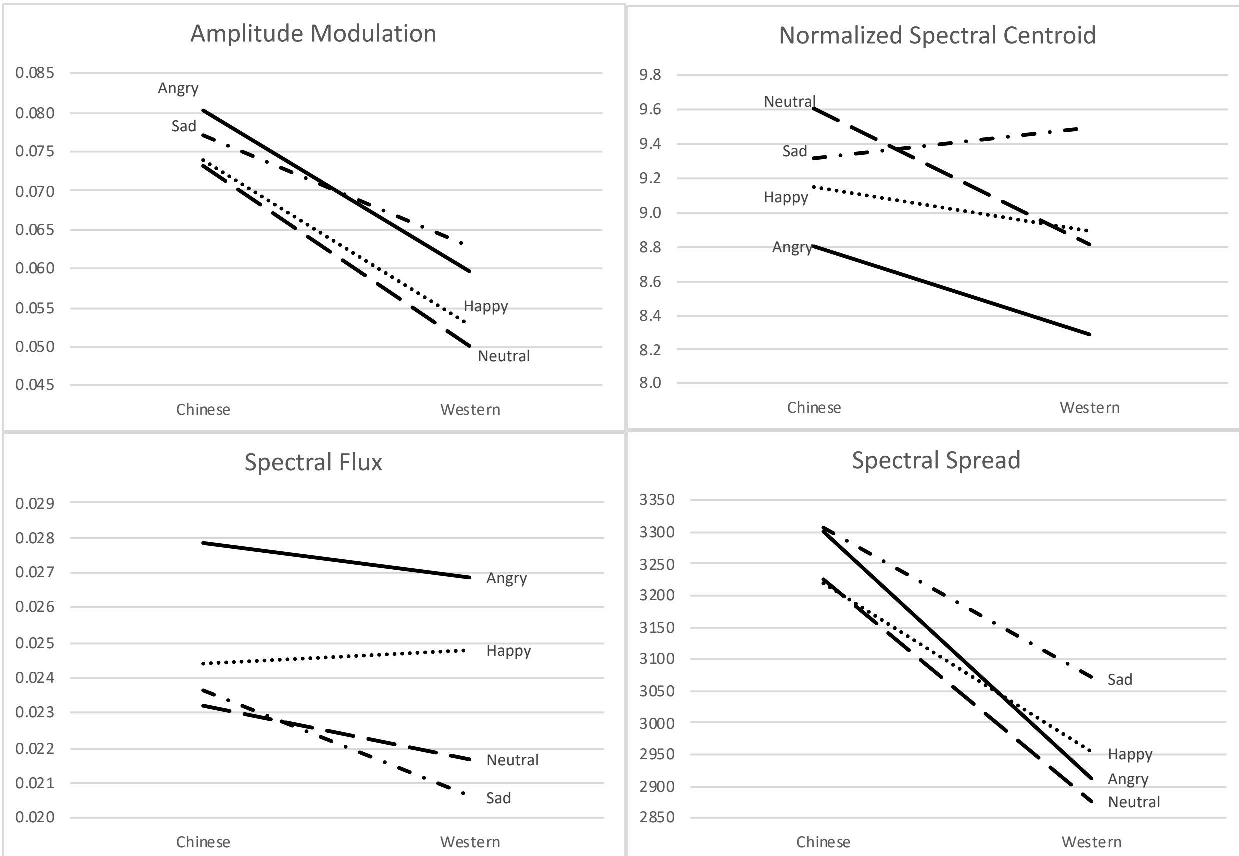


*Figure 3-15b*. Comparison between performer's instrument culture for different audio descriptors.

Performers in the different musical traditions appear to express the emotion intents somewhat differently as can be seen in the differences on each of the acoustic correlates in Figures 3-15a and b. It is interesting to note that in between the Angry and Happy emotion intents, the three acoustic correlates that were significantly different in the pairwise comparisons—log attack time, attack slope, and spectral flux—had slightly greater differences between the Happy and Sad emotion intents in the Chinese instruments than in the Western instruments. Even though there were greater differences between the two emotion intents by the

Chinese performers as compared to the Western performers, the Happy emotion intent on Chinese instruments was less accurately identified by the listeners. It is possible that the manipulations of timbre by the performers from the Chinese musical tradition deviated or even went in opposition to how the majority of listeners regardless of their musical background, would utilize timbre. Because of this, listeners overall judgements of the Happy emotion intent by Chinese performers were less accurate.

The differences in effective duration and temporal centroid were bigger among the different emotion intents for the Western instruments than for the Chinese instruments. This might mean that there is a bigger difference in how Western performers used articulation (legato or staccato) and tempo variations to express emotion intents as compared to performers on Chinese instruments. There is also a greater variation in the spectral centroid over the different emotion intents by the Western performers than the Chinese performers, indicating a greater degree of differentiation of brightness and darkness in timbre when expressing different emotion intents.

Energy modulations were much lower for the Western performers as compared to the Chinese performers, perhaps indicating less use of vibrato, and less change in the amplitude of each note. The spectral spread is also consistently greater in the Chinese performers as compared to the Western performers across all the different emotion intents, indicating a richer spectrum in the performances on the Chinese instruments. The generally higher values for the spectral spread of the Chinese instruments may, however, indicate aspects of the inherent timbre of Chinese instruments, as compared to Western instruments, and not be due to performers' manipulations.

### 3.2.4 Listener's judgement of emotion intent

The stimuli were next grouped according to the emotion intent listeners selected, regardless of whether it was in concordance with the performer's emotion intent. Each stimulus was labelled according to the emotion intent that garnered the highest number of responses. ANOVAs were conducted for the audio descriptors for each emotion intent across listener groups. Comparing across emotion intents, there were significant differences in all the acoustic descriptors.

Similar to performers' decisions in timbre manipulation, listeners also judged shorter attack times and steeper attack slopes as more likely to be an Angry or Happy emotion intent, and longer attack times and gentler attack slopes as more likely to be Sad. The attack times and attack slopes of the Angry and Happy emotion intents are also much closer whereas those of the Sad emotion intent are distinctly farther apart. This trend was similarly perceived by the different groups of listeners, regardless of whether it was being performed by Chinese or Western instruments. The attack characteristics therefore seem to be salient perceptual features that are used consistently across all groups of listeners and performers.

The values for effective duration and temporal centroid also appear to be significantly different between the Sad emotion intent as compared to either the Angry or Happy emotion intents. It appears that the listeners made similar choices in judging the articulation for the different emotion intents as the performers.

In terms of frequency modulation, listeners judged lower frequency modulations as Sad and higher as Angry. This is similar to the manipulations of frequency modulation by performers. Chinese musician listeners judged a lower amplitude modulation as a Sad emotion intent whereas Western musician and nonmusician listeners judged the largest amplitude modulations as Sad instead. From the performers' data, it appears that amplitude modulations for Chinese instruments are generally higher than those of Western instruments. It is therefore uncertain whether differences in amplitude modulation provide a reliable source of information for listeners or whether the variance in amplitude modulations simply reflects the inherent nature of the instruments.

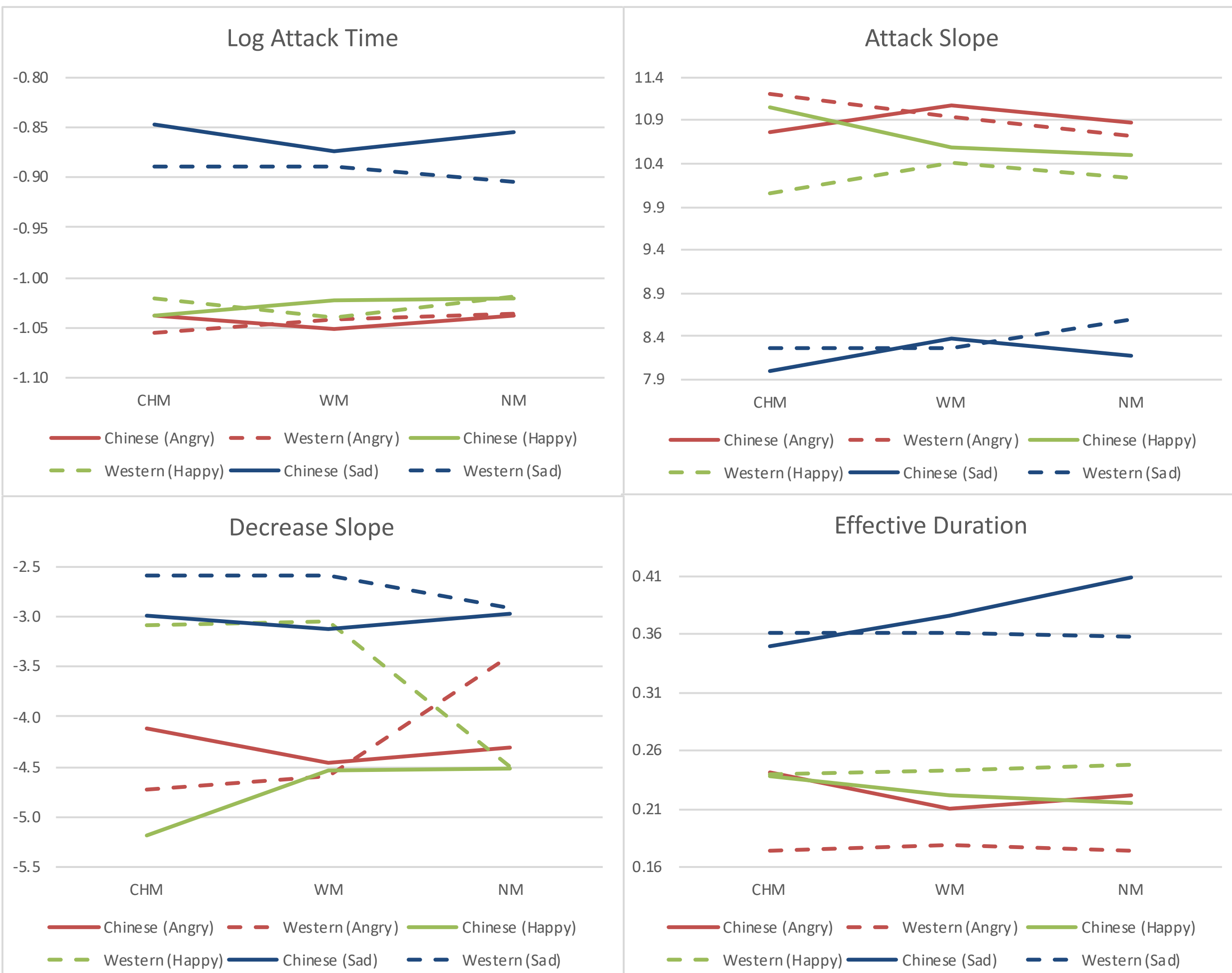


*Figure 3-16a*. Comparison between listener's musical backgrounds and performer's instrument culture for different audio descriptors.

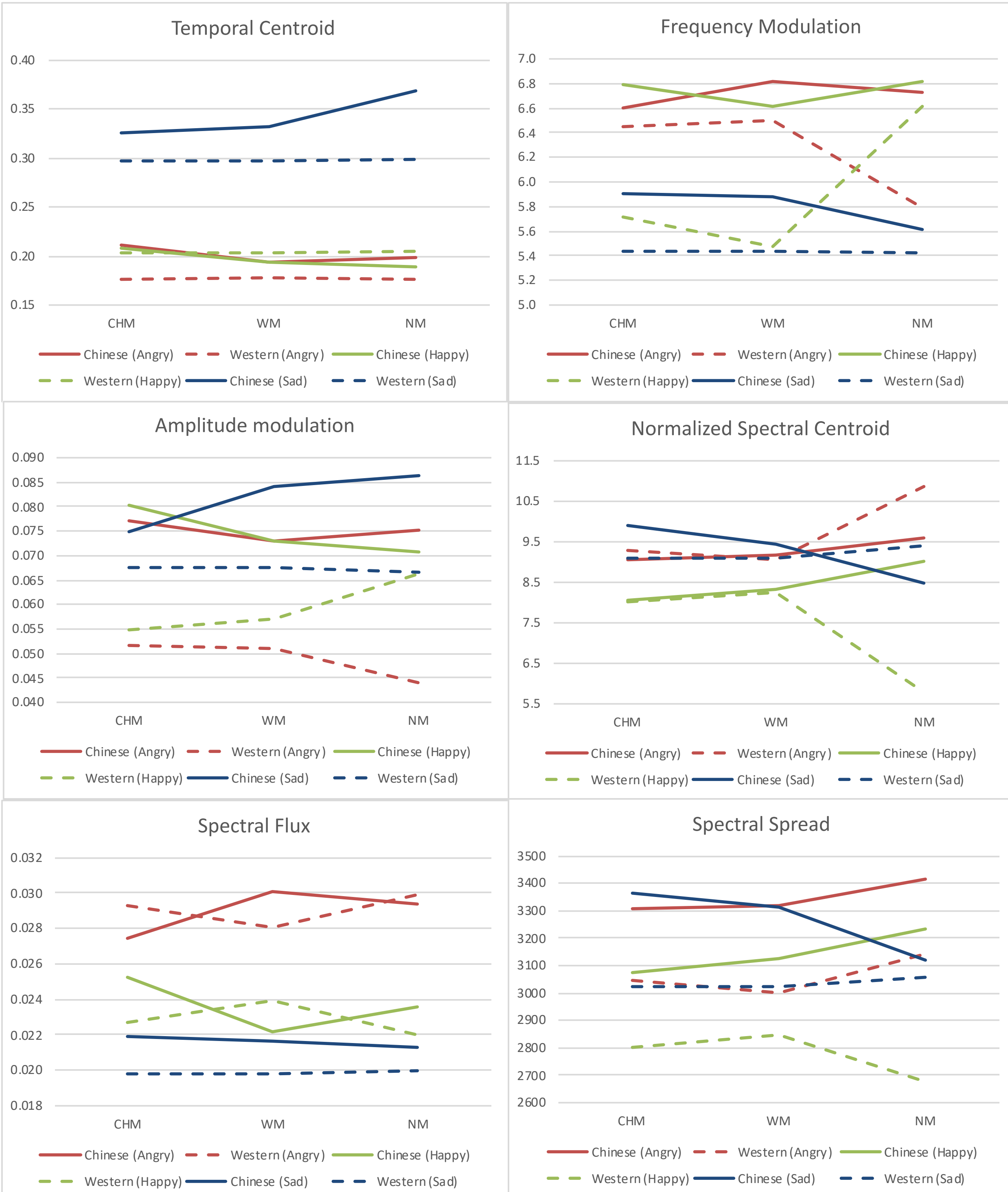


*Figure 3-16a*. Comparison between listener's musical backgrounds and performer's instrument culture for different audio descriptors.

The values for spectral flux in listeners' judgements also corresponded to the performer's manipulation of it for the various emotion intents. The Angry emotion intent had the highest spectral flux while the Sad emotion intent the lowest.

An interesting divergence occurs between performers' manipulations and listeners judgements on the acoustic correlate of spectral centroid. Performers used the brightest timbre for a Sad emotion intent and the darkest for an Angry emotion intent. Listeners however, judged the darkest timbre as Happy rather than Angry. Interestingly, nonmusician listeners had the largest variance between the different emotion intents in terms of the spectral centroid. Listeners who were musicians, both Chinese and Western, seem to have utilized the difference in spectral centroid less in their judgements of emotion intent as compared to nonmusicians.

# CHAPTER 4

## Discussion

This study aims to look at how timbre is manipulated when performers express different emotion intents, and how listeners might utilize these manipulations to interpret the various emotion intents being communicated. In addition, by comparing groups of musicians trained in different musical traditions, this study seeks to answer other related questions: Does the manipulation of timbre differ from one musical tradition to the next? If so, is it learned specifically within a particular musical tradition?

If expressive timbre rules are learned, there will be differences in how performers trained in different musical traditions utilize timbre to express different emotion intents. Listeners from different musical traditions may also understand the expressive intents differently. It was therefore hypothesized that participants in Experiment 2 who were Chinese musicians would be more accurate at judging the emotion intent of an excerpt played on a Chinese instrument and participants who were Western musicians would be more accurate at judging an excerpt played on a Western instrument, whereas nonmusicians would likely be the least accurate in both.

Results showed that although the nonmusicians did perform the least accurately among the three groups of participants, the hypothesis concerning accuracy differences between the Chinese and Western musicians was not confirmed. Chinese musician listeners were, on the whole, more accurate at judging emotion intents than their Western counterparts. Although training in a particular musical tradition likely contributes to differences in how timbre is utilized, socio-cultural factors may play a role as well. Due to constraints in time and resources, this study could not control for certain factors. For instance, both groups of musicians playing the excerpts in Experiment 1 were from Singapore. Certain rules or constraints within these performers' renditions may stem from socio-cultural influences in their lives, in addition to what they had learned through training in a specific musical tradition. Another factor has yet to be explained: in Experiment 2, Western musician and nonmusician listeners alike were all from Montréal, whereas the Chinese musician listeners were from Singapore. This situation would not allow for a fully factorial experimental design that could tease out effects of musical training per se and that may have contributed to differences in listening strategies (such as rules regarding timbre use for expression of certain emotions). At the same time, even though all the Western musician listeners had been screened to ensure no prior exposure to Chinese music, all the Chinese musician listeners had had varying degrees of exposure to Western music. Due to the extensive presence of Western art music in many places around the world, including Singapore, it would be very unlikely to find an individual trained in Chinese music who has had little or no exposure to Western art music.

## 4.1 Communication of Emotion Intent

### 4.1.1 Timbre differences inherent to instrument's mode of sound production

Different modes of sound production create an inherent difference in the kind and range of timbre manipulation they afford a performer. Bowed chordophones, which provide a wide range of possibilities in timbre manipulation, have the potential to express the many subtleties for communicating emotion intents. Plucked chordophones, due to their impulsive mode of sound production, may afford a smaller range in timbral control for the expression of emotion intent. Performers might utilize other means in addition to timbral manipulation in different ways for a successful communication of musical intent, but because the design of this study attempts to minimize other factors, it possibly resulted in a lower accuracy in the communication of emotion intent for this category of instruments. Aerophones perform significantly less accurately, even though they seem to afford equal or even more timbral manipulation possibilities compared to bowed chordophones given that they are controlled directly by the performer's breath, as opposed to the control having to be transferred from a performer's hands to the bow before reaching the strings.

Chinese instruments appeared to be more effective than Western instruments at expressing an Angry emotion intent. This may be in part due to a difference in musical training or other extraneous socio-cultural factors that gives a more nuanced utilization of expression for an Angry emotion intent.

Different instrument categories were also found to have differences in their accuracy at expressing different emotion intents—bowed chordophones are the best at expressing Angry and Sad emotion intents but are the worst at the Happy emotion intent. It may be that the expression of a Happy emotion intent relies less on learned rules in timbral manipulations, timing, and expressive intonation, but with other musical parameters such as notated melodic and rhythmic elements contributing to a larger extent in its expression. This may be the reason why bowed chordophones, which afford a great range of timbral manipulation perform less accurately as this aspect no longer provides enough information for listeners. In addition, listeners might be influenced by the frequency of use of bowed chordophones in expressing melancholic music and therefore have a lower number of successful judgements on the Happy emotion intent. Also, instruments such as the flute (both Chinese and Western) may be much more often used for depicting joyous and cheerful musical intents and therefore elicit more accurate responses for the Happy emotion intent. A possibility for this might be the cultural implications these particular instruments have. Although there have not been systematic studies on the use of certain instruments in association with certain emotions, a study by Gene Behrens and Samuel Green (1993) concluded that emotional content of music may be a function of its cultural connotations—certain instruments were associated more with certain emotions.

The communication of emotion intent also does not depend solely on timbre. The design of this experiment aims to minimize other factors such as notated rhythm and melody by keeping it constant for the different emotion intents. Performers, however, were still able to control for the expressive timing and intonation in their playing. The difference in overall accuracy for the different emotion intents may mean that in this complex interaction of different musical parameters contributing to musical communication, the degree of contribution of each and their interactions differ across the emotion intents.

Three-way interaction effects between instrument culture, instrument category, and emotion intent suggests that different emotion intents require different kinds of timbral manipulation, which are more easily performed on some instruments than others. At the same time, training in the different musical traditions may also elicit different ways of using timbre in the expression of emotion intents. Some of the rules that are learned for a particular emotion intent may be common across musical traditions, which explains the high level of accuracy in judgement in the Chinese bowed chordophone on a Sad emotion intent, for example. Other rules could be very specific within an instrument culture on a particular instrument, hence the low proportion of correct responses in judgements of emotion intent in the Chinese bowed chordophone on a Happy emotion intent, for instance. There may also be aspects of the timbre inherent to an instrument that makes it better at communicating some emotion intents but not others. The harsh attacks provided by the bow and the vibrating membrane of the *erhu,* for instance, afford more nuances in the expression of an Angry emotion intent.

#### 4.1.2 Differences between listener groups

Even though the Western musician listeners did not share the socio-cultural influences the Chinese musician listeners had with the performers, it appears that there may be certain timbre rules that are learned and are heavily used in the expression of an Angry emotion intent, and which are reflected in the better accuracy in the musician listeners. Having both musical training and socio-cultural factors playing a part, the Chinese musician listeners—who were all from Singapore—had much more information available for decoding the emotion intent of the performances and therefore performed better than the other two groups of listeners from Montréal.

Various reasons might account for this difference between the emotion intents. Although tempo, timing and expressive intonation are still within the performer's control and may be utilized to differing extents by the performers for different emotion intents, other variances in melodic and rhythmic parameters are minimized. The notated pitch, tonality, and rhythm of each excerpt remain constant over different emotion intents. The communication of a Happy emotion intent may utilize more of these melodic and rhythmic parameters present within the notation such as melodic movement, or rhythmic patterns, in addition to timbre, timing, and expressive intonation. When these are kept constant, the success of communicating a Happy emotion intent is therefore reduced. Although both the Western musician and nonmusician listeners perform better than chance for the Happy emotion intent, it is likely that their accuracy was not significantly different from each other because both these groups of listeners rely on the same information that resides more within the notation than in expressive performance. The Chinese musician listeners, with the added advantage of shared socio-cultural factors with both the groups of Chinese musician performers and Singapore-based Western musician performers, may have access to more cues for their judgement of this emotion intent.

With the notated melodic and rhythmic cues controlled for, the greater accuracy for the Sad intent may imply that these cues are not crucial for its accurate identification. The lack of significant difference between the Western musician and nonmusician groups in this case might mean that socio-cultural cues contribute to a large extent in the communication of both the Happy and Sad intents. It might also be possible that many cues used in expressing a Sad intent occur commonly enough such that musical training is not crucial for its identification.

#### 4.1.3 Timbre parameters learned in musical communication

The significant difference in the accuracy of emotion intent judgements between listeners trained in the Chinese and Western music traditions in the performances on Chinese instruments point towards the likelihood that the role of timbre in the communication of emotions in music is learned. Chinese musician listeners were more accurate at stimuli performed on Chinese aerophones, whereas Western musician listeners were better at stimuli performed on Western aerophones. This result might imply that for aerophones, musical training plays an important role in the communication of emotion intent in music.

For the bowed chordophones, however, all three groups of listeners were better at the Western instruments than the Chinese instruments. The fact that excerpts played on Western bowed chordophones were more accurately rated than those played on Chinese bowed chordophones across all the listener groups could be due to the larger affordances for timbre manipulation on the violin. The greater difference between Western than the Chinese bowed chordophones in the nonmusicians' judgements as compared to other listener groups might imply that there may be some very specific aspects of timbral manipulation on bowed chordophones that are learned within the Chinese music tradition and that have very different rules as compared to those of the Western music tradition. As a result the Chinese bowed chordophones elicit a lower proportion of correct judgements of emotion intents in both the Western musician and nonmusician listener groups.

Western musician listeners and nonmusician listeners both perform much worse on a Western plucked chordophone (the guitar) as compared to the Chinese plucked chordophone. This result is interesting as both of these groups of listeners should be more familiar with the guitar than the *pipa* and have more exposure to Western art music than Chinese music. It is uncertain whether this has to do with certain specificities in the way this particular performer expresses emotion intents or if it is common both to all guitar performances and to the way timbre manipulation for expressing emotion intent is learned in guitar performance.

That Chinese musician listeners judge emotional intents conveyed on Western instruments more accurately than do Western musician listeners suggests in turn that socio-cultural factors involving processes of acculturation, or the learning of traditions from another culture (Redfield, Linton, & Herskovits, 1936), play a part in musical communication as well.

The Western musician and nonmusician listeners on the other hand were all from Montreal and therefore are likely to have different socio-cultural influences from the Singapore-based Chinese musician listeners and the two groups of performers. Because the Western musician listeners share a similar set of socio-cultural influences with the nonmusician listeners, it is possible that the differences in judging emotion intents between these two groups can likely be attributed to aspects of timbre that are learned through musical training and practice.

In addition, Western musician listeners' poorer accuracy in judging some emotion intents in performances on Chinese instruments (as compared to nonmusician listeners) may suggest that there are specific rules for the manipulation of timbre within a particular music tradition for certain emotion intents. These rules, which are specific to one musical tradition and not another, result in trained musicians being less accurate in the judgement of emotion intent for excerpts from the other culture because they do not conform to the rules that these musicians have learned in their own culture.

The Angry emotion intent expressed by Chinese instruments had more accurate responses as compared to that expressed by Western instruments, regardless of the listener group. On the other hand, the Happy emotion intent expressed by Chinese instruments had less accurate responses as compared to that expressed by Western instruments, regardless of the listener group. There might be some timbral manipulations that are being used for the Happy emotion intent in the Chinese music tradition that goes against the rules of how it is supposed to be expressed within the Western music tradition. There might also be some attribute of the inherent timbre of the Chinese instruments that somehow made them less successful in communicating a Happy emotion intent. There is also the possibility that in the Western music tradition, the communication of a Happy emotion intent involves a larger contribution of musical parameters such as melodic and rhythmic aspects found within the notated music, factors which have been minimized in this study. With the decrease in pertinent information for the communication of a Happy intent, Western musicians therefore fared much worse. The opposite could be true for the Angry intent. A certain amount of the timbral manipulations and other expressive techniques used by the Chinese performers for the Angry emotion intent might communicate the emotion similarly to listeners regardless of enculturation or musical training. The higher proportion of correct responses within the group of Chinese musician listeners, however, points to the likelihood of some of the manipulations used by the Chinese music performers to being attributable to the enculturation and training within a musical tradition.

#### 4.1.4 Acoustic correlates of timbre utilized in communicating emotion intent

Performers were observed to consistently utilize different dimensions of timbre such as the attack time and slope, energy modulations, and spectral and spectrotemporal properties in expressing different emotion intents. At the same time, listeners mostly made use of a similar set of acoustic correlates of timbre in decoding the performers' intents. Only spectral centroid was different between the performers' manipulations and the listeners' judgements. Although spectral centroid has been found to be a perceptually salient correlate of timbre from previous research, it is interesting to note that listeners and performers are not in concordance as to how it varies according to different emotion intents. There were also significant differences between Chinese and Western performers' use of particular correlates of timbre. The Chinese performers, for instance, had the highest amplitude modulation for an Angry emotion intent whereas the Western performers had the highest for a Sad emotion intent. The Chinese performers also utilized a large spectral spread for an Angry emotion intent, the same intent having the lowest spectral spread in the Western performers. Taken together with evidence showing Chinese musician listeners judging excerpts played by Chinese performers more accurately than their Western counterparts, the data above suggest that these different uses of the various acoustic correlates of timbre are indeed learned.

A few interesting observations emerged in the acoustic analyses of the recorded sounds. Firstly, the Angry and Happy emotion intents share many acoustic properties. It is interesting to note, however, that listeners were least accurate in making judgements of a Happy emotion intent, but much more accurate at an Angry one. Because an Angry emotion intent shares many acoustic parameters with a Happy emotion intent, they are likely to sound quite similar. This may be the reason why the Happy emotion intent has the least proportion of correct judgement s in Experiment 2. Interestingly, it is the Happy intent that garnered the lowest accuracy. It appears then that when timbre characteristics indicate that a stimulus is not likely to be of a Sad emotion intent, listeners will more often judge it to be Angry rather than Happy. In a normal listening environment, there are usually many different musical parameters that contribute to helping a listener judge the emotion intent. It may be that there are usually several other factors responsible for a Happy emotion intent that are much more perceptually salient, such as melodic movement, harmonic content, and so on, whereas for an Angry emotion intent, timbre manipulations might count more (than melodic or harmonic content) in judgements. Therefore, when other musical parameters are minimized, listeners tend to perform less accurately on judgements of the Happy emotion intent. The distinctiveness in the acoustic correlates of the Sad emotion intent likely explains why this emotion intent is the most accurately identified by all the listeners.

Secondly, nonmusicians seem to utilize the brightness dimension of timbre (correlated with spectral centroid) much more than either Chinese or Western listeners. They also judged a bright sound as carrying an Angry emotion intent, whereas a mellower one was considered as Happy. Performers, on the other hand, used a low spectral centroid when expressing an Angry emotion intent and a relatively higher one for Happy. I wonder if this divergence may be due to listeners responding partially based on their *induced* emotions in addition to the *perceived* emotions they had been instructed to focus on. The Happy intent appears to be the most difficult one to judge accurately, and listeners may be utilizing other ways to help them glean more information, one of which may be the kinds of emotions the music *induces* in them.

Finally, the values of the acoustic correlates based on performers' emotion intent seem to have a high degree of concordance with those based on musician listeners' responses, whereas those of nonmusician listeners diverged more. Once again, this result suggests that trained musicians strongly concur on aspects of timbre that represent certain emotion intents. The result also points strongly towards the possibility that there are rules in the use of timbre for the communication of emotions in music and that these rules are learned via formal musical training and practice.

### 4.2 Role of Timbre in Communicating Emotion Intent

The results suggest that various timbre properties outlined here are utilized in communicating emotion intents. However, the magnitudes of each of these properties and their combinations vary with different emotion intents. Musical training also plays a part in the performers' and listeners' choices of which properties contribute to an emotion intent and to what extent they do so, both in the performer and the listener.

Although it can be seen that in addition to timbre, other musical parameters are also involved to varying degrees, this study also suggests that timbre plays an important and integral role in the communication of emotion intent. Many acoustic properties, such as characteristics of the attack, appear to have a high concordance between the performers' manipulations and the listeners' judgements of emotion intents. However, there are other acoustic features that differ either in their use between the different musical traditions of the performers or in their contribution to perception among the different groups of listeners.

**4.3 Complexities of Musical Communication**

Based on previous studies on the communication of emotions in music, and the presence of difference in response to the different excerpts used in this study, melodic and rhythmic factors are likely to play a part in communicating emotion intent. This communication involves many different musical, as well as extramusical, elements interacting in a complex way. These interactions of the different musical parameters (melody, rhythm, harmony, dynamic variations, and so on) all contribute to varying degrees to the communication of an emotion intent. At the same time, the extent to which each parameter is involved in musical communication, as well as the combinations of parameters employed, likely depends on the musical tradition in which a performer or listener is enculturated and involves not just musical elements learned through training in a particular musical tradition, but also socio-cultural factors specific to certain societies (languages used, ways of expression, ways of life, etc.). However, exactly how much and how they contribute to emotion perception in music is outside the scope of this study.

Although it has not been possible to filter out the above-mentioned influences in this study of the communication of emotion, the musical elements have been controlled for as much as possible. The correspondence is not perfect, however, given the complexities of musical communication described above. However, with structural features of each excerpt being held constant, the high degree of accuracy in correspondence between performers and listeners indicates significant reliance on timbre cues and certain other performance nuances such as expressive tuning, timing, and dynamics, for the communication of emotion intent. While it may not be possible within this study to tease out the contributions of each of these factors, timbre is certainly implicated in the process of communicating emotion intents in music. Differences in the responses between the listener groups also suggest an aspect of the use of timbre that is learned within a particular musical tradition, which implies that within the complexities of musical communication, there is an aspect of timbre's function that depends on rules that are acquired through specific musical training.

# CHAPTER 5

## Conclusion

As a multidimensional construct, timbre is characterized by a combination of different perceptual components that in many cases can be linked to audio descriptors. Acoustic analysis of the recorded excerpts reveals that all the audio descriptors selected for this study are utilized in different combinations and to differing degrees in the expression of the various emotion intents. All in all, these results suggest that timbre plays an important role in the communication of emotion intents in music.

### 5.1 Contributions

Rules about the use of timbre appear to be learned in musical training and practice. Therefore, an understanding of these rules will be helpful not only as a hermeneutic tool, but also in pedagogy, and other aspects of music scholarship. An understanding of the timbre dimensions listeners use to decode musical intents could also contribute to increasing our understanding of music perception and cognition.

### 5.2 Limitations

As mentioned earlier, there are several limitations of this study due to time and resource constraints. All the performers in Experiment 1 had to be recruited from Singapore so as to be able to prepare a set of stimuli in time for Experiment 2, which required a group of listeners from Singapore who were trained in Chinese music. These circumstances created a confounding factor—a variable that could not be controlled for and that lead to spurious associations—that was not originally intended. At the same time, it was easy to find participants who have never been exposed to Chinese music, but because of the ubiquity of Western art music (especially in Singapore), it was not possible to gather a group of listeners trained in Chinese music who had not been exposed to Western music. The resulting impact can only be reduced by finding Chinese listeners who have as little training as possible in Western art music; almost all of the Chinese music listeners had at least a few years of lessons in Western art music.

Only one musician for each instrument was used for both Experiment 2 and the acoustic analyses so that the duration of Experiment 2 would be reasonable for the participants. However, fielding different musicians on the same instrument would provide data as to whether they utilize aspects of timbre similarly (or not). Also, as only a single musician for each instrument was selected for acoustic and perceptual analysis, there is no basis to confirm whether their performance is characteristic of the majority of performers on that instrument or otherwise. Further analyses of the other recordings would shed light on this issue.

### 5.3 Future Directions

As a complex multidimensional construct, timbre does not possess a single physical domain that can be easily mapped onto a single perceptual domain. Research has shown the salience of several acoustic properties related to timbre perception. Many studies have investigated characteristics of timbre that influence listeners' perceptions of musical intents. This study has attempted to fill a lacuna in understanding the function of timbre in the communication of emotion intents in music. However, many aspects and questions that arose along the way will require further clarification and/or additional work.

As mentioned earlier, acoustic and perceptual analyses have only been performed on a single musician for each instrument. Recordings of the other performers, which have not been used in this study, could be analyzed in order to shed further light on whether the timbre manipulations are indeed consistent across different performers of the same instrument and whether there are some differences seen in the acoustic analyses that are due to unique individual characteristics.

In addition to timbre, melodic, rhythmic, musical dynamics and other factors are likely to contribute to the communication of emotion intents in music, both structurally and expressively. It has not been possible to look at each of these factors and how they contribute within the scope of this study, but this will be an interesting area to look into for future research.

Certain instruments may also be closely associated with certain emotions because they have been historically used in certain contexts or because of their inherent sound quality. Whether or not listeners' judgements in this study could be a factor due to frequent associations with certain tropes in instrumental usage, resulting in the listeners being distracted away from basing their judgements of emotion intent solely on timbre, may therefore be an interesting question to consider for future research.

Excerpts recorded by a group of performers trained in Western art music from Montréal would be helpful in teasing out whether certain elements in timbral manipulation that are due to socio-cultural factors influenced the group of performers from Singapore. At the same time, a group of listeners trained in Western art music from Singapore could also provide additional data to clarify whether and how socio-cultural factors function as a covariate in this study.

Another area that can be examined is how these timbre dimensions are organized over time. While this study has taken out the individual notes and computed an average over the entire excerpt on the different dimensions of timbre, it has not looked at whether any combinations or progressions of timbre change occur over the excerpt. Consistent occurrences of these combinations or progressions might further elucidate rules that can be used in timbre manipulations, particularly as concerns timbral expressivity in music performance.

The model for discrete emotion categories was used in this study in order to simplify the experimental design for an already very complex question of timbre perception and function in music. However, as emotions do not always fall into discrete categories in real life and especially music-evoked emotions, an exploration of the communication of emotion intent using different affect models may increase the ecological validity of the findings. It may also illuminate subtleties in how timbre varies along different affect dimensions.

Musical meaning also involves much more than just emotions in music, and therefore, other aspects of musical meaning also need to be studied. Musical semiology, another complex field of study intrinsically linked to musical communication, must also be considered when discussing the functions of timbre. Timbre is likely to play an integral part in the processes of signification. How timbre plays a role in communication and what aspects of timbre serve this purpose will prove interesting and valuable in furthering our knowledge and understanding of this understudied musical parameter.

## Appendix 1: List of excerpts used for the Pre-Experiment

| Stimulus | Name of film | Sound-track # |
|---|---|---|
| Practice 1 | Modern Times | 12 |
| Practice 2 | The Rainmaker | 7 |
| Excerpt 01 | Lethal Weapon 3 | 8 |
| Excerpt 02 | The Alien Trilogy | 9 |
| Excerpt 03 | Cape Fear | 1 |
| Excerpt 04 | Oliver Twist | 15 |
| Excerpt 05 | The Fifth Element | 19 |
| Excerpt 06 | Hannibal | 1 |
| Excerpt 07 | The Untouchables | 8 |
| Excerpt 08 | The Fifth Element | 17 |
| Excerpt 09 | Man of Galilee | CD1 #2 |
| Excerpt 10 | JFK | 8 |
| Excerpt 11 | The Alien Trilogy | 5 |
| Excerpt 12 | The Rainmaker | 3 |
| Excerpt 13 | Oliver Twist | 8 |
| Excerpt 14 | The Omen | 9 |
| Excerpt 15 | The Portrait of a Lady | 3 |
| Excerpt 16 | The Portrait of a Lady | 9 |
| Excerpt 17 | The English Patient | 18 |
| Excerpt 18 | Big Fish | 15 |
| Excerpt 19 | Angel Heart | 4 |
| Excerpt 20 | Shakespeare in Love | 3 |
| Excerpt 21 | Shine | 10 |

| **Stimulus** | **Name of film** | **Sound-track #** |
|---|---|---|
| Excerpt 22 | Pride & Prejudice | 1 |
| Excerpt 23 | Oliver Twist | 2 |
| Excerpt 24 | Juha | 2 |
| Excerpt 25 | Modern Times | 3 |
| Excerpt 26 | Modern Times | 5 |
| Excerpt 27 | Modern Times | 9 |
| Excerpt 28 | City Lights | 2 |
| Excerpt 29 | City Lights | 7 |
| Excerpt 30 | City Lights | 11 |

*Table 2*. List of excerpts used for the Pre-Experiment.

## Appendix 2: Variance of listener responses for each excerpt in Pre-Experiment

| Excerpt # | Variance | Rank |
|---|---|---|
| **1** | 4.671 | 27 |
| **2** | 4.891 | 28 |
| **3** | 2.460 | 13 |
| **4** | 3.783 | 20 |
| **5** | 2.837 | 18 |
| **6** | 5.045 | 29 |
| **7** | 1.057 | 2 |
| **8** | 4.351 | 25 |
| **9** | 3.946 | 21 |
| **10** | 1.131 | 3 |
| **11** | 3.999 | 21 |
| **12** | 2.456 | 12 |
| **13** | 1.833 | 7 |
| **14** | 4.337 | 24 |
| **15** | 4.307 | 23 |
| **16** | 1.647 | 6 |
| **17** | 1.928 | 8 |
| **18** | 1.579 | 5 |
| **19** | 2.710 | 17 |
| **20** | 1.983 | 9 |
| **21** | 4.413 | 26 |
| **22** | 0.824 | 1 |
| **23** | 2.493 | 14 |
| **24** | 4.296 | 22 |
| **25** | 1.431 | 4 |
| **26** | 2.598 | 15 |
| **27** | 2.295 | 11 |
| **28** | 2.706 | 16 |
| **29** | 2.147 | 10 |
| **30** | 3.358 | 19 |

*Table 3*. Variance and ranks of excerpts used for the Pre-Experiment.

## Appendix 3: Excerpts used in Experiments 1 and 2


*Figure 4-1*. Excerpt 1: The Untouchables (8).


*Figure 4-2*: Excerpt 2: JFK (8).

*Figure 4-3*. Excerpt 3: Modern Times (3).

*Figure 4-4*. Excerpt 4: Pride and Prejudice (1).

*Figure 4-5*. Excerpt 5: Big Fish (15).

## References


Adorno, T. W. (1976). *Introduction to the sociology of music*. US: Seabury Press.

Askenfelt, A. (1986). Measurement of bow motion and bow force in violin playing. *Journal of the Acoustical Society of America, 80*(4), 1007-1015.

Audacity Team. (2016). Audacity®. Version 2.2.0. Audio editor and recorder. Retrieved from http://audacityteam.org/

Barthet, M., Depalle, P., Kronland-Martinet, R., & Ystad, S. (2010). Acoustical correlates of timbre and expressiveness in clarinet performance. *Music Perception, 28*(2), 135-154.

Barthet, M., Depalle, P., Kronland-Martinet, R., & Ystad, S. (2011). Analysis-by-synthesis of timbre, timing, and dynamics in expressive clarinet performance. *Music Perception, 28*(3), 265-278.

Behrens, G. A., & Green, S. B. (1993). The ability to identify emotional content of solo improvisations performed vocally and on three different instruments. *Psychology of Music, 21*(1), 20-33.

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.

Canazza, S., De Poli, G., Rinaldin, S., & Vidolin, A. (1997). Sonological analysis of clarinet expressivity. In M. Leman (Ed.), *Music, gestalt, and computing: Studies in cognitive and systematic musicology*, (pp. 431 - 440). Berlin: Springer-Verlag.

Canazza, S., De Poli, G., Rodà, A., & Vidolin, A. (1997). *Analysis by synthesis of the expressive intentions in musical performance.* Paper presented at the International Computer Music Association 1997, Thessaloniki, Greece.

Canazza, S., De Poli, G., & Vidolin, A. (1997). Perceptual analysis of the musical expressive intention in a clarinet performance. In M. Leman (Ed.), *Music, gestalt, and computing* (pp. 441-450). Berlin: Springer-Verlag.

Clarke, D. (2003). *Musical autonomy revisited*. New York: Routledge.

Conover, W. J., & Iman, R. L. (1981). Rank transformations as a bridge between parametric and nonparametric statistics. *The American Statistician, 35*(3), 124-129.

Cook, N. (2001). Theorizing musical meaning. *Music Theory Spectrum, 23*(2), 170-195.

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. (2011). A comparison of the discrete and dimensional models of emotion in music. *Psychology of Music, 39*(1), 18-49.

Ehresman, D., & Wessel, D. (1978). Perception of timbral analogies. *Rapport IRCAM, 13*.

Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion, 6*(3-4), 169-200.

Fan, Y. Q. [范玉琪]. (2014). “Gao Shan Liu Shui” in the timbre of guzheng and music aesthetics [以《高山流水》为例浅谈古筝音色与美学的关系]. *Xin Shi [心事]*(8), 202.

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.

Geertz, C. (1973). *The interpretation of cultures*. New York: Basic Books.

Grey, J. M. (1977). Multidimensional perceptual scaling of musical timbres. *Journal of the Acoustical Society of America, 61*(5), 1270-1277.

Hannon, E. E., & Trainor, L. J. (2007). Music acquisition: effects of enculturation and formal training on development. *Trends in cognitive sciences, 11*(11), 466-472.

Hockett, C. F. (1960). Logical considerations in the study of animal communication. In W. E. Lanyon & W. N. Tavolga (Eds.), *Animal sounds and communication*. Washington, DC: American Institute of Biological Sciences.

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

Juslin, P. N. (1997). Emotional communication in music performance: A functionalist perspective and some data. *Music Perception, 14*(4), 383-418.

Juslin, P. N. (2013). From everyday emotions to aesthetic emotions: Towards a unified theory of musical emotions. *Physics of Life Reviews, 10*(3), 235-266.

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin, 129*(5), 770-814.

Juslin, P. N., & Timmers, R. (2010). Expression and communication of emotion in music performance. In P. N. Juslin & J. Sloboda (Eds.), *Handbook of music and emotion: Theory, research, applications* (pp. 453-489). Oxford: Oxford University Press.

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

Kamenetsky, S. B., Hill, D. S., & Trehub, S. E. (1997). Effect of tempo and dynamics on the perception of emotion in music. *Psychology of Music, 25*(2), 149-160.

Kivy, P. (2002). *Introduction to a philosophy of music*. Oxford: Clarendon Press.

Konečni, V. J. (2008). Does music induce emotion? A theoretical and methodological analysis. *Psychology of Aesthetics, Creativity, and the Arts, 2*(2), 115-129.

Krumhansl, C. L. (1989). Why is musical timbre so hard to understand? In S. Nielzén & O. Olsson (Eds.), *Structure and perception of electroacoustic sound and music* (pp. 43-53). Amsterdam: Excerpta Medica.

Kruskal, J. B., & Wish, M. (1978). *Multidimensional scaling*. London: Sage.

Langer, S. K. (1948). *Philosophy in a new key (1942)*. New York: The New American Library.

Martin, F. N., & Champlin, C. A. (2000). Reconsidering the limits of normal hearing. *Journal of the American Academy of Audiology, 11*(2), 64-66.

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods, 44*(2), 314-324.

McAdams, S., & Cunibile, J.-C. (1992). Perception of timbral analogies. *Philosophical Transactions of the Royal Society London Series B, 336*(1278), 383-389.

McAdams, S., & Goodchild, M. (2017). Musical structure - Sound and timbre. In R. Ashley & R. Timmers (Eds.), *The Routledge companion to music cognition* (pp. 129-140). London: Taylor & Francis.

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.

Mead, G. H. (1934). *Mind, self and society*. Chicago: University of Chicago Press.

Meyer, L. B. (1956). *Emotion and meaning in music*. Chicago: University of Chicago Press.

Meyer, L. B. (1989). *Style and music: Theory, history, and ideology*. Chicago: University of Chicago Press.

Meyer, L. B. (1994). Emotion and meaning in music. In R. Aiello & J. A. Sloboda (Eds.), *Musical Perceptions*. Oxford: Oxford University Press.

Morrison, S. J., & Demorest, S. M. (2009). Cultural constraints on music perception and cognition. *Progress in Brain Research, 178*, 67-77.

Nattiez, J.-J. (1990). *Music and discourse: Toward a semiology of music*. New Jersey: Princeton University Press.

Nattiez, J. J. (2003). Musiques du XXe siècle. In *Musiques : Une encyclopédie pour le XXIe siècle*. Arles: Actes Sud.

Patel, A. D. (2010). *Music, language, and the brain*. New York: Oxford University Press.

Paus, T., Zatorre, R. J., Hofle, N., Caramanos, Z., Gotman, J., Petrides, M., & Evans, A. C. (1997). Time-related changes in neural systems underlying attention and arousal during the performance of an auditory vigilance task. *Journal of cognitive neuroscience, 9*(3), 392-408.

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.

Piekut, B. (2014). Actor-networks in music history: Clarifications and critiques. *Twentieth-Century Music, 11*(2), 191-215.

Redfield, R., Linton, R., & Herskovits, M. J. (1936). Memorandum for the study of acculturation. *American Anthropologist, 38*(1), 149-152.

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

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review, 110*(1), 145-172.

Sandel, G. (n.d.). *Definitions of timbre* (compiled by G. Sandell). Retrieved from http://acousticslab.org/psychoacoustics/PMFiles/Timbre.htm

Scherer, K. R. (2004). Which emotions can be induced by music? What are the underlying mechanisms? And how can we measure them? *Journal of New Music Research, 33*(3), 239-251.

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.

Smith, B. K. (1995). PsiExp: An environment for psychoacoustic experimentation using the IRCAM musical workstation *Society for Music Perception and Cognition Conference'95*. Berkeley, CA: University of California, Berkeley.

Spencer, H. (2015). The origin and function of music. In J. Shepherd & K. Devine (Eds.), *The Routledge reader on the sociology of music* (Vol. 2). NY: Routledge. (Original work published 1854)

Terwogt, M. M., & Van Grinsven, F. (1991). Musical expression of moodstates. *Psychology of music, 19*(2), 99-109.

Tillmann, B., & McAdams, S. (2004). Implicit learning of musical timbre sequences: statistical regularities confronted with acoustical (dis) similarities. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(5), 1131-1142.

Wobbrock, J. O., Findlater, L., Gergle, D., & Higgins, J. J. (2011). The Aligned Rank Transform for nonparametric factorial analyses using only ANOVA procedures. In *Proceedings of CHI 2011 – 29th conference on human factors in computing systems* (pp. 143-146).

Zangwill, N. (2011). Music, essential metaphor, and private language. *American Philosophical Quarterly, 48*(1), 1-16.

Zhao, D. Y. [赵冬艳]. (2017). Analyzing the aesthetics and use of timbre in vocal music [音色在声乐演唱中的美学特质及运用分析]. *Northern Music [北方音乐], 37*(7), 59.