Perception of affective intentions in music: Timbre cues and differences in musical cultures

Lena Heng

Department of Music Research
Schulich School of Music
McGill University
Montréal, Canada

August 2023

A thesis submitted to McGill University in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Contents

Abstract

Perception of affective intentions in music is a complex, yet commonplace, phenomenon that takes place in everyday life. It could be influenced by factors such as listeners' experience with different musical traditions and the style of the music. This dissertation explores the perception of affective intentions in music, investigating how differences in musical cultures influence the use of acoustic and musical features in the listening process, and proposes a music narrative framework in which appraisal processes contribute to understanding musical meanings and perceived affective intentions.

The first experimental study compares the differences between listeners with training in Chinese and Western musical traditions and nonmusicians in their perception of affective intentions of recorded excerpts interpreted with a variety of intended affects by performers on instruments from these two cultures. Results demonstrate the role of musical training in listeners' decoding of affective intentions. Acoustic analyses of the stimuli also found consistent and purposeful use of temporal, spectral, and spectrotemporal attributes in listeners' judging of affective intentions.

The second experimental study sought to expand this investigation to explore the dynamic processes involved in listening. A lengthy piece of Chinese orchestral music was used and participants were asked to provide continuous ratings on emotional intensity, valence, and arousal. Functional data analysis is used to compare differences in listener groups under the assumption that these continuous responses reflect a smooth variation, thereby treating each curve as arising from a single process. Time series analysis is also used to explore how each listener group reacts to the various acoustic and musical features over the course of the music. From the two experimental studies, it appears that acoustic and musical cues that influence valence responses are likely to be culturally learned, whereas the cues that influence perceived arousal and emotional intensity responses are based more on universal, culturally independent response mechanisms. An examination of the different combinations of cues involved in the ratings of affective intentions suggests that the listener groups might also process these cues in qualitatively different ways.

As it appears that acoustic and musical features are not utilized in a simple, direct, and linear way in the perception of affective intentions, the final part of this dissertation proposes a music narrative framework that attempts to incorporate processes of appraisal in understanding and judging perceived affective intentions in music.

This research uses various statistical analyses to disentangle the large number of complex variables that influence the listening process and to examine the dynamic relationships between sonic features and perceptual processes. It synthesizes current understanding of the cognitive processes underlying perceived musical affect and expands this by exploring continuous real-time listener responses to a lengthy piece of Chinese orchestral music. It also extends current understanding of timbre functions and the way learning and experience influence music perception, demonstrating this by examining different listener groups and using instruments and music outside of Western classical music practice. It advances knowledge about learning and perceptual processes involved in understanding musical meanings and affective intentions and highlights the complexity of how acoustic and musical features contribute to this process.

Résumé

La perception des intentions affectives en musique est un phénomène complexe, mais courant dans la vie quotidienne. Elle pourrait être influencée par des facteurs tels que l'expérience des auditeurs avec différentes traditions ou styles de musique. Cette thèse explore la perception des intentions affectives dans la musique, en étudiant la façon dont les différences de cultures musicales influencent l'utilisation des phénomènes acoustiques et musicales dans le processus d'écoute, et propose un cadre narratif musical dans lequel les processus d'évaluation contribuent à la compréhension des significations musicales et des intentions affectives perçues.

La première étude expérimentale compare les différences entre des participants formés aux traditions musicales chinoises et occidentales et des non-musiciens dans leur perception des intentions affectives d'extraits musicaux incluant divers affects voulus par des musiciens utilisant des instruments de ces deux cultures. Les résultats démontrent le rôle de la formation musicale dans le décodage des intentions affectives. Les analyses acoustiques des stimuli ont aussi révélé une utilisation cohérente et intentionnelle des attributs temporels, spectraux et spectrotemporels dans le jugement des intentions affectives.

La deuxième étude visait à explorer les processus dynamiques impliqués dans l'écoute. Un long morceau de musique orchestrale chinoise a été utilisé et les participants ont été invités à fournir des évaluations continues sur l'intensité émotionnelle, la valence et l'activation émotionnelles. L'analyse des données fonctionnelles est utilisée pour comparer les différences entre les groupes en partant du principe que ces réponses continues reflètent une variation régulière, traitant ainsi chaque courbe comme résultant d'un processus unique. L'analyse des séries temporelles est également utilisée pour explorer la façon dont chaque groupe réagit aux diverses caractéristiques acoustiques et musicales sur la durée du morceau. D'après les deux études expérimentales, il apparaît que les propriétés acoustiques et musicales qui influencent les réponses de valence sont susceptibles d'être culturellement apprises, alors que les propriétés qui influencent les réponses perçues d'activation et d'intensité émotionnelle sont davantage basées sur des mécanismes de réponse universels et culturellement indépendants. Une analyse des différentes combinaisons de propriétés impliquées dans les évaluations des intentions affectives suggère que les groupes d'auditeurs pourraient également traiter ces caractéristiques de manière qualitativement différente.

Comme il apparaît que les caractéristiques acoustiques et musicales ne sont pas utilisées de manière simple, directe et linéaire dans la perception des intentions affectives, la dernière partie de cette thèse propose un cadre narratif musical qui incorpore les processus d'évaluation dans la compréhension et le jugement des intentions affectives perçues dans la musique.

Cette recherche utilise diverses analyses statistiques pour examiner les relations dynamiques entre les caractéristiques sonores et les processus perceptifs. Elle synthétise la compréhension actuelle des processus cognitifs sous-jacents à l'affect musical perçu et l'approfondit en explorant les réponses continues des participants écoutant un long morceau de musique orchestrale chinoise. Elle élargit également la compréhension actuelle des fonctions du timbre et de la manière dont l'apprentissage et l'expérience influencent la perception de la musique, le démontrant via l'examination de différents groupes d'auditeurs ainsi qu'en utilisant des instruments et oeuvres différents de la musique classique occidentale. Cette recherche fait progresser les connaissances sur les processus d'apprentissage et de perception impliqués dans la compréhension des significations musicales et des intentions affectives et souligne la complexité des propriétés acoustiques et musicales qui y contribuent.

Acknowledgments

“It is good to have an end to journey toward; but it is the journey that matters, in the end.”

— Ursula K. Le Guin, *The Left Hand of Darkness*

Acknowledgments terrify me, not because I am ungrateful, but because I am terrified I might miss someone out as there is just such an incredibly large number of beings who have given so generously and made such a huge impact on me. Coming close to the end of my PhD studies as I am writing this, I am reminded of the quote by one of my favourite authors, that it is the journey that matters. Indeed, looking back, this journey has been an amazing one and I have many people to thank for it.

First and foremost, I would like to express my utmost gratefulness to my supervisor, Stephen McAdams. He is the most uncompromising and meticulous researcher, a wonderful mentor, and an immensely patient teacher. His optimism and faith gave me the opportunity to move across continents and pursue my interest in research. His open-mindedness also allowed me to freely discuss my ideas, and his support and enthusiasm enabled me to shape some of these preposterous ideas into actually good ones. I would also like to thank Robert Hasegawa for insightful discussions about music, and for always taking the time out of his very busy schedule to help me with any questions I have. Thank you to Bennett Smith for the wonderful wizardry with all things technical, hikes, lunchtime discussions of Star Trek and Doctor Who, and chocolates.

Past and present members of the MPCL, you have all made my time at McGill awesome. There is such a wide variety of talents here that I have my knowledge

expanded and horizons broadened with every interaction. Iza Korsmit and Yuval Adler, it was fun sharing our PhD journeys every step of the way! Erica Huynh and Jade Roth, beer, cats, foosball! Huang Yifan, Zhu Linglan, Corinne Darche, Joshua Rosner, Behrad Madahi, Andrés Gutiérrez Martínez, Ben Duinker, I always enjoy being at the lab because of all of you! Special thanks to Lindsey Reymore who is always generous with her time, offering help with music analysis and writing, and Kit Soden for getting the beautiful Chinese instruments and orchestration page up on TOR at such short notice. Also shoutout to Aurélien Antoine for advice on coding and help with everything French. My French is still terrible but I am working on it!

I am very grateful to the various institutions that have provided financial support for various aspects of my PhD studies and this dissertation project. McGill for my scholarship, and the Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT) for the Inter-Centre Research Exchange. The research was also supported by grants from the Canadian Social Sciences and Humanities Research Council (SSHRC) and the Fonds de recherche du Québec—Société et culture (FRQ-SC), as well as a Canada Research Chair awarded to Stephen McAdams.

Finally, I have to thank the countless people, past and present, who have supported, encouraged, influenced, and changed me. There are way too many for me to name, but without every single one of these people, I would not be the person I am today. Last but not least, thank you M, for sharing all the journeys with me.

Contribution of authors

This is a manuscript-based thesis. The following chapters are based on manuscripts prepared for submission or have been submitted already to peer-reviewed journals.

1. Heng, L. and McAdams, S. (2022). Timbre's function in the perception of affective intentions: Differences between musical traditions. [Manuscript submitted to *Musicae Scientae*]. (Chapter 2).

2. Heng, L., Wei, C., and McAdams, S. (2023). Continuous response in music listening: Training in different musical traditions influence perception of affective intentions. [Manuscript in preparation]. (Chapter 3).

3. Heng, L., and McAdams, S. (2023). A music narrative framework for affect perception in music listening. [Manuscript in preparation]. (Chapter 4).

Stephen McAdams was the supervisor of this thesis and provided guidance in the experimental design, statistical analyses, and the interpretation of the results. He also provided funding for the compensation of experimental participants. Bennett Smith programmed all the interfaces for the experimental studies. Claire Wei was a McGill undergraduate research student in Psychology whom I co-mentored with Stephen McAdams. Under my supervision, she contributed to the data collection and statistical analysis for Chapter 3. As a principal author, I was responsible for the background research, devising the experimental paradigms, analyzing the data, interpreting the results, and writing the above-listed manuscripts as well as all other parts of this thesis.

List of Figures

List of Tables

Part I

Introduction

Engaging in music is a common activity most people do regularly. This can involve physically more dynamic activities such as singing, playing an instrument, or dancing, to activities that demand less physicality such as listening to music as an audience. Regardless of the musical activity, it should not be too difficult to agree that the aspect of processing a series of sonic vibrations through our perceptual systems is a central one in most musical activities. Music is organized sound. As Davies (2019) argues, it is this organization that allows one to hear sounds as music; someone deriving pleasure in merely hearing the sensuous properties of sounds could not be said to be enjoying it as music per se. Music then, is a set of sounds that follow certain principles of order, according to certain conventions. An understanding of these conventions and the implications of the principles of order will be necessary for sounds to be heard as music.

Main Ideas and Concepts

Musical Meanings

Musical meanings emerge from the interpretation of organized musical information. They are inseparable from the performers' and/or composers' expression and the listeners' recognition of the musical intentions. Similarly, Juslin and Timmers (2010) define the communication of affective intentions in music as involving on the one side, the intention to express particular emotions by the performer and on the other, by the listener recognizing the emotion that is expressed. Communication therefore involves both the expressive intent and the receptive position. The discourse surrounding musical meanings and communication in music has been complex, with differing perspectives from philosophers such as Langer (1948), who put forward theories of music being the precursor to language, to psychologists such as Patel (2010), who argues that we cannot simply understand music as we do another form of language. Not only are there conventions with regards to the constraints and strategies used in various styles and traditions of music, this process of interpreting musical information is dependent on its progression over time. In addition to the sound qualities in themselves, the relationship between these sound qualities provide musically pertinent information that listeners have to perceive and process (Thoresen, 2015).

Not all musical meanings emerge from the same aspect of a musical source. Meyer (1956) discusses "designative meaning," in which a stimulus may be meaningful because it indicates or refers to something which is different from itself in kind, and "embodied meaning" in which a stimulus may acquire meaning because it indicates or refers to something which is like itself in kind. Patel (2010) refers to these as intramusical and extramusical meanings, and other scholars and researchers have also discussed these different kinds of meanings extensively. Hanslick (1854) for instance, thinks that the study of musical aesthetics should be rooted in musical structure instead of the reactions of the listeners, making him effectively a formalist studying intramusical meanings. On the other side, Ratner (1980), Agawu (1991), and Hatten (2004) assert that topics provide subjects for musical discourse—they believe that extramusical meanings are part and parcel of understanding music. Kivy (1980, 2002) and Davies (2019) believe that music is capable of expressing emotions by virtue of its form, putting them at the intersection of intramusical relations eliciting extramusical affective responses. Meyer (1956) acknowledges the evocation of mental imagery of nonmusical phenomena in music listening, Sloboda (1985) discusses narrativity in musical thought, and Kivy (2002) uses the term "wordless drama" in interpretive writing of instrumental music; these are all different forms of extramusical references to music. Koelsch (2011) attempts to sum up these different perspectives, focusing on the processing of meaning as it emerges from the interpretation of musical information. Musical meanings, he believes, can "emerge from extra-musical sign qualities, from intra-musical structural relations, from musicogenic effects, from the establishment of a unified coherent sense out of 'lower-level' units, and from musical discourse" (p. 90).

Meyer (1994) posits that even though meanings are mental facts, they are not arbitrary because our experience of the world has created in us an apprehension of relationships existing in the objective world, and this is the basis upon which our mental facts are built. Meyer believes that meaning arises "when an individual becomes aware, either affectively, or intellectually, of the implications of a stimulus in a particular context" (p. 9). It would appear then that an awareness of implications plays an important role in meaning-making. If the implications within the context of a stimulus are involved, meaning-making will not only require auditory processing responses but also an awareness of certain expectancies. Extracting information from a stimulus means that the individual will have to compare the stimulus with an understanding of previous occurrences, compute the probabilities of occurrences, and then see if the expectant responses are satisfied, delayed, or blocked totally. Meyer (1989) also suggests musical styles provide particular musical constraints for which musical features are selected and used. There has to be a shared understanding of how certain musical features might be associated with extramusical or intramusical elements and how the relationships with other features pan out over the course of a piece of music. This understanding occurs only when there is a knowledge of the style in question, whether implicitly or explicitly. At the same time, although styles differ in the music found among different communities of people, periods in history, or even between sets of compositions by the same composer, what remains constant are 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" (Meyer, 1994, p. 7). The process of composition comprises a series of choices made within some set of constraints. Constraints can be physical or psychological, deriving from principles governing perception and cognition of musical patterns, but they can also be culture-dependent, deriving from rules that have become ingrained through repeated use. Style involves choosing within some set of constraints, and musical meanings arise when a listener understands the implications involved in this choice. Since the creation of musical meanings involves the element of conscious decisions in selecting and organizing sonic elements within a set of constraints, and the comprehension of musical meanings occurs when these implications are understood, discussing musical meanings necessarily involves considering the set of constraints and decisions that are available in organizing musical materials, as well as the cognitive processes involved in comprehending the implications of this organization.

It would appear then that there could be a high degree of misconstrued understandings of musical meanings when one listens to the music of an unfamiliar musical tradition. However, as Thompson and Balkwill (2010) proposed in their cue-redundancy model, listeners across different cultures are able to appreciate affective qualities of unfamiliar music by attending to these commonalities, whereas listeners who are familiar with a musical style should find it easier to decode emotional meaning in that music because they can draw from both culture-specific as well as common psychophysical cues. Thompson and colleagues (2022) also discuss how musical events carry a whole contextual framework and examine the roles of structure, self, and source appraisal in making sense of these musical events. As such, a listener's cultural background and experience play an important part in their comprehension of music. Most studies exploring cross-cultural differences in music perception have studied populations from different geographical locations (e.g., Margulis et al., 2019; Balkwill et al., 2004). Even though comparing differences in music perception between participants with different musical experiences from different population can provide information about the role of enculturation in different musical traditions, it is often difficult to tease apart influences due to linguistic and socio-cultural differences from those of musical experience and expertise.

Affective Intentions

Affective intention is an important and salient notion in musical communication. Russell (2003) observes that emotions are a dominant aspect of human life and an important topic in psychological research. It is, however, an extremely complex field of study with many different points of view (Russell, 2003). With regards to the communication of affective intent in music, three main issues appear to be pertinent here: the different models of affect and emotions, the locus of emotion, and the types of musical features and the ways they carry information.

Models of Affect and Emotion

In the field of emotions and affect, psychologists hold differing views on the types of models that can best describe them. The most often used and cited ones range from Ekman (1992) who argues for the existence of “universal basic emotions”, to Shaver and colleagues (1987) who propose the concept of emotion prototypes, to Russell (1980), who proposed the two-dimensional (activation-pleasantness) circumplex model of affect, to Schimmack and Grob (2000), whose three-dimensional model describes the core affect in terms of pleasure, energy, and tension.

Barrett (2006) also proposes a different model for understanding emotions. She points out that many contemporary discrete models of emotion are guided by the assumption that emotions are entities. These assumptions, however, have not been well grounded in the available empirical evidence. Instead of emotions being entities of the mind, she sees them as emergent phenomena that react to and change with the context. An experience of emotion results from “conceptualizing a very basic form of affective responding during the act of categorization, where the categorization of affect is guided by knowledge about emotion that is acquired from prior experience, tailored to the immediate situation, and designed for action” (p. 21). An experience of emotion, in other words, arises from a veridical sensation before the consciousness registers it as a concept, and an emotion is an object of consciousness.

These different models of affect and emotion have been used in various experimental studies on music, but few have systematically studied how the different models might tap into similar or different aspects of emotion communication. It is also unclear which aspects of acoustic and musical features are involved. Often, these studies have implied that the discrete and dimensional models of emotions are highly convergent although only a few have explicitly studied their convergence (e.g., Eerola & Vuoskoski, 2011; Vuoskoski & Eerola, 2011). Cespedes-Guevara and Eerola (2018) provide evidence demonstrating a constructionist attribution of music emotion perception and argue that "it may be more parsimonous to assume that music communicates fluctuations of affect which can be mapped onto many possible meanings via associative mechanisms" (p. 14). They propose the perception of emotions in music to consist of an active process of meaning construction: musical acoustic cues signal variations in levels of arousal and valence which may be conceived into various discrete categories of meanings as these variations are differentiated by a listener. The musical structure may also allow a listener to privilege particular meanings and discount others. In other words, music perception might involve attributions of affect that could be mapped onto different possible meanings through associative mechanisms. It would appear then that listeners make sense of the music they hear in different ways depending on the context. An awareness of the implications of a stimulus plays an important role in meaning-making.

Locus of Emotion

Another important aspect when studying affective responses is the perennial dichotomy between induced and perceived emotions. This issue is also an important one in music affect studies. A handful of empirical studies have acknowledged the complexity of the relationship between these emotion loci (e.g., Evans & Schubert, 2008). Gabrielsson (2001) distinguishes between emotion perception and induction in empirical studies of music and emotions, clarifying the difference: the perception of emotional expression "without necessarily being affected [by it]—is mainly a perceptual-cognitive process" (p. 124), whereas emotions induced by music are listeners' emotional response to it—an emotional reaction has taken place. The line between these two, however, is often blurred, and although it is possible to discuss them independently, they might influence each other in the actual processes of perceiving an expressed emotion and of having an emotion being induced.

In the discourse around induced emotions in music, Scherer (2004) believes that there are different mechanisms through which music can evoke emotions, and also that the emotions induced by music may be very different from emotions experienced in real life. Konečni (2008), on the other hand, argues that the body of research supporting induced emotions in music is not convincing. He believes that even though it is possible for basic emotions to be induced, there likely has to be a personal association with the emotion-inducing event. Even as music arouses emotions, the fact that music is a form of art that is abstract and does not appear to have much direct implications to everyday concerns (Juslin, 2013) leads one to wonder if the types of affect and emotions psychologists have been studying are the same set of affect and emotions that are associated with music. The types of emotions music induces are likely to involve an overlapping but different subset of emotions from those that humans experience more generally. Perceived emotion is another aspect of this complex question. There are some everyday emotions such as disgust, which music might be unable to communicate, and so perceived emotions in music are also likely to involve a different subset of emotions. Juslin (2013) writes about aesthetic judgment in his theory of musical emotions which might be implicated to a greater extent in perception than in induction. Cespedes-Guevara (2021) builds a constructionist model in explaining musical emotions and emphasizes the importance of appraisal mechanisms. Lennie and Eerola (2022) propose the CODA (Constructivistly-Organized Dimensional-Appraisal) model that also emphasizes appraisal in the judgment of affect. They also believe that it is the assessing of goal relevance that distinguishes between perceived and induced emotions.

Narrative content and musical meaning

When an abstract stimulus is encountered, humans often tend to create narratives in order to make sense of it. With an abstract art form such as music, narratives are frequently and effortlessly generated (Margulis, 2017). It will be reasonable to assume that affective content can be easily associated with the narrative content formed from music, and that it is also likely affective intentions perceived in music can influence the narrative content that is being generated. However, there appears to be still a lacuna in research on the relationship between narratives generated, and affective intentions perceived in music.

Acoustic and Musical Features Involved in Communicating Affective Intentions

An important prerequisite of communication is the presence of both expression and recognition (Juslin & Timmers, 2010). 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. Mead (1934) 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 when the gesture made has the same meaning for the individual who makes it and the individual who responds to it. The notion of expression therefore 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 that on one side, the performer needs to have an awareness of the processes that are entailed in the sound production that may elicit particular emotional intentions, and on the side, the listener has to decode these intentions in a way that is congruent with what is expressed by the performer.

In terms of communicating affective intentions in music, a few factors are pertinent. Firstly, in musical cultures that have a tradition of codifying musical works as repeatable entities (Western classical music being one, Chinese music being another), there is the framework of the composed musical structure, elements usually "represented by designations in the conventional musical notation, such as tempo markings, dynamic markings, pitch, intervals, mode, melody, rhythm, harmony, and various formal properties” (Gabrielsson & Lindström, 2010, pp. 367–368). Secondly, there are elements that are realized in performance (Juslin & Timmers, 2010, p. 458), and finally, the aspect of the listeners’ perception of the intended expression.

Spencer (2021) proposed that vocal music, and by association instrumental music, is intimately related to vocal expression of emotions. Juslin and Laukka (2003) conducted a meta-analysis on vocal expression of emotions and musical expression to compare their similarities and differences in the communication of emotions. They found that studies of vocal expression and music performance have converged on the conclusion that encoders (speakers and performers) can communicate basic emotions to decoders (listeners) with above-chance accuracy for the five basic emotion categories: anger, sadness, happiness, fear, and love. Cross-cultural vocal communication of emotions was also found to be accurate, although within-culture accuracy was slightly better. The results are therefore consistent with Thompson and Balkwill’s cue-redundancy model. Acoustic cues used in communication are probabilistic and there is a redundancy in the available cues. More than one way of using the cues might lead to a similarly high level of decoding accuracy.

Eerola and colleagues (2012) studied 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. The fact that such brief isolated sounds were capable of communicating affective intentions points to the role timbre plays on the surface acoustic properties of a sound. This is communicated through the inherent property of an instrument, the range of timbre manipulations afforded to it, and the conscious effort of a performer manipulating the sound.

Eerola and colleagues (2009) attempted to predict affective ratings in music from audio using multivariate regression models and found that several factors estimated through the computation of novelty curves can be used to predict emotional ratings: timbral elements of spectral centroid, spectral spread, and spectral entropy; harmonic characteristics such as key clarity, harmonic change, and measure of “majorness”; registral cues; rhythmic characteristics including clarity of pulsation and rhythmic periodicity; articulation; and musical structure.

A study by McAdams and colleagues (2017) attempted to relate acoustic descriptors with affective ratings of perception and to model the affective qualities of musical instrument sounds. They found that different combinations of audio descriptors make major contributions to the emotion dimensions from Schimmack and Grob’s (2000) three-dimensional model of affect, which suggests that these affective dimensions involve different, though overlapping, combinations of acoustic properties. Valence is “more positive with lower spectral slopes, a greater emergence of strong partials, an amplitude envelope with a sharper attack, and an earlier decay; [... tension arousal is higher with] brighter sounds, more spectral variation, and gentler attacks; [... and energy arousal is greater with] brighter sounds, higher spectral centroids, slower decrease of spectral slope, and a greater spectral emergence” (p. 17).

Acoustic descriptors of a sound contribute not only to static information in the communicative process, but also interact in their relations with other sounds that are heard in complex ways. Researchers have examined the role of timbre in building and releasing of tension (Paraskeva & McAdams, 1997), the function of timbre in orchestral gestures for emotional valence (Goodchild et al., 2019), and the possibility of learning timbre sequences (Tillmann & McAdams, 2004). All these studies provide evidence to support timbre as a strong structuring force in music. However, because of the complexity of musical sequences, as well as their evolution in time, the contributions of acoustic descriptors and how they evolve in music perception is more difficult to study.

Continuous response of affect perception in music

Music and the associated affective responses exist in time, and it stands to reason that a study of listeners' responses to music perception should be time-dependent. Despite this, collecting time-dependent, continuous responses to music is a less common approach in the study of emotional responses to music. As Schubert notes, "only a small fraction has focused on the fundamental nature of time in music and the emotion it produces"(2010, p. 223). Many studies have implicitly assumed that post-performance response, collecting participants' responses after the conclusion of the stimulus, are comparable to assessing continuous responses over time. This, however, might not be an accurate reflection of participants' responses. Duke and Colprit (2001) demonstrate that participants' global evaluations reflect peak moments, rather than an average of the continuous response. It is also not possible to determine moment-to-moment changes in global assessments of lengthier stimuli.

Some issues arising from the collection of continuous self-report responses require specific methodological resolutions. Problems of synchronization, large data sets and overfitting, and interpretation need to be addressed. Researchers exploring listeners' continuous responses to music have utilized descriptive approaches by ‘eyeballing’ graphs and looking for unfolding patterns, as well as inferential approaches such as functional data analysis (e.g., McAdams, et al., 2004; Vines, et al., 2005) and time series analysis (e.g., Bailes & Dean, 2012; Dean & Bailes, 2010; Dean & Bailes, 2011; Schubert, 2004). A combination of these methods will be useful in exploring continuous time-varying responses to a lengthy piece of music in real time.

Main Thesis

This thesis explores the process of creation of musical meanings during listening, and how listeners with different musical backgrounds might process and interpret the sonic information in different ways. Constraints found within different musical traditions are likely to diverge in several areas, even if there may be many other areas of convergence. Listeners experienced in different musical traditions will have different sets of decisions available to them in this process of comprehending the implications of the sonic materials.

Most research on music listening from different cultures employs participants from different geographical locations (e.g., Balkwill et al., 2004; Balkwill & Thompson, 1999; Bowling et al., 2012; Cowen et al., 2020; Laukka et al., 2013; Thompson & Balkwill, 2010). The differences observed suggest that musical experience does influence the music listening process, but it is difficult to untangle how much learning a particular musical tradition influences this processing, and how much is due to linguistic and other socio-cultural factors. Expertise in a musical tradition and linguistic and socio-cultural factors might also interact to further complicate these differences. Although there are individual differences among people, and micro-cultures within the same geographical location, including participants with different musical training from the same geographical location attempts as much as possible to reduce the differences due to language and socio-cultural factors so that any that are due to musical experience and learning can be more clearly observed.

The fact that musical sounds exist in time also has to be taken into account when studying the process of meaning-making in music listening. Even short sound events happen over time. Acoustic features evolve within this short time period and a great deal of important information is communicated even in short sonic events. One of the salient timbral qualities listeners perceive in single instrumental tones, for instance, is attack time (Grey, 1977; Krumhansl, 1989; McAdams et al., 1995)—time being an obviously crucial aspect involved here. As a perceptual attribute, timbre is also not a direct mapping of acoustic properties of a sound event, but rather, connected to the sound event by a "complex sequence of information processing steps in the human auditory system" (Siedenburg & McAdams, 2017, p. 2). In musical relationships occurring over longer durations, it is also clear that an important fundamental property is that it "requires time to exist [and] any definition of music must include this property, even if by implication" (Schubert, 2010, p. 223). All these mean that temporal context plays an important role in influencing listeners' responses. "An identical physical stimulus may be perceived differently, depending on the context" (Vines et al., 2005, p. 137). A dynamic evaluation of the relationship between acoustic and musical elements on the one hand with real-time listener perceptions on the other will be very helpful in aiding the understanding of the listening process and how musical communication takes place.

Research Questions

This thesis seeks to clarify the following questions:

1. Do differences in musical training play a role in influencing how listeners perceive affective intentions in music? Will different acoustic and musical features appear more salient because of differences in how and what aspects of the musical sound are emphasized in different musical traditions? Will the same acoustic and musical features be salient but interpreted in different ways? Will acoustic and musical features be used in combination in different ways by listeners from different musical traditions?

2. How do listeners' perceived affective intentions change dynamically over a lengthy piece of music? What acoustic and musical features might be involved in listeners' perceptions?

3. What might be the processes that underlie the perception of affective intentions? Are there different mechanisms that modulate and/or mediate this judgment and how might they be implicated dynamically over the course of a lengthy piece of music?

Modern Chinese Art Music and Chinese Orchestral Music

The stimuli used in the experimental studies in this dissertation include those of Chinese instruments, and Chinese orchestral music. This thesis also explores differences in musical training in groups of listeners with no formal musical training, formal training in Western classical music, and formal training in Chinese music, all residents of Singapore. Due to the differences in instrument construction, aesthetic ideals, and history of development, the timbre of Chinese instruments, as well as orchestration techniques and performance aesthetics are different from those of Western instruments and orchestral music. A brief description of the Chinese orchestra and its development will be outlined in this section for readers who might be less familiar with Chinese music and the Chinese orchestra.

Chinese music consists of many different genres, types of instruments, ensemble make-up, and even differing philosophies on music-making. The Chinese orchestra is a very recent invention, dating back to less than a century ago, with its development very much influenced by Western art music that has entered China since the nineteenth century (Chan, 2003, pp. 14–17). The term modern Chinese art music will be used here to describe this modern Chinese music that has been developed, including Chinese orchestral music, as well as solo or chamber pieces composed for the instruments found in the Chinese orchestra. Before this, the various types of music present in China included different forms of folk and ancient classical music. Although there are very few musicians trained solely in the ancient classical tradition now, folk traditions follow a very different trajectory. Even though there is a divide within professional music training institutions—folk musicians usually do not go through the same kind of institutionalized training as do the musicians in modern Chinese art music—there is still a very strong connection between modern Chinese art music and folk music. The folk music of China has a long history and at the same time, it is still very much active and constantly evolving. Throughout history, folk music has been a big part of the Chinese musical culture and could be one of the most important forms of traditional music that influences the development of modern Chinese art music.

The Western orchestra started to develop as early as 1600. With the development in construction of string instruments by the end of the seventeenth century, the Baroque orchestra started to grow from this core of bowed string instruments (Spitzer & Zaslaw, 2004). By the Classical period though, woodwinds and brass had become stable constituents of the orchestra, and as Adler (2002) observes, as instruments within the orchestra become fixed, the ways of treating each "note, chord, timbre, and nuance became an integral part of the composition [and] it was necessary to codify the art of orchestration" (p. 5). What this implies then is that the sounds and the use of the timbres of the instruments have been codified into a set of abstract knowledge structures within the minds not just of composers and musicians versed in Western classical music, but also implicitly of listeners who have been enculturated in this musical tradition. Although how each sound and the combinations of them may be used are extremely diverse and varied, there are certain stylistic constraints that are present and which will influence the perception of these instrumental sounds.

Similarly, due to the way in which the Chinese orchestra and its instruments have developed, a very different set of abstract knowledge structures and sound universes may arise. There are numerous types of folk instrumental ensembles all around China, and ensembles from different regions use different instruments, play music of very different styles, and have very different functions for their music. The first recorded attempt at "orchestrating" a traditional melody for an ensemble was by a "silk and bamboo" ensemble in 1920. This ensemble is so named because of the materials used in the instruments, silk for the bowed and plucked string instruments and bamboo for the wind instruments. In 1935, the folk music ensemble of the former Central Broadcasting Corporation was formed, originally meant to perform folk music from the southern parts of China and so contained only silk and bamboo instruments from that region. It was later expanded to include music from the northern parts of China, bringing along various kinds of percussion and other wind instruments (Chan, 2003, pp. 14–15). As opposed to the Western orchestra, which developed gradually from a core of strings with other instruments slowly being added, evolving the sound of the orchestra in the process, instruments of the Chinese orchestra were put together in a relatively short period of time, with each of their sounds and playing techniques idiomatic of the folk instrumental ensembles they originated from. Since the 1950s, there has been ongoing research and development into instrument construction and modification of the acoustic properties of the instruments. The sounds of the instruments commonly used in the Chinese orchestra are now very different from when they were originally used in their traditional folk ensembles. However, the influence from the sounds of the traditional folk ensembles are never too distant from the Chinese orchestra, and there are complex interactions between the modern idiomatic sound of the instrument within the orchestra and the idiomatic sound within the various traditional folk ensembles. Although contemporary musicians trained in modern Chinese art music may not be trained in the performance of folk music, they are, in the course of their musical training, taught the fundamentals and theoretical rules underlying the structural and stylistic features of folk music. Like the Western orchestra, the Chinese orchestra is made up of numerous instruments with different modes of sound production and different timbral universes. A list of some commonly found instruments in the Chinese orchestra, their characteristics, and some orchestration techniques for these instruments can be found in the following link: https://timbreandorchestration.org/resources/instruments/ensembles/chinese-orchestra.

In addition to factors that are influenced by the construction of the instruments, there is also that of performance aesthetics. Performance aesthetics may have been shaped over time with higher values placed on certain types of sounds over others because of the function of music in society, the types of meanings intended to be communicated, and even the locations in which the music might be played. Modern Chinese art music draws influences from the Western classical music tradition notably in composition and orchestration techniques. Composers, musicians, and listeners would have learned a certain set of knowledge structures with regards to the sonic characteristics derived from these techniques. At the same time, performance aesthetics are also highly constrained by the instrument—the materials and the construction of an instrument means that certain types of sounds are just not possible on a particular instrument, whereas others can be manipulated with a high degree of control.

This combination of the way in which an instrument is built, the performance aesthetics, and the role of the instrument both as a solo instrument and within an orchestra or an ensemble contribute to the range of sounds that are desired within a musical tradition. Familiarity with particular universes of timbre also influences a listener's perception of different instrumental sounds (Siedenburg, 2017; Siedenburg et al., 2016) and the relation of their timbre universes to affective communication.

A few recent studies have attempted to expand the understanding of the timbre of Chinese instruments. Zhang and Xie (2017) conducted multidimensional scaling of listeners' perceptual ratings of dissimilarity of Chinese instrumental sounds. They had listeners rate the dissimilarities between pairs of 18 different sounds of plucked, bowed, and wind instruments that may be found in the modern Chinese orchestra.

The set of sounds was analyzed with the MIR Toolbox and four of the acoustic descriptors—spectral centroid, low energy rate, log-attack time, and brightness—were found to correlate significantly with the perceptual data. Spectral centroid measures the spectral centre of gravity and gives an indication of the brightness of a sound. Low energy rate shows the percentage of frames that have less than average energy, and brightness measures the amount of energy above 1.5 kHz. Log-attack time takes the log of the duration between the start and end of an attack and reflects the sharpness of an attack. As compared to the timbre spaces for Western instruments where spectral centroid and attack time were the two most heavily weighted perceptual dimensions, this timbre space for Chinese instruments saw low energy rate and brightness as the main contributors. While low energy rate and brightness are both related to the spectral centroid, they also describe slightly different aspects of the acoustic signal. It appears then that Chinese instruments might have certain special timbre characteristics that differ from Western ones, or that there is a confounding factor of listener characteristics. The group of participants for this experiment were all Chinese university students who might have different experiences with music as compared to the participants from the other timbre space experiments. Wang, Wei, and Yang (2021) also found that both cultural and musical backgrounds influence emotion perception in music, and the acoustic features implicated in this are slightly different in the different listener groups. In another study, Wang, Wei, Heng, and colleagues also found that “musical elements related to timbre, register, and dynamics features are culturally specific” (Wang, Wei, Heng, et al., 2021, p. 11).

Methods of Data Analyses

The Timbre Toolbox originated from Peeters and colleagues (2011) who created computations to extract acoustic descriptors of sound signals. These descriptors are derived from the diverse set of audio descriptors found in the literature on speech analysis, perception of musical timbre, sound classification, and music information retrieval. The current revised version of the Timbre Toolbox was reprogrammed by Kazazis and colleagues (2021). The MIRtoolbox (Lartillot, 2022) is another toolbox run in the MATLAB environment with an integrated set of functions dedicated to the extraction of musical features from audio files. These two toolboxes are used to derive the set of acoustic and musical features found in the stimuli used within the experimental studies in this dissertation.[1]

Being a complex auditory attribute or set of attributes of a perceptually fused sound event, the audio features are likely to be strongly collinear. Novel methods of analyses also have to be explored to study processes that evolve over time. Partial least-squares regression (PLSR) is able to analyze "data with strongly collinear (correlated), noisy, and numerous X-variables, and also simultaneously model several response variables" (Wold et al., 2001, p. 109). A Varimax rotation is then performed on the resultant space. The Varimax rotation is a statistical technique that maximizes the variance shared among items. The squared correlation of items related to one factor is increased, while the correlation on any other factor is decreased so that the interpretation of the loadings of items are simplified, and the factors upon which the

[1] These descriptors used were developed from studies of Western musical instruments and may be limited in scope when applied to tasks such as automatic classification of instruments from other musical traditions. However, their use in these studies serves to provide a description of the various aspects of the sound signal and are not for automation tasks.

data load can be more specifically identified (Dilbeck, 2017).

In functional data analysis, the data are treated as a function rather than as individual vectors of observations at different time points. The main assumption in this method is that the set of observations within a function originates from the same process. For data that are collected continuously, the number of observations is too numerous for conventional methods of statistical analyses to be performed. Treating the set of observations from one participant as a single function allows for these types of analyses without problems of overfitting (Ramsay et al., 2009). Time series analysis is another analytical method useful for data that are taken sequentially in time and that are autoregressive in nature (Box et al., 2016). Impulse response functions are useful for studying the impact a variable has on a system by analyzing the reactions that arise from these shocks. In this case, the various acoustic and musical cues are taken as shocks and listener responses as the reactions to these shocks. Vector autoregressive (VAR) models are often used for forecasting, although it may be unclear which set of responses truly reflect the system as the same underlying VAR is used to compute all the impulses. To address this, structural vector autoregressive (SVAR) models apply restrictions so that unique impulse responses can be derived for each shock to a system. Impulse response functions can then be modelled to show how listeners respond to changes in each acoustic and musical feature.

Thesis Outline

This dissertation contains five chapters. Chapters 2–4 are self-contained manuscripts of three research articles and Chapter 5 provides a summary discussion.

Chapter 2 investigates how timbre functions as a carrier of information for communicating affective intentions in music. This chapter details an experimental study in which three groups of listeners with different musical backgrounds (musicians trained in Chinese and Western music and nonmusicians) were presented with phrases, measures, and single notes of recorded excerpts interpreted with different affective intentions. They had to make global judgments of the affective intentions within a two-dimensional affective space. Listeners trained in Chinese music appeared to be the most accurate, followed by listeners trained in Western music, and nonmusicians were the least accurate. This could be because the Chinese music tradition emphasizes different aspects of the musical sound as compared to the Western music tradition, and listeners trained in Chinese music have a larger number of acoustic cues available for their judgments. Acoustic features influencing listeners' perceptions were explored, and it appears that the different listener groups share more acoustic features they use to decode arousal than is the case with valence. It is likely that perceived arousal calls for more universal, common cues in decoding, whereas perceived valence depends more on learned, culture-specific cues.

Chapter 3 extends this exploration of perceived affective intentions by studying the continuous affective responses of listeners to a piece of Chinese orchestral music. The same three groups of listeners (with mostly different participants) were compared. Functional data analysis on the continuous responses found significant differences in certain sections of the piece, and valence responses diverged much more than arousal or emotional intensity responses between the three listener groups. Impulse response functions with time series analysis found certain acoustic and musical features likely influencing listeners' perceived affective intentions. Similar to the previous study

(Chap. 2), the features influencing perceived arousal were more similar among the three listener groups than those influencing perceived valence. It is also observed that the features do not always relate in the same manner in different regions of the piece.

As it appears that acoustic and musical features are not utilized in a simple, direct, and linear way in the perception of affective intentions, Chapter 4 incorporates a framework in which perceived affective responses in music derive from and interact with listener experience and knowledge and the narratives putatively constructed in the listening process. Four different sections that express high-arousal/negative-valence (H–), low-arousal/negative-valence (L–), low-arousal/positive-valence (L+), and high-arousal/positive-valence (H+) according to listeners' ratings were analyzed separately, and a close analysis of the acoustic and musical features implicated in each of these different sections, together with a study of the compositional techniques used, are explored to examine how these are utilized as listeners respond to music with varying affective and (implied) narrative intentions. Appraisal mechanisms likely influence each of these processes, and the resulting affective intentions are a dynamic product of complex interactions within the entire framework.

Finally, Chapter 5 summarizes and integrates the overall findings in this dissertation and discusses how they contribute to the understanding of perceived affective intentions in music perception and the influences and interactions of timbral and musical features with experience and learning. It concludes with directions for future research and remarks on the role of the function of timbre and musical features in the dynamic process of music perception.

Part II

Timbre's function in the perception of affective intentions: Effect of enculturation in different musical traditions

This chapter is based on the following research article:

> Heng, L. and McAdams, S. (2022). Timbre's function in the perception of affective intentions: Effect of enculturation in different musical traditions. [Manuscript submitted to *Musicae Scientae*].

Abstract

Timbre has been identified as a potential component in the communication of affect in music. Although its function as a carrier of perceptually useful information about sound source mechanics has been established, less is understood about whether and how it functions as a carrier of information for communicating affect in music. To investigate these issues, listeners trained in Chinese and Western musical traditions and nonmusicians were presented with phrases, measures, and individual notes of recorded excerpts interpreted with a variety of affective intentions by performers on instruments from the two cultures. Results showed greater accuracy and more extreme responses in Chinese musician listeners and lowest accuracy in nonmusicians suggesting that musical training plays a role in listeners' decoding of affective intention. Excerpts were also analyzed to determine acoustic features that are correlated with timbre characteristics. Temporal, spectral, and spectrotemporal attributes were consistently used in judging affective intent in music, suggesting purposeful use of these properties within the sounds by listeners. Examination of listeners' use of acoustic features reveals a greater sharing of features between Chinese and Western musician listeners compared to nonmusicians. Acoustic features used for decoding affect also appeared to differ more between listener groups for valence than for arousal. How timbre is utilized in musical communication appears to be implicated differently across musical traditions and valence responses seem to be more culture-specific and arousal responses more similar across cultures.

Keywords: timbre, musical affect, acoustic features, cross-cultural differences

Introduction

Communicative behaviour is shared by conscious living beings (Dance, 1970). The triadic relation between an object or idea, the reference to this object/idea, and the receiver of the notion of this object/idea has been widely acknowledged (e.g., Mead, 1934; Pierce, 2014). Others have brought this idea into the field of communication in music (e.g., Thoresen, 2015). Even so, there are still lacunae in the study of this process of signification in musical communication. As Nattiez (1990) believes, language should not be taken as the model of signification—it is simply one amongst many—and music provides another model with such signification potential.

Musical meanings, vary widely across cultures (Cross, 2009). In musical cultures with a tradition of codifying musical works as repeatable entities (Western classical music being one, Chinese music being another), musical meanings have to consider performers and composers' expression and the listeners' recognition of intentions in the music. In the context of this study, musical meanings will be broadly construed as information that is carried by music. The principles of auditory perception enable a listener to parse and process elaborate and abstract sonic information and allow for the communication of complex information. At the same time, Tomlinson (1984) reminds musicologists of the importance of interpreting the cultural context from which the particular form of music arises. Putting this another way, this information has to be interpreted through "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 these 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. 1). Meyer uses the term "cultural noise" to refer to “disparities which may exist between the habit responses required by the musical style and those which a given individual actually possesses” (p. 16). Because of this, listeners with different musical backgrounds and experiences may interpret the same piece of music in different ways. Understanding musical meanings has to take into account the poietic source that derived from a process of creation and the psychological processes and cultural backgrounds of listeners in reconstructing this "message" (Nattiez, 1990).

Although Juslin and Timmers (2010) primarily discuss affective intentions in music, their definition of affect communication is applicable to communication in the more general sense. Conceptualizing it as a variant of Brunswik’s lens model, Juslin and Timmers (2010) examine the communicative process where intentions from the performer(s) are expressed by means of a set of cues that are probabilistic and partly redundant, and these are then interpreted by the listener.

Affective intention in music, as an umbrella term covering all aspects of evaluative states with regards to musical communication, is an important and salient notion that may be considered an aspect of musical meaning. Communication of affective intentions involves “both a performer’s intention to express a specific emotion and recognition of this emotion by a listener” (Juslin & Timmers, 2010, p. 455).

In the field of emotions and affect, psychologists hold differing views on the types of models that can best describe emotions and affect. The most often used and cited of these models range from Ekman (1992), who argues for the existence of “universal basic emotions”, Shaver and colleagues’ (1987), who proposed the concept of emotion prototypes, Russell (1980) who came up with a two-dimensional circumplex model of affect, to Schimmack and Grob (2000) whose three-dimensional model describes the pleasure-energy-tension of core affect. Experimental studies on music, affect, and emotion have drawn on these different models. Eerola and Vuoskoski (2011) compared the discrete and dimensional models and found Russell's (1980) circumplex model to be sufficient in explaining listeners' judgments of perceived affect in music. Even in instances where listeners ascribe categorical or prototypical emotions to the music they hear, there are variations and nuances to these categories. It can therefore be assumed that regardless of whether listeners place the music they hear into specific emotion categories or otherwise, an understanding of the affective intentions according to a dimensional model is likely present, although the number of dimensions may vary. This study does not aim to evaluate the likelihood of the discrete versus dimensional model. In the interest of parsimony and based on evidence that the two-dimensional model of valence and arousal provides sufficient explanatory power (Eerola & Vuoskoski, 2011), this model will be used in the current study.

It has been shown that there are numerous ways in which music can communicate and elicit affect and emotions (Juslin & Laukka, 2004). Various dimensions of musical sound carry important information for the communication of affective intentions. These include the composed musical structure, elements usually "represented by designations in the conventional musical notation, such as tempo markings, dynamic markings, pitch, intervals, mode, melody, rhythm, harmony, and various formal properties" (Gabrielsson & Lindström, 2010, pp. 367–368), and elements that are realized in performance (Juslin & Timmers, 2010, p. 458). McAdams (1989) discusses the constraints of these dimensions on their potential for bearing form. Well-known bearers of form include pitch and duration, and the contributions of these to the interpretation of musical meanings have frequently been studied. Timbre also has a strong potential for bearing form, although there has been less extensive work on this musical dimension.

Timbre has been shown to carry perceptually useful information about sound source mechanics, and although it has been identified by music perception scholars as a component in the communication of affect in music (e.g., Eerola & Vuoskoski, 2011; Eerola et al., 2012; McAdams et al., 2017), there is still a lot to uncover about how it functions as a carrier of information for such communication. Timbral characteristics of an instrument have been found to influence listeners' perceptions of emotion (Hailstone et al., 2009). Studies have also shown the influence of timbre manipulation in performance on perceived emotion (Juslin & Timmers, 2010). A recent definition by McAdams (2019) considers timbre as a "complex auditory attribute, or as a set of attributes, of a perceptually fused sound event in addition to those of pitch, loudness, perceived duration, and spatial position. . . [and] is also a perceptual property, not a physical one" (p.23). Even though timbre is a psychophysical attribute, and it is the perception of a sound that defines timbre, we must not underestimate the importance of the physical acoustic properties that give rise to timbre perception. As surface acoustic properties carry important information for perception, a systematic approach to sound analysis that is oriented towards human perception (Peeters et al., 2011) and its relation to the communication of affective intentions in music will greatly aid the understanding of timbre's function in musical communication.

With regards to attributing affective intentions to musical sounds, Scherer and Oshinsky (1977) systematically manipulated certain acoustic parameters to study listeners' ratings on both discrete emotions and dimensional affective intentions. Other researchers since then have also found consistent mappings of acoustic cues to affective communication (e.g., Schimmack & Grob, 2000; Eerola et al., 2012; Bowman & Yamauchi, 2016; McAdams et al., 2017; Soden et al., 2019). There is much overlap in the acoustic dimensions that seem to be involved in carrying these affective intentions, suggesting that affective content is not just related to a single acoustic dimension, but also communicated through the complex combinations and interactions of several acoustic parameters. Juslin (2000), and Scherer and colleagues (2011) applied the modified lens model to the process of expression and perception. In these lens models, specificities or uniformity in cues do not have to be present in decoding. In other words, listeners may utilize many different strategies or attend to different cues in flexible ways and still be successful at decoding intentions.

Thompson and Balkwill (2010) argue that musical experiences are constrained in important ways in several aspects, including the environment, the structure of the auditory system, and the nature of perception and cognition. Commonalities allow for cross-cultural understanding of affective intention in music, whereas culture-specific understanding is built upon these constraints through processes of enculturation. Their cue-redundancy model proposes that listeners across different cultures are able to appreciate affective qualities of unfamiliar music by attending to these commonalities, whereas listeners who are familiar with a musical style should find it easier to decode emotional meaning in that music because they can draw from both culture-specific as well as common psychophysical cues. Balkwill and Thompson (1999) investigated the perception of emotion in listeners from different cultural backgrounds. They had listeners from Western culture and familiar with the Western tonal system listen to Hindustani music and identify the dominant emotion present in each piece. They found that listeners were sensitive to emotions in music from an unfamiliar tonal system, and that common psychophysical cues provide sufficient information for decoding when culture-specific cues are unavailable. Thompson and colleagues (2022) also discuss how musical events carry a whole contextual framework and the roles of structure, self, and source appraising in making sense of these musical events. As such, a listener's cultural background and experience play an important part in their comprehension of music.

Studies have revealed some differences between listeners from different cultures—Chinese and Western—in the multidimensional space obtained from rating dissimilarities of instrument sounds (e.g., Zhang & Xie, 2017). Although these differences might have been due to the different sets of instruments used (Chinese vs. Western instruments), the different dimensions obtained from the multidimensional scaling could also imply a focus on different aspects of a sound by different groups of listeners (McAdams et al., 1995). Wang, Wei, Heng, and colleagues (2021) also found that both cultural and musical backgrounds influence emotion perception in music. Egermann and colleagues (2015) found that listeners from different cultures (Pygmy and Western) had more similar arousal responses when presented with Western and Pygmy music, and differed more in their valence responses. They suggested that "music-induced arousal responses appeared to be based on rather universal, culturally independent response mechanisms" (p. 8). The current study aims to extend previous research and answer the following questions:

1. Do differences in musical training affect the agreement between the listeners' perceptions of performers' affective intentions, and do they influence the ways in which acoustic cues are used?

2. What sound properties are implicated in different affective intentions and in

the information provided by varying amounts of musical context?

We will use three groups of listeners: two groups trained in either Chinese or Western musical traditions and a group of nonmusicians. We hypothesize that listeners trained in the Chinese music tradition would be more accurate than the other two groups in excerpts interpreted by Chinese instruments, because they would have the largest number of culture-specific cues available to them. Listeners trained in the Western music tradition will be more accurate than the other two listener groups in excerpts interpreted by Western instruments. Both musician listener groups would also be more accurate in the excerpts of music that are not from the musical tradition they are familiar with as compared to nonmusicians, as they will have cues available to them from formal learning and greater experience with music more generally. We also hypothesize that there will be an overlap of acoustic cues used for decoding affective intentions and that the cues used by the two musician groups will have more overlap than those used by the nonmusician group.

To answer the question of agreement between expressed and perceived affective intentions, an ANOVA will be performed on listeners' response accuracy based on their agreement concerning the affective quadrant of the expressed intention. Post hoc analyses will compare differences between groups. Multivariate analyses of variance (MANOVA) will be conducted to determine the effects of listener group and instrument on the dependent variables arousal and valence to study the intensity of listeners' responses. The relationship between acoustic features and valence and arousal ratings will be explored with partial least-squares regression (PLSR).

Method

Listening Experiment

Participants

To investigate these issues, three groups of listeners from Singapore with different musical backgrounds were recruited for listening experiments: Chinese musicians (CHM) (n = 30; M years of musical training = 12.0 years; SD musical training = 2.98; 20 females and 10 males; M age = 26.1 years; SD age = 5.73), Western musicians (WM) (n = 30; M years of musical training = 12.13 years; SD musical training = 7.40; 19 females and 11 males; M age = 22.67 years; SD age = 8.22), and nonmusicians (NM) (n = 30; M years of musical training = 0.2 years; SD musical training = 0.41; 14 females and 16 males; M age = 23.23 years; SD age = 6.15). Participants were recruited from the same geographical location to reduce the effects of socio-cultural and linguistic differences. Musicians had to meet the criterion of having more than five years of formal musical training in either a Chinese or Western music tradition, and nonmusicians had to have less than a year of formal training in any type of music. There was no significant difference between the number of years of musical training between the CHM and WM listeners, $t(58)$ = -0.09, p = .93. None of the WM listeners had any prior training in Chinese music, whereas some CHM listeners had received formal instruction in Western music (4 out of 30 CHM participants). The length of formal training these CHM participants had on a Western instrument varied between one to three years and all CHM listeners self-identified as being more proficient in Chinese music than Western music. All participants had casual exposure to both Chinese and

Western art music; these two types of music are ubiquitous in Singapore. 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 (International Organization for Standardization, 2004; Martin, Champlin, et al., 2000). They were compensated for their participation. This study was certified for ethical compliance by McGill University's Research Ethics Board II.

Stimuli

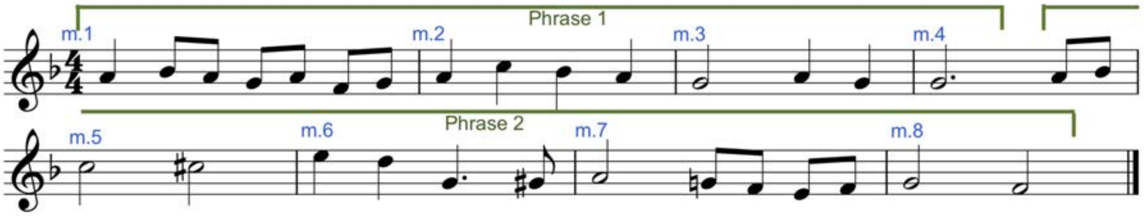
To ensure that the duration of the experiment was reasonable, it was decided that only one excerpt of music would be used. Because of this, it was important that the selected excerpt not bias any particular group of listeners with regards to its style or genre. The excerpt also had to be able to carry affective intentions effectively and have the potential to communicate different affective intentions comparably well. The excerpt selected is a melody taken from an anthology of cue sheets for silent film compiled by George West (1920). There is an ambiguity in the modality of the excerpt—although it starts off predominantly in the major mode, it traipses into the minor mode before returning to major. Selecting an excerpt from film music also reduces bias towards the melody for any listener group as it is assumed that all three groups would have similar exposure to Hollywood movies and silent films.

One professional musician for each of the Chinese instruments (*dizi*, *erhu*, and *pipa*) and Western instruments (flute, violin, and guitar) was recruited for the recording, with instruments from both traditions including an aerophone, a bowed chordophone, and a plucked chordophone, respectively. The two-dimensional model of valence and arousal (Russell, 1980) was explained to the performers, and they were asked to interpret the excerpt of music in performance with five different affective intentions: low-arousal and negative-valence (L–), high-arousal and negative-valence (H–), high-arousal and positive-valence (H+), low-arousal and positive-valence (L+), and neutral (N). All the instruments performed the excerpt at the notated octave and the durations of the excerpts ranged from 11 s to 45 s. The recordings were conducted in a room with soundproof insulation appropriate for sound recording, using a Zoom H4n Handy Recorder (Zoom Corporation, Tokyo, Japan) with the recorder similarly placed 1 m from all the instruments. Input level was adjusted for each instrument. After the recording, each clip was digitally normalized so that their peak amplitudes were –1 dB relative to full scale.

Individual notes (30 in total) from the recorded stimuli were extracted by looking for the note onsets in Audacity® version 2.3.0 (2018). Each note comprises the duration from its onset to the next note's onset. The eight measures and two phrases were extracted in a similar way, each measure or phrase being the onset of the first note of that measure or phrase up to onset of the first note of the following measure or phrase. The entire excerpt can be found in Figure 2.1, as well as its division into measures and phrases.

Figure 2.1

Musical Except Used with Measures being Labelled m.1 to m.8 and Phrases Labelled as Phrase 1 and Phrase 2.



Procedure

All of the listeners took part in two experimental sessions conducted at least a week apart. As the stimuli used for both experiments were obtained from the same recordings, this delay between the first and second experiments was to reduce any memory effects. Experiment 1 involved participants first listening to individual notes extracted from the recorded excerpts and making judgments about each stimulus' perceived affective intention within a two-dimensional affective space of valence and arousal (Russell, 1980). These notes were interpreted with a variety of affective intentions by performers on Western and Chinese instruments within the full musical context. Experiment 2 involved participants listening to measures and phrases of these same recorded excerpts in separate blocks of trials and again making global judgments of the affective intention in the 2D space. Schubert (1999) examined the validity and reliability of a 2D emotion-space interface and found that it had good semantic resolution, was intuitive to use, and exhibited high reliability and validity. Participants made ratings by positioning a cursor within a two-dimensional interface on the computer screen using a mouse (valence on the horizontal axis, arousal on the vertical axis). Valence and arousal values were coded with a resolution of .01 and ranged from –1 to +1. Participants were given verbal instructions, and questions about the two-dimensional scales and perceived affect were posed by the experimenter to ensure they understood what was required. There were four practice trials at the beginning of each experiment to familiarize participants with the interface and for them to clarify any doubts they had about the procedure.

Both Experiments 1 and 2 took about 45 minutes each. Half of the participants were randomly assigned to Experiment 1 first, and the other half were assigned to Experiment 2 first. The order of all trials within each experiment was randomized. The experimental sessions were run with the PsiExp computer environment (Smith, 1995). Sounds were stored on a Macintosh laptop running OSX (Apple Computer, Inc., Cupertino, CA) and were presented over Sennheiser HD 280 Pro headphones (Sennheiser Electronic GmbH, Wedemark, Germany). Participants were seated individually in a quiet room and sound levels were set to a comfortable listening level for all participants. They were instructed not to adjust the level of the sound during the experiment. All statistical analyses were performed using R statistical software (R Core Team, 2022).

Results

The listening experiment was a mixed-measures design with one between-subject and four within-subject measures. There were three listener groups in the between-subject measure: musicians trained in the Chinese music tradition (CHM), musicians trained in the Western classical music tradition (WM), and nonmusicians (NM). The four within-subject measures were:

1. Four affective intentions intended by performers: l low-arousal/negative-valence (L–), low-arousal/positive-valence (L+), high-arousal/negative-valence (H–), and high-arousal/positive-valence (H+);
2. Two instrument cultures: Chinese instruments and Western instruments;
3. Three instrument categories: aerophones, bowed chordophones, and plucked chordophones; and
4. Three context levels: Note, Measure, and Phrase.

We will first consider listener accuracy in affective intention identification and then present group differences in the valence and arousal ratings. In recording the stimuli, performers were only instructed to interpret an excerpt as much as possible in each affective intention, with the expectation that they would target the extremes as much as possible in their performance. It is to be expected, however, that not all notes carry the same amount of a particular affective intention as the expression also depends on the note's position in a phrase, the instrument in question, and individual variations within the performer. In addition, for the listening experiment, it was assumed that when a participant positions their cursor at a position within a particular quadrant of the valence-arousal space, they perceive the stimulus to have that affective intention, even if the position is close to the origin. During the instruction phase, they were also reminded that they should position their cursor exactly on 0 if they felt that the stimulus expressed neither positive nor negative valence and was neither high nor low in arousal. Therefore, even though there might be arguments for exploring gradations in decoding accuracy, for the purposes of this study decoding accuracy was determined by whether a participant placed the cursor in the quadrant intended by the performer. A more detailed exploration of listener responses with respect to how intense or extreme their ratings are follows in the section on group differences. Subsequently, acoustic analyses were conducted to determine differences between listener groups in the use of acoustic cues underlying perceived affective intention.

Listener Accuracy

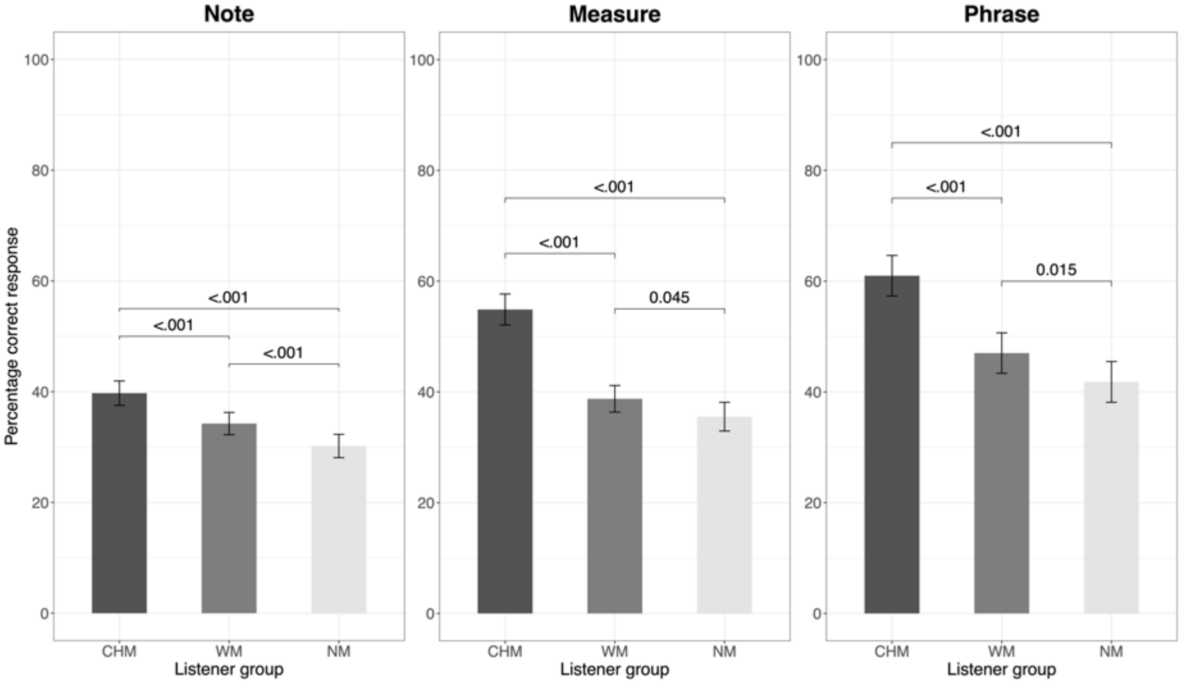
To test our hypotheses concerning differences between listener groups in their ability to perceived the affective intentions intended by the performers, we first measured their response accuracy. Listeners' responses were first individually coded into the emotion-space quadrant (L–, H–, H+, or L+) they fell within in the two-dimensional valence and arousal space. Responses in the same affective intention as the one intended by the performer were coded as correct. The same process was carried out for the responses on single notes, measures, and phrases. The percentage of correct responses for each affective intention was then computed for each participant. Due to violations of homoscedasticity assumptions, we performed the aligned rank transform procedure (Wobbrock et al., 2011) on the data before applying a 5-way mixed model ANOVA on the transformed data. The between-subjects variable is the listener group, and the four within-subject variables are the performer's affective intention, instrument culture, instrument category, and context level. We then carried out post hoc comparisons based on a linear model aligned and ranked on the factors (Elkin et al., 2021). The Bonferroni-Holm method was used to adjust the *p*-values to ensure that Type I errors are not inflated due to the multiple comparisons.

Corrected p-values are reported.

The five-way interaction was significant, $F(24,\ 6177) = 3.78$, $p < .001$, $\eta_p^2 = .01$, although its effect size is very small. There is a significant main effect of context (note, measure, phrase), $F(2,\ 6177) = 7145.76$, $p < .001$, $\eta_p^2 = .70$. The percentage of correct responses increases as musical context increases for the listener. Post hoc comparisons indicate that the accuracy for the phrase context is significantly higher than the measure context, $t(6177) = 60.174$, $p < .001$, and the accuracy for the measure context is significantly higher than the note context $t(6177) = 59.372$, $p < .001$.

In order to further examine the effects of listener group, affective intentions, instrument culture, and instrument category, four-way ANOVAs were performed separately for each context level. Tables A1–A3 in Appendix A show the aligned rank transformed ANOVA results separately for note, measure, and phrase accuracy. There is a significant main effect of listener group in all three contexts, $F(2,\ 87) = 62.24$, $p < .001$, $\eta_p^2 = .59$ for notes; $F(2,\ 87) = 82.64$, $p < .001$, $\eta_p^2 = .66$ for measures; and $F(2,\ 87) = 50.45$, $p < .001$, $\eta_p^2 = .54$ for phrases. Figure 2.2 plots the percentage of correct responses by each listener group for notes, measures, and phrases, with post hoc comparisons between listener groups. CHM performed more accurately than both WM and NM, and WM were significantly more accurate than NM in all three context levels.

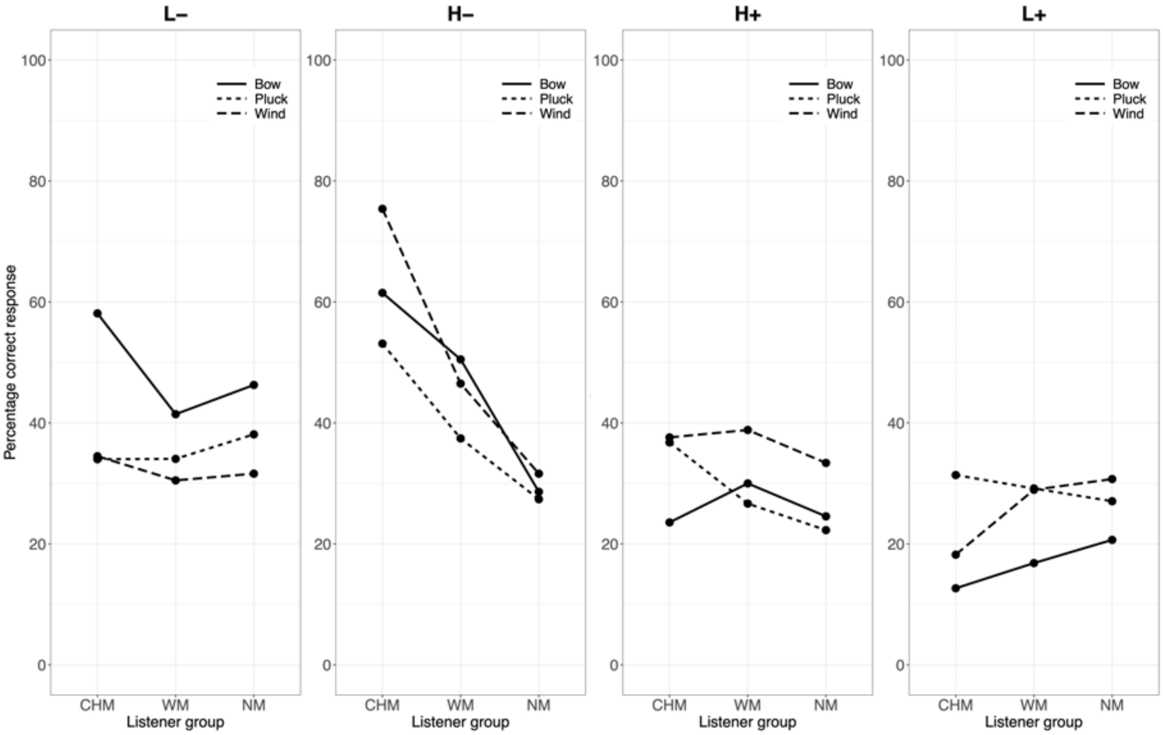
Figure 2.2
Percentage Accurate Responses for Each Listener Group with Standard Error of the Mean



The four-way interaction revealed small effect sizes. It was significant for the measure context, $F(12,\ 2001) = 8.15$, $p < .001$, $\eta_p^2 = 0.047$, and the phrase context, $F(12,\ 2001) = 3.06$, $p < .001$, $\eta_p^2 = 0.018$, but not the note context. Chinese instruments elicited a significantly greater number of accurate responses as compared to Western instruments in the note context, but no differences were observed in the measure or phrase contexts. There was no interaction between instrument culture and listener group in all three context levels. Differences in accuracy also occur more on an instrument level, rather than the culture the instrument belongs to. These data were therefore collapsed across instrument culture in subsequent analyses and discussion. Simple effects analyses reveal interesting interaction effects. To further address this question of differences between listener groups, we examined the three-way-interactions of listener group, affective intention, and instrument category which were significant for all three contexts. The percentage of correct responses for each affective intention over all three contexts is shown in Figures 2.3–2.5. We will present the factors in turn.

Figure 2.3

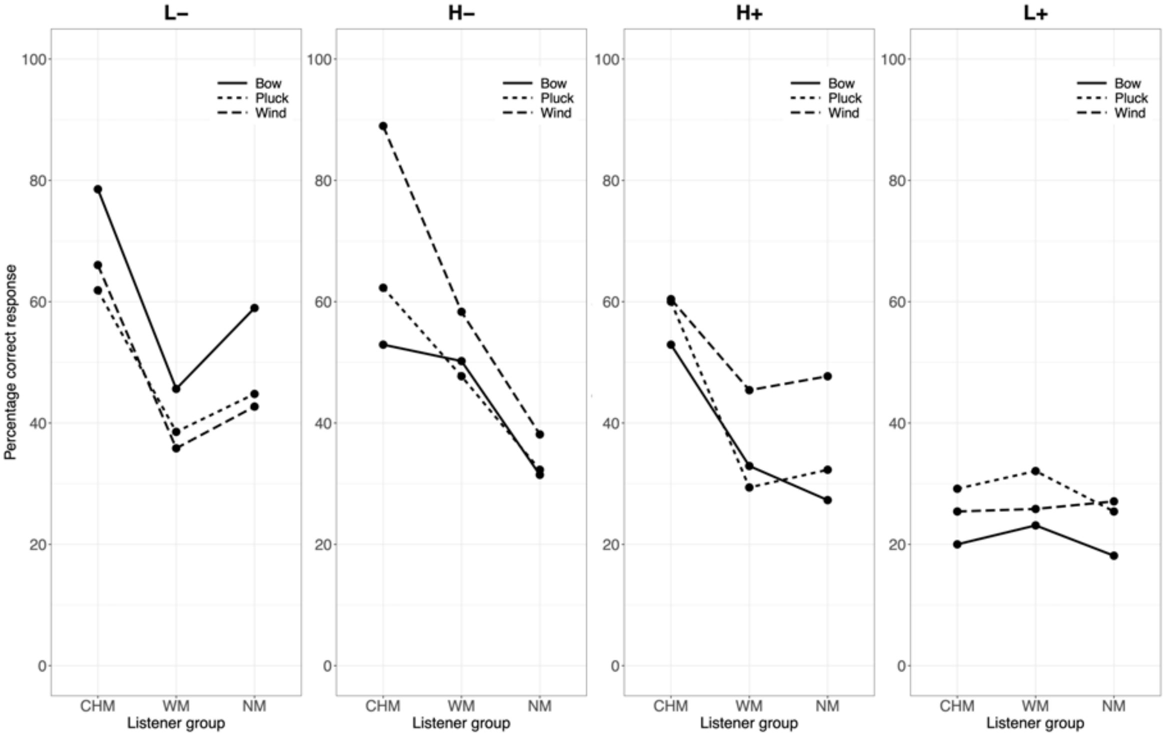
Three-way Interaction for the Single-Note Condition (Listener Group × Affective Intention × Instrument Category)



Note. Affective intentions of the panels from left to right: low-arousal/negative-valence (L–), high-arousal/negative-valence (H–), high-arousal/positive-valence (H+), low-arousal/positive-valence (L+).

Figure 2.4

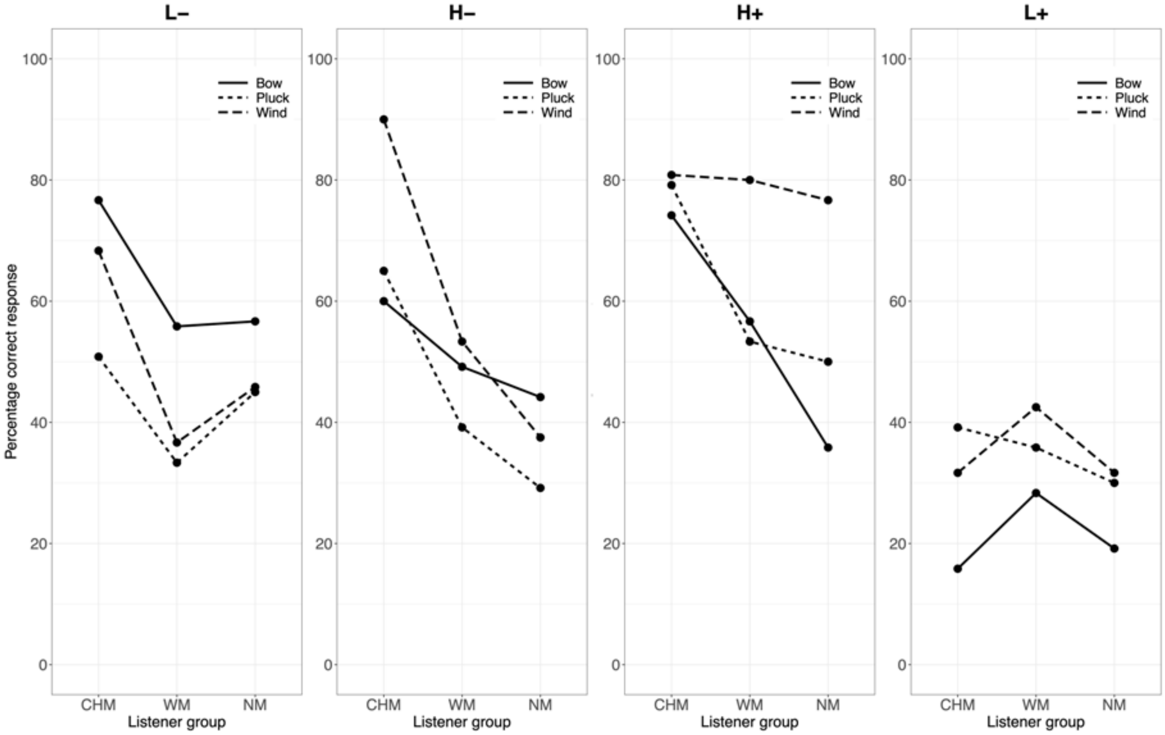
Three-way Interaction for the Measure Condition (Listener Group × Affective Intention × Instrument Category)



Note. Affective intentions of the panels from left to right: low-arousal/negative-valence (L–), high-arousal/negative-valence (H–), high-arousal/positive-valence (H+), low-arousal/positive-valence (L+).

Figure 2.5

Three-way Interaction for the Phrase Condition (Listener Group × Affective Intention × Instrument Category)



Note. Affective intentions of the panels from left to right: low-arousal/negative-valence (L–), high-arousal/negative-valence (H–), high-arousal/positive-valence (H+), low-arousal/positive-valence (L+).

Instrument Culture

Chinese instruments elicited a significantly greater number of accurate responses as compared to Western instruments in the note context, but no differences were observed in the measure or phrase contexts. There was no interaction between instrument culture and listener group in all three context levels. It also appears that differences in accuracy occur more on an instrument level, rather than the culture the instrument belongs to. This data was therefore collapsed across instrument culture in subsequent analyses and discussion.

Instrument Category

Bowed instruments elicit a higher percentage of accurate responses from all listeners as compared to plucked or wind instruments for the L– affective intention in all three contexts (Appendix A, Table A4). The difference in accuracy between plucked and wind instruments is statistically significant in none of the three contexts for L– and L+. There is no significant difference between bowed and plucked instruments for H– and H+ in all three contexts and listeners decode the affective intentions expressed by wind instruments significantly more accurately than bowed instruments and plucked instruments in all contexts except for the note context where bowed and wind instruments were not significantly different. In the H+ affective intention, wind instruments elicit the largest number of accurate responses from all listeners compared to the other instrument categories across all three contexts. Lastly, for the L+ affective intention, stimuli from bowed instruments had the least percentage of correct responses across all three contexts, but there is no significant difference between plucked and wind instruments (Appendix A, Table A4). It appears that performances on bowed instruments are most successful at communicating affective intentions that are negatively valenced and low in arousal, but not very successful at positively valenced low-arousal ones. Performances on wind instruments on the other hand appear to be able to carry all affective intentions effectively except for L–, but especially so for positively valenced high-arousal. These differences could be related to the mode of sound production of the instruments: the actions of bowing, plucking, or blowing having certain affordances and constraints in modifying various acoustic dimensions.

Listener Group

CHM appear to perform better than the other two listener groups for the negatively valenced affective intentions (L– and H–). In the note context, no significant differences between the listener groups are found for the two positively valenced affective intentions (L+ and H+). Differences start to appear with increasing context—CHM perform more accurately than the other two listener groups for H+ in both measure and phrase contexts. No differences are observed between WM and NM for H+. Finally, no significant differences between the three listener groups are found for L+ across all three contexts (Appendix A, Tables A5–A7).

Affective Intention

L+ has the lowest percentage of accurate responses across all participants over all three contexts. H+ elicits significantly less accurate responses in the note and measure context but outperforms the other affective intentions in the phrase context (Appendix A, Table A8). It is possible that a substantial portion of the H+ affective intention is communicated in the note-to-note relationships and less through the quality of sounds of individual notes. On the contrary, H– elicits the highest percentage of accurate responses in the note context. In the measure and phrase contexts, although H– still outperforms H+ and L+, its percentage of accurate responses is not significantly different from that of L–. It is likely that sufficient information about H– can be communicated over individual notes, and it does not require greater contextual information as does H+ to accurately communicate its affective intention.

In the single-note context, only the L+ for wind instruments shows a significantly higher percentage of accurate responses by NM than CHM, $t(87) = –2.75$, $p < .05$, or WM, $t(87) = –2.81$, $p < .05$. None of the other contexts were significantly different in the number of accurate responses between the three listener groups for this affective intention. The low number of accurate responses could be due to L+ sharing acoustic and musical features that made listeners confuse it easily with other affective intentions.

Taken together, these results paint a picture of complex interactions between musical training, the affective intention, and the type of instrument used to express the affect. Bowed instruments appear generally better at communicating low-arousal negatively valenced affective intentions whereas wind instruments communicate high-arousal positively valenced affective intentions most effectively. Although CHM are generally more accurate than both WM and NM, and WM are more accurate than NM, there are some instances where this does not hold true.
Low-arousal/positive-valence appears to be difficult to decipher accurately, and musical training does not seem to give listeners an advantage in accuracy with this affective intention. Related to the affective intention in this quadrant, peace and tenderness have been found to be confused with sadness or melancholy in music perception studies (e.g., Balkwill & Thompson, 1999; Gabrielsson & Lindström, 2010). Although the present study has specifically used dimensional scales for affect,

the same confusion might still occur as acoustic and musical features might be shared between low-arousal stimuli with positive and negative valence.

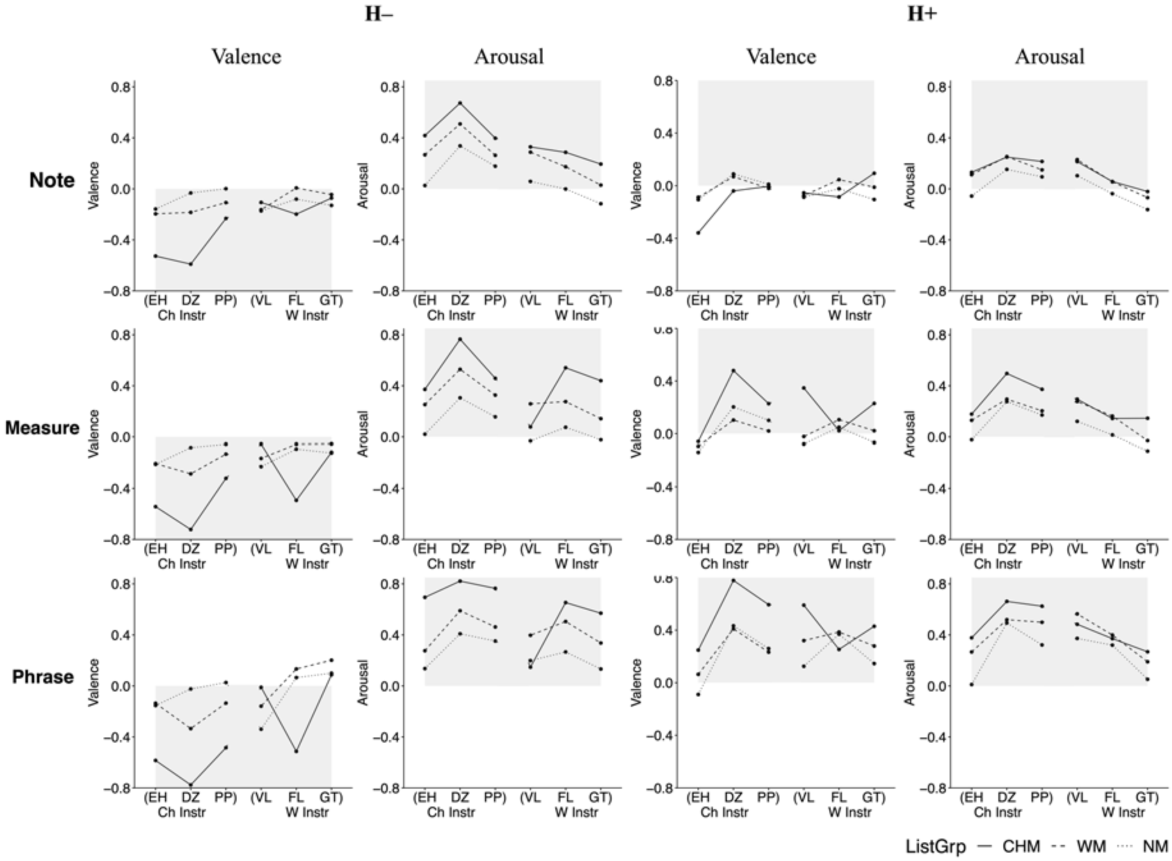
Group Differences in Valence and Arousal Ratings

Two-way mixed multivariate analyses of variance (MANOVA) were conducted separately for each context (note, measure, phrase) and affective intention (L–, L+, H–, H+) to test our hypotheses concerning the effects of listener group and instrument on the dependent variables arousal and valence, taking into account the combined effect of the dependent variables on the independent variables. This allows for a more detailed exploration of listener reactions with respect to the intensity of their responses. No consistent patterns are observed in the interaction effects of instrument culture or category and so instrument is taken as the independent variable here (with six levels) as it is likely that differences are related to the six individual instruments rather than the culture or category to which they belong. Figures 2.6 and 2.7 plot the average arousal and valence for each instrument in each affective intention and for the note, measure, and phrase contexts. Shaded areas indicate the affect intended by the performer. Arousal appears generally accurate with the data points more or less staying on the correct side of the space (shaded areas). Valence, however, is somewhat more ambiguous; positively valenced conditions are sometimes confused for negative valence. With increasing musical context, valence appears to become more differentiated and becomes more accurate for the high-arousal conditions. Low-arousal/positive-valence (L+), however, continues to be confused with low-arousal/negative-valence (L–) even with increased context. Ratings also get more extreme with increased context as listeners likely become more confident of their responses and cues become less ambiguous.

The modified ANOVA-type statistic (MATS) was calculated because assumptions of multivariate normality and covariance homogeneity might not be met in the multivariate data obtained in this study. In this statistic, *p* values were obtained by a parametric bootstrap approach proposed by Konietschke and colleagues (2015) for general MANOVA, and inferences of significance were based on quantiles of resampling distributions (Friedrich & Pauly, 2018). The main effects of listener group and instrument, as well as the interaction effect of listener group with instrument are all found to be statistically significant, indicating that both factors and their interaction are affected by the combined dependent variables. Therefore post hoc ANOVAs were carried out separately for valence and arousal (Appendix A, Tables A9–A11) to examine the results in detail.

Figure 2.6

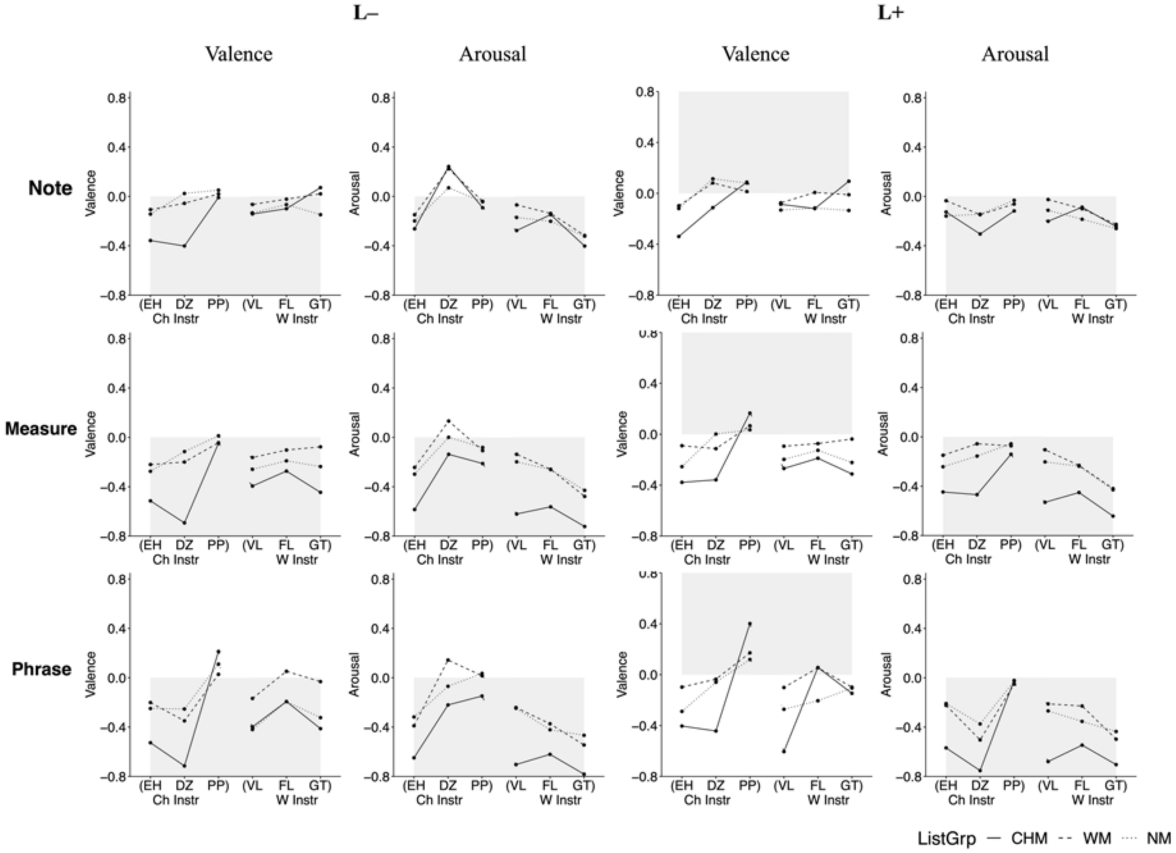
Valence and Arousal Ratings for High-arousal Conditions



Note. Shaded area shows intended affect.

Figure 2.7

Valence and Arousal Ratings for Low-arousal Conditions



Note. Shaded area shows intended affect.

High-Arousal/Negative-Valence

The main and interaction effects are all significant except for the listener group × instrument interaction for arousal in the note context. CHM consistently have the highest arousal ratings followed by WM, and NM have the lowest arousal ratings in the note context. With increasing context, this pattern is still observed, except for CHM ratings for the violin, which decrease with increasing context. CHM might be utilizing certain acoustic features prominent in individual notes as cues for high arousal. Although these acoustic features are present in the same amount in the increased contexts, they may become less important as CHM focus on other musical aspects found in the relationship between notes, which might be more salient in the violin than in other instruments. CHM also consistently gave the most negative valence ratings for Chinese instruments and the flute in all three contexts. It appears that this affective intention is communicated very successfully by these instruments to CHM.

High-Arousal/Positive-Valence

There is no significant interaction effect of listener group with instrument for arousal in any context for this affective intention. Across all instruments, CHM tended to give the highest arousal ratings and NM the lowest. For valence, CHM gave significantly more negative ratings for erhu than did WM and NM in the note context. With increasing context, however, CHM valence ratings became significantly different for all instruments except the flute. It appears that for H+ increasing musical context on the *erhu*, *dizi*, *pipa*, violin, and guitar, but not the flute, gives additional information for CHM.

Low-Arousal/Negative-Valence

For L–, main effects and interactions are all statistically significant except for the listener group × instrument interaction for arousal in the phrase context. CHM tend to give lower or similar valence and arousal ratings compared to the other two listener groups. With increasing context, however, CHM arousal ratings decrease whereas those for WM and NM remain high. Increased musical context appears to provide the pertinent information for this affective intention for CHM more so than it does for the other two groups.

Low-Arousal/Positive-Valence

There are significant main and interaction effects for all three contexts in L+. All the instruments except *pipa* appear to have great difficulty communicating the positive-valence in this affective intention. The average ratings of all three groups of listeners are in the negative range and do not appear to increase even with increasing musical context. All of the instruments are successful in communicating the low-arousal aspect, but only the *pipa* appears to be slightly successful at communicating positive valence in L+.

Partial Least Squares Regression of Acoustic Features with Affect Ratings

Using the revised Timbre Toolbox (Kazazis et al., 2021) implemented in the MATLAB environment, temporal, spectral, and spectrotemporal descriptors of individual notes were analyzed. Based on hierarchical clustering analyses done by Peeters and colleagues (2011), 13 acoustic descriptors that represent the different clusters were selected. These acoustic descriptors included medians (central tendencies) and interquartile ranges (IQR, variability over time) of spectral centroid, spectral flatness, and RMS envelope, as well as the median for noisiness, harmonic spectral deviation, spectral variation, temporal centroid, frequency and amplitude of energy modulation, and log attack time. These calculations were performed on individual notes. For measures and phrases, the average value of the descriptor across all notes included in the measure or phrase was used.

To test our hypothesis concerning the overlap of acoustic features used in affective intention perception, a partial least-squares regression (PLSR) was performed to examine the relationship between acoustic features and different listener groups' ratings of valence and arousal. Each listener group was examined separately in order to see if there are differences in the acoustic features that might predict their ratings of perceived affective intentions. A 10-fold cross-validation model was applied to the PLSR model with training on 9 subsets and testing on the remaining one to estimate prediction error.

The number of latent variables was selected by taking the point where the drop in root mean square error of prediction score after cross-validation was no longer significant. A Varimax rotation was performed to simplify the interpretation of the factors. This yields different numbers of latent variables for the different listener groups. The percentage of variance explained by each component is shown on the loading plots. Figures 2.8–2.10 visualize the loadings of the components for valence and arousal in CHM, WM, and NM listeners. Table 2.1 lists the factors contributing to each principal component with their loadings. Only factors with loadings of 0.2 or greater are listed.

Figure 2.8

Valence and Arousal Loadings for Chinese Musicians)

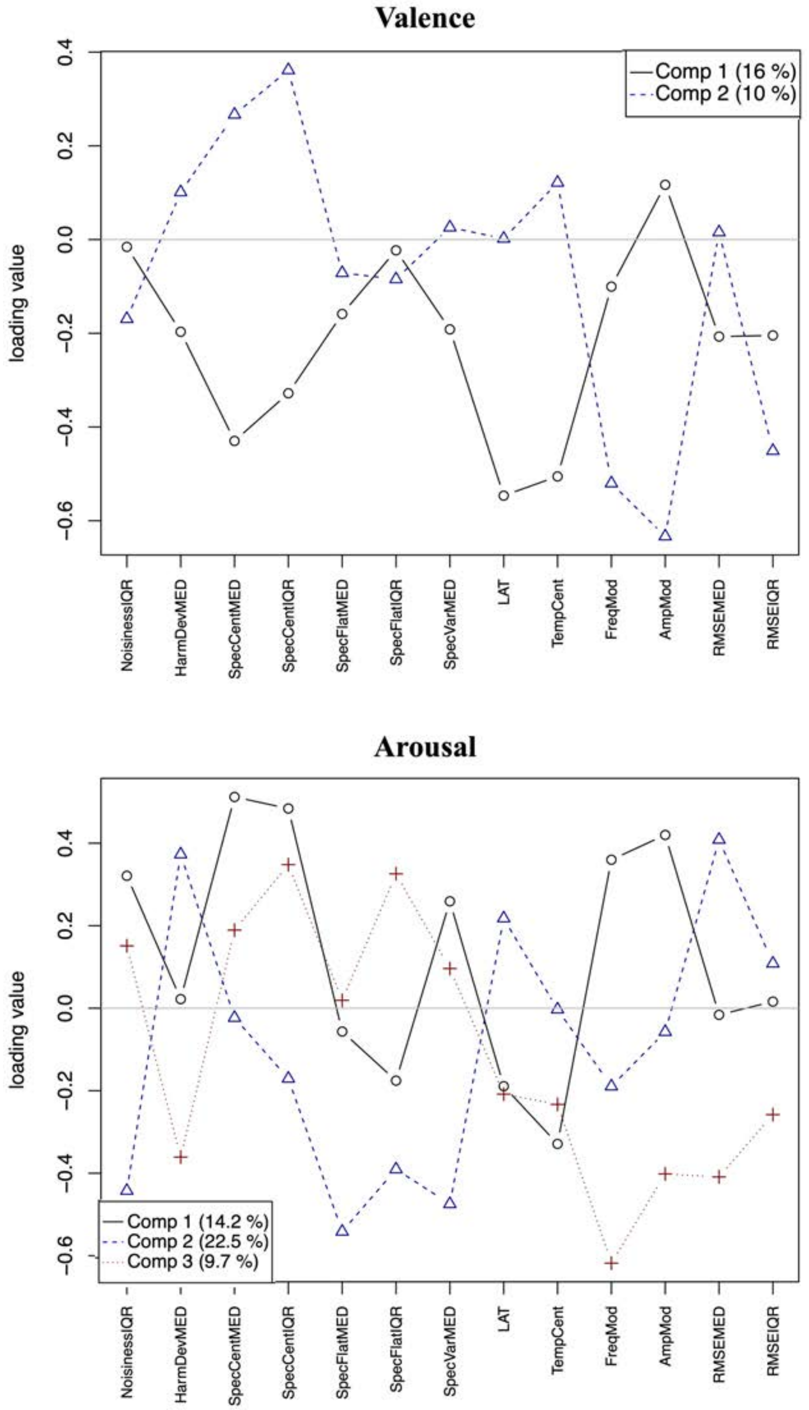


Figure 2.9

Valence and Arousal Loadings for Western Musicians

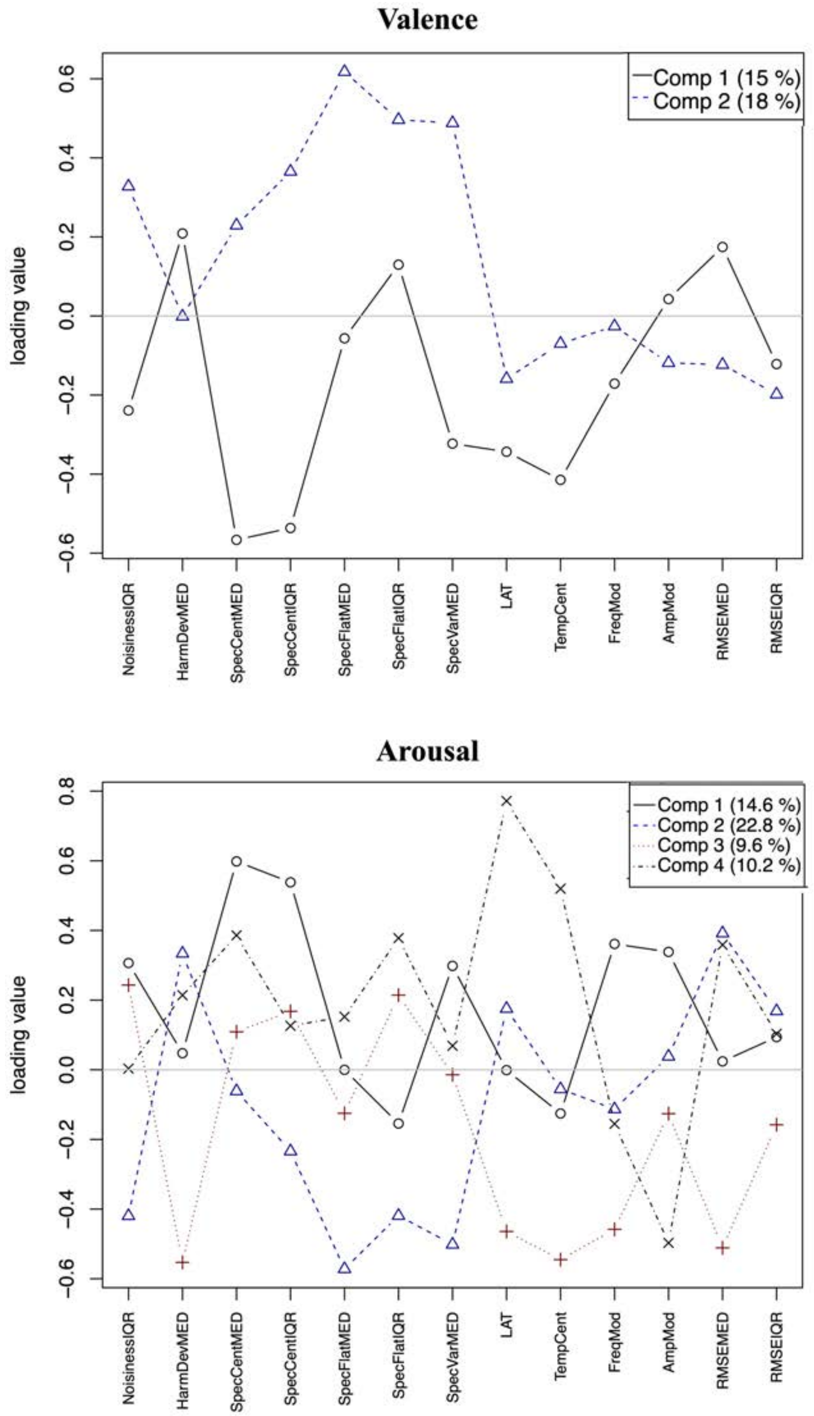


Figure 2.10

Valence and Arousal Loadings for Nonmusicians

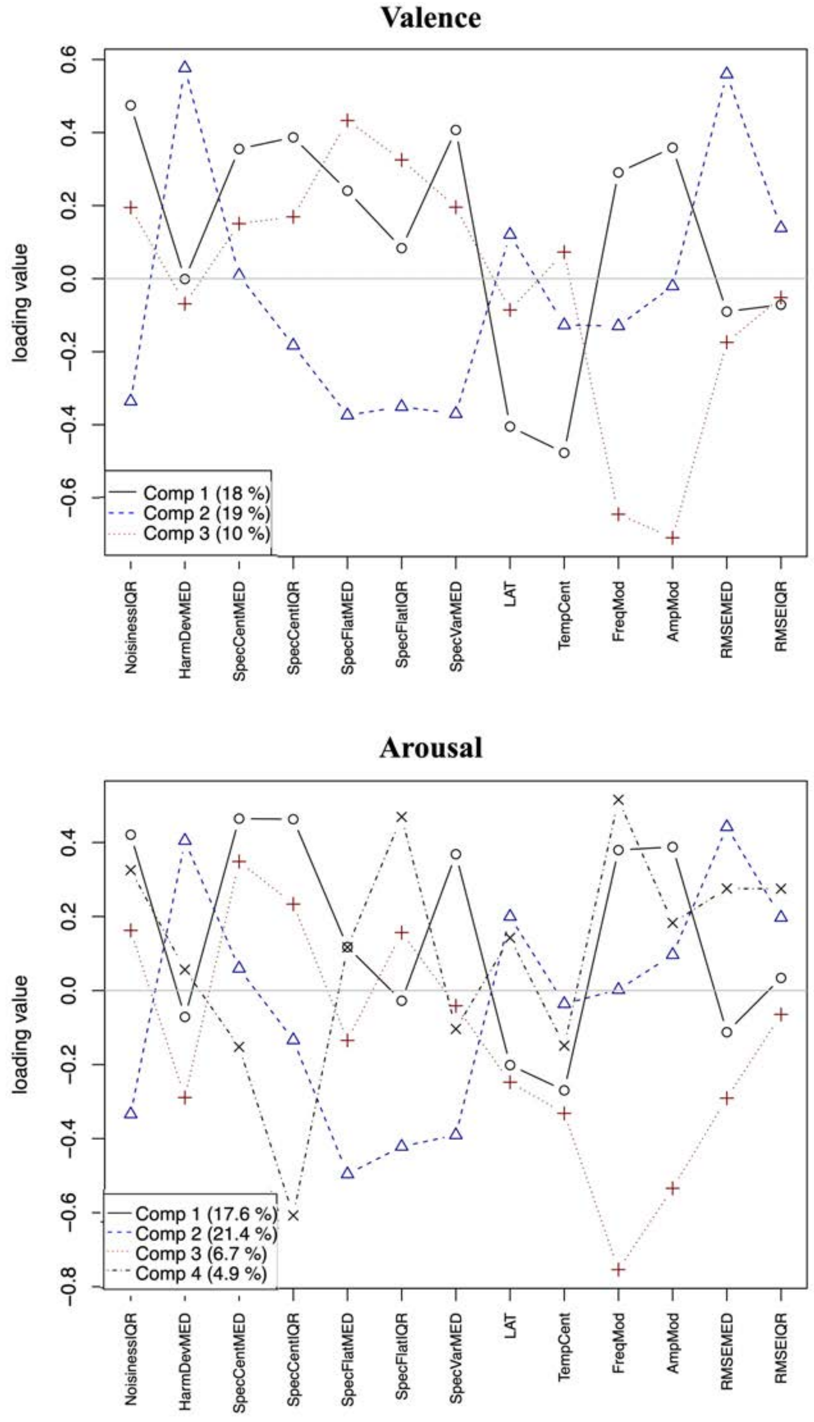


Table 2.1

Loadings for Valence and Arousal on Each Principal Component for CHM, WM, and NM

		Valence		**Arousal**	
CHM	PC1	Spectral centroid (med)	–0.430	Noisiness (IQR)	0.321
		Spectral centroid (IQR)	–0.328	Spectral centroid (med)	0.512
		Log attack time	–0.547	Spectral centroid (IQR)	0.484
		Temporal centroid	–0.505	Spectral variation (med)	0.259
		RMS energy (med)	–0.207	Temporal centroid	–0.329
		RMS energy (IQR)	–0.205	Modulation frequency	0.360
				Modulation amplitude	0.420
	PC2	Spectral centroid (med)	0.267	Noisiness (IQR)	–0.442
		Spectral centroid (IQR)	0.362	Harmonic deviation (med)	0.373
		Modulation frequency	–0.520	Spectral flatness (med)	–0.541
		Modulation amplitude	–0.633	Spectral flatness (IQR)	–0.390
		RMS energy (IQR)	–0.451	Spectral variation (med)	–0.475
				Log attack time	0.218
				RMS energy (med)	0.408
	PC3			Harmonic deviation (med)	–0.361
				Spectral centroid (IQR)	0.348
				Spectral flatness (IQR)	0.326
				Log attack time	–0.208
				Temporal centroid	–0.233
				Modulation frequency	–0.618
				Modulation amplitude	–0.402
				RMS energy (med)	–0.409
				RMS energy (IQR)	–0.258

Continued on next page

Table 2.1 – continued from previous page

		Valence		Arousal	
WM	PC1	Noisiness (IQR)	–0.239	Noisiness (IQR)	0.306
		Harmonic deviation (med)	0.209	Spectral centroid (med)	0.598
		Spectral centroid (med)	–0.566	Spectral centroid (IQR)	0.538
		Spectral centroid (IQR)	–0.536	Spectral variation (med)	0.298
		Spectral variation (med)	–0.323	Modulation frequency	0.361
		Log attack time	–0.343	Modulation amplitude	0.338
		Temporal centroid	–0.415		
	PC2	Noisiness (IQR)	0.328	Noisiness (IQR)	–0.420
		Spectral centroid (med)	0.229	Harmonic deviation (med)	0.334
		Spectral centroid (IQR)	0.365	Spectral centroid (IQR)	–0.234
		Spectral flatness (med)	0.618	Spectral flatness (med)	–0.572
		Spectral flatness (IQR)	0.496	Spectral flatness (IQR)	–0.419
		Spectral variation (med)	0.488	Spectral variation (med)	–0.502
				RMS energy (med)	0.392
	PC3			Noisiness (IQR)	0.243
				Harmonic deviation (med)	–0.553
				Spectral flatness (IQR)	0.214
				Log attack time	–0.464
				Temporal centroid	–0.545
				Modulation frequency	–0.458
				RMS energy (med)	–0.511
	PC4			Harmonic deviation (med)	0.214
				Spectral centroid (med)	0.386
				Spectral flatness (IQR)	0.378
				Log attack time	0.772
				Temporal centroid	0.520
				Modulation amplitude	–0.497
				RMS energy (med)	0.359

Continued on next page

Table 2.1 – continued from previous page

		Valence		**Arousal**	
NM	PC1	Noisiness (IQR)	0.475	Noisiness (IQR)	0.421
		Spectral centroid (med)	0.355	Spectral centroid (med)	0.464
		Spectral centroid (IQR)	0.387	Spectral centroid (IQR)	0.463
		Spectral flatness (med)	0.241	Spectral variation (med)	0.368
		Spectral variation (med)	0.407	Log attack time	–0.201
		Log attack time	–0.405	Temporal centroid	–0.270
		Temporal centroid	–0.477	Modulation frequency	0.380
		Modulation frequency	0.291	Modulation amplitude	0.388
		Modulation amplitude	0.359		
	PC2	Noisiness (IQR)	–0.336	Noisiness (IQR)	–0.334
		Harmonic deviation (med)	0.577	Harmonic deviation (med)	0.405
		Spectral flatness (med)	–0.374	Spectral flatness (med)	–0.496
		Spectral flatness (IQR)	–0.351	Spectral flatness (IQR)	–0.421
		Spectral variation (med)	–0.370	Spectral variation (med)	–0.390
		RMS energy (med)	0.560	Log attack time	0.200
				RMS energy (med)	0.442
	PC3	Spectral flatness (med)	0.433	Harmonic deviation (med)	–0.289
		Spectral flatness (IQR)	0.325	Spectral centroid (med)	0.349
		Modulation frequency	–0.645	Spectral centroid (IQR)	0.234
		Modulation amplitude	–0.709	Log attack time	–0.248
				Temporal centroid	–0.332
				Modulation frequency	–0.754
				Modulation amplitude	–0.534
				RMS energy (med)	–0.291
	PC4			Noisiness (IQR)	0.325
				Spectral centroid (IQR)	–0.608
				Spectral flatness (IQR)	0.469
				Modulation frequency	0.515
				RMS energy (med)	0.276
				RMS energy (IQR)	0.275

Valence

Chinese Musician Listeners. Two components appear to explain the loadings for valence (26.0% total variance explained) in CHM (Figure 2.8, top panel). Spectral centroid median and IQR, log attack time, temporal centroid, RMS energy median and IQR load negatively on the first component explaining 16% of the variance for valence. Modulation amplitude, modulation frequency, and RMS energy IQR load negatively and spectral centroid median and IQR load positively onto the second component, which explains 10% of the variance for valence.

The acoustic descriptors characterize different aspects of a sound. The mapping from the acoustic space to a semantic description of the sounds' qualities could be somewhat ambiguous, but some authors have found terms that appear to be relatively consistent (Wallmark & Kendall, 2018; Zacharakis et al., 2012). We will attempt to describe the quality of the sounds as reflected by how the combination of acoustic descriptors maps onto adjectives commonly found in the timbre semantics literature. It appears that CHM listeners perceive darker, softer, and less sustained sounds as more positively valenced. However, if the sounds are bright but combined with less vibrato and little change in dynamics, CHM also perceive them as communicating a positive valence.

Western Musician Listeners. For WM, two components also appear to explain the loadings for valence (33.0% total) (Figure 2.9, top panel). The first component explains 15% of the variance: noisiness IQR, spectral centroid median and IQR, spectral variation median, log attack time, and temporal centroid load negatively onto it. Noisiness IQR, spectral centroid median and IQR, spectral flatness median and IQR, and spectral variation median load positively onto the second component, which explains 18% of the variance for valence.

Similar to CHM, a darker and less sustained sound with less variation spectrally communicates positive-valence. When the sound is bright and combined with large spectral variation, and is high in noise content, WM perceive it as positively valenced. It appears that the acoustic features loading onto the first principal component for valence in WM are very similar to those of CHM. However, in the second principal component, a different combination of features inform WM of a positive-valence. This accounts for the shared similarities in some of the valence responses with CHM but divergences in others.

Nonmusician Listeners. Three components explain the loadings for valence (47.0% total) in the NM listeners (Figure 2.10, top panel). The first component explains 18% of the variance, with noisiness, spectral centroid median and IQR, spectral variation median, modulation frequency and amplitude loading positively and log attack time and temporal centroid loading negatively onto it. Harmonic deviation median and RMS energy median load positively and noisiness, spectral flatness median and IQR, and spectral variation median load negatively onto the second component which explains 19% of the variance. The third component explains 10% of the variance, and spectral flatness median and IQR load positively and modulation frequency and amplitude load negatively onto it.

NM listeners appear to share the least number of acoustic features with the other two groups in valence perception. Although they perceive noisiness, brightness, and large spectral variation as communicating a positive-valence, they also combine this set of features with a less sustained sound and large amount of vibrato. This particular combination differs from CHM who use the combination of a large amount of vibrato with brightness to imply a negative valence. It also differs from WM whose combination of less sustained sounds with lower brightness and spectral variation implies a positive valence. Altogether, these results point towards more shared commonalities in how the two musician groups utilize combinations of acoustic features in understanding expressed affective intentions, which is reflected both in the accuracy and pattern of responses of the three listener groups.

Arousal

Chinese Musician Listeners. Three components appear to explain the loadings for arousal (46.4% total) in CHM (Figure 2.8, bottom panel). Noisiness IQR, spectral centroid median and IQR, spectral variation median, modulation amplitude, and modulation frequency load positively, and log attack time and temporal centroid load negatively onto the first component, which explains 14.2% of the variance. Noisiness IQR, spectral flatness median and IQR, and spectral variation median load negatively, and harmonic deviation median, log attack time and RMS energy median load positively onto the second component which explains 22.5% of the variance. Finally, harmonic deviation median, modulation frequency, modulation amplitude, RMS energy median and IQR, log attack time, and temporal centroid load negatively, and spectral centroid IQR and spectral flatness IQR load positively onto the third component which explains 9.7% of the variance for arousal.

The various components reveal different aspects of the combinations of features. A brighter sound with more spectral variation, noisiness, and vibrato points towards higher arousal for CHM. A brighter sound with more vibrato that is also less sustained might be perceived as higher in arousal. A louder sound that is less noisy also implies higher arousal.

Western Musician Listeners. Four components appear to explain the variance for arousal (57.2% total) in WM (Figure 2.9, bottom panel). The first component that explains 14.6% of the variance has noisiness IQR, spectral centroid median and IQR, spectral variation median, and modulation frequency and amplitude loaded positively onto it. Noisiness, spectral centroid IQR, spectral flatness median and IQR, and spectral variation median load negatively, and RMS energy median and harmonic deviation median load positively onto the second component, which explains 22.8% of the variance. The third component explains 9.6% of the variance and the fourth explains 10.2% of the variance. Harmonic deviation median, log attack time, temporal centroid, modulation frequency, and RMS energy median load negatively onto the third component. Spectral centroid median, spectral flatness IQR, log attack time, temporal centroid, and RMS energy median load positively, and modulation amplitude loads negatively onto the fourth component.

Brighter and noisier sounds with more vibrato and spectral variation seem to be perceived as higher in arousal, whereas sounds that are loud but have less noise content are also perceived as high in arousal. Shorter, less sustained sounds with little vibrato communicate high arousal as well, but the combination of loudness, brightness, and more sustained sounds also communicates high arousal.

Nonmusician Listeners. Four components explain the loadings for arousal (60.6% total) in the NM listeners (Figure 2.10, bottom panel). The first component explains 17.6% of the variance and has noisiness, spectral centroid median and IQR, spectral variation median, and modulation frequency and amplitude loading positively, and log attack time and temporal centroid loading negatively onto it. Harmonic deviation median and RMS energy median load positively, and noisiness IQR, spectral flatness median and IQR, and spectral variation median load negatively onto the second component, which explains 21.4% of the variance. Harmonic deviation median, log attack time, temporal centroid, modulation frequency and amplitude, and RMS energy median load negatively and spectral centroid median and IQR load positively onto the third component which explains 6.7% of the variance. Finally, spectral centroid IQR loads negatively and noisiness, spectral flatness IQR and modulation frequency, and RMS energy median and IQR load positively on the fourth component which explains 4.9% of the variance.

Similarly to the other groups of listeners, a brighter sound with high noise content, substantial vibrato, and large spectral variation generally indicates a higher arousal. A combination of loudness, less noise content, and less spectral variation also communicates high arousal similarly for all three groups of listeners. Brighter, shorter, less sustained sounds with little vibrato also imply high arousal for all three groups.

Summary

Results from the PLSR demonstrate clearly that acoustic features do not contribute singly to the perception of valence or arousal. Many acoustic features load positively onto one principal component but negatively onto another. This is likely due to interactions between the different acoustic features, with the affective intentions being elicited by combinations of acoustic features rather than single ones alone. It also appears that the acoustic features contributing to valence are much more influenced by musical training as compared to those that contribute to arousal. CHM and WM share a greater number of similar acoustic features they use to decode valence, whereas NM differ much more on the acoustic features that contribute to their perception of valence. Arousal on the other hand appears to be elicited with more similar combinations of acoustic features for all three groups.

A larger percentage of the variance can be explained by acoustic features for arousal than for valence, with 26.0%, 33.0%, and 47.0% for valence and 46.4%, 57.2%, and 60.6% for arousal in CHM, WM, and NM, respectively. This result suggests that in addition to surface acoustic features, other factors could be at play in communicating perceived valence to listeners, more so than for perceived arousal. Similarly, a larger percentage of the variance is explained by acoustic features for NM, followed by WM, and the least for CHM. It appears there are aspects in the sound that might not be captured by these acoustic descriptors but that contribute to informing listeners about the perceived affective intention of a musical sound, and the ways in which these aspects of a sound are utilized are influenced by musical training.

Discussion

This study examined the following questions:

1. whether differences in musical training affect the accuracy of listeners'

perception of affective intentions, and whether they influence the ways in which acoustic cues are used; and

2. what sound properties are implicated in different affective intentions and in the amount of information provided by varying amounts of musical context.

Differences in Listeners' Responses

Plots of the average ratings of perceived affective intentions for each group of listeners show a general trend of spreading out to the extremes as context increases from single notes to measures to phrases. Arousal also tends to be more accurate than valence although the pattern is different for each instrument and each group of listeners. These results are consistent with Egermann and colleagues' (2015) finding, that arousal responses tend to be based more on culturally invariant mechanisms.

Participants' responses show a significant increase in accuracy as the amount of musical information increases, with a large main effect size for context, and their responses also became more differentiated. The three listener groups were also significantly different, and this main effect showed a large effect size. The reason for this difference could be because of a different emphasis on the use of timbre in the musical training of the Chinese as compared to the Western musical tradition, which allows CHM to be sensitive to minute changes in the way timbre is being manipulated in the expression of affective intentions. Thompson and Balkwill's (2010) cue-redundancy model states that "listeners who are familiar with a musical style should find it relatively easy to decode emotional meaning in that music, because they can draw from both culture-specific and psychophysical cues" (p. 766). This would mean that WM would have greater decoding accuracy for stimuli interpreted by Western instruments and CHM greater decoding accuracy for stimuli interpreted by Chinese instruments. However, this is not observed and there does not seem to be a cultural advantage for WM—CHM consistently perform more accurately than WM and NM regardless of the instrument tradition. Both Chinese orchestral music and Western classical music are ubiquitous musical traditions in Singapore and CHM, WM, and NM will have similar amount of casual exposure to these musics in their everyday life. The only difference then would appear to be formal training and possibly the conscious choice of listeners to listen to one particular type of music over another. The majority of CHM also did not have any formal training in Western music—this difference is unlikely due to CHM having a broader range of casual musical experiences as compared to WM. In addition to that, it seems here that CHM have access to more cues informing affective intention than WM or NM because of their greater sensitivity to timbral cues. In the single-note context, musical relations are not as available for the listener, and the response is based almost entirely on the perceived timbre of the individual note. CHM might have been more sensitive to the nuances of timbre variation from the emphasis on it in their musical training whereas knowledge of musical relations habituated by musical training provides additional cues for both the musician groups with increasing musical context. Although Chinese instruments are marginally better than Western instruments in eliciting accurate responses about their expressed affective intention in the note condition, there are no differences between the instrument traditions in the other two contexts. In a similar way, performers on Chinese instruments in this study might place greater emphasis in the timbral features within each note in expressing affective intentions.

Accuracy varies across the different affective intentions. The listener group × affective intention interaction shows a medium effect size. CHM generally perform more accurately and have more extreme responses than the other two listener groups in high-arousal/negative-valence intentions. Susino and Schubert (2017) proposed a stereotype theory of emotion to attempt to explain how perceptions of emotions might be filtered through a listener's stereotype of the encoding culture. East Asians are stereotyped to be significantly less emotionally expressive than European Americans (Adam & Shirako, 2013). This could partially explain why WM perceived less of the stimuli performed by Chinese instruments as high-arousal/negative-valence if they had a belief that Chinese music might be more "anger-reticent". In addition, it is also likely there is a highly nuanced set of timbre characteristics for this affective intention that is used in the Chinese music tradition, which CHM understand and are able to use effectively in their listening responses. The low-arousal/positive-valence intention appears to be the least accurate in all the three listener groups, and CHM have the least accurate responses, a response pattern that is in contrast with all the other affective intentions in which CHM usually are the most accurate. It could be that low-arousal/positive-valence intentions share timbre and musical characteristics with other affective intentions, making it difficult to distinguish them from the others, not only in the single-note context but also with added musical context. This is consistent with other studies (e.g., Balkwill & Thompson, 1999; Gabrielsson & Lindström, 2010) that found tenderness and peace to be correlated with sadness.

Certain instrumental timbres might be better at expressing one affect compared to others, possibly due not only to the characteristics of the instrument, but also its connotations and use in particular musical tropes. For example, Behrens and

Green (1993) found that for Western listeners, the violin was able to express sadness more accurately than other instruments. This appears to be the case in this study as well with the violin and erhu eliciting the highest percentage of correct responses in the low-arousal/negative-valence intention. Hailstone and colleagues also noted that “the perception of emotion conveyed by a melody is affected by the identity or timbre of the musical instrument on which it is played” (2009, p. 2152). Composers appear to understand this implicitly, selecting particular instruments over others more frequently to express certain emotions (e.g., Huron et al., 2014; Schutz et al., 2008). In addition to affordances and constraints of instruments for the expression of affective intentions, performer variability could play a role as well. It is not possible in the context of this study to explore variability between performers, or to control for individual variation in performances on different instruments. In order to reduce influence from this as much as is possible, all the excerpts were interpreted by musicians who perform regularly on a professional basis. This may, however, also allow for idiosyncratic interpretations of the affective intentions on a particular instrument.

Effects of Timbre on Perception of Valence and Arousal

Results of the PLSR analysis of acoustic descriptors with valence and arousal ratings help to demonstrate the contributions of these features to listeners’ perceptions of valence and arousal. Rather than features contributing to an affective intention independently, it appears that the perception of affective intentions is influenced by a combination of various acoustic features. Features also do not necessarily map onto affect perception in the same direction. A particular acoustic feature might map in different directions for particular affective intentions depending on other features it interacts with.

Consistent with Egermann and colleagues' (2015) findings, it also appears that listeners' use of these features for valence perception is very much influenced by musical training and enculturation. In contrast, the three groups of listeners share more of the same combinations of acoustic features that they use to decode arousal levels. An understanding of what acoustic features contribute to the communication of arousal may be more culturally invariant, whereas those that communicate valence could be more culturally learned.

Additionally, it appears that other factors aside from the surface acoustic features play a greater role in explaining the variance for perceived valence than arousal. Similarly, a larger amount of the variance can be explained by the surface acoustic features for NM and the least for CHM. Even though all participants are presented with the same set of acoustic features, listeners from the different groups use differing amounts and types of cues for decoding. When these features align in production and comprehension with the intended affect, listeners have greater success at recognizing the intended affective intention. This can be observed clearly here in the greater accuracy in judgments and the way combinations of acoustic features used for valence by CHM and WM overlap much more than they do with NM. However, even though NM use very different acoustic features, it does not render them totally unsuccessful at decoding affective intentions. There is a redundancy of cues in musical communication, and even with several misaligned cues, a certain degree of emotion recognition is still possible.

Conclusion

The function of timbre in communicating affective intention in music is highly complex. There are many interactions between the acoustic features that make up the quality of a sound. With increasing musical context, the relationship between these acoustic features, and other musical parameters such as pitch relations, implied harmonies, rhythmic and metrical characteristics becomes involved in the listening process as well. Listeners from different musical traditions show different accuracies in the perception of affective intention, suggesting not only that musical training plays a role in listener responses, but that musical traditions also play an important part. Within the context of this study, CHM are consistently more accurate than both WM and NM, and this may be related to the differences between Chinese and Western musical traditions with regards to the use of timbre in expression. With an emphasis on certain means of timbral expression for particular affective intentions, CHM are likely to be more sensitive to nuanced differences during music listening. Although this gives CHM an advantage in terms of more cues at their disposal for accurate decoding, the abundance and redundancy of cues means that there is still a sufficient number of common psychophysical cues so that NM are not entirely inaccurate in emotion recognition.

The acoustic features that inform arousal responses appear to be more similar across the three listener groups than those that inform valence responses. This strongly suggests that conventions regarding the function of timbre in valence perception are learned through explicit musical training whereas those for arousal might be more culturally invariant. Acoustic features also do not contribute singly to particular intentions, but rather, listeners perceive the entirety of the timbre of a sound through various combinations of acoustic features.

There might be a slight asymmetry in participant expertise with a few Chinese musician listeners having had formal instruction in Western music but no Western musician listeners having any formal instruction in Chinese music. However, the small number (4 out of 30 CHM participants) and the relatively short duration of formal training they had on a Western instrument is unlikely to contribute significantly to the results. Nonetheless, it might be helpful in future studies to examine if there is a difference between Chinese musician listeners without formal training in Western music and those who do. It also has to be acknowledged that participants were recruited based on the duration of formal instruction they had in a particular musical tradition. This might not take into account the differences individuals have with regards to the musical conventions they have absorbed implicitly through exposure. Validated measures examining participants' musical knowledge might be adapted for future studies.

There was only one excerpt used in the experiment, and only one performer for each instrument. Different excerpts of music could be utilized, and more performers recruited for each instrument in future studies. This would provide a better control with regard to particular affective tendencies a melody might intrinsically carry. A few different performers on each instrument would also provide better control for individual variability in performance.

No continuous response was elicited from listeners with respect to changes in affective intentions over the course of the excerpt in this study. Although the comparisons across responses for notes, measures, and phrases indicate that musical context provides increasing cues for understanding, future studies could study continuous responses to better understand the function of timbre over the course of the music. Listeners with formal training in other musical cultures can also be recruited to explore how affective intentions and timbre might be perceived by listeners from different musical traditions.

This study demonstrates the role that cultural and formal musical learning play in timbre perception. The use of Chinese instruments also provides a different perspective in studying music perception as most studies have utilized proto-musical materials or music from the Western classical repertoire and with Western participants and instruments. This study also recruited participants from the same geographical location but who have different formal musical training, thus reducing differences that might be due to socio-cultural or linguistic factors.

Acknowledgments

We thank CIRMMT for funding an Inter-Centre Research Exchange for LH, Dr. Dorien Herremans (Singapore University of Technology and Design) for hosting LH for the Inter-Centre Research Exchange and for providing support in data collection. This research was supported by grants to SMc from the Canadian Social Sciences and Humanities Research Council (895-2018-1023) and the Fonds de recherche du Québec—Société et culture (017-SE-205667), as well as a Canada Research Chair (950-223484). We also thank Bennett K. Smith for programming the experimental interface and Marcel Montrey for advice in statistical analyses.

Part III

Continuous response to perceived affective intentions in music listening: Differences in musical training

This chapter is based on the following research article:

> Heng, L., Wei, C., and McAdams, S. (2023). Continuous response in music listening: Training in different musical traditions influence perception of affective intentions. [Manuscript in preparation].

Abstract

Perception of affective intentions in music is a complex, yet commonplace, phenomenon that takes place in everyday life. It could be influenced by listeners' experience with different musical traditions and the style of the music. This study aims to explore the continuous responses of listeners as they listen to a piece of Chinese orchestral music and how their musical backgrounds influence affective responses, as well as the acoustic and musical features that each group of listeners utilizes in the perception of different affective intentions. Three groups of listeners (trained in Chinese and Western art music traditions and nonmusicians; $n = 30$ per group) were presented with a piece of music (930 s long) and responded continuously on a unidimensional emotional intensity scale and a two-dimensional valence and arousal interface in separate blocks. Functional data analysis is used to compare differences between listener groups as these continuous responses reflect a smooth variation. Time series analysis is also used to explore how each listener group utilizes the various acoustic and musical features over the course of the music. Results show significant differences between listener groups over different sections of the music. Valence responses diverge more than arousal or emotional intensity responses. Although emotional intensity and arousal responses appear to be very similar over the course of the music, it appears that listeners utilize different sets of acoustic and musical features in their ratings. No single acoustic or musical feature communicates a clear and distinct type of information, but rather, features add on to or interact with other elements to provide listeners with a more nuanced picture in their understanding. The perception of affective intentions in music listening is a complex process which is influenced by the degree of familiarity listeners have with a musical tradition, the narrative content implicated in the music, and the complex sonic environment created by the composer's creation and the musicians in their interpretation in performance.

Keywords: timbre features, musical features, musical affect, continuous response, cross-cultural differences

Introduction

Listening to music is a common activity people engage in. While music could be enjoyable purely based on its sensuous properties, Davies argues that only when we hear it as organized sounds will we be hearing it as music.(Davies, 2019). Musical meaning emerges from the interpretation of organized musical information. Juslin and Timmers (2010) primarily discussed affective intentions in music, but their definition of communication in music is applicable to communication in the broader sense. They described communication as involving “both a performer’s intention to express a specific emotion and recognition of this emotion by a listener" (p. 455). Extending this to Davies’ idea then, musical meaning emerges when the listener interprets organized musical information in a performance. Consensus of musical meanings among different listeners however, is not a given fact. While there can be general agreement on aspects of a piece of music, interpretations can also differ widely.

Scholarly opinions on how the process of signification takes place in music understanding differ. Shepherd and Wicke (1997) argue that the “concept of the ‘signifier’ is ... inappropriate in approaching questions of signification in music” (p. 117) because there exists “a technology or instrumentality of articulation” (p. 119) in the process. Rather than a particular sound or combination of sounds pointing towards a particular meaning the way language does, the qualities within the sound shaped by its articulation play a big role in the range of meanings they carry. Thoresen (2015) theorizes about different types of signs present in musical sound and how these play different roles in the process of signification. Although he acknowledges that sounds function differently in words than in music, he does not explicitly differentiate them, focusing on how articulation is instrumental in contributing to musical meaning the way Shepherd and Wicke do. He does, however, emphasize the sign-interpreting aspect in music semiosis, implying in a way the importance not of a single, direct sound-to-meaning relationship, but of interpretations arising from combinations of the way sounds and aspects of the qualities of sounds come together. Patel (2010) also believes that we cannot simply understand music as we do another form of language. Even so, the narrative nature of music frequenctly appears when listeners are asked to recount what they imagine when they listen to music. Margulis demonstrates the ease with which narratives are generated in music listening (Margulis, 2017), as well as several commonalities that emerge among them, especially among listeners from the same culture (Margulis et al., 2019). Whether or not intended by the music, narrative content appears to be a form of musical meaning that is easily elicited in music listening.

Affective Intentions in Music

Affective intention in music, as an umbrella term covering all aspects of evaluative states with regards to musical communication, is an important and salient notion that may be considered another important aspect of musical meaning. There are different models used to describe affect and emotions. Commonly cited ones include Ekman (1992) who argues for the existence of “universal basic emotions”, Shaver and colleagues’ (1987) who proposed the concept of emotion prototypes, Russell (1980) who advanced a two-dimensional circumplex model of affect (arousal and valence), and Schimmack and Grob’s (2000) three-dimensional model of pleasure-arousal-tension of core affect. “Listeners’ perception of emotional expression—to perceive an expression of, say, sadness in the music without necessarily being affected oneself—is mainly a perceptual-cognitive process" (Gabrielsson, 2001, p. 124). This is to be differentiated from *induced* affect in music listening in which listeners have a change in psychological state due to the effects of music listening.

Experimental studies on music, affect, and emotion have drawn on these different models. Eerola and Vuoskoski (2011) for instance, compared the discrete and dimensional models and found Russell's (1980) circumplex model to be sufficient in explaining listeners' judgments of perceived affect in music. Schubert (2004) also explored the relationship between musical features and perceived emotion using the two-dimensional arousal-valence emotion space, using quantitatively coded musical features to predict perceived emotion in music.

In addition to pleasure and arousal, Reisenzein (1994) also suggests that the different states of emotion vary in terms of both quality and quantity. In addition to the category of emotion, there is also how intense it is. He attempted to disentangle this concept of intensity and how it relates to the valence and arousal model in participants' conceptions of emotion terms. He found that there are a substantial number of subjective states rated by people that appear to conform to the fact that emotion intensity is determined by *both* valence and arousal. On the other hand, there is also a strong proponent amongst psychological theories of emotions that "intense emotions are *quantitatively* [emphasis added] different from normal emotions in that their intensity is determined by the person's level of arousal" (Rickard, 2004, p. 372).

To the contrary, Warrenburg (2020) suggests a continuum of emotion models instead where one end consists of basic emotion models and the other social construction models. Lying along this continuum are various models claiming the importance of appraisal processes and to psychological construction of meanings. The process of appraisal, for instance, is a key concept in the CODA (Constructivistly Organized Dimensional-Appraisal) model proposed by Lennie and Eerola (2022). Appraisal constructs relevance and meaning between an organism and their (musical) environment, which is influenced by a listener's experience and knowledge. Affective intentions in music are therefore "active engagements involving a wide range of dynamically interacting trajectories" (Schiavio et al., 2017, p. 800).

Cross-cultural Differences in Music Perception

Meyer postulates that although meaning resides in the mind, there is still a certain level of objective consensus (Meyer, 1994). In his earlier book on style and music, he writes: "Style is a replication of patterning, whether in human behavior or in the artifacts produced by human behavior, that results from a series of choices made within some set of constraints" (1989, p. 3). Musical style therefore is the organization of music within a range of possibilities such as can be comprehended by one who has an understanding of these patterns of organization (Meyer, 1994). Although styles differ in the music found among different communities of people, periods in history, or even between sets of compositions by the same composer, what remains constant are 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" (Meyer, 1994, p. 7). What Meyer terms "cultural noise" might be present because of "disparities which may exist between the habit responses required by the musical style and those which a given individual actually possesses” (p. 16).

Music psychologists Thompson and Balkwill (2010) also believe that musical experiences are constrained in important ways in several aspects, including the environment, the structure of the auditory system, and the nature of perception and cognition. Commonalities allow for cross-cultural understanding of affective intention in music whereas culture-specific understanding is built upon these constraints through processes of enculturation. Their cue-redundancy model proposes that listeners across different cultures are able to appreciate affective qualities of unfamiliar music by attending to these commonalities, whereas listeners who are familiar with a musical style should find it easier to decode emotional meaning in that music because they can draw from both culture-specific and common psychophysical cues. Egermann and colleagues (2015) found that listeners from different cultures (Pygmy and Western) had more similar arousal responses when presented with Western music, and differed more in their valence responses. They suggested that “music-induced arousal responses appeared to be based on rather universal, culturally independent response mechanisms” (p. 8). Similarly, valence is also found to be influenced more by enculturation and explicit musical training (Heng & McAdams, 2022). A listener’s cultural background and experience play an important part in their comprehension of music because appraising these musical events is dependent on a whole contextual framework of social, structural, source, and identity roles (Thompson et al., 2022).

Musical and Acoustic Features Influencing Music Perception

Different dimensions in music carry important information for the communication of musical affect and meaning. The contributions of pitch, timing, and harmony have frequently been studied. Timbre, as a "complex auditory attribute, or as a set of attributes, of a perceptually fused sound event in addition to those of pitch, loudness, perceived duration, and spatial position. . . [and] is also a perceptual property, not a physical one" (McAdams, 2019, p. 23), has also been shown to carry perceptually useful information.

A great deal of previous work has explored the relationship between musical and acoustic features with perceived affect in listeners. Krumhansl (1998) explored how musical structures and the amount of emotion might be related. Musical structures such as pitch relations are perceptually salient and provide important information for listeners (e.g., Krumhansl & Kessler, 1982; Gabrielsson & Lindström, 2010). The perception of consonance also plays an important role in the music listening process—"consonant tone combinations tend to be perceived as pleasant, stable, and positively valenced; dissonant combinations tend conversely to be perceived as unpleasant, unstable, and negatively valenced" (Harrison & Pearce, 2020, p. 216). Eerola and colleagues (2009) attempted to predict affective ratings in music from audio using multivariate regression models and found that timbral elements of spectral centroid, spectral spread, and spectral entropy; harmonic characteristics such as key clarity, harmonic change, and measure of "majorness"; registral cues; rhythmic characteristics including clarity of pulsation and rhythmic periodicity; articulation; and musical structure estimated through the computation of novelty curves can be used to predict emotion ratings. Various acoustic descriptors have also been found to be combined in different ways to provide cues for decoding affective intentions and depending on features that interact together, a particular acoustic feature might map onto affect perception in different directions (Heng & McAdams, 2022). Developmental studies also “suggest that early recognition of emotions in music relies on perceptual mechanisms that detect variations in arousal in vocalizations, such as tempo and loudness, but discrimination of discrete emotions depends on learning culture-specific cues such as mode” (Cespedes-Guevara & Eerola, 2018, p. 6). Taken together, learned musical cues are likely to influence listeners’ perceived valence more than arousal.

Dean and Bailes (2010) examined acoustic influences on listeners’ perception of change, arousal, and valence. Listeners’ perception of continuous change were found to be strongly influenced by sound intensity. They also investigated spectral flatness as an acoustic feature that could potentially influence listeners’ perception because it is influenced by every spectral component and thus provides a global description of timbre. They also attempted to model spectral centroid as a predictive feature for valence and arousal (Dean & Bailes, 2011). Both spectral centroid and flatness, however, did not appear to have direct overall influence on listeners’ perception of valence and arousal for the piece they studied.

Many complex acoustic and musical features potentially contribute and interact in informing listeners of what music might be trying to express. An understanding of how listeners use these acoustic and musical features in listening will be very valuable in increasing our knowledge of how they contribute to auditory perception. If listeners with different musical backgrounds respond differently, it can imply that in addition to musical features like harmony, rhythmic characteristics, and so on, timbral features can also be learned.

Continuous Response to Music Listening

Another important, fundamental property of music is also the fact that it "requires time to exist [and] any definition of music must include this property, even if by implication" (Schubert, 2010, p. 223). This also means that temporal context plays an important role in influencing listeners' responses. "An identical physical stimulus may be perceived differently, depending on the context" (Vines et al., 2005, p. 137). A dynamic evaluation of the relationship between acoustic and musical elements on the one hand with real-time listener perceptions on the other will be very helpful in aiding the understanding of the listening process and how musical communication takes place.

Although these continuous responses might violate assumptions of stationarity as subsequent values are dependent on preceding ones, Vines and colleagues (2005) suggest that "understanding this nonstationarity may be crucial to understanding the phenomenon [of human communication]" (p. 136). Functional data analysis assumes listeners' continuous responses "reflect a smooth variation [. . .] that could be assessed, in principle, as often as desired, and is therefore a [. . .] *function*" (Ramsay et al., 2009). The collection of ordered data points for each participant are fitted onto a single curve. In other words, instead of considering each datapoint as a discrete observation, functional data analysis treats each curve as arising from a single process and allows for statistical analyses as such.

Research Questions and Hypotheses

It is likely that listeners with different musical backgrounds will respond differently to the same piece of music. The type and the amount of difference will be an interesting area for investigation. Emotional intensity as defined in this study refers to how strong an emotion is, regardless of the type of emotion it is. It will also be interesting to see if listeners' judgments of emotional intensity relate to the strength of both valence and arousal or only arousal alone. A previous study found that increasing musical context provides more information for listeners to make judgments about perceived affective intentions (Heng & McAdams, 2022). Collecting listeners' continuous response to music can help further understand how musical context provides information for listeners. In addition, tracking when these responses converge and when they diverge will also be possible and will open up new avenues for understanding the processes of affect perception in music listening. In a piece of music with a strong programmatic and narrative intention, listeners' perception are likely to be coloured by the narrative content they imagine—their appraisal of the acoustic and musical cues in the music might be influenced by their idea of the narrative content.

Specifically, we hypothesize that

1. listeners with different musical backgrounds will respond differently in terms of perceived affect;

2. valence responses between different listener groups will diverge more than arousal responses;

3. acoustic factors within the musical sound will relate more directly to perceived arousal;

4. musical cues that have been previously learned, such as modal structures,

metrical relations, and so on, exert a greater influence on listeners' perceived valence ratings more than on their arousal ratings; and

5. a listener's perceived emotional intensity reflects more their perception of the distance away from the origin in the two-dimensional valence-arousal model, than its relation to the single dimension of arousal.

Finally, the acoustic and musical features potentially influencing listeners' responses will be explored.

Method

Participants

Three groups of listeners with different musical backgrounds were recruited from Singapore. By having only participants from Singapore, effects of differences in language and other socio-cultural factors can be held relatively constant so that any differences can be attributed more directly to musical training. The first group were musicians trained in the Chinese music tradition (henceforth CHM) ($n = 30$; M years of musical training = 11.2 years; SD musical training = 5.49; 15 females and 15 males; M age = 35.5 years, SD age = 8.88). The second group were musicians trained in the Western music tradition (henceforth WM) ($n = 30$; M years of musical training = 11.9 years; SD musical training = 5.92; 17 females, 11 males, and 2 unspecified; M age = 27.4 years, SD age = 10.23). The final group were nonmusicians (henceforth NM) ($n = 30$; M years of musical training = 0.16 years; SD musical training = 0.29; 15 females and 15 males; M age = 35.5 years, SD

age = 12.30). All participants self-reported no problems with hearing.

Stimuli

The Chinese orchestra is a relatively modern invention that utilizes traditional Chinese instruments and combines a composition aesthetic derived from the rich folk and traditional musics around China, as well as composition and orchestration techniques borrowed from Western orchestral music. A Chinese orchestral work composed by Wang Chenwei (2021) was used for the listening experiment. The work utilizes Chinese instruments but was written with modes and metrical concepts borrowed from Arabic music and a programmatic content based on Aleppo and the Syrian war. The composer has clearly delineated the piece into seven sections and has explicitly detailed the narrative content for each section. The piece opens with a quiet and solemn section, “Hymn from the Ancients.” A grand and majestic section depicting an ancient and prosperous city follows, aptly named “Citadel on the Hill.” The next section “Bustle of the Souq” is joyous and busy, contrasting with the subsequent highly tensed section, “Flames of Resentment.” Another intense section follows, “Crossfire of the Armies.” The intensity decreases and a desolate “Tears of the Rubble” follows. The piece concludes hopefully, the final section titled “Path to the Future.” This piece was selected for several reasons. Firstly, it has a rich programmatic and narrative content as defined by the composer, so that affective intentions might be more easily elicited. CHM are expected to be more familiar with the instrumental timbres as compared to the other two listener groups. If familiarity with instrumental timbres and orchestration styles influences perception of affective intentions, we would expect to observe a difference between CHM and the other two groups. With formal musical training, CHM and WM are likely to have internalized a repertoire of formalized musical rules and commonly used tropes. Even though this piece of music uses modes and meters that are less familiar to both groups of musician listeners, they might interpret it with respect to the musical traditions they are familiar with. This would set them apart from the nonmusicians who would likely respond more to the surface acoustic features of the music as compared to the musicians. Instead of selecting a piece of Chinese orchestral music that uses more conventional Chinese musical forms and elements, using one with Arabic musical features will ensure that it does not bias any particular musician group. A third reason for selecting this piece of music is that it is relatively less well-known to the listeners in Singapore as it was commissioned and premiered by the Taipei Chinese Orchestra and only published in a recording by the orchestra in December 2021. The recording was provided by the composer with a sample rate and bit depth of 48kHz/24-bit (wav, stereo). The entire piece of music is 930 seconds in duration.

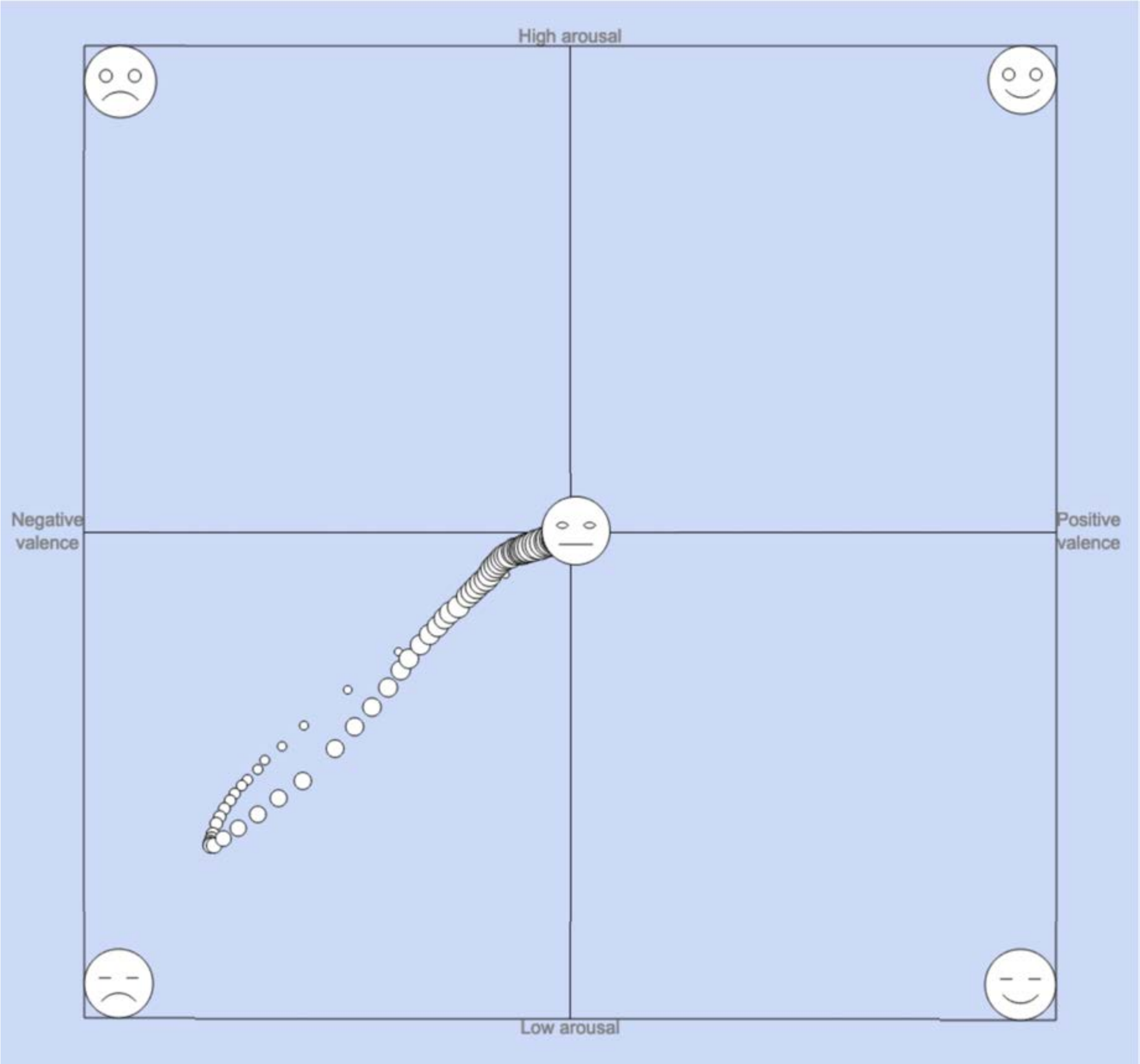
Procedure

Participants were each given a unique link for the experiment where they could access the experimental task online. They were asked to use headphones for the listening task and to complete the experiment on a computer in a quiet space. They first had to click on a box to acknowledge having read and accepted a written informed consent. Participants were then presented with three clips of music corresponding to the lowest, medium, and highest dynamic levels that they might encounter in the music and asked to adjust their volume level so that it was comfortable and they could still hear the softest clip clearly. They then proceeded to the instructions. First, the difference between perceived and induced affect was explained and participants had to respond to questions to ensure they understood the difference. They were then reminded that they were supposed to only think about perceived affective intentions and not how the music made them feel in this experiment. Next, the experimental interface was introduced with a unidimensional scale for emotional intensity and a two-dimensional scale for valence and arousal. For the unidimensional emotional intensity ratings, participants moved the mouse left and right continuously to indicate changes in perceived emotional intensity they thought the music was trying to express as the music progressed (from “low intensity” on the left to “high intensity” on the right). It did not matter where on the screen they placed their cursors vertically—only the horizontal values were mapped. This also ensured that there would not be any discontinuity in their responses. Participants were able to move their cursors throughout the entire space for the two-dimensional valence (positive-negative) and arousal (low-high) ratings, and a face emoji evolved to indicate the perceived emotion (Figure 3.1). This interface was adapted from Nagel and collegues’ (2007) EMuJoy interface, which was developed to collect continuous real-time data of perceived emotions in music. The horizontal and vertical positions were continuously captured throughout the entire piece of music. A practice excerpt of another piece of Chinese orchestral music by the same composer, lasting 39 seconds was presented for participants to get used to the interface and the tasks.

The music was then presented for the actual experimental task. All participants performed the emotional intensity rating task first and the two-dimensional valence-arousal rating task after that. Participants could choose to take a break in between the two tasks. After completing the experimental tasks, participants had to fill out a demographic questionnaire gathering information on their age, sex, years and type of musical training, and amount of time spent on musical activities. This was then followed by a short debrief, and participants were also encouraged to email the experimenters if they wished to further clarify any other thing. Each participant was compensated with SGD$10 for their participation, which took from 40 minutes to an hour.

Figure 3.1

Interface for Valence and Arousal Ratings



Note. Face emojis are overlaid on the four extreme corners here to show how it changes according to its position. Only one face emoji was present on the interface in the experiment.

Data Analysis

Listeners' continuous responses were collected at a rate of 2 Hz over the entire 930 seconds of the music. This gave 1861 observation points per participant for each scale of emotional intensity, valence, and arousal. This sampling rate is used because listener responses appeared to lag behind causal musical events by between around one to three seconds and based on Nyquist's sampling theorem, the sampling rate should be twice that of the highest expected frequency (Schubert, 2010). The emotional intensity scale ranged from 1 to 9, and the valence and arousal scales ranged from -1 to +1.

Functional Data Analysis

The first aim of this study was to explore whether different musical backgrounds contributed to differences in emotional intensity, valence, and arousal responses. To answer this question, functional data analysis was applied to the data. In functional data analysis, all the observation points on one scale for each participant were treated as one single function instead of discrete independent observations. This allows for conventional statistical methods such as regression analyses to be carried out without problems of overfitting. The data were smoothed using b-spline functions of order 6 and a roughness penalty of 0.001. As listeners might have different response speeds to the same audio stimulus, phase variation might occur between the listeners' curves. Since the variation between listeners' responses to the music is of interest, it is critical that phase variation be removed so that differences between listener responses on the same point in the music could be more accurately explored. Rather than using only specific points for alignment, values throughout the curve are registered. The rationale here is that if the curves differ only in terms of amplitude variation, then their values will be proportional throughout time t. Each point on the registered curve plotted against the target curve should approach a straight line that tends to pass through the origin. This time-warping function was therefore fitted onto the curves after smoothing, and analyses were performed on the registered curves. Mean curves (solid lines) with ±1 standard deviation (dotted lines) for each listener group were plotted (Appendix B Figures B1–B9).

A functional regression model is then estimated, after which an F statistic is computed for the regression. This version of the statistic differs from the classic F statistic in the manner in which it normalizes the numerator and denominator sums of squares (Ramsay et al., 2009). With this functional F statistic, the regions in which the listener groups diverged in their responses could be seen. Post-hoc functional t-tests were performed on the three listener groups to compare their responses over the entire piece of music. This also allows us to explore whether the three listener groups varied in similar ways and at similar temporal regions across all three scales or whether differences in their responses varied among the different scales used.

Time Series Analysis

With this set of data, which are taken sequentially in time and which are likely autoregressive in nature, time series analysis proves to be a very useful technique that can explore the relationships between acoustic and musical features with listener responses. Responses are averaged over each listener group, giving a mean set of observations each from CHM, WM, and NM groups for emotional intensity, valence, and arousal.

The process of transforming the variables appropriate for analysis follows the procedure outlined by Dean and Bailes (2010). Acoustic and musical features are likely to influence listeners' perceptions but not the other way round and so perceptual responses are taken as the endogenous variables and acoustic and musical features as exogenous ones. Listener responses are likely non-stationary in each moment—the process does not remain in statistical equilibrium and the probabilistic processes change over time (Box et al., 2016). The perceptual response is thus likely to be influenced by preceding moments. Therefore, the endogenous variables first have to be made stationary.

Autocorrelation and partial autocorrelation functions for each of the endogenous series are first determined in order to set the lag amount for stationarizing. The stationarized series are then tested with the Augmented Dickey-Fuller Generalized Least-Squares test and the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test for stationarity. All endogenous series showed stationarity after one differencing step. The exogenous series are then transformed in the same way as the endogenous series. Granger causality tests determine plausible exogenous predictors. With these, ARIMAX (autoregressive integrated moving average procedure with an exogenous variable) models are built, and the best model is selected based on minimizing the Akaike Information Criterion (AIC).

Impulse response functions represent the reactions of endogenous variables to "shocks" hitting the system (Durlauf & Blume, 2008) and are useful for studying the impact a variable has on a system. Vector autoregressive (VAR) models are used in forecasting by capturing the co-evolution of the variables (Lütkepohl, 2005). However, it is “often not clear which set of impulse responses actually reflects the ongoings in a given system [as] different sets of impulses can be computed from the same underlying VAR” (p. 357). Structural vector autoregressive (SVAR) models apply restrictions that allow the identification of impacts to a system, and unique impulse responses can be derived for each shock to a system. This is done by modelling the instantaneous relations between the observable variables directly so that a model with uncorrelated residuals can be found. This is necessary so that when impulse responses are constructed, only those that actually reflect what is happening in the system can be highlighted. Impulse response functions are constructed on these SVAR models. Each of the acoustic features is modeled as input and the listeners’ emotional intensity, valence, and arousal responses are reactions to shocks in the system. With this, we can see what the lag response in the endogenous variable is with respect to an input from the exogenous one.

Acoustic and Musical Features

Using the revised Timbre Toolbox (Peeters et al., 2011, revised by Kazazis et al., 2021) implemented in the MATLAB environment (Mathworks, Natick, MA), the entire piece of music was analyzed for several acoustic features. The acoustic descriptors selected for analysis were spectral centroid, spectral crest, spectral decrease, spectral flatness, spectral flux, spectral kurtosis, spectral roll-off, spectral skewness, spectral slope, spectral spread, spectral variation, RMS energy, and zero-crossing rate. A short-term Fourier transform with the power spectrum estimation is applied to the audio file. A Hann window of length 50 ms with a 50% overlap between successive frames was used.

A recent study by Harrison and Pearce (2020) found that the MIRToolbox's roughness measure is the best candidate for estimating consonance when only audio information is available. This function uses Sethares's roughness model. In Sethares's (1998) model, the frequency of each component is multiplied by its amplitude, and its dissonance with all other peaks is computed based on Sethare's dissonance curve (1993). The average of the dissonances of all these pairs of peaks is taken as a measure of roughness.

The *mirmode* function from the MIRToolbox estimates the modal strength of the music. It uses Gomez's key profile matrix, which utilizes a combination of machine learning with the key estimation algorithm proposed by Krumhansl and Kessler (1982) and returns a value between –1 and +1: the closer it is to –1, the more minor it is, and the closer it is to +1, the more major. The onset detection function from a VAMP plugin created by the Centre for Digital Music in Queen Mary, University of London, for the SonicVisualiser (2010) is used to estimate note density. As the piece of music utilizes modes and metrical structures found in Arabic music, modal strength and note onset density might not be sufficient in explaining listeners' perceptions. In order to explore whether there is also a relationship between listeners' perceptions of affective intentions and the Arabic modes and metrical structures used, the *maqāmāt* and *iqā'āt* used by the composer were also noted in the sections they appeared in.

Modes and Metrical Structures in Arabic Music

In their introduction to their book on Arabic music, Farraj and Shumays (2019) remind readers that Arabic music encompasses a wide range of musical traditions and genres. The composer of this piece of music has also made a detailed study of the modes and meters used and has detailed his usage very explicitly, both within the score and in his explanatory notes accompanying the score. A very brief explanation about *maqāmāt* and *iqā'āt* will follow here in order to explain the musical analyses and descriptions carried out in this study.

Maqām and jins. The Arabic *maqām* is a "system of scales, habitual melodic phrases, modulation possibilities, ornamentation norms, and aesthetic conventions" (Farraj & Shumays, 2019, p. 4). A *jins* (plural *ajnās*) is a scale fragment of a *maqām* (plural *maqāmāt*). They are the basic melodic unit, and the *maqām* can be understood as a pathway between many *ajnās* (Farraj, 2018). Each *jins* conveys a distinct mood or character and interestingly, Farrag and Shumays state that the "mood of each *jins* is subjective, and no study in ethnomusicology has found any evidence that it is perceived in the same way by all (or even majority of) listeners" (2019, p. 193). So although the *jins* possess affective content, what this affective content is might mean something very different to different listeners. Although Arabic scales are rooted in the Pythagorean tuning system, there are many intervals that cannot be described using simple harmonic ratios. Farraj and Shumays believe that whereas harmonic relationships are based on physical reality, it is in fact the aesthetic and cultural choices of Arabic musicians throughout history that established the framework used in Arabic scales instead of it being driven by purely mathematical rationale. For the purposes of this work however, the composer has specified that a quarter-tone interval is sufficient. While this may not be an accurate reflection of Arabic musical practices, the intention of the work is not an imitation but an original composition inspired by Arabic musical elements. The list of *ajnās* used in the work are provided in Appendix B, Figure B19.

Iqā'. *Iqā'* (plural *iqā'āt*) is used to describe a rhythmic cycle. *Iqā'āt* are made up of two different basic building blocks, the dum and tak, onomatopoeias derived from the sound produced on membranophones. Similarly to the *maqāmāt* and *ajnās*, although the composer has incorporated some specific *iqā'āt* in the work, they are not meant to be a true indication of, but rather an inspiration from Arabic musical practices. The list of *iqā'āt* used in this work can also be found in the Appendix B, Figure B20.

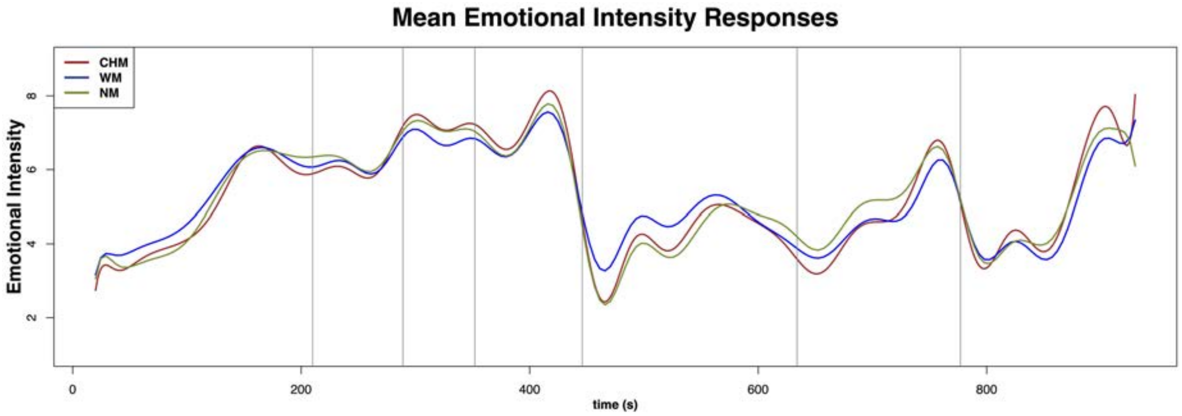
Results

Participants' responses were captured for the entire 930 s of the music. However, only responses from 20 s to 930 s were included in the analysis. This is to prevent the initial cursor movement from influencing the results as the cursor. The cursor always started from the middle of the interface, its position might not accurately reflect the perceived affective intention at the start of the music. Schubert (2013) cautions about treatment of the initial segment of continuous response. There is not much contrast in the first 200 s of this piece of music, and the exclusion of this 20 s simply removed the initial dip in the three scales of emotional intensity, valence, and arousal. He also explores the afterglow effect in which participants may not all return the cursor to the centre at the conclusion of the music. In this study, participants were not instructed, nor were they expected to do so, and the position at which their cursor left off was assumed to be the perceived affect at the conclusion of the piece. However, the standard deviations show an increase at the end of the piece, which could be due to the afterglow effect. On the other hand, it might also be a unique artifact of this piece of music. Towards the end, there is a drop in dynamic level, a decrease in the number of instruments, and a reduction in the overall note density, before a sudden dramatic build up to an intense, loud climactic conclusion.

The means of the responses of each listener group on emotional intensity, valence, and arousal were plotted from 20 s to 930 s (Figures 3.2–3.4, respectively). Appendix B, Figures B1–B9 plot the mean and standard deviation curves of the responses on emotional intensity, arousal, and valence for each group of listeners. A few observations can be made from these mean curves. Firstly, it appears that the shapes of the mean emotional intensity curves are very similar to those of arousal. Listeners likely interpret the strength of emotional response in a similar way to that of the arousal response.

Figure 3.2

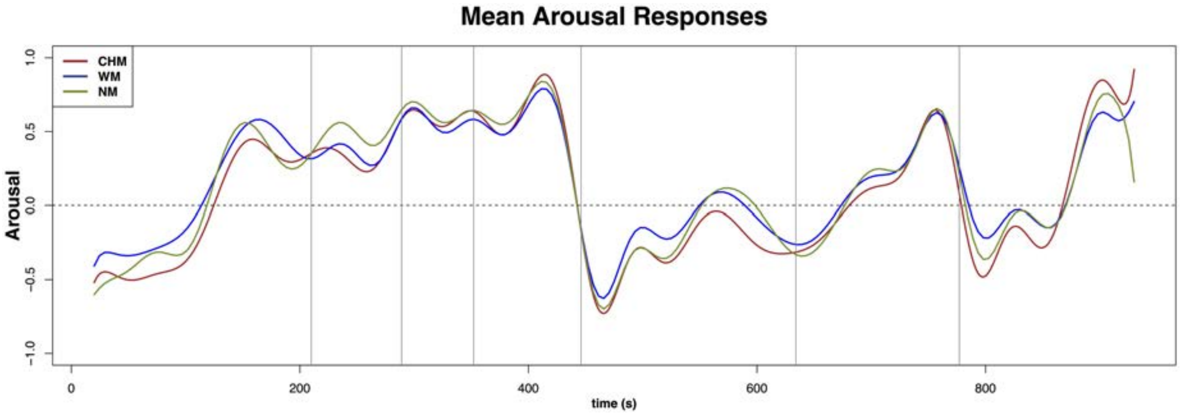
Mean Emotional Intensity for Each Listener Group



Note. Vertical lines indicate temporal position of section divisions delineated by the composer.

Figure 3.3

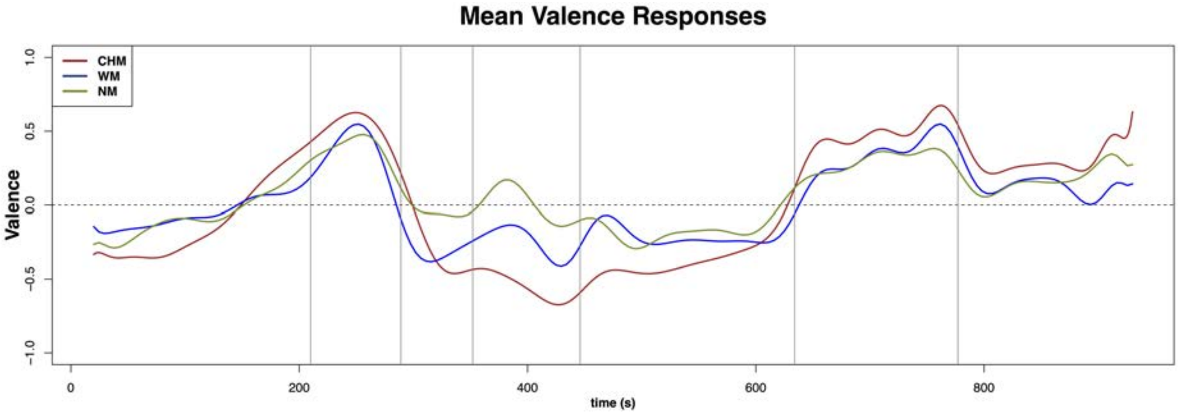
Mean Arousal for Each Listener Group



Note. See Note in Figure 3.2.

Figure 3.4

Mean Valence for Each Listener Group



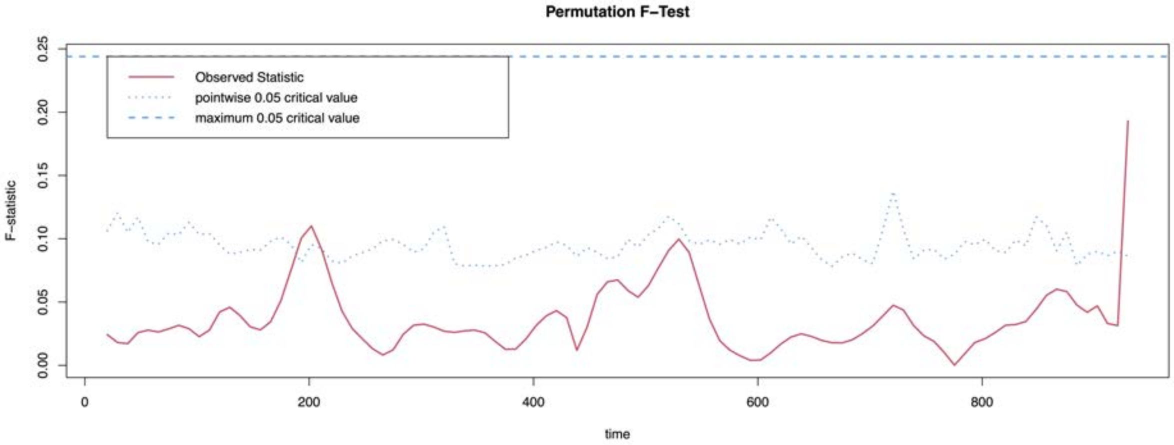
Note. See Note in Figure 3.2.

The mean curves for valence showed more divergence between the listener groups than those of emotional intensity and arousal. This was especially prominent in the middle of the piece, from around 300 s to 500 s. Valence responses of CHM were more extreme than the other two listener groups compared to emotional intensity or arousal. This could be due to CHM being more confident about their responses in general, due to the timbre and performance practice of a Chinese orchestra being the most familiar to this group of listeners.

Functional F-tests were performed to explore the differences between the three listener groups for emotional intensity, valence, and arousal. Figures 3.5–3.7 show the functional F statistic for emotional intensity, arousal, and valence, respectively. The pointwise critical F value of .05, seen as a blue dotted curve, is calculated for participants' responses over the 910 s included in the analysis, and the maximal critical F value is seen as a straight blue dashed line over the entire function. The red line indicates the functional F statistic across time. The pointwise critical F value will be used as a significance criterion here because listeners' responses vary according to the context of the acoustic and musical parameters, and because this study investigates time-varying changes that are influenced by varying contextual cues rather than a single unitary response for the entire piece of music. So temporal regions in which the F curve exceeds the pointwise critical curve are take as significantly different among listener groups. These functional F-tests confirm the pattern observed in the mean curves – there were less regions that are significantly different between the three listener groups in emotional intensity and arousal, and many more regions of difference in valence. Although the shapes of the mean emotional intensity and arousal curves are very similar, they did not share the same regions in which listener responses diverged significantly between groups. Post hoc functional t-tests were performed to compare these differences and explore the directions in which the listener groups diverged.

Figure 3.5

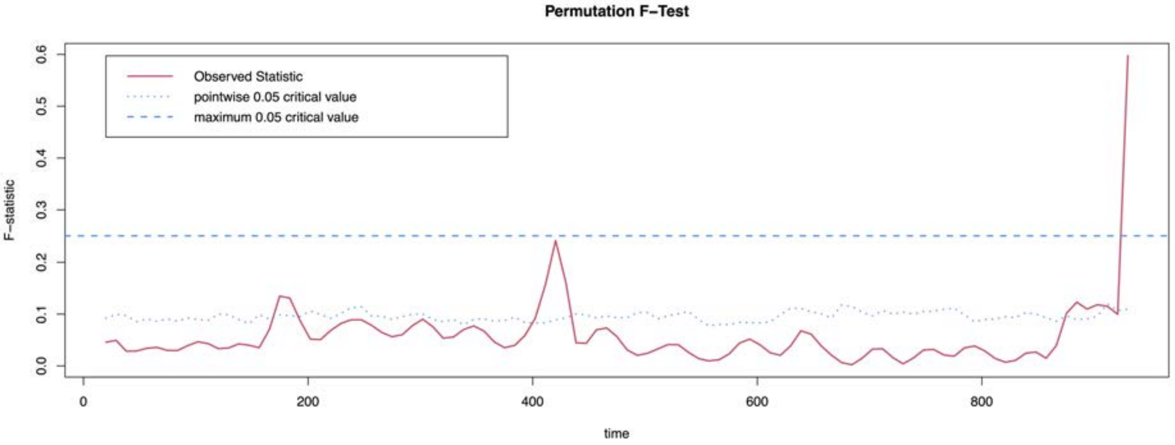
Functional F-Test for Emotional Intensity



Note. Blue dotted curve shows pointwise critical F value of .05, straight blue dashed line shows maximal critical F value. Red line indicates the functional F statistic across time.

Figure 3.6

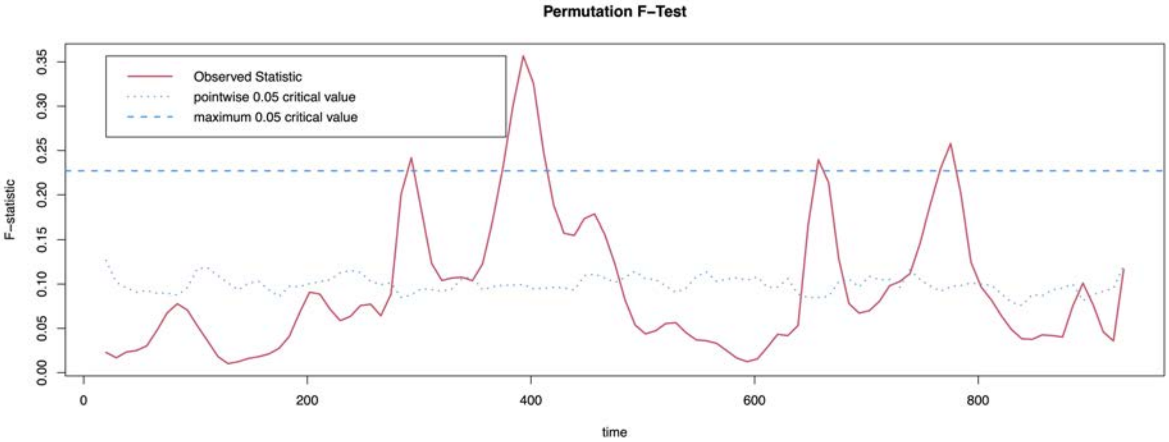
Functional F-Test for Arousal



Note. See Note in Figure 3.5.

Figure 3.7

Functional F-Test for Valence



Note. See Note in Figure 3.5.

Emotional Intensity and Arousal

A visual examination revealed close similarities between the emotional intensity and arousal mean curves. The regions where arousal was close to zero corresponded to regions where the emotional intensity was around 5, the midpoint of the emotional intensity scale. The peaks, troughs, and directions of the slopes also match closely for both sets of curves. Rather than emotional intensity reflecting the distance from the origin, it appears that emotional intensity relates directly to the level of arousal. A low arousal, rated as −1 on the arousal scale is considered as having a low emotional intensity of 1, and a high arousal, rated as +1 is considered as having a high emotional intensity of 10. The three groups of listeners differed in less regions for emotional intensity than for arousal. Especially prominent is a region between around 400-430 s where the groups differed in arousal but not in emotional intensity

(Figures 3.5 and 3.6). The narrative content of the piece in this region is marked by the idea of conflict and war, both in the indications of sectional descriptions by the composer—“Crossfire of the armies”—and by the onomatopoeic representations with loud, constant, pounding drums, and heavy dissonances. CHM appear to have the highest arousal ratings, which continue to stay high even as the music starts to decrease in intensity in anticipation of the following section, “Tears of the rubble”. Their arousal ratings dropped the most rapidly at the end of this section to converge with that of the other two listener groups when the following section begins. From a visual observation, the standard deviation of the emotional intensity and arousal responses for each listener group are very similar, indicating also a similarity between the three listener groups along these two dimensions.

Valence

A visual examination of the emotional intensity curves with the valence curves did not reveal any clearly observable correspondences. Peaks and troughs of valence responses did not appear to be related to the peaks and troughs of emotional intensity. The standard deviation of valence responses for each listener group shows a smaller variance in CHM and WM, and a larger variance for NM (Appendix B, Figures B7–B9). There is more agreement within the musician listener groups for valence responses as compared to the nonmusician group. Valence responses between 270 s to 480 s showed significant differences in the F-test among the three listener groups (Figure 3.7). From 270 s to 315 s, CHM rated valence significantly more positively than WM. From 315 s to 480 s, CHM’s valence responses continued becoming more negative while WM and NM either stopped decreasing or started to increase in their valence ratings. For the rest of this section, CHM had the lowest valence ratings, NM the highest, and WM fell in the middle.

The composer has clearly delineated the region at 289 s, separating it sonically with a loud punctuated chord and narratively by labelling it “Flames of Resentment”, a stark contrast to the previous section “Bustle of the Souq”. Musically, the section leading to this region has dense unisons on the plucked and bowed strings over the rhythmic pattern from *iqā’* Yuruk Semai. This section begins with loud fanfare-like unisons in winds utilizing various *ajnās* including *jins* Sabā Zamzam, *jins* Nikriz, *jins* Ḥijāzkar, *jins* Kurd, *jins* ’Ajam, and *jins* Nahāwand before the instrument density drops off, leaving only a solo *sanxian* over the cello, double bass, and bass *sheng* in the low register. These fanfare-like and quieter sections alternate two more times with various *jins* Ḥijāzkar being passed. Following that, an ascending scalar pattern coupled in thirds and built from the *ajnās* Ḥijāz and Sabā Zamzam creates a noisy and almost dissonant effect. This link provides a video showing listeners’ valence and arousal responses with annotations of the associated *maqāmāt* and *iqā’āt* over the course of the music: https://mpcl.music.mcgill.ca/supplementaryMaterials/HengThesis2023/VA_modes_rhyt.mp4.

Although there was a decrease in the valence ratings for all three listener groups, WM showed the greatest decrease. CHM started off this section with a higher valence than WM and NM, but their ratings also very quickly decreased. Although NM also showed a general decrease in valence ratings, their rate of change was the least among the three groups. From about 350 s onwards, the valence ratings of both musician groups were in the negative region, whereas NM’s valence ratings became positive.

Between 360 s to 420 s, NM rated their perceived valence significantly higher than both CHM and WM, and WM rated theirs significantly higher than CHM. *Maqām* Nahāwand is used here and it might be heard as a minor mode to listeners more familiar with Western music than Arabic music. The instrumentation also creates a dark timbre. However, it appears that modal and timbral content did not contribute as much in communicating a negative valence to NM, and less to WM to CHM. Musically at this section, it is likely the increasing dissonances, timbral quality, and the expectation of a negative narrative content led CHM to continue decreasing in their valence ratings, but this did not occur for the other two groups.

The next region where the three groups diverged significantly happens between 650 s to 680 s. CHM perceived the valence to be significantly higher than WM and NM—the bi-metric rhythm in the handdrum versus the melody introduces an interesting change to the general mood compared to what happened before. The instruments used also create a clearer, brighter sound. These features possibly contribute to the perception of a more positive-valence in CHM.

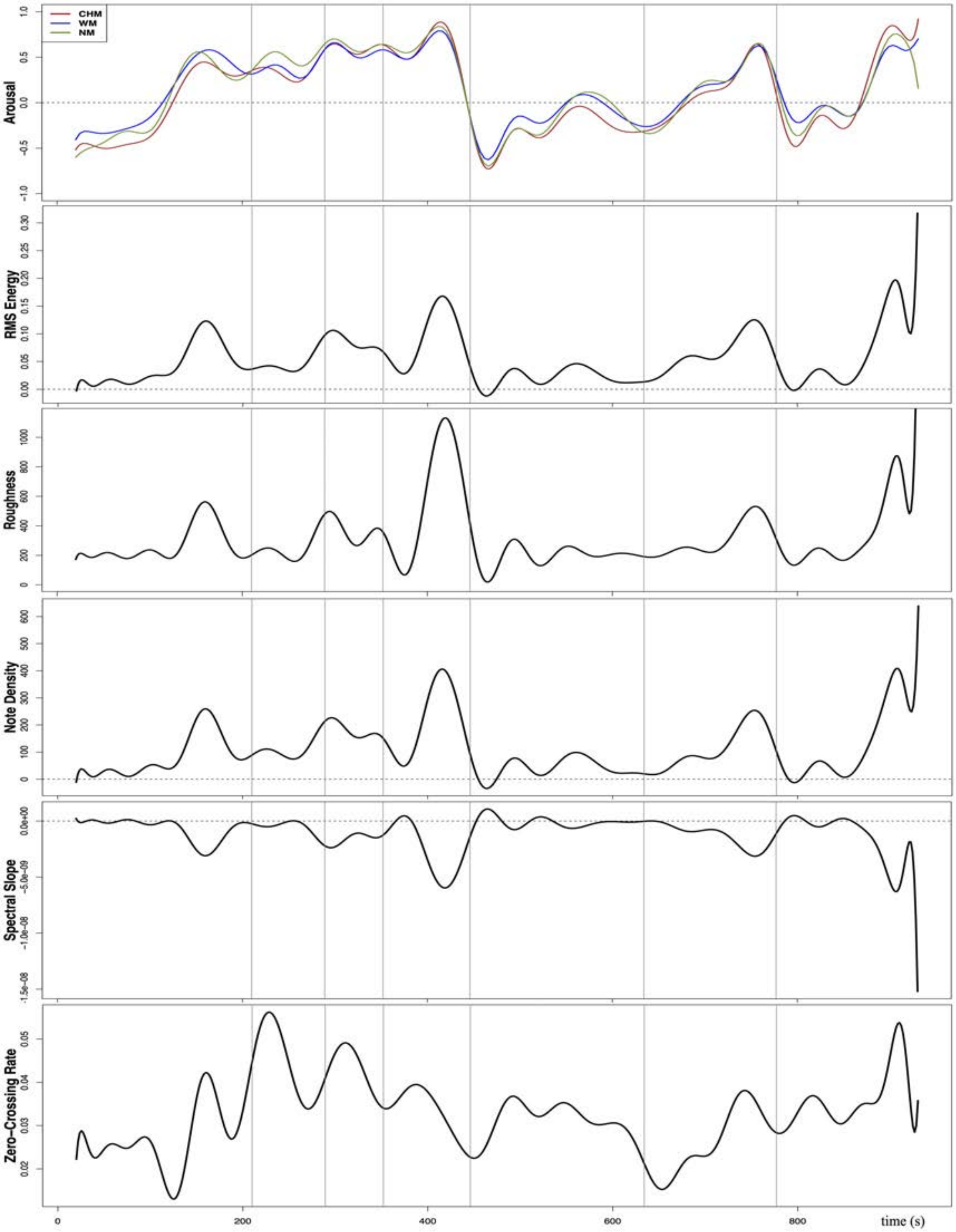
The final region where valence responses of the listeners were significantly different occurs between 740 s to 780 s. Here NM rated the perceived valence significantly lower than that of CHM and WM. There is a large swell into dense unisons and stable melodic tendencies, which fall on the fifth and sixth degrees of the scale. This stability is likely sensed by musician listeners much more than by nonmusicians.

Affective Responses with Acoustic and Musical Features

Mean curves for each listener group are compared with the curves of the acoustic and musical features over the entire piece of music. As the shape of the emotional intensity curves are very similar to that of arousal, Figure 3.8 shows only the mean arousal curves of the three listener groups together with RMS energy, roughness, note onset density, spectral slope, and zero-crossing rate curves as they vary over the piece of music. RMS energy measures the root mean square of frame energy of the signal, an indication of the intensity level of the music. Roughness is used as a measure of dissonance perception, and note onset density provides information about the general textural density of the music, measuring both the number of instruments sounding together at the same time, as well as the amount of activity that is happening. Spectral slope measures the amount of decrease of the spectrum and is an important aspect of timbre; the slope is directly related to the resonant characteristics of a sound and may be implicated in recognition. Spectral slope is most pronounced when the energy in the lower frequencies is much greater than the energy in the higher frequencies. Zero-crossing rate shows the number of times signal crosses the zero axis in a given time window and tends to be small for periodic sounds and large for noisy ones. Note that although the magnitude of changes do not correspond exactly, the peaks and troughs of these curves match very closely, especially for RMS energy, roughness, and note density. Spectral slope varies negatively with emotional intensity and arousal responses, the troughs of the spectral slope curve matching the peaks of the emotional intensity and arousal curves, and vice versa.

Figure 3.8

Mean Curves of Arousal with RMS Energy, Roughness, Note Density, Spectral Slope, and Zero-Crossing Rate



To further explore the effects of these acoustic and musical features, impulse response functions were modeled to study their contributions to listeners' valence and arousal responses. Listeners' responses were collected at a rate of 2 Hz, and the measured values of the acoustic and musical features were averaged over successive 0.5 s windows. After transforming the responses into stationary processes, the same transformations were applied to all the acoustic and musical features and ARIMAX models were built. The best models were selected based on the lowest AIC. After the VAR model is built, structural restrictions are imposed, and impulse response functions are constructed with these SVARs. Figure 3.9 shows an example of the impulse response plot for note density on arousal responses for CHM, WM, and NM. The full set of impulse response functions can be found in Appendix B, Figures B10 to B18. The y-axis reflects the amount of unit change in response to each acoustic or musical feature, and the x-axis shows the number of lags taken for the reaction to reach a certain level. Each lag is 0.5 s. Although listeners could be familiar with particular musical tropes and styles and anticipate certain responses before a particular stimulus is presented, it is difficult to untangle the complex contributions from anticipatory responses and reactionary responses. In these analyses, we will therefore only focus on responses that occur after the presentation of the sound and so a restriction that each impulse response starts from zero was applied. As we are more interested in the immediate reaction engendered by particular acoustic and musical features on listeners' responses, rather than the smaller background contributions they might have, the additional restriction that the impulse response converges to zero after a period of time was also imposed. The red dotted lines show the 95% confidence interval for each impulse response. Only impulse responses with the 95%

confidence intervals breaching the horizontal axis are considered to have an influence on listeners' responses.

Figure 3.9
Impulse Response Functions of Note Density with Arousal Responses of Chinese Musician, Western Musician, and Nonmusician Listeners

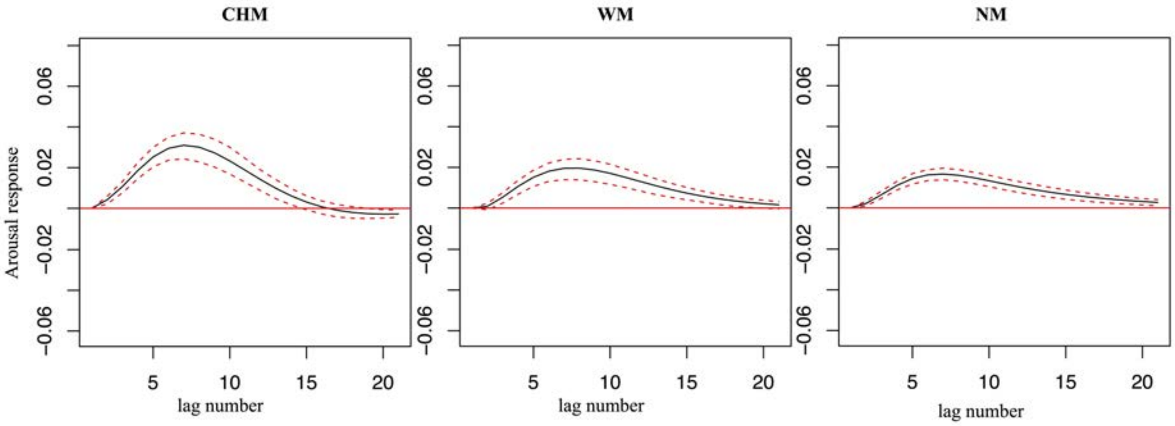


Table 3.1 shows the acoustic and musical features that influence each group of listeners in their perceived emotional intensity, arousal, and valence responses across the entire piece of music. All three groups of listeners appear to utilize similar features in their perception of emotional intensity, although the two musician groups show greater magnitude than the nonmusicians in their impulse response curves.

These features point towards a more even distribution of energy across partials, a noisier, louder sound with greater harmonic dissonance, and a denser texture that listeners use to rate a higher emotional intensity. The three listener groups also appear to use similar features that they use in their perception of arousal. WM and NM can be seen to utilize the same features, which are also the same ones

Table 3.1

Acoustic and Musical Features Influencing Listeners' Perception of Emotional Intensity, Arousal, and Valence

<table>
<tr><th colspan="3">Emotional intensity</th></tr>
<tr><td>Spectral Slope –
RMS Energy +
Zero-Crossing Rate +
Roughness +
Density +</td><td>Spectral Slope –
RMS Energy +
Zero-Crossing Rate +
Roughness +
Density +</td><td>Spectral Slope [–]
RMS Energy +
Zero-Crossing Rate [+]
Roughness [+]
Density +</td></tr>
<tr><th colspan="3">Arousal</th></tr>
<tr><td>Spectral Slope –
RMS Energy +
Zero-Crossing Rate +
Roughness +
Density +
Spectral Flatness +
Spectral Flux [+]
Spectral Kurtosis [–]</td><td>Spectral Slope –
RMS Energy +
Zero-Crossing Rate +
Roughness +
Density +</td><td>Spectral Slope –
RMS Energy +
Zero-Crossing Rate +
Roughness +
Density +</td></tr>
<tr><th colspan="3">Valence</th></tr>
<tr><td>Spectral Centroid [–]
Zero-Crossing Rate [+]</td><td>Spectral Slope [+]
RMS Energy [–]</td><td>Zero-Crossing Rate [+]
Density [+]</td></tr>
</table>

Note. Only features whose 95% C.I. of impulse response functions breach 0 are listed. A positive sign indicates that when values of a feature increase, listeners' ratings on the particular affective scale rise whereas a negative sign indicates an increase in the value of a feature leading to a decrease in ratings. Square brackets indicate a peak response of less than ±0.01.

implicated in their emotional intensity responses. In addition to all the features utilized by WM and NM, spectral flatness, spectral flux, and spectral kurtosis also featured in CHM's perception of arousal.

Spectral flatness is an indication of the "peakiness" of the spectrum—sinusoidal components produce a more peaky spectrum, and white noise gives a flat spectrum. In an orchestral setting, a dense texture, with instruments spanning many different registers, and complex chordal structures could also have a higher spectral flatness, as compared to a thinner texture and a less complex harmonic structure. Spectral flux is a measure of how quickly the power spectrum of a signal is changing. Many changing instruments, changes in instrumental combinations, or changes in the playing techniques of the instruments could create a high spectral flux, compared to a slow passage with less changes and contrasts in sound. Spectral kurtosis measures the flatness of the spectrum around its centroid and higher kurtosis might indicate a section of music where the pitches of the instruments are clustered more closely together, or it might indicate a thinner texture in which only very few instruments are playing a very clear melodic line.

A visual comparison of the mean valence curves with the acoustic and musical feature curves did not yield any clear correspondences. It appears to be more difficult to attribute any clear and strong influence of any of the features with listeners' perceived valence responses. Impulse response functions corroborate this, showing not only that valence responses appear to have the lowest number of acoustic and musical features that influence its perception, but also the very small contributions of these features. CHM and WM do not share any features in the perception of valence, but zero-crossing rate appears to influence both CHM and NM. Spectral centroid is a measure of the spectral centre of gravity. Sounds with higher spectral centroids are perceived as brighter, and for CHM a slightly darker but noisier sound points towards a more positive valence. More tonal, darker, and softer sounds indicate a more positive valence for WM, and noisier sounds with denser textures are perceived as more positively valenced for NM.

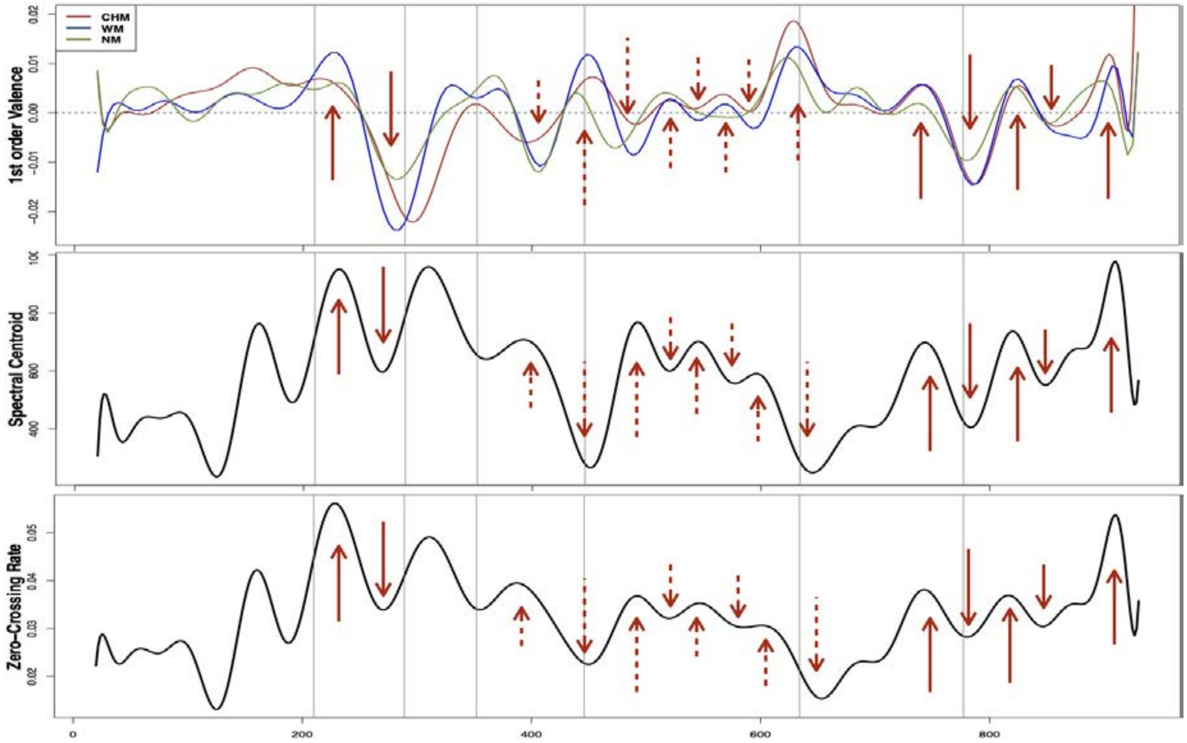
Derivatives of Listener Responses

Analyzing derivatives is an important component of functional data analysis (Levitin et al., 2007) and being able to visualize these changing values (of velocity and acceleration) can reveal patterns of energy exchange that were not obvious in the raw position data (Vines et al., 2005). These derivatives can reflect the rate of change of perceptual responses over time, highlighting regions of rapid change, as well as static regions of non-change. The first derivative for valence with respect to the acoustic and musical features is first explored. From visual observation, only two features appear to show a relationship with the first derivative for valence. Figure 3.10 shows these two features, the spectral centroid and zero-crossing rate, with the first derivative of valence. While the magnitude of the values do not correspond exactly, peaks and troughs of these acoustic features aligned closely to the peaks and troughs found in the first derivative mean curves of valence.

The regions marked by the solid arrows in Figure 3.10 show the peaks and troughs of the spectral centroid and zero-crossing rate curves matching those of the rate of change of listeners' perceived valence. The five regions marked with dashed arrows have clear peaks in the rate of change of perceived valence matching the troughs of spectral centroid and zero-crossing rate curves, and vice versa. Looking simply at the acoustic features, it may appear that the rate of change of perceived valence does not relate in a consistent manner. However, when the affective content of the music is taken into account, this relationship is clearly not random. The regions

Figure 3.10

First Derivative of Valence Responses with Spectral Centroid and Zero-Crossing Rate



Note. Solid arrows show peaks and troughs going in the same direction, dashed arrows showing peaks and troughs in opposite directions.

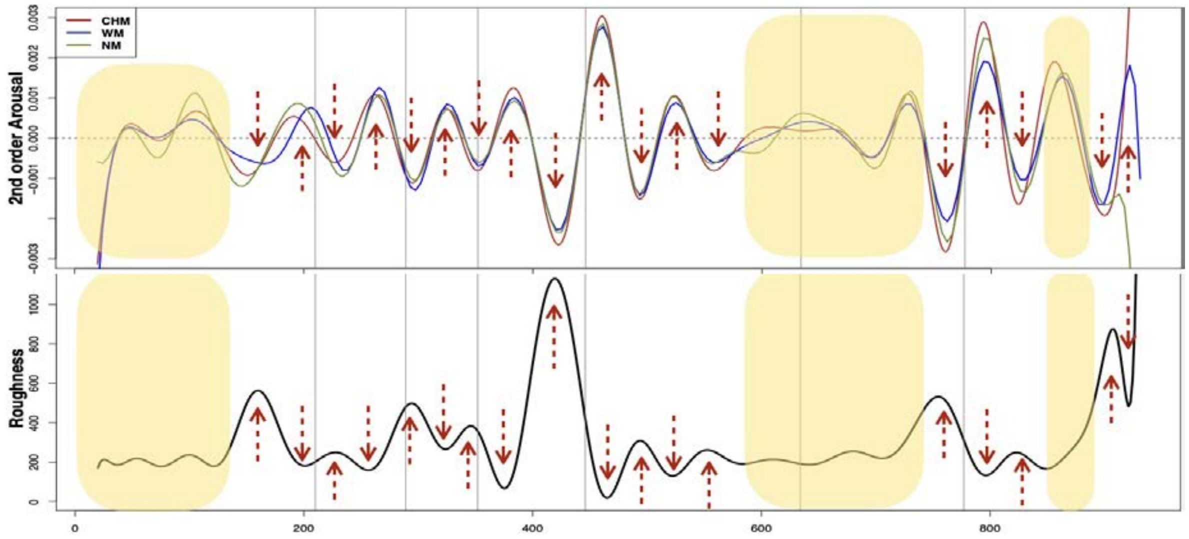
marked by solid arrows are regions where the music is primarily positively valenced, whereas those marked by dashed arrows are regions in which the music carries a negative valence (roughly 300-700 s in Figure 3.4).

As the perception of consonance plays an important role in influencing listeners' perceived affect, the relationship between roughness and listeners' responses was explored. There did not appear to be a large consistent or direct relationship between roughness and listener responses on the three scales, but when the second derivative or acceleration of arousal is examined, a clear relationship can be observed.

Figure 3.11 plots the second derivative of mean arousal curves with roughness. Each of the peaks and troughs of roughness line up with the troughs and peaks of arousal acceleration—when roughness increases, listeners decelerate in their rate of arousal change until a peak in roughness, and then as roughness decreases, listeners' rating changes accelerate. Three regions highlighted in yellow on Figure 3.11 stand out where no clear relationships between the curves can be seen. Similarly to what is happening in the first derivative of valence and spectral centroid and zero-crossing rate, different things are happening musically in these regions. In these three regions, the tempo is very slow or metrically free. It appears then that in regions that are marked with a clear sense of meter, listeners do respond in a consistent way with respect to a perception of roughness.

Figure 3.11

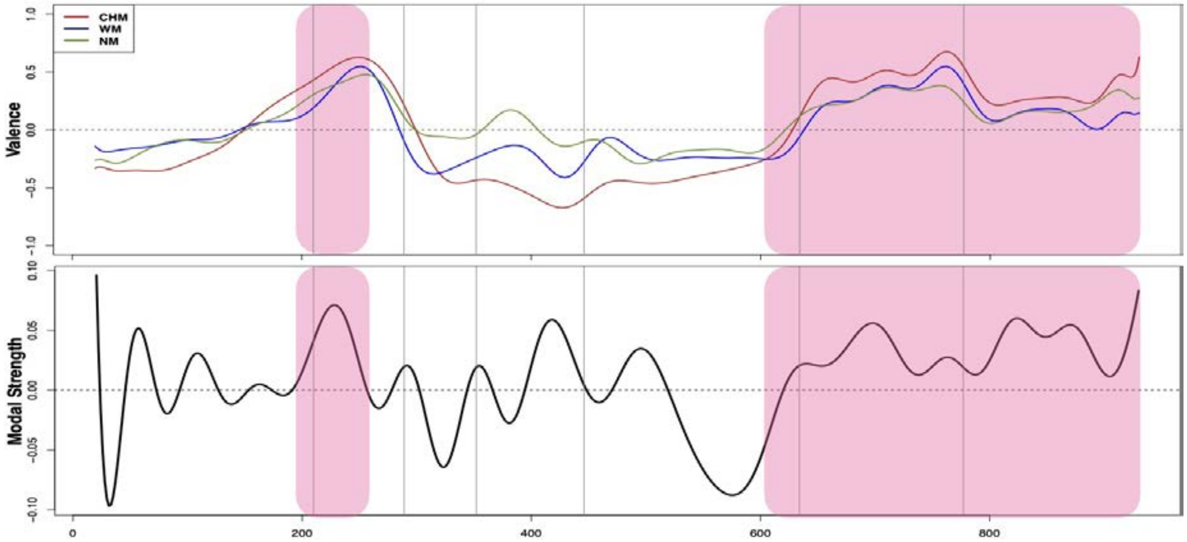
Second Derivative of Arousal Responses with Roughness



Note. Dashed arrows showing peaks and troughs in opposite directions. Areas shaded in yellow show metrically free regions.

Figure 3.12

Valence Responses with Modal Strength



Note. Highlighted areas show regions where an increase in “majorness” corresponds to more positive-valence ratings and vice versa.

The modal strength of this piece and its relation to listeners’ valence responses was explored. Although Arabic modes are used in this piece, it is hypothesized that the listeners are familiar with the concept of major and minor tonality but not with Arabic modes and therefore might base their judgments on how close the music sounds to a major or minor tonality, rather than on the Arabic modes and the affect they might elicit. There were two general regions where an increase in “majorness” corresponded with an increase in listeners’ valence ratings, one around 200–250 s, and another bigger region from 600 s to the end (Figure 3.12). There are no clear trends that can be observed where the slopes of the curves move in the same direction, the correspondences present here are simply the side of the horizontal axis the curves fall on. In other words, it does not appear that listeners perceive a more positive-valence with a stronger “majorness”, simply that a characteristic of being major indicates in general a positive-valence. There are, however, many regions in which this correspondence does not apply, and no general characteristics can be discerned in the regions where this correspondence occurs and regions in which it does not.

Discussion

This study set out to examine the following hypotheses: H1) listeners with different musical backgrounds will respond differently in terms of perceived affect; H2) valence responses between different listener groups will diverge more than arousal responses; H3) acoustic factors within the musical sound will relate more directly to perceived arousal; H4) musical cues that have been previously learned, such as modal structures, metrical relations, and so on, will exert a greater influence on listeners' perceived valence ratings than on their arousal ratings; and H5) a listeners' perceived emotional intensity ratings will reflect their perception of the distance away from the origin in the two-dimensional valence-arousal model rather than varying in correlation with the arousal dimension. The results appear to support the first three hypotheses proposed in this study although there were complex factors influencing each of them. Evidence for the fourth hypothesis is somewhat mixed; both valence and arousal responses do appear to be influenced by learning, but the aspects that are learned are less clear. There is no clear correspondence of learned musical cues with perceived affective responses, but there are likely complex interactions with other features, affective intentions, and the narrative content of the music. The fifth hypothesis was not supported as the mean curves for emotional intensity and arousal shared many similarities.

H1. Listeners With Different Musical Backgrounds Respond Differently

Mean curves for valence showed more divergence between the listener groups than those of emotional intensity and arousal. A large region in the middle of the piece between 270–480 s showed significant differences in the F-test among the three listener groups (Figure 3.7). In general, CHM gave more extreme valence responses than the other two groups. The two musician listener groups also had a smaller variance in their valence responses as compared to the nonmusician group, indicating a greater within-group consensus for perceived valence. Even though the three groups of listeners diverged significantly in their perception of affective intentions for certain temporal regions, their divergence was mostly in terms of quantity rather than quality, except for a region in the middle of the piece (355-409 s, Figure 3.4) with high-arousal/negative-valence. For the other divergences, the mean valence and arousal for the three groups of listeners remained on the same side of the space (upper or lower for arousal, left or right for valence) and differed only in magnitude.

The composer clearly delineated the work into seven sections, each having a distinct programmatic content and narrative idea. These narratives relate closely to specific affective intentions. The piece opens with a slow, sombre section titled "Hymn of the Ancients" with low-arousal and negative-to-neutral valence. "Citadel on the Hill" follows, a grand and majestic depiction of the height of political stability and prosperity, carrying with it a high arousal and relatively neutral valence. The third section titled "Bustle of the Souq" portrays a busy marketplace with plenty of activity going on and conveys a happy, joyous atmosphere, which can easily be related to a high-arousal/positive-valence affective intention. The fourth and fifth section, “Flames of Resentment,” and “Crossfire of the Armies” contrast the previous high-arousal/positive-valence section with high arousal and negative valence. The penultimate section is titled “Tears of the Rubble”, a low arousal negatively valenced section, and the final section “Path to the Future” has largely a low-arousal/positive-valence affect before making a dramatic close with high-arousal/positive-valence. Each of these sections relate closely to specific affective intentions. Although this study did not set out to investigate whether participants’ perceived affective intentions accurately corresponded to the affective intentions the composer intended to express in the music, it does appear that this communication was quite successful when listeners’ responses were compared to the intended affect—the valence and arousal curves (Figures 3.3 and 3.4) corresponded largely with the composer’s intentions. There was generally consensus among participants, with the standard deviations relatively consistent and not too large in all three listener groups. Emotional intensity and arousal responses also had more agreement among participants than did valence responses.

Listeners rated the emotional intensity on their first presentation of the piece of music. With no information and expectation of what the music would be, all three groups might have been equally uncertain as to what to expect. Their responses likely reflect their surface interpretation of the acoustic and musical features in the music, and all three groups utilize a similar set of features to inform them of the perceived emotional intensity of the music. Spectral slope, RMS energy, zero-crossing rate, roughness, and note onset density can be seen to relate directly to listeners’ perceived emotional intensity responses, and so these could be surface features that are able to communicate particular types of information quickly and effectively. These features do not depend as much on prior musical experience or expertise in a particular style of music.

In the second presentation of the music, all listeners had to rate their perceived valence and arousal. By this second presentation, listeners will have already formed some expectations. The listeners' expectations might have influenced them to focus on particular aspects of the musical sound and ignore others. The aspects of the sound a listener focuses on may be related to their musical background. The Brunswickian lens model that is adapted by Juslin (2000) for musical communication suggests that multiple probablistic cues are available. Thompson and Balkwill's (2010) propose common and culture-specific cues listeners have when listening to music from various cultures. As can be seen from the impulse response functions, CHM have a number of cues that are related to their arousal responses: spectral flatness, spectral flux, and spectral kurtosis, in addition to those that were also found to relate to emotional intensity responses. The arousal responses of WM and NM, on the other hand, were influenced by the same set of features as their emotional intensity responses. CHM who are more experienced with the instruments and compositional techniques used in Chinese orchestral music might have had an idea of which features figure more prominently in the communication of particular intentions and therefore have more information available for their judgments. These additional pieces of information might work in tandem with the expectations CHM have formed, thereby giving them a clearer idea of the affective intentions in the music.

Familiarity with a musical tradition also plays a role in tempering listener responses in other ways. For instance, the common trope of a dramatic ending that is expected by musician listeners may be less expected for the nonmusician listeners. When the sudden drop to a *pianissimo* occurred towards the ending of the piece, the perceived arousal responses of CHM and WM dropped slightly, but rose again immediately to end on a high arousal. These two groups of listeners anticipated a return to a loud and majestic close and therefore kept their arousal responses higher than those of the NM.

H2: Valence Responses Diverge More Than Arousal Responses

Divergences of perceived affective intentions between listener groups are more pronounced for valence than emotional intensity or arousal responses, indicating a greater cultural-specificity for perceived valence. Even in regions where the three listener groups showed significant differences in their valence responses, the differences were more in terms of magnitude than of emotional quality (e.g., 152–190 s; 289–352 s). The amount of divergence was not constant throughout the music but varied in regions with different affective intentions. The regions with low-arousal/negative-valence were the most agreed upon in the quality of perceived affective intention for all three listener groups (0–109 s and 446–534 s). Differences were greatest for the region with high-arousal/negative-valence (355–409 s), differing in both magnitude and quality of affect. Differences in perceived affective intentions were varied in regions of high-arousal/positive-valence (124–289 s, 649–777 s, and 885–930 s) and low-arousal/positive-valence (634–649 s and 777–871 s).

In all the regions where listener responses diverged significantly, all three listener groups converged in the quality of their perceived emotional intensity and arousal ratings, but this is not always the case for their valence ratings. In the region between 355–409 s, the three groups of listeners not only differed significantly in the magnitude of their valence ratings, NM also rated valence positively, whereas the valence ratings for CHM and WM were negative. This similarity in arousal responses but divergence in valence responses between listener groups suggests that valence response is likely to be culturally learned. Perceived arousal and emotional intensity responses, on the other hand, are likely to be based more on universal, culturally independent response mechanisms and therefore show a greater agreement among the listener groups. This is consistent with Egermann and colleagues' (2015) findings that arousal responses are more universal, and valence responses are more culturally learned.

Emotional intensity and arousal responses were also found to be elicited by a similar group of acoustic and musical features in all three listener groups. Spectral slope, RMS energy, zero-crossing rate, roughness, and note onset density were found to influence all three groups of listeners in their perception of emotional intensity and arousal. It appears that these features were able to communicate particular straightforward and simple types of information quickly and effectively and did not depend as much on prior musical experience or expertise with a particular style of music. They might also be less dependent on any narratives formed by the listeners.

H3. Acoustic Factors Relate More Directly To Perceived Arousal Than Valence

The impulse response functions demonstrate several acoustic and musical features that relate directly to emotional intensity and arousal responses. They also tend to show a greater influence on emotional intensity and arousal responses than the features that relate to valence responses do. Dean and Bailes (2010) showed that intensity influences listeners perception of change in music. Intensity is also directly implicated with RMS energy, which can be seen to influence perceived emotional intensity and arousal in this study. In addition, the presence of a higher note density indirectly reflects a larger number of instruments used, which directly influences listeners' perception of emotional intensity and arousal. Zero-crossing rate indicates noisiness, a higher level of which also indicates a higher perceived emotional intensity and arousal. Roughness, as a measure of consonance, is seen to be more directly related to the perception of emotional intensity and arousal than to that of valence. Greater harmonic dissonance informs listeners of a higher emotional intensity and arousal. Spectral slope, as an indication of the shape of the spectral envelope, also influences listeners' perception of emotional intensity and arousal. As opposed to the more direct influence to perceived arousal from surface acoustic cues, cues that influence valence responses function in more complex ways. Even though listeners are relatively successful in their perceived valence responses when these are compared with the composer's intentions, there were very few cues that can be seen to directly influence these responses. The impulse response functions of the cues that relate to valence responses were much smaller in peak value compared to the cues that influence emotional intensity and arousal responses. More cues were found to relate directly to perceived arousal responses than to valence responses.

However, several acoustic and musical features can also be seen to interact in complex ways with perceived affective intentions. Spectral centroid and zero-crossing rate, for instance, interact with the expressed valence intentions of the piece of music to influence listeners' rate of change in perceived valence. As shown in Figure 3.10, where the musical narrative is perceived to have a negative valence, the peaks and troughs align with troughs and peaks in the first derivative curve for valence indicating that when listeners perceive an increase in brightness and noisiness, they decrease in their rate of change of valence, and when they perceive a decrease in brightness and noisiness, they increase in their rate of change for valence. On the other hand, when the musical narrative is hopeful and joyous, a local trough in the spectral centroid and zero-crossing rate corresponds also to a local trough in the rate of change of perceived valence. For positively valenced sections, increases in brightness and noisiness lead to an increase in the rate of change of valence and vice versa.

Roughness also influences listeners' perception of arousal in complex ways. Peaks and troughs of roughness align with the troughs and peaks of the acceleration of arousal ratings in sections of the music that have a clear metrical component (Figure 3.11). Consonant tone combinations have been found to be perceived as positively valenced (Harrison & Pearce, 2020), so it is interesting to see a relationship of roughness with arousal rather than valence. Although modern Chinese orchestral music is very much influenced by Western classical music, it is not uncommon for harmony and its associated consonance and dissonance relations to function with different rules from that of Western common-practice harmony. For instance, depending on context, the dissonances caused by seconds (major or minor) might imply different qualities and/or quantities of intensity and therefore be associated with arousal. In the piece of music used in this study, the composer does not often invoke common-practice functional harmony in structuring his music. Listeners are likely to grasp a sense of harmonic style used here and have an expectation about how the harmonies function in this piece of music, which then serves to inform their affective responses. With harmony free from the connection to a major/minor tonality, consonance/dissonance implications for perceived valence are further removed. What remains then are the implications of roughness for how intense and/or high in arousal the music is.

Finally, modal strength demonstrates only a very slight relationship with perceived valence responses. This might be due to a few reasons. As the piece is written with modes inspired from the Arabic musical traditions, this could pose a problem with the tonality measure in MIRToolbox. Even though the computation of the frequency of each note with regards to a major or minor mode might be accurate, this might not be a true reflection of how listeners perceive the modality as the Arabic modes are fundamentally very different from the major/minor modes. In addition, listeners might also have been aware of the very different modal framework used in this piece and therefore were not consciously basing their judgments on how “major” or “minor” it sounded. The Arabic modes used in the piece have also been labelled as they occur over the piece of music, but it does not appear that there is any consistent relationship between the modes and listener responses—the same *maqāmāt* and *ajnās* occur without any discernable patterns in places with different valence and arousal content. The unfamiliarity of the Arabic musical tradition to these listeners meant that there were no learned implications of affective intentions of the Arabic modes available for them.

H4. Learned Cues Exert Influences On Both Perceived Valence and Arousal

Acoustic and musical features influencing listeners' perception of affective intentions were found to be related to familiarity with a musical tradition, indicating that learning plays a role in the cues that are available for listeners. Impulse response functions show a larger number of features influencing CHM responses to arousal than was the case for WM or NM. In addition to spectral slope, RMS energy, zero-crossing rate, roughness, and density influencing arousal responses for all three groups of listeners, spectral flatness, spectral flux, and spectral kurtosis also influence CHM's perceived arousal. Being familiar with the timbre of the instruments used and the stylistic features of this genre of music allowed this group of listeners to be more sensitive to the various nuances that contribute to the expression of particular affective intentions and therefore gave them more access to the culture-specific cues for decoding the intentions.

Along with learned cues that provide specific information regarding perceived affective intentions, learning likely also plays a role in narrative formation as "particular mappings between sound pattern and story are highly dependent on enculturation" (Margulis et al., 2019, p. 6). It can be seen here that there are complex interactions between acoustic features and affective, musical, and narrative content.

The three listener groups diverged in their affective responses in certain regions of the music. As mentioned previously, their divergences were more in terms of quantity than quality, but with CHM providing the most extreme responses followed by WM, and lastly NM (e.g., 152–352 s; 446–593 s; 634–777 s). The more extreme responses suggest that CHM are more confident in their responses as they have more information available to them for making judgments on affective intentions. WM, with knowledge in the Western classical music tradition, also had available to them more in-depth knowledge of musical structures, orchestration techniques, and so on, as compared to NM. This knowledge also provided them with a greater amount of information as compared to NM for making affective judgments.

The perception of affective intentions in music listening is therefore a complex process that is influenced by the degree of familiarity listeners have with a musical tradition, the narrative content implicated in the music, and the complex sonic environment created by the composer's work and the musicians' interpretation in performance. The perception of emotional intensity, arousal, and valence taps into different aspects of commonalities and culture-specific understanding of listeners. Several acoustic features for instance, influence the perception of emotional intensity and arousal more directly and similarly across all three listener groups, whereas others influence only CHM and not the other two listener groups. The more extreme and accurate valence responses of CHM when conpared with the composer's intentions also point towards the implications of culture-specific understanding.

H5. Highly Similar Emotional Intensity and Arousal Responses

A listener's perceived emotional intensity does not appear to reflect their perception of the distance away from the origin in the two-dimensional valence-arousal space, but rather, is more related to the single dimension of arousal. A high degree of similarity can be observed between the shapes of the emotional intensity and arousal mean curves, but not between the shapes of the emotional intensity and valence mean curves (Figures 3.2–3.4). Emotional intensity was not necessarily high when valence was either very positive or very negative. It is, however, always high when arousal is high and low when arousal is low. There is also considerable overlap in the features influencing listeners' perception of emotional intensity and arousal. All three listener groups use spectral slope, RMS energy, zero-crossing rate, roughness, and note onset density similarly in their perception of both emotional intensity and arousal. There was much less overlap in the features that influence emotional intensity and those that influence valence, again pointing towards the similarity in the dimensions of perceived emotional intensity and arousal.

Conclusion, Limitations, and Future Directions

Music listening involves not only numerous rapid and complex mechanisms in processing sonic materials, these real-time processes also usually occur and evolve over a certain period of time. This study tracks listeners responses and their changes over the course of a lengthy piece of music and explores some acoustic and musical features that might be implicated in the perception of affective intentions. Although there are several acoustic and musical features that appear to influence perceived affective intentions, many others do not show any direct relationships. Valence responses also appear to be less directly related to the acoustic and musical features explored in this study. However, when narrative and musical contexts are taken into consideration, a different pattern emerges. Listeners appear to take into account the affective and narrative content of what is happening in the music and respond in different ways to the accompanying acoustic features. With enculturation in a musical tradition, listeners are likely to be more sensitive to changes in the music that communicate particular affective intentions. However, even without having formal training in a particular musical tradition, there are enough shared cues that provide information for decoding. As espoused by the cue-redundancy model (Thompson & Balkwill, 2010), learning and experience in a musical tradition provide listeners with more available cues for musical communication. Perceived affective intentions in music are therefore influenced by many factors including listeners' experience of a musical tradition, the acoustic features present in the music, musical content such as harmony, tonality, rhythm, etc., and the narrative content of the piece. More importantly, these factors do not implicate the perception of affective intentions independently; complex interactions exist between these features, the musical and narrative context, and prior knowledge and experience that communicate nuanced affective intentions to a listener.

Emotional intensity and arousal responses appear to be elicited from more universal, common cues, than valence responses. When listeners are presented with an unfamiliar piece of music for the first time, the narrative content is likely not yet fully formed. They might therefore focus on acoustic and musical features that convey information for them to make decisions to the expressed affective intentions more directly. With an additional presentation of the music, there is likely to be a clearer expectation of not only the musical things to expect, but the narrative content of the music may also be more clearly formed. This may be even more prominent for music in which the narrative and programmatic content is especially emphasized, as in the case of the music used in this study. Valence especially appears to be a complex dimension. The narrative content likely influences how listeners perceive the valence of the music to a large extent.

This study examined a single piece of music, so although results appear to show that listeners do respond differently based on their musical backgrounds and that certain acoustic and musical features relate to perceived affective intentions, it is difficult to generalize to all music listening. A few different pieces of music, in different musical styles and traditions will have to be studied as well to see if these responses to acoustic and musical features can be generalized, although one might reasonably presume each piece of music to have its own specificities in terms of acoustic and musical features it employs.

While functional data analysis allows for curve estimation from discrete datapoints, as well as derivatives of this functional data, and therefore has the potential to study changes over time, it is more difficult to infer correlations between a set of curves in response to a single continuous variable. Therefore, although a detailed visual exploration reveals correspondences between acoustic and musical features and perceived arousal for instance, it is difficult to establish a reliable statistical inference on how the correlation of the features with perceived affective intentions changes over the course of the music. Time series analysis is used to supplement this analysis, and impulse response functions are used to visualize the influences of these acoustic and musical features. Although this approach sheds light on how listeners' perceptions react to the various acoustic and musical features, it provides a generalization over the entire length of music and is not able to show changes over the course of the music such as where the importance of a particular feature might change in different parts of the music. Other useful techniques of functional data analysis such as functional principle component analysis can also be used to further explore the complex relationships that could underlie the perception of affective intentions in continuous music listening.

This piece of music is intended to communicate a wide range of different affective intentions and narrative content. It appears that different acoustic and musical features are implicated in the sections that have different narrative content and affective intentions. Analyses exploring how listeners respond in these different sections will be helpful to further understand this complex process of music perception.

Music listening is a common human activity but at the same time, is a very complex behaviour. Increasing our understanding of how listeners perceive complex, lengthy, continuous sounds is important in untangling auditory perception and how acoustic and musical features are used. It also illuminates how learning plays a role in influencing the way listeners process sounds and how it can shape perception. In addition to understanding how humans process musical sounds, this can also illuminate the ways in which everyday sounds and language are processed, as well as how aspects of the sounds can convey affective intentions in human communication.

Acknowledgements

We thank composer WANG Chenwei for providing the music used for the experimental stimulus. This research was supported by grants to SMc from the Canadian Social Sciences and Humanities Research Council (895-2018-1023, 435-2021-0224) and the Fonds de recherche du Québec—Société et culture (2017-SE-205667), as well as a Canada Research Chair (950-223484). We also thank Bennett K. Smith for programming the experimental interface.

Part IV

A music narrative framework for affect perception in music listening

This chapter is based on the following research article:

> Heng, L., and McAdams, S. (2023). A music narrative framework for affect perception in music listening. [Manuscript in preparation].

Abstract

The perception of affective intentions when listening to a piece of music is influenced by numerous complex factors. A previous study (Heng et al., 2023) explored the dynamic processes involved in listening to a lengthy piece of Chinese orchestral music. Listeners were relatively accurate in judging the affective intentions when their valence and arousal responses were compared with the composer's intentions. However, the acoustic and musical features were not utilized in a simple, direct, and linear way. Listeners with different musical backgrounds (Chinese music, Western music, and nonmusicians) shared many features they used to decode perceived arousal but utilized features differently for perceived valence. This study therefore sought to examine separate sections of the piece with different perceived affective intentions, in order to obtain a clearer picture of how acoustic and musical features are utilized differently by listeners with different musical backgrounds and in sections of the music with different affective intentions. A music narrative framework that attempts to incorporate processes of appraisal in understanding and judging perceived affective intentions in music is proposed.

Keywords: affect perception, continuous music listening, narratives, acoustic cues, musical cues, cross-cultural.

Introduction

People regularly engage with music through listening, performing, or any other form of musical activity. Music has the potential to represent and express a variety of meanings. In instrumental music, consensus concerning these musical meanings is not straightforward. Although certain aspects of music appear to evoke very similar meanings in some people, others are less easily agreed upon. Patel (2010) rightly argues that we cannot simply understand music as we do another form of language. As a departure from language, rather than a particular sound or combination of sounds pointing towards a specific meaning, the qualities within the sound shaped by its articulation play a big role in the range of meanings they carry in music (Shepherd & Wicke, 1997). These qualities might include aspects carried in musical relations such as pitch height and contour, harmonic connotations of major/minor modalities and tension/resolution, metrical and rhythmic structures, as well as acoustic and timbral features. Musical meanings can be created in a variety of ways, including associative mechanisms to extra-musical materials, structural relationships within the music, physiological effects elicited by music, amongst others. For instance, a comprehension of the associations of particular sounds or sound sequences to external ideas and/or emotions provides a listener with musical meanings. An understanding of tonal processes and an expectation of what comes next also provides particular musical meaning. This means that there has to be an affective or intellectual understanding of the musical sounds and their implications. Such implications are not arbitrary. Within a musical tradition, or a certain style of music, there are particular constraints in which certain musical elements can be selected, used, and combined together. There is also a shared understanding of how certain musical features might be associated with extra-musical or intra-musical elements, and how the relationships with other features pan out over the course of a piece of music. This understanding occurs only when there is a knowledge of the style in question, whether implicitly or explicitly. In addition to the particularities of a certain musical style, there are also universal processes shared by humans because of the way the auditory system and the human brain are organized. The creation of musical meanings therefore involves the element of conscious decisions in selecting and organizing musical elements within a set of constraints and the comprehension of musical meanings occur when these implications are interpreted.

Among the many types of meanings music can generate, narrative content stands as a prominent and frequently occurring one. Humans have a proclivity to narrativize abstract stimuli. Instrumental music, as a vehicle of abstract sonic information, is a strong contender for narrativization. Margulis (2017) demonstrates the ease with which listeners generate narratives, especially so with more contrasting music that they are familiar with and enjoy. Margulis and colleagues also found that people generate similar stories to particular musical excerpts within the same culture, but less so between different cultures. This finding suggests that the narratives people form as they listen to music are not random, but reliably influenced by the music they heard (Margulis, Wong, et al., 2022). However, even though it is common for listeners, regardless of their culture, to generate narratives during music listening, the relationship between the characteristics of the musical sound and the narratives that are generated varies between cultures (Margulis et al., 2019).

Affects and emotions are another type of meaning music easily elicits. Affective intentions in musical communication cover all aspects of evaluative states and are an important and salient notion implicated in musical meanings. Huron (2001) believes that although there is no testable hypothesis of why music has such a ubiquitous presence in humans, there is suggestive evidence that music may be important for mood regulation and synchronization, and social bonding. It can rouse and pacify, as well as allows one to feel good. Schubert (2009) argues that the function of music is to produce dissociation or suspension of disbelief, such as inducing pleasure, demonstrating the importance of affect in music. This ability of music to represent or express affective intentions is likely one of the main reasons music motivates so many people to devote their time to it. This process of communication requires the element of expression, an intention to bring across a certain emotion, as well as the element of recognition, an understanding of the emotion that is being expressed. “Music’s potential to convey referential information is [also] separate from the issue of whether the music is the result of felt emotion or a sending intention or both” (Juslin & Timmers, 2010, p. 455). In other words, for communication to take place, the listener simply has to comprehend what the intended emotions are without having to actually feel them. Gabrielsson (2001) distinguishes between emotion perception and emotion induction in which induced emotions are listeners’ emotional response to music—an emotional reaction has taken place; and perceived emotions are mainly perceptual cognitive processes—listeners do not necessarily have to be affected by it, although the border between the two is somewhat blurred. Lennie and Eerola (2022) also believe that the “distinction between perceived and induced emotions could be as simple as assessing a stimulus’ goal-relevance” (p. 16). Listeners obviously are able to tell apart the feelings that are induced from those that they understand the music to be expressing. Although both perceived and induced affect can be influenced by knowledge and experience, they are likely to be implicated differently. Autobiographical memories, for instance, likely play a greater role in influencing the feelings induced in listeners when they hear a particular piece of music. Explicit knowledge in tonal relations, on the other hand, could inform perceived emotions more than induced ones.

Juslin (2013, 2019) provides a framework of various mechanisms influencing both induced and perceived emotions in music. Included in this framework are aesthetic judgments that integrate the other mechanisms and are affected by internal and external factors in the music. Social and cultural influences, prior experiences and knowledge, and contextual information within the music, as well as in the music listening process, are implicit in this mechanism. Cespedes-Guevara's (2021) constructionist model emphasizes the process of appraisal and attention in affective responses to music. Similarly, Lennie and Eerola (2022) propose the Constructivistly Organized Dimensional-Appraisal (CODA) model that also places importance on appraisal facilitating a "music-elicited emotional episode through the ongoing construction of relevance and meaning between an organism and their (musical) environment" (p. 14). Cespedes-Guevara and Eerola (2018) provide a constructionist account of attribution of emotional meanings to music. Drawing from various constructionist theories, they propose that the perception of emotions in music consists of an active process of meaning construction: musical acoustic cues signal variations in levels of arousal and valence which may become "differentiated and categorized into discrete meanings in a conceptual act [...] musical structures afford certain meanings to be privileged over alternative ones" (p. 13). They also believe that music perception might involve attributions of affect that could be mapped onto different possible meanings through associative mechanisms. It would appear then that listeners make sense of the music they hear in different ways depending on the context, and an awareness of the implications of a stimulus plays an important role in meaning-making. It is therefore possible that for a piece of music with an explicit narrative content and clear changes in affective intentions in relation to the content, listeners might utilize different strategies in their listening process. Their appraisal of the affective content expressed by the music is also likely to be influenced by the narrative content.

A Music Narrative Framework for Perceived Musical Affect

Even though inducing and perceiving affect in music likely involves many shared mechanisms and pathways, the pathways and mechanisms that are most active might be different. In affect perception, the focus or goal-orientation is less on the effects the music has on an individual's personal sensations and well-being, but more on what information the music is bringing across, what is being *represented* in and by the music. In this case, it follows that narratives that are created will be an important element for the perception of meanings and affective intentions.

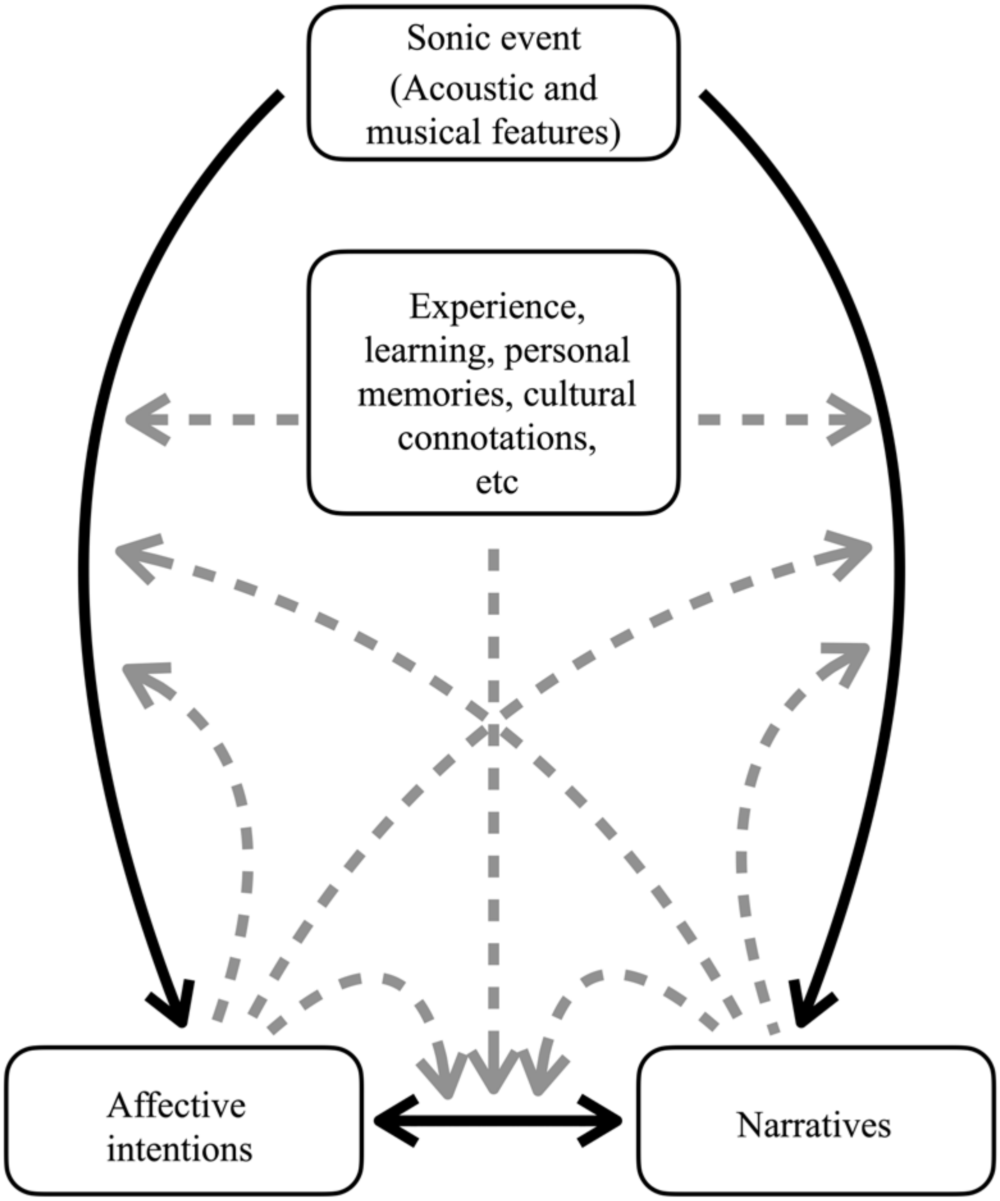
Margulis, Williams, et al. (2022) showed that listeners imagine narratives as they listen to the music rather than concoct the entire narrative after the music has concluded. This continuous and dynamic processing of narrative content means that there is a potential that different narrative content influences the appraisal of the musical affect differently, and that different acoustic or musical elements might be implicated in different ways for the appraisal.

Most studies on perception in music, however, have utilized only short sounds, or elicit only a single response for a longer section of music, and only a few studies have attempted to investigate the changes in listeners' perception of affective intentions over the course of a lengthy piece of music. Bailes and Dean (2012) studied the evolving pattern of listener perceptions in contemporary electroacoustic compositions and found that sound intensity partially explained perceptions of arousal, but enculturated experience and stylistic preferences likely influenced valence responses more. In another study, Dean and Bailes (2014) reported a reflection of the large-scale musical structure in nonmusicians' continuous perceptions of change, suggesting that "structural contrasts may be perceived even without great familiarity with the constituent musical components", and that "small-scale surface events (such as changes in acoustic intensity) can be perceived as continuous change" (pp. 107–108). They suggest that "the comparative analysis of different segments of the perception of affect in relation to concomitant continuous change, ... may in turn reveal distinctions in their qualitative and quantitative predictive relations" (p.108). This continuous tracking of listener responses and a comparison of how their responses differ in regions of music with varying narrative content will also help illuminate this process of affective perception more clearly.

To clarify this process of affective perception in music, we can benefit by incorporating a framework that integrates narrative formation, expectations, and appraisal. Figure 4.1 summarizes this proposed framework.

The perception of a sonic event involves several appraisal mechanisms. Initial affective intentions might be elicited based on appraisal of the surface acoustic and musical features. This is likely to be modulated and/or mediated by individual

Figure 4.1
Music Narrative Framework



Note. Black solid lines indicate appraisal mechanisms, grey dashed lines indicate influences on appraisal mechanisms.

experiences, personal memories, learning, and socially and culturally learned knowledge. These features might also generate narrative content that is similarly influenced by prior experience. The narratives created might provide additional information to influence the appraisal of perceived affective intentions and vice versa. Each of these may also form feedback loops in which they influence the appraisal mechanisms—narrative content that is formed might feed back into the appraisal of the sonic event and influence new narratives; affective intentions that are formed might feed back into the appraisal of the sonic event to influence further affective intentions that are perceived. They also feed forward to each other—narrative content formed could feed into the appraisal of the sonic event to influence perceived affective intentions; affective intentions perceived could feed into the appraisal of the sonic event to influence the creation of new narratives.

Acoustic and Musical Cues in Perception of Affective Intentions

Gabrielsson and Juslin (2003) reviewed studies on perceived emotional experiences in music and summarized the variables related to perceived emotional expression. They found tempo, mode, rhythm, melodic shape, and harmony to be among the most often studied features. However, in addition to these features, several aspects of timbre have been found to be important to the communication of affective intentions in music as well. In their review of the expression and communication of emotion in music, Juslin and Timmers (2010) also found that cues such as tempo, sound level, timing, intonation, articulation, timbre, vibrato, and pauses are most frequently studied.

A recent study by Micallef Grimaud and Eerola (2022) investigated a variety of musical and acoustic cues for their contribution to emotional expression in music. In addition to a perceptual task, they also had participants perform a production task in order to explore whether there are differences in these two approaches in terms of how musical affect might be communicated. Although both approaches yielded mostly similar results, they emphasized the importance of exploring both production and perception approaches in the study of musical affect because differences in control might mean an emphasis on certain cues in production but less emphasis in perception and vice versa. Even so, there is evidence that the communication of musical affect can be relatively successful. Following Balkwill and Thompson's (1999) cue-redundancy model, that "there is often redundancy in how a piece reinforces a specific emotion" (p. 45), this could be due to the redundancy of musical cues that provide enough information to listeners even though performers and listeners may not always place equal emphasis on the same cues. In their discussion of future studies to investigate narrativization in music, Margulis and colleagues (2019) suggest that it might be worthwhile to also explore cultures defined in other ways than geography. If listeners from the same geographical region, who are relatively similar in their socio-cultural backgrounds but trained in different musical traditions, demonstrate differences in their perception of affective intentions in music, then particular musical cultures might place different emphases or have different sets of rules in utilizing and organizing acoustic and musical elements.

The continuous response of listeners as they listen to a piece of Chinese orchestral music was explored in a previous paper (Heng et al., 2023). The piece of music utilizes Chinese instruments but was written with modes and metrical concepts borrowed from Arabic music and a programmatic content based on Aleppo and the Syrian war. It has a rich narrative content that is explicitly detailed by the composer, Wang Chenwei. The content is reflected in the performance indications, expressed through the compositional process, and shaped in performance. From the listeners' responses, it appears that sections of this piece of music that are explicitly described by the composer as having particular narrative content are more or less accurately associated with listeners' perceived affective intentions.

Functional data analysis was used to explore differences in the perception of affective intentions in music by listeners with different musical backgrounds from the same geographical location: Chinese musicians, Western musicians, and nonmusicians (CHM, WM, and NM, respectively), all from Singapore. Participants responded continuously to the music in real time, rating their perceived emotional intensity on a unidimensional scale in a first listening, and then their perceived arousal and valence on a two-dimensional space in a second listening. The shapes of arousal and emotional intensity responses were very similar to each other, but differed for valence responses. Impulse response functions modelled over the entire piece of music revealed particular acoustic and musical features influencing listeners' perception of affective intentions. All listeners appear to utilize RMS energy, roughness, note onset density, spectral slope, and zero-crossing rate in their ratings of emotional intensity and arousal. In addition, CHM also appear to utilize spectral flatness, spectral flux, and spectral kurtosis for their arousal ratings. Valence ratings however, only appear to be weakly influenced by a couple of features, which vary across the listener groups and explain less variance than the features predicting arousal ratings.

Although modelling impulse response functions of acoustic and musical features with listener responses over the entire piece of music demonstrates the possible influence of some features, a visual study of the functional plots of these features with listener responses reveals additional differences in the relationship of the features and listener's affective responses in different regions of the piece. The peaks and troughs of spectral centroid and zero-crossing rate for instance align with the peaks and troughs of the rate of change of listeners' valence responses in some regions of the music, and with troughs and peaks in others. Music analysis reveals that these different regions have different affective intentions and narrative content. In order to examine how acoustic and musical features might be implicated differently in sections with different narrative and associated affective content, listeners' average responses for sections of music expressing high-arousal/negative-valence, low-arousal/negative-valence, low-arousal/positive-valence, and high-arousal/positive-valence were analyzed separately, and compared with the corresponding acoustic and musical features of that section. Demonstrating the ways different features are implicated in each of the sections will suggest that listeners' perceptions could be influenced by their appraisal of the intended affective intention. Among other things, this appraisal might arise from narratives listeners form.

This paper therefore expands from the previous paper (Heng et al., 2023) by exploring each of the sections with different affective intentions. Impulse response functions to show acoustic and musical features being used differently in sections with different affective intentions and narrative content can corroborate what was observed in the visual exploration. As the emotional intensity ratings appear to be very similar to those of arousal, this paper will only explore the data from listeners' perceived valence and arousal responses with respect to sections of the piece with different narrative intentions and associated affective intentions. A close analysis of the acoustic and musical features implicated in each of these different sections, together with a study of the compositional techniques used might help to uncover how these are utilized as listeners respond to music with varying affective and narrative intentions. In addition to studying the correspondence between acoustic and musical features with listeners' perceived affective responses, an exploration of the variations in features, as well as an analysis of the music also provide insights into how the acoustic and musical cues "produced" by the composer and performers are used in the creation of narratives and affective intentions.

Method

This study provides a selective and more detailed analysis of part of the same dataset that was collected in a previous study (Heng et al., 2023). For clarity, the procedure for data collection will be outlined here. For details, refer to Heng and colleagues (2023).

Participants

Three groups of 30 participants with different musical backgrounds were recruited from Singapore. The first group were musicians trained for an average of 11 years in the Chinese music tradition (henceforth CHM). The second group were musicians trained for an average of 12 years in the Western music tradition (henceforth WM). The final group were nonmusicians with less than one year of training throughout their lives (henceforth NM). All the participants self-reported no problems with hearing. They read and accepted an online written consent form and were compensated for their participation. This study was certified for ethical compliance by McGill University's Research Ethics Board II.

Stimuli

A Chinese orchestral work composed by Wang Chenwei (2021) was used for the listening experiment. This work uses Chinese orchestral instruments but borrows modes and metrical concepts from Arabic music and has a programmatic content based on Aleppo and the Syrian war. This piece was selected for a few reasons. As the experiment aims to elicit perceived affective intentions from participants, selecting a piece that has a rich programmatic and narrative content is justified. The composer describes the narrative content intended for each section in detail, and very explicitly explains how he attempted to express the intentions. Very clear affective intentions are also associated with these narratives that occur through the piece of music. Listeners with expertise in the Chinese music tradition should be more familiar with the instrumental timbres and orchestration styles. Having formal musical training means that CHM and WM are likely to have internalized a repertoire of formalized musical rules and commonly used tropes. Finally, none of the participants reported having heard this piece of music before. The recording was provided by the composer with a sample rate and bit depth of 48kHz/24-bit (wav, stereo). The entire piece of music is 930 seconds in duration.

Four sections in which listeners average ratings fall primarily within one of the four quadrants in the arousal-valence space were selected for detailed analysis The third section is titled “Bustle of the Souq” and the associated affective intention is clearly animated, joyous, and lively, a high arousal and positive valence (H+). The next two sections “Flames of Resentment” and “Crossfire of the Armies” have the associated affective intention of high arousal and negative valence (H–). The sixth section “Tears of the Rubble” expresses low arousal and negative valence (L–), and the final section “Path to the Future” expresses low arousal and positive valence (L+).

Background of Musical Styles Used in Study

The Chinese Orchestra. The Chinese orchestra is a relatively modern creation that utilizes traditional Chinese instruments and combines a composition aesthetic derived from the rich folk and traditional musics around China, as well as composition and orchestration techniques borrowed from Western orchestral music.[2]

Maqām and jins in Arabic music. The Arabic *maqām* is a system of scales, melodic phrases, modulations, and ornamentations (Farraj & Shumays, 2019). A *jins* (plural *ajnās*) is a scale fragment of a *maqām* (plural *maqām’āt*) and the basic melodic unit (Farraj, 2018). Each *jins* conveys a distinct mood or character. Although the *jins* possess affective content, this affective content might mean something very different to different listeners ((Farraj & Shumays, 2019)). For the purposes of this piece, the composer has specified that a quarter-tone interval approximation is sufficient. Although this may not be an accurate reflection of Arabic

[2] A list of commonly used Chinese orchestral instruments, their characteristics, and some orchestra examples can be found in this link: https://timbreandorchestration.org/resources/instruments/ensembles/chinese-orchestra

musical practices, the intention of the work is not an imitation but an original composition inspired from Arabic musical elements.

Iqā' in Arabic music. *Iqā'* (plural *iqā'āt*) is used to describe a rhythmic cycle. *Iqā'āt* are made up of two different basic building blocks, the dum and tak, onomatopoeias derived from the sound produced on membranophones. Similarly to the *maqāmāt* and *ajnās*, while the composer has incorporated some specific *iqā'āt* in the work, it is not meant to be a true indication of, but rather an inspiration from, Arabic musical practices.

Procedure

Participants were given a unique link and accessed this browser-based experimental task online. During the instruction phase, the experimental interface was introduced, with a unidimensional scale for emotional intensity and a two-dimensional scale for valence and arousal. Participants had to respond by moving their cursor continuously throughout the entire piece of music. All participants rated the emotional intensity on the first presentation of the piece of music and valence-arousal on the second. The entire experiment took between 40 minutes to an hour for all the participants.

Data Analysis

Acoustic and musical features of the music were analyzed using the revised Timbre Toolbox (Peeters et al., 2011; revised by Kazazis et al., 2021) implemented in the MATLAB environment (Mathworks, Natick, MA). The acoustic descriptors selected for analysis were spectral centroid, spectral crest, spectral decrease, spectral flatness, spectral flux, spectral kurtosis, spectral roll-off, spectral skewness, spectral slope, spectral spread, spectral variation, RMS energy, and zero-crossing rate. A short-term Fourier transform with the power spectrum estimation was applied to the audio file. A Hann window of 50 ms duration with a 50% overlap between successive frames was used. MIRToolbox's roughness measure was used to estimate consonance, the mirmode function was used to estimate the modal strength of the music (Lartillot, 2022), and the onset detection function from a VAMP plugin created by the Centre for Digital Music at Queen Mary University of London for SonicVisualiser (2010) was used to estimate note density. All statistical analyses were performed using R statistical software (Version 4.2.0; R Foundation for Statistical Computing, Vienna, Austria).

Time Series Analysis and Impulse Response Functions

Participant responses were collected sequentially in time and are likely autoregressive in nature. Time series analysis is an appropriate technique for exploring the relationship of acoustic and musical features with perceptual responses. Responses are averaged over each listener group, CHM, WM, and NM, for emotional intensity, valence, and arousal.

In the previous paper (Heng et al., 2023), impulse response functions were used to explore acoustic and musical features that influenced listeners' responses over the entire piece of music. A few features such as note onset density, RMS energy, roughness, and spectral slope, were found to consistently influence listeners' perceptions. Listeners perceived an increase in arousal with an increase in RMS energy for instance, and this holds true for this particular acoustic feature, across the entire piece of music regardless of the type of affective intention. What is less certain, however, is whether the other features that do not appear to influence listeners' perceptions of affective intentions when taken over the entire piece of music would in fact contribute significantly to particular affective intentions but not others. Functional data analysis showed descriptively that certain acoustic and musical features do not relate to perceived affective intentions in the same directions in the different sections of the music that have different affective intentions. In order to clarify this issue, listeners' responses over each selected section of music need to be taken as separately determined time series.

The piece of music used in this study has a rich programmatic and narrative content. It is divided into seven sections, each with a clear explicit description the intended narrative by the composer, which also corresponds to specific types of affective intentions. From the listeners' arousal and valence responses, these sections are perceived relatively accurately. Most of the responses are within the same affect quadrant, and the differences are more in the extremity of the responses—CHM gave the most extreme responses, followed by WM, and NM's responses were the least extreme. Four sections were selected and the observations for the participants of each listener group were averaged and treated as a unique time series.

At each moment in time, the perceptual responses of listeners are likely influenced by preceding moments. Differencing renders the processes more stationary and reduces the degree of autocorrelation (Box et al., 2016). The differencing process essentially takes the change from one moment to the next and uses this rate of change in statistical inferences. Listeners' responses are taken as endogenous variables and the acoustic and musical features as exogenous variables. The endogenous variables are stationarized and tested with the Augmented Dickey-Fuller Generalized Least-Squares test and the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test for stationarity. If they did not show stationarity after one differencing step, a second differencing step was performed and the differenced version was tested for stationarity again. The exogenous series were then transformed in the same way as each endogenous series against which they were to be compared.

Vector autoregressive (VAR) models are used to capture the dynamic properties of variables (Lütkepohl, 2005). Given that it is not clear which set of impulse responses actually reflect the behaviour in a given system, structural vector autoregressive (SVAR) models apply restrictions that allow the identification of impacts to a system, and unique impulse responses can be derived for each shock to a system. In order to explore how each of the acoustic and musical features influence listeners' perceptions of affective intentions, these impulse response functions derived from the SVAR models for each of these features were obtained for each of the four sections of music.

Results

The previous paper detailed the impulse response functions of listeners' perceived emotional intensity, arousal and valence for each acoustic and musical feature over the whole piece. Spectral slope, RMS energy, zero-crossing rate, roughness, and note onset density were found to influence arousal responses of all three groups of listeners. In addition, spectral flatness, spectral flux, and spectral kurtosis also affected CHM perceived arousal responses. On the other hand, there were less features that relate to listeners' perceived valence responses. Spectral centroid and zero-crossing rate influence CHM valence responses, spectral slope and RMS energy influenced WM, and zero-crossing rate and note onset density influenced NM.

Derivatives obtained from the functional data revealed relationships that varied in different directions in different sections of the music. In regions that are negatively valenced, the peaks and troughs of rate of change of perceived valence matched the troughs and peaks of spectral centroid and zero-crossing rate curves, and in regions that are positively valenced, peaks and troughs vary in the same direction. In regions that have a clear sense of meter, peaks and troughs of the second-order derivative of perceived arousal match the troughs and peaks of the roughness curve, but no relationship is observed in the sections that are very slow or metrically free. As it appears likely that perceived affective intentions vary in different ways with acoustic and musical features depending on what is happening in the music, sections with different affective intentions will be analyzed separately in this paper.

Impulse response plots for the influence of each acoustic and musical feature on emotional intensity, arousal, and valence for each group of listeners are now modelled for each of the four sections of music independently. As an example, Figure 4.2 shows one such impulse response plot for RMS energy on arousal responses for CHM in each of the excerpts corresponding to one of the four affective quadrants. The y-axis shows the amount of change in the endogenous variable in response to a unit of change in the exogenous variable. The x-axis shows the number of lags (each lag is 0.5 s) taken for the reaction to reach a certain level. The red dashed lines plot the 95% confidence intervals for the functions. Only impulse responses with their 95% confidence intervals breaching zero are considered to significantly influence listeners' responses.

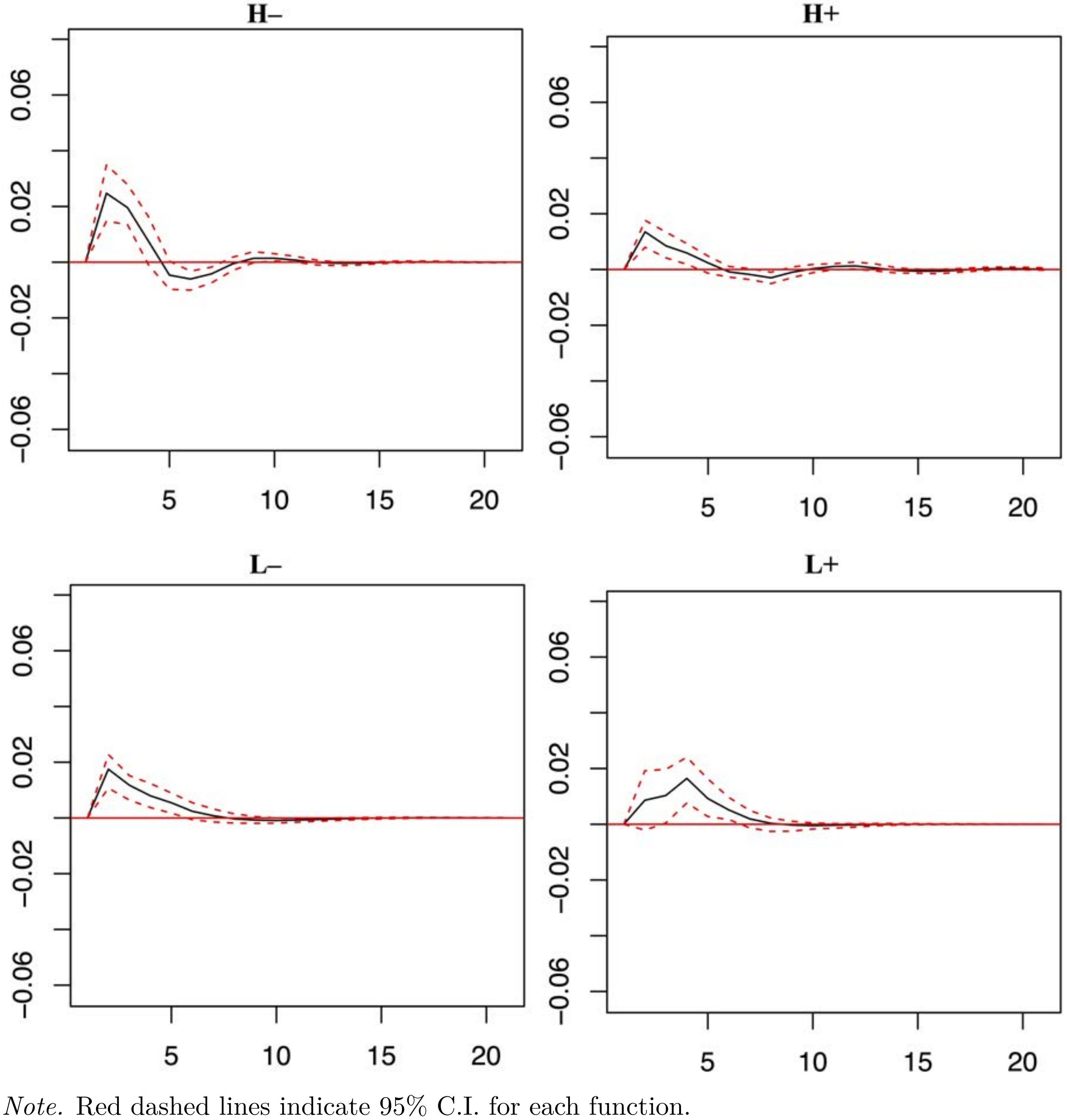
In all the tables that follow, a positive sign indicates that when values of a feature increase, listeners' ratings on the particular affective scale rise, whereas a negative sign indicates an increase in the value of a feature leading to a decrease in ratings. −/+ indicates a fluctuation from negative to positive, +/− indicates a fluctuation from positive to negative, suggesting a flux in the values that correspond to an increase in ratings. Square brackets indicate a peak response of less than ±0.01 and are treated as contributing only a marginal influence. There is evidence that listeners are able to perceptually differentiate a variety of different audio descriptors (e.g., McAdams, 2019; Stilp et al., 2010), although some features might be perceptually more dominant than others (Kazazis, 2020). Acoustic features also do not all influence perceived affect in direct relationships, but may combine in various ways and in different directions (e.g., Eerola et al., 2012; Heng & McAdams, 2022; Heng, et al., 2023). Therefore it is not surprising that close to 85% of the significant impulse response functions have a peak response of less than ±0.01—many features on their own contribute only marginally to perceived affective intentions.

Arousal

Table 4.1 shows the acoustic and musical features that significantly influence CHM, WM, and NM perception of arousal for each section with the corresponding affective intention. RMS energy can be seen to play a role in all four affective

Figure 4.2

Impulse Response Functions of RMS Energy with Arousal Responses of Chinese Musician Listeners in the Four Affective Quadrants



Note. Red dashed lines indicate 95% C.I. for each function.

quadrants for CHM, although its relation is somewhat nuanced. Higher RMS energy relates to higher arousal ratings in the H+, L–, and L+ affective intentions, but in H–, it is the aspect of energy fluctuation—dynamics that first increase, then immediately decrease—that relates to higher arousal ratings. RMS energy features in three of the affective intentions for WM and only two for NM, and for these two other groups of listeners, the relationship is more straightforward—an increase in RMS energy relates to an increase in arousal ratings. Note onset density also figures prominently here. In WM, note onset density appears to influence the perception of arousal in all affective intentions. With the exception of H–, a higher density relates to higher perceived arousal. In H–, however, WM appear to be influenced by the changes in density. Note onset density is significant in H–, H+, and L– for CHM; in H– and H+, it is the changes in density that relate to perceived arousal, whereas in L–, a denser texture is related to an increase in arousal ratings. Note onset density plays a role for NM only in the H– and L– affective intentions, with a higher density signifying a higher perceived arousal in both excerpts.

Table 4.1

Acoustic and Musical Features Influencing Listeners’ Perception of Arousal

CHM	
H– (“Flames of Resentment” and “Crossfire of the Armies”)	**H+ (“Bustle of the Souq”)**
Spectral flux + Spectral roll-off [+] RMS energy +/– Zero-crossing rate + Note onset density –/+	Spectral flatness – Spectral flux [–] Spectral spread – RMS energy + Zero-crossing rate [+] Note onset density –/+
L– (“Tears of the Rubble”)	**L+ (“Path to the Future”)**
Spectral flux [+] Spectral slope – Spectral spread [+] RMS energy + Zero-crossing rate [+] Note onset density [+]	RMS energy +
WM	
H– (“Flames of Resentment” and “Crossfire of the Armies”)	**H+ (“Bustle of the Souq”)**
Spectral flux [+] Note onset density [–/+]	Spectral flux [–] Spectral roll-off [+] RMS energy [+] Note onset density [+]
L– (“Tears of the Rubble”)	**L+ (“Path to the Future”)**
Spectral kurtosis [+] Spectral skewness [–] Spectral spread [+] RMS energy [+] Note onset density [+]	Spectral slope [+] RMS energy [+] Mode [+] Note onset density [+]

Table 4.1 (cont.)

Acoustic and Musical Features Influencing Listeners' Perception of Arousal

NM	
H– ("Flames of Resentment" and "Crossfire of the Armies")	**H+ ("Bustle of the Souq")**
RMS energy [+] Note onset density [–/+]	None
L– ("Tears of the Rubble")	**L+ ("Path to the Future")**
Spectral slope [+] Note onset density [+]	Spectral centroid [–] Spectral slope [+] RMS energy [+]

The other features are less consistent amongst the listener groups and affective intentions. Spectral flux is implicated in the negatively valenced sections for the two musician groups. In both H– and L– sections, a greater spectral change informs the musician listeners of a higher arousal. Additionally for both musician groups, lower spectral flux indicates higher arousal if it is found in the H+ section. Spectral flux, however, is not seen to be related to NM's perception of arousal.

Spectral roll-off can be seen to influence both musician groups as well, although it is implicated in different sections. For CHM, a higher spectral roll-off, indicating more richness and brightness, is rated as higher in arousal if it happens in the H– section. WM, on the other hand, judge a higher spectral roll-off as higher in arousal in the H+ section.

Spectral spread provides information about the spread of the spectrum around its mean value, indicating richness of a sound. A greater spread is perceived as a higher arousal for both musician groups in the L– intention, and less spread is perceived as higher in arousal in the H+ affective intention for CHM.

Modal strength, an indication of how major a section of music is, appears to influence only WM, and only in the L+ intention. The more major the music sounds, the higher WM perceive the arousal to be.

Spectral slope measures the amount of decrease of the spectrum toward higher frequencies and provides information on the brightness—the steeper the spectral slope, the lower the high-frequency energy. Spectral slope might influence the richness of a sound as well. This feature is implicated only in the L– affective intention for CHM, and a richer and brighter sound is rated as higher in arousal. Spectral slope is also implicated in the L– intention for NM, but in contrast to CHM, they perceive a richer and brighter sound as lower in arousal. In addition, spectral slope influences both WM and NM in their arousal responses for L+. A thinner, darker sound is perceived as higher in arousal when it is heard in the L+ section for WM and NM.

Valence

Table 4.2 shows the acoustic and musical features influencing perceived valence for the three groups of listeners. As compared to arousal, there are fewer features that appear to play a statistically significant role. For CHM, spectral variation and note onset density play a role in the H– affective intention, and RMS energy and zero-crossing rate in the H+ affective intention. No features were found to relate significantly for the two low-arousal affective intentions. WM utilize RMS energy, zero-crossing rate, and note onset density in their valence responses for the H– affective intention, spectral roll-off and note onset density for H+, and spectral roll-off for both the low-arousal affective intentions. Finally for the NM, spectral kurtosis, spectral slope, RMS energy, and note onset energy are implicated in the H– affective intention, spectral centroid in H+, note onset density in L–, and spectral flux in L+.

Table 4.2
Acoustic and Musical Features Influencing Listeners' Perception of Valence

CHM	
H– ("Flames of Resentment" and "Crossfire of the Armies")	**H+ ("Bustle of the Souq")**
Spectral variation [+] Note onset density [–]	RMS energy [+] Zero-crossing rate [+]
L– ("Tears of the Rubble")	**L+ ("Path to the Future")**
None	None
WM	
H– ("Flames of Resentment" and "Crossfire of the Armies")	**H+ ("Bustle of the Souq")**
RMS energy – Zero-crossing rate [–] Note onset density [–]	Spectral roll-off [+] Note onset density [+]
L– ("Tears of the Rubble")	**L+ ("Path to the Future")**
Spectral roll-off [–]	Spectral roll-off [–]
NM	
H– ("Flames of Resentment" and "Crossfire of the Armies")	**H+ ("Bustle of the Souq")**
Spectral kurtosis [+] Spectral slope [–] RMS energy [–] Note onset density [+]	Spectral centroid [–]
L– ("Tears of the Rubble")	**L+ ("Path to the Future")**
Note onset density [+]	Spectral flux [+]

Note onset density, reflecting the density of the musical texture, appears to be important in the high-arousal/negative-valence affective intention for all three listener groups. However, the direction of influence is different between the musician and nonmusician listeners. CHM and WM increase their valence ratings when the texture becomes less dense, whereas the opposite happens for NM.

Spectral roll-off, providing aspects of information about the richness and centricity of the spectrum, appears to play an important role for WM, appearing in H+, L–, and L+ affective intentions, but is not implicated for the other two listener groups. Within WM, the direction of spectral roll-off's influence is also different in high- versus low-arousal affective intentions. In a high-arousal/positive-valence context, a brighter sound could indicate a more positive valence, whereas in the low-arousal contexts, a darker sound appears to indicate a more positive valence.

RMS energy can be seen to influence the perception of valence in H– for both WM and NM where a softer sound indicates a more positive valence. RMS energy is not implicated in the H– intention for CHM, but it relates to their valence ratings for H+: a louder sound implies a more positive valence.

Zero-crossing rate, as an indication of noisiness, is found to be related to valence responses in the H+ affective intention for CHM, and the H– intention for WM. A noisier sound is perceived as more positively valenced when the affect is high arousal and positively valenced for CHM. For WM, a noisier sound is perceived as more negatively valenced when it is heard in the H– section.

Spectral variation and spectral flux, both measurements of the change in spectral shape, are seen to influence CHM's perception of valence in H– and NM's perception of valence in L+. Musical sounds with more dynamically changing spectra are perceived as more positively valenced in H– for CHM and in L+ for NM.

Finally, there also appear to be many more features implicated for the H– affective intention for NM than for the two musician groups. Spectral kurtosis, a measure of the spectral shape influences how rich a sound is, and the thinner the sound, the more positively valenced NM rate the H– section. Spectral slope indicates how quickly the higher partials fall off, which also contributes to the richness of a sound, as well as its brightness. In a H– section, the less rich and bright a sound is, the more positively NM will rate the valence. Spectral centroid is related to the auditory brightness of a sound and is also a salient timbral attribute widely found in many perceptual studies. However, in terms of affective implications, it appears to influence only NM in a H+ section of music. The darker the sound, the more positively valenced NM will rate the music to be.

Discussion

It is common for people to form narratives when they listen to music. In this piece, in which the composer explicitly intends a particular narrative content, it is likely that he also utilizes certain techniques to ensure that these narratives are easily elicited. Although the study requires participants to rate their perceived arousal and valence with respect to the music, the corresponding acoustic and musical features that influence their responses may also be influenced by the narrative content of the music.

Acoustic and Musical Features with Affective Responses in Different Regions

If the music narrative framework shown in Figure 4.1 holds true, the same acoustic and musical features might elicit different affective intentions because of influences of narrative content on the appraisal process. Feedback processes also mean that it might be worthwhile to consider the features for valence and arousal in combination. Even though an impulse response function demonstrates that a particular feature likely influences the perception of arousal, for instance, feedback processes may mean that it is also implicated in valence responses. Considering features for valence and arousal in combination, therefore, might provide a fuller picture of what might be implicated when listeners perceive particular affective intentions. Table 4.3 consolidates these features together. Although one might theorize that perceived affective intentions could also exert influence on the appraisal mechanisms in narrative formation, this is not demonstrable in this study as participants were not asked explicitly about any narratives they formed during the listening process.

In his synopsis of the piece of music, the composer provided a very detailed outline of the narrative content of this work. Seven major sections are delineated with contrasting narrative content. The piece opens with “Hymn from the Ancients”, a quiet, sparse plaintive melodic line taken from the 1400 BC Hurrian Hymn No. 6 that was found on a cuneiform tablet excavated near Aleppo. The next section labelled “Citadel on the Hill” is a grand orchestral tutti that depicts the Citadel of Aleppo, symbolic of the city’s political might. Even though the participants were not given any information, the characteristics of the musical sound provided some cues to this

Table 4.3

Combined Acoustic and Musical Features Influencing Listeners' Perception of Arousal and Valence

CHM	
H– ("Flames of Resentment" and "Crossfire of the Armies")	**H+ ("Bustle of the Souq")**
Spectral flux + Spectral roll-off [+] RMS energy +/– Zero-crossing rate + Note onset density –/+ Spectral variation [–] Note onset density [+]	Spectral flatness – Spectral flux [–] Spectral spread – RMS energy + Zero-crossing rate [+] Note onset density –/+ RMS energy [+]
L– ("Tears of the Rubble")	**L+ ("Path to the Future")**
Spectral flux [+] Spectral slope – Spectral spread [+] RMS energy + Zero-crossing rate [+] Note onset density [+]	RMS energy +
WM	
H– ("Flames of Resentment" and "Crossfire of the Armies")	**H+ ("Bustle of the Souq")**
Spectral flux [+] Note onset density [–/+] RMS energy + Zero-crossing rate [+] Note onset density [+]	Spectral flux [–] Spectral roll-off [+] RMS energy [+] Note onset density [+]
L– ("Tears of the Rubble")	**L+ ("Path to the Future")**
Spectral kurtosis [+] Spectral skewness [–] Spectral spread [+] RMS energy [+] Note onset density [+] Spectral roll-off [+]	Spectral slope [+] RMS energy [+] Mode [+] Note onset density [+] Spectral roll-off [–]

Table 4.3 (cont.)

Combined Acoustic and Musical Features Influencing Listeners' Perception of Arousal and Valence

NM	
H– ("Flames of Resentment" and "Crossfire of the Armies")	**H+ ("Bustle of the Souq")**
RMS energy [+] Note onset density [–/+] Spectral kurtosis [–] Spectral slope [+] RMS energy [+] Note onset density [–]	Spectral centroid [–]
L– ("Tears of the Rubble")	**L+ ("Path to the Future")**
Spectral slope [+] Note onset density [+]	Spectral centroid [–] Spectral slope [+] RMS energy [+] Spectral flux [+]

Note. A positive sign indicates that an increase in the value of the feature leads to an increase in the perceived affect of the affective context it falls within.

narrative and historical content. The following sections present a descriptive analysis of the musical elements that develop throughout the piece. For clarity, this link provides a video showing listeners' valence and arousal responses with annotations of the associated *maqāmāt* and *iqā'āt* over the course of the music: https://mpcl.music.mcgill.ca/supplementaryMaterials/HengThesis2023/VA_modes_rhyt.mp4.

High-arousal/positive-valence

The descriptive title provided by the composer for the high-arousal/positive-valence section reads "Bustle of the Souq", the third section in this piece. Transitioning from the previous grand and majestic section into something with a livelier pace is about 20 s of constant plucked eighth notes that change into sixteenth notes. Higher strings gradually enter with tremolos followed by the tambourine and handdrum, increasing the activity level. This serves to bring an awareness of an increase in activity to listeners, even before the following section begins. "Bustle of the Souq" opens with only the tambourine and handdrum, playing the *iqā'* or rhythmic cycle, Yuruk Sama'i. This is characterized by a regular pattern of six quarter notes and has a dance-like lilt to it. The section prior is titled "Citadel on the Hill" and the music is more stately and majestic with a metrical pattern using the *iqā'* Muhajjar that repeats in a lengthy cycle of 14 beats. This shorter, more repetitive, and dance-like rhythmic pattern in "Bustle of the Souq" that contrasts with the previous stately rhythm also contributes to communicating a livelier and more animated atmosphere. A primary instrument in this piece is the *sanxian*, a three-stringed lute with the bridge sitting on a stretched snakeskin. With the first entry of the *sanxian* melody in this section, the composer uses the *maqām* 'Ajam which, to listeners who are unfamiliar with Arabic but familiar with Western modes, sounds to be in a major key. Margulis, Miller, et al. (2022) found change, departure, and constrast to constitute a salient dimension in narrative responses. The contrast created by the composer therefore signals a change in the narrative content, as well as providing indication of some potential narratives (e.g., this is no longer something majestic; it might be lighter, less serious in nature, etc.).

Only spectral centroid appeared to play a small role in NM perception of the H+ affect. More features are involved for the two musician groups. Lower spectral flux and higher RMS energy inform these two groups of musician listeners similarly of a higher arousal with a more positive valence. Note onset density also appears to play a role, although for CHM, it is the changes in density, and for WM it is a higher density that influences the perception of higher arousal and more positive valence. Spectral flatness, spectral spread, and zero-crossing rate are also involved for CHM—noisier, and possibly richer sounds suggest higher arousal and more positive valence. In addition to less spectral changes, louder sounds, and a denser texture, brighter sounds also appear to inform WM of this affective intention. This is in contrast to NM who interpret a darker sound as H+. There are more overlaps in the features used by the two musician groups in this affective intention. Musical training could be influential in providing certain types of information that aid appraisal processes for this affective intention, which is reflected in the shared features that relate significantly to CHM and WM perception here.

High-arousal/negative-valence

This H– section follows directly from the H+ section. This affective intention lasts over the next two sections titled by the composer: "Flames of Resentment" and "Crossfire of the Armies". Even before the end of the H+ section, the composer introduced several new *maqām'āt*, some of them with pitches outside the 12-tone equal temperament these groups of participants are familiar with, and others with intervals that sound more "foreign" such as augmented seconds. These likely signal the start of change for listeners, and in the final 5 s of the H+ section, the composer overlays the *jins* Sabā Zamzam with Ḥijāzkar to create an atonal scale. All the instruments play in two unison groups a third apart with a large crescendo swell, providing a very clear indication of a change in valence for the listeners. The composer clearly delineates this section titled "Flames of Resentment" from "Bustle of the Souq" that came before with a loud punctuated chord that uses the entire orchestra to clearly distinguish the two sections. With this, it is likely that new narrative content is formed in contrast to that of the previous section. The first part of this H– section alternates between unison in the winds and a sparser texture with *sanxian* melody. The use of *jins* Sabā Zamzam with its characteristic augmented second interval and Nikriz, which sounds similar to a minor mode, gives this section a darker connotation, especially for listeners familiar with the tonality system of Western music. The composer's choice of using a large group of *suonas*, double-reed Chinese wind instruments, means that the timbral universe of the *suonas* dominates this unison section, even as the higher-pitched and brighter sounding *dizis* (Chinese bamboo flute) augments them and adds a certain harshness and brightness to the sonority. Interestingly, even though all the instruments are playing the same pitch class at varying octaves, this section sounds dense rather than sparse or open. This could be due to the timbre of the instruments involved and a combination of their sounds, which could combine to produce slightly inharmonic partials with a high amount of noise. The loud fanfare-like unisons in winds then drop off, leaving only a solo *sanxian* over the cello, double bass, and bass *sheng* in the low register. These fanfare-like and quieter sections alternate two more times. Following that, an ascending scale pattern coupled in thirds, creates a noisy and almost dissonant effect. This section sounds heavier in general, a stark contrast especially when it comes immediately after a lighter, more joyous H+ section that has a regular dance-like rhythm throughout.

The next part of this high-arousal/negative-valence section is named "Crossfire of the Armies" and uses several onomatopoeic sound associations to invoke this image. Ostinato eighth notes are overlaid with an asymmetrical rhythmic pattern, the *iqā'*

Aqsāq, nine quavers in a 2-2-2-2-1 grouping. The *maqām* Nahāwand used here sounds like a minor mode. This however quickly moves into a melody using fragments of *maqām* Kurd over dissonant ostinati eighth notes. It starts soft, and a lengthy crescendo, together with gradually increasing pitch, creates a highly tensing affect. This tension does not let off, but instead, C minor and F# minor triads are stacked together harmonically, while a combination of *jins* Zamzam and Ḥijāz melodically continues this high tension. A lengthy roll on the timpani and Chinese drum finally brings the music from a *fortississimo* to a *pianissimo* to end this section.

The H– affective intention stands out as the one in which there was greatest divergence between the listener groups. CHM responses were the most extreme (and accurate) in relation to what the composer envisioned. RMS energy is similarly implicated in the responses of all three listener groups. It is, however, also implicated in the other affective intentions. While providing some form of information regarding the perceived affect, RMS energy alone might not be sufficient for listeners to differentiate between affective intentions that are similar in arousal level. Note onset density is also implicated in H– for all three listener groups. However, the interpretation of this feature might be somewhat more complicated. For all three groups, impulse response functions showed that changes in note density provided information for a higher arousal level. However, it is a denser texture that informs CHM and WM about a more negative valence. This means that within this section of music, in moments where the density changes inform these musician listeners about a high arousal, but at the same time when moments of high density appear, it differentiates the section as a negatively valenced section as opposed to a positive one. NM appeared to interpret a dense texture differently from the two musician groups, rating a less dense texture as being more negatively valenced.

Spectral flux also plays a role in H– for the two musician groups. The more spectral changes there are, the greater CHM and WM rated it as higher in arousal and more negatively valenced. In H+, it is a lower spectral flux that appears to contribute to a higher arousal rating. If narrative content and perceived affective intentions feed into mechanisms that appraise acoustic features for affective intentions, it would seem that with a combination of other acoustic and musical features, as well as a narrative content that has developed, spectral flux comes to be appraised as contributing to a high arousal, differentiating this from a positively valenced narrative content and/or affective intention. Similarly to the H+ section, there are more overlaps in the features that inform the two musician listener groups of a H– affect, as compared to the nonmusician group. More crucially, it can also be observed that the two musician groups are highly nuanced in their appraisal process, using spectral flux in opposite ways when the mechanism for perceiving affective intentions is modulated by their understanding of the narrative content.

Low-arousal/negative-valence

A quiet desolate start opens the next section, with low-arousal/negative-valence, titled “Tears of the Rubble”. The sparsity of this whole section compared to the immediately preceding noisy and dissonant H– section is another stark constrast the composer created. He uses *maqām* Sabā, a popular *maqām* in the Arabic repertory that is used to signify melancholia and sadness and describes the segment as sounding “like the wind blown sand”. A solo melody on the *sanxian* is answered by deep and slow counter melody in the cello and double bass. A few other instruments in running thirty-second notes are slowly added to the background before a calm, tranquil ending on an open fifth. Slow tempo, soft dynamics, and general sparseness of the texture have usually been associated with sadness and melancholia. This is a section that is accurately perceived by all listener groups. However, CHM gave the most extreme responses, followed by WM, whereas the responses of NM hovered closer to zero. It is evident that the affective intention has been successfully communicated by the composer through his use of various musical elements and orchestration techniques. On top of that, these intentions are likely augmented by listeners' understanding of the various acoustic cues found here. NM utilize the least number of cues, with a more muffled sound and less clear harmonic sounds likely informing them of a negative valence and low arousal. More acoustic cues appear to be in play for WM and CHM. Both these groups of listeners agree on a thinner, softer sounds, and a sparser texture that indicates a low-arousal negatively valenced affect. In addition to these, WM tend to perceive a darker more muffled sound as an indication that the music has low arousal and negative valence, somewhat similarly to that of NM. CHM, on the other hand, perceive a cleaner sound with less spectral changes in addition to the characteristics they share with the WM. Spectral flux is implicated in both L– and H+ affective intentions, but they influence perception in different directions. Less spectral changes indicate low arousal and negative valence in the L– section but high-arousal/positive-valence in the H+ section. The feature in itself will communicate conflicting information to listeners, but rather, the usefulness of this feature in communicating affective intentions is when it is taken in combination with other acoustic and musical features. Spectral roll-off is also found to be an influencing feature in L–, L+, and H+ affective intentions for WM. How this feature influences WM perception, however, appears to work in opposite directions in H+ and L–. In the H+ section, greater richness and brightness contribute to higher arousal and more positively valenced ratings. In contrast, in L–, greater richness and brightness inform WM of a lower arousal and more negative valence. Spectral roll-off appears to be another cue here that does not function independently, but instead depends on the combination with other acoustic and musical features to provide information on affective intention. In addition to combinations with other cues to clarify the affective intention, listeners might also be appraising these features in different ways because of the narrative content they have already formed as they listened to the music. In addition to spectral flux, which appears to be modulated by narrative content when being utilized in the perception of a L– affective intention, spectral spread also appears to be implicated for CHM. WM on the other hand use spectral roll-off in different ways when their mechanism for affect perception is modulated by their understanding of the narrative content.

Low-arousal/positive-valence

The final section is titled “Path to the Future”. Most of this section has low arousal and positive valence, and the arousal rises only close to the ending of the piece. Even though all the listener groups accurately perceive the affective intention of this section, a substantial portion of their responses were closer to zero than for the other sections. Although the music does not elicit extreme responses in this affective intention, it does appear that listeners are able to agree on the quality of the affect the music expresses. The low-arousal positively valenced part of this section is played by the solo *sanxian*, which, with the instrumentation, already differentiates this section timbrally and musically from all the other sections. Only RMS energy appears to contribute to CHM's perception of affective intention of this section, with a softer dynamic level implying a more extreme low-arousal positively valenced response. Narratively, however, the music has moved through the sections labelled "Crossfire of the Armies" and "Tears of the Rubble", the affective intentions of both sections being accurately decoded by CHM. A familiarity with the highly programmatic and narrative-rich content found in much of Chinese music likely provides clear cues to this group of listeners of a final triumphant ending, a common musical trope in many pieces of Chinese orchestral music. With only a solo *sanxian*, this section does not possess the majestic and heroic musical gestures signifying a triumphant close, but it is likely clear to CHM that the valence at least is becoming much more positive than that of the previous two sections. It thus appears likely that CHM use their understanding of the narrative content of the music to guide them in making this perceptual decision for affective intention.

Both WM and NM utilize spectral slope and RMS energy to inform them of this low-arousal/positive-valence section. A softer sound likely informs the low arousal, but when combined with the other acoustic features listeners from each of these groups used, a slightly different characteristic of the musical sound appears to be implicated. For WM, a darker, quieter sound together with a hint of minor modality informs this affective intention, whereas for NM, it is a brighter but more muffled sound with greater spectral changes over time.

Music Narrative Framework in the Perception of Affective Intentions

As Dean and Bailes (2014) noted in their study, the "continuous response perception of change time series are composed of distinct segments, many of which reflect structural elements" (p.108). The formation of narratives and affective intentions perceived over the music could also contribute to structural divisions and create distinct segments that may influence perceptual processes. The music narrative framework proposes ways in which learned experiences, imagined narrative content, and perceived affective intentions modulate and/or mediate the appraisal mechanisms that feed into further creation of narratives, and perceived affect. It has been shown that people generate narratives easily in response to instrumental music (Margulis, Wong, et al., 2022). These narratives are also created during the listening process and not simply formed only at the end of an entire listening (Margulis, Williams, et al., 2022). Listeners presented with a lengthy piece of music then likely go through a process of narrative generation and also update the narrative as new sonic information is encountered. The narratives generated could inform listeners of the intended affect. At the same time, affective intentions could be perceived directly from the sonic input, which goes on to inform listeners' narrative generation. The way in which acoustic and musical features in this study relate in different combinations and directions to perceived affective responses provides support for the notion that listeners do not simply perceive a particular affective intention with particular values of acoustic and musical features through associative mechanisms alone. In line with what Cespedes-Guevara (2021) proposes, both associative mechanisms and more complex processes of appraisal occur so that listeners are able to differentiate different nuances of affective intentions. Perceived affective intentions require the listener to make sense of what intended emotions are represented by the music without having to actually feel them and it stands to reason that narratives play a role in influencing them to a greater extent than other processes like brain-stem reflexes, emotional contagion, entrainment, and so on, which may play a greater role in induced affect in music (Juslin, 2013).

Each of the listener groups appear to utilize different acoustic and musical features in their ratings of perceived affective intentions. There are some commonalities: RMS energy, for instance, is a feature that influences the perceived arousal of all three groups of listeners, and in the same direction regardless of the affective intention. This means that a higher RMS energy informs listeners of a higher arousal and vice versa. It is therefore a feature that is not influenced much by experience and expertise in particular musical traditions. Narratives formed will also be less likely to influence the appraisal of this feature. These types of features form the set of common cues that provide relatively similar amounts of information to listeners regardless of their musical backgrounds (Balkwill & Thompson, 1999; Thompson & Balkwill, 2010). Other features such as spectral flux, spectral spread, and spectral roll-off are implicated in more complex ways. An increase in the value of spectral flux, for instance, points towards a higher arousal in H– but a lower arousal for H+ for both CHM and WM. A smaller spectral spread indicates lower arousal and more negative valence in the L– section, but a higher arousal and more positive valence in the H+ section for CHM. Even with these conflicting cues, however, listeners do not appear to confuse these affective intentions—the ratings of their perceived affective intentions usually align with the composer's intended ones. It is likely therefore that the features do not function independently, but rather, similarly to other studies on global affective responses to music (e.g., Eerola et al., 2012; Heng & McAdams, 2022), it is the combination of cues that provide information on the affective intention of the music.

The formation of narratives could also play an important role in clarifying the perceived affective intentions. These imagined narratives may influence the appraisal processes for the sonic information (Figure 4.1). When narratives point a listener towards a particular affective intention, the appraisal processes for the acoustic and musical features that enter the auditory system will be modulated by this information, and the cues will be utilized accordingly. In addition, experience, learning, and cultural knowledge also play a role in modulating these appraisal processes by selecting and focusing on different aspects of the musical sound and inferring what these sounds imply. An expectation of a dramatic ending, for instance, modulates what CHM and WM infer from a drop in dynamic level towards the ending. This understanding that the temporary drop in dynamics is used to create a greater contrast for the loud and grand closing chord means that the two musician groups did not end the piece with a drop in perceived arousal level. NM on the other hand did not have this piece of knowledge modulating their appraisal process and were the only group that ended the piece with a decrease in arousal level. NM also have fewer features that appear to function in opposite directions for sections with different affective intentions. Although they might form similar narratives to the other listener groups in the process of music listening, it is likely that musical training provided experience and knowledge about certain ways in which acoustic and musical features are frequently combined and used in music. This knowledge influences appraisal processes of the musician listeners, and NM were less able to appreciate the various nuances created by the features that are influenced by different narrative content.

It is also interesting to note that impulse response functions in fact do not show any features directly influencing CHM perception of valence for L– and L+. When the responses of perceived affective intentions of CHM are compared with the composer's intentions, however, they appeared to be accurate. Not only that, CHM also provided the most extreme responses, suggesting they are more confident of their perceived affective intentions than the other two listener groups. If acoustic and musical features are only implicated in a direct way, with particular features or combinations of features influencing perception in the same direction, then it will be difficult to account for why some features appear to function in opposite directions, and why CHM were able to be accurate and confident about their responses even when it does not look like they use any acoustic or musical cues for their judgment of valence in some affective intentions. However, with feedback mechanisms proposed in the music narrative framework, the perception of valence could feedback to influence the appraisal of arousal responses and vice versa. This would explain the difficulty in finding strong and consistent features that influence valence perception in many studies. Even though no cues can be found to significantly contribute to valence responses directly, the perception of valence is still accurate because of acoustic and musical features that influence perceived arousal, as well as narratives that modulate and/or mediate appraisal mechanisms.

Conclusion

Musical meanings and affective intentions are complex, and numerous pathways and mechanisms work in conjunction during the listening process. Narrative content and affective intentions are only one aspect of this, albeit a prominent one in much of music listening. In musical cultures in which listening (without active participation in the music-making activity) is a predominant aspect, the formation of narratives might be of great importance for musical engagement. This study explores the perception of affective intentions in this mode of musical engagement and attempts to frame narratives and affective intentions within appraisal processes that are influenced by affective intentions feeding back into the formation of new affective intentions, affective intentions feeding forward into the formation of narratives, narratives feeding back into the formation of new narratives, and narratives feeding forward into the formation of affective intentions, as well as by prior experience and learning.

The strong narrative nature intended by the composer in this piece of music meant that many compositional decisions were specifically designed for this specific goal. From the way listeners' responses align with the intended affective intentions, it appears that the composer (and performers) are successful in communicating these intentions to the listeners. This study did not explicitly ask for narratives that listeners formed during the listening process, so it is uncertain if the correspondences are only with the affective intentions or with both the affective intentions and the narrative content. The narratives formed could be static, unitary events. They could also be a continuously changing event. It would be very interesting and useful to incorporate functional data to analyze how narrative events relate to acoustic and musical features. If listeners are asked to explicitly indicate boundaries where they form new narratives, time series analyses can be applied to track continuous changes within each set of narratives, their perceived affective intentions, and the relation with features in the music. Future studies could study different pieces of music, different musical styles, in listeners with different musical and cultural backgrounds. Listeners could also be explicitly probed with regards to the narratives they form over the course of the music. Pieces of music with different levels of narrative intent could also be studied to explore if such a music narrative framework applies to music that was not intended to be programmatic or narrative in nature.

Humans have the ability to make sense of complex, abstract information by taking multiple cues in combination and organizing and processing them in a way that creates coherent sense. Music is an example of this activity that humans engage in, and an exploration of this phenomenon is helpful in illuminating other ways in which complex and abstract information unfolds in perception and cognition. A framework that helps to explain appraisal mechanisms in perceived musical affect contributes to understanding this complex communicative process. Studying a lengthy piece of music and listeners' real-time, continuous responses also illuminates the dynamic processes of music perception and cognition.

Acknowledgements

We thank composer WANG Chenwei for providing the score and recording used for the experimental stimulus. This research was supported by grants to SMc from the Canadian Social Sciences and Humanities Research Council (895-2018-1023, 435-2021-0224) and the Fonds de recherche du Québec—Société et culture

(2017-SE-205667), as well as a Canada Research Chair (950-223484). We also thank Bennett K. Smith for programming the experimental interface.

Part V

Influences on perceived affective intentions in music

Overview

The final chapter of this dissertation summarizes the main findings from the two experimental studies and one conceptual paper with re-analysis of some of the data and attempts to consolidate these to aid in understanding how listeners make use of different acoustic and musical features in the perception of affective intentions in music. It also explores how experience and expertise in different musical traditions influence the ways in which these features are utilized. This dissertation explores the creation of musical meanings during the process of music listening and how learning might influence this process. As a dynamic activity, the process of listening takes in a plethora of complex sonic information, processes it in real time, and creates a range of meanings through different mechanisms including signification and appraisal.

Chapter 1 provided the theoretical concepts related to musical meanings, perception of affective intentions, common and culture-specific cues in communication, and analytical methods adopted in the experimental studies. A brief introduction to the Chinese orchestra and Chinese music was also provided. Chapter 2 reported on an experimental study investigating how timbre functions as a carrier of information for communicating affective intentions in music, how experience and expertise in different musical traditions influence the ways these acoustic features might be utilized, and how affective response depends on the amount of musical material encountered. Chapter 3 reported on another experimental study that explored the continuous perceived affective responses of listeners with different musical backgrounds. The use of a lengthy piece of music allowed for the examination of this dynamic process of music listening, as well as how different aspects of the music might influence this process. Chapter 4 was a conceptual paper that laid out a framework incorporating appraisal mechanisms with the sonic event, listener knowledge, affective intentions, and constructed narratives in the music listening process based on a re-analysis of some data from Chapter 3. This fifth and final chapter ties these together and discusses their implications in music listening.

Differences in Musical Training

One of the aims of this dissertation is to examine how differences in musical training influence the perception of affective intentions in music. To this end, it is seen from the first experimental study that listeners who have formal training in music performed more consistently and accurately in judging expressed affective intentions, supporting Thompson and Balkwill's (2010) cue-redundancy model. Familiarity with how particular affective intentions are usually produced in performance likely mean that the musician listeners are sensitive to more cues when it comes to decoding perceived affective intentions. However, we also found that listeners trained in the Chinese music tradition responded more consistently and accurately in judging expressed affective intentions regardless of the instrument tradition in which the excerpt of music is performed. The different emphases placed on the function of timbre could mean that listeners trained in Chinese music might be more sensitive to nuances in timbral manipulation in performance, whereas listeners trained in Western music focus more on other musical aspects.

The second experimental study found that when the continuous ratings of affective intentions for each listener group were averaged, Chinese musician listeners tended to be more extreme in their responses. One reason could be that this group of listeners is more confident and consistent about their judgments. When the piece of music was analyzed for its intended affect based on the composer's indications, we also found that the Chinese musician (CHM) listener group had the most congruent judgments with the composer's intentions, followed by the Western musician (WM) listeners and then the nonmusicians (NM) with the least congruence. Results from the two experimental studies paint a consistent picture supporting the idea that familiarity with the musical style and the role timbre plays in musical expression provides more information for a listener to make decisions about the intended musical affect. That some musical cultures emphasize particular aspects of the music also suggests that listeners who are familiar with that musical culture are likely to pay more attention to those aspects of the music. Chinese music, for instance, places heavy emphasis on how each phrase of music has to be produced when expressing particular intentions. Listeners trained in this musical tradition are therefore also more likely to place greater emphasis on how the timbre is being produced when attempting to decode intentions during music listening.

Acoustic Descriptors and Musical Features Utilized in Perceived Affective Intentions

The first experimental study examining listeners' responses to musical sounds of varying durations found that acoustic features do not contribute independently or in a linear or unidirectional way to perceived affective intentions. Rather, a combination of different acoustic features act together to provide the information for listeners. There are also more similar combinations of features shared by all the listener groups in arousal perception than in valence perception. This finding is consistent with Egermann and colleagues (2015) study where they found that listeners' valence responses are very much influenced by musical culture and training, but arousal responses are based on more universal psychophysical cues. There is also support for the cue-redundancy model here. When acoustic features align in production and comprehension with the intended effect, there will be greater accuracy in the perceived affective responses. In the first study, it is seen that combinations of acoustic features used for decoding affective intentions by CHM and WM overlap much more than they do with NM. However, even though NM use very different acoustic features, it does not render them totally inaccurate at decoding affective intentions. There is a redundancy of cues in musical communication, and even with several misaligned cues, a certain amount of decoding accuracy is still possible.

The second experimental study explored listeners' time-varying continuous responses to perceived affective intentions in music. Impulse response functions show again that no single acoustic or musical feature communicates a clear and distinct type of information, but rather, features add on to or interact with other elements to provide listeners with a more nuanced picture in their understanding. It also appears that emotional intensity and arousal responses are elicited from more universal, common cues than valence responses. Different acoustic and musical features are found to be implicated in the sections that have different narrative content and affective intentions, suggesting that appraisal of the musical content plays an important role in influencing perceived affective intentions.

Appraisal Mechanisms and Narratives

Several studies on musical emotions have pointed towards appraisal mechanisms as providing an important function in affective responses to music (Lennie & Eerola, 2022; Cespedes-Guevara, 2021; Cespedes-Guevara & Eerola, 2018). Listeners make sense of the music they hear in different ways depending on the context and an awareness of the implications of a stimulus within the musical sound. Margulis and colleagues have demonstrated the prevalence of narratives formed during music listening (Margulis, Miller, et al., 2022; Margulis, Wong, et al., 2022; Margulis et al., 2019; Margulis, 2017). If narratives and appraisal mechanisms are commonplace, it follows that the perception of affective intentions would likely be influenced by the narrative content listeners create during their listening process, and narratives created would be influenced by the perceived affective intentions. The music narrative framework that is proposed explains the ways in which learned experiences, imagined narrative content, and perceived affective intentions modulate and/or mediate appraisal mechanisms and continually influence the creation of narratives, as well as perceived affect.

Contributions

As a complex but commonplace human behaviour, music listening involves many processes happening in real time. Information from many different aspects of the sonic event contributes and interacts in a variety of ways with contextual information and a listener's knowledge and experience. This dissertation attempts to disentangle the large number of complex variables that influence the listening process and to examine the dynamic relationships between sonic features and perceptual processes.

The first experimental study finds that the group of listeners trained in Chinese music is more accurate than the group of listeners trained in Western music and the group of nonmusicians, when it comes to judging the affective intention expressed by a performer in an excerpt. Chinese musicians were also more accurate regardless of the culture of the instrument playing the excerpt. This may be due to the greater emphasis placed on timbral manipulations in performance to convey particular affective intentions in Chinese music. An exploration of the acoustic features found a few components influencing listeners' perception of affective intentions. These components comprise combinations of a number of acoustic descriptors, suggesting that features do not contribute singularly, or in a linear fashion, to inform perceived musical affect. Instead, it is the specific combinations of a number of acoustic features that influence this complex process. There are also many more features that are similarly used by all the three groups of listeners for arousal responses than is the case for valence responses. This is consistent with studies showing that there are more cues associated with arousal responses, and that there are small or nonexistent associations with valence responses (e.g., Bänziger et al., 2013; Egermann et al., 2015).

The second study explores the continuous response of listeners as they listen to a piece of Chinese orchestral music and how their musical backgrounds might influence affective responses. It also attempts to examine the dynamic process in which acoustic and musical features are utilized by each group of listeners in their perception of different affective intentions. Results show that the listener groups differed significantly over certain sections of the music. Similarly to the first experimental study, valence responses also diverge more than arousal or emotional intensity responses, again suggesting that the perception of valence is learned within different musical traditions, whereas arousal and emotional intensity responses draw on more universal, common psychophysical cues. Similarly to the first study, acoustic and musical features do not contribute independently, but rather combine and interact with one another. In addition, depending on the musical and affective context, some features relate to listener responses in different directions, suggesting that listeners are very nuanced in their perception of affective intentions in the music.

In Chapter 4, a music narrative framework is proposed to attempt to incorporate processes of appraisal in understanding and judging perceived affective intentions in music. When a piece of music is divided into sections of different affective intentions, we observed that the acoustic and musical features that appear to influence listeners' perceptions behave in different ways. This again suggests that features do not contribute singularly, nor in a linear, unidirectional fashion, to listeners' perception of affective intentions. Depending on the context of the music, similar features might provide different information. It appears then that there is likely a process of appraisal in which listeners' prior knowledge and contextual understanding interact with the sonic cues available to provide a set of reliable information regarding what the music is trying to express. The proposed music narrative framework therefore reconciles these complex interactions with appraisal mechanisms that modulate and/or mediate sonic cues, created narrative content, and perceived musical affects, and provide an explanation for how learning and experience influence music perception.

This dissertation uses various statistical methods for complex, high dimensional multivariate, and dynamic time-varying data. These analytical methods are less often used in the study of music and listeners' perception but may prove to be valuable resources in this field, as real-world music listening behaviour is associated with numerous complex and dynamic factors. Additionally, the use of Chinese instruments and Chinese orchestral music provides a different perspective in the study of music perception. Most studies have utilized proto-musical materials or music from the Western classical repertoire with Western participants. Other studies exploring differences in musical cultures have participants from different geographical regions. As there may be many other factors in addition to differences in musical training that influence listeners with different socio-cultural backgrounds, it might be more difficult in studying participants from different geographical regions to differentiate factors that arise from musical training from those that are due to differences in socio-cultural experience. In a similar strand, Margulis and colleagues (2019) suggest that it might be worthwhile to explore cultures defined in other ways than geography. The studies with listeners from the same geographical region, who are relatively similar in their socio-cultural backgrounds, but trained in different musical traditions demonstrate that there are differences in the perception of affective intentions in music, providing evidence that particular musical cultures might place different emphases or have different sets of rules in utilizing and organizing acoustic and musical elements. The perception of affective intentions in music listening is a complex process that is influenced by the degree of familiarity listeners have with a musical tradition, the narrative content implicated in the music, and the complex sonic environment created by the composer and the musicians in their interpretation in performance.

Limitations and Future Directions

The two experimental studies in this dissertation explore differences between listeners with contrasting musical experiences: a group with formal training in the Chinese music tradition, a group with formal training in the Western classical music tradition, and a group of nonmusicians. Although it serves to demonstrate differences in the ways these three groups of listeners utilize acoustic and musical features in their perception of affective intentions in music, future studies could explore listeners with experience and knowledge in other musical traditions.

The second study is an online, browser-based listening experiment. As compared to in-person experiments, there is less control with respect to the type of equipment participants use and the environment in which they are listening to the musical stimuli. Instructions before the start of the listening task ensure that participants use headphones, not complete the experiment with a mobile device, and have the volume adjusted to a comfortable and adequate level. The variance in participants' responses also appeared acceptable. However, in-person listening experiments should be conducted in the future to ensure that the data from online experiments are replicable in in-person ones.

This dissertation also examines perceptual processes in music listening and calls for methods and techniques from various disciplines. As an emic researcher trained in both Western and Chinese music, there is an advantage of a greater understanding of the context and approaches to music and listening when working with participants from Singapore. The analytical methods provide potentially different ways of studying complex, time-varying responses and could be useful in other fields. With evidence that the participants in these studies can be relatively accurate in their perception of affective intentions in music, future studies could expand this with other listeners from more diverse populations. Different pieces of music could be explored, in the Chinese music tradition as well as other musical traditions.

Although many different acoustic descriptors can characterize various aspects of a sound, this dissertation has also demonstrated that listeners perceive the combination of these different descriptors instead of separate, independent dimensions in their perception of affective intentions in music. Timbre as an important component in the perception of musical sounds is shown to be utilized in purposeful ways by all the three groups of listeners in these studies, and in an even more nuanced way by listeners trained in Chinese music. This implies that the way timbre functions in musical communication can be taught. Future studies could be performed on listeners trained in different musical traditions. An exploration of these different musical traditions and their teaching and performance practices could be related to the differences in the ways timbre is used in the perception of affective intentions for listeners in these other traditions.

Studying the differences in listeners trained in different musical traditions but with similar linguistic and socio-cultural backgrounds makes it easier to explore factors related to musical training independently from those related to language and socio-cultural differences. How timbre can be learned through formal musical training can be demonstrated more definitively. Future research could be extended to examine timbre perception in both neurodivergent and neurotypical populations, and whether learning can influence both populations in terms of the acoustic cues they attend to.

Music listening involves many complex cognitive and perceptual mechanisms. Appraisal mechanisms could underlie some of these processes and allow listeners to understand affective intentions from complex sonic cues. The final part of this dissertation proposes the music narrative framework that could explain how listeners make sense of the music they hear. Future studies could look into the actual narrative content that is formed and relate it to the perceived affective intentions, as well as explore how they interact with each other and with prior knowledge and experiences in the appraisal processes.

As a complex, dynamic, and time-varying phenomenon, the research in this dissertation offers an avenue of examining the auditory processes of humans, as well as how learning and experience play a role in modulating complex cognitive processes. It demonstrates the value of an interdisciplinary approach in studying complex human behaviours, combining in-depth understanding of different musical traditions, emic knowledge of the population of participants, with experimental methods used in psychology, and statistical techniques from various fields that have gradually found their way into the behavioural sciences.

References

Adam, H., & Shirako, A. (2013). Not all anger is created equal: The impact of the expresser's culture on the social effects of anger in negotiations. *Journal of Applied Psychology*, *98*(5), 785–798. https://doi.org/10.1037/a0032387

Adler, S. (2002). *The study of orchestration* (3rd ed.). WW Norton & Company.

Agawu, V. K. (1991). *Playing with signs: A semiotic interpretation of classic music.* Princeton University Press.

Bailes, F., & Dean, R. T. (2012). Comparative time series analysis of perceptual responses to electroacoustic music. *Music Perception*, *29*(4), 359–375. https://doi.org/10.1525/mp.2012.29.4.359

Balkwill, L.-L., & Thompson, W. F. (1999). A cross-cultural investigation of the perception of emotion in music: Psychophysical and cultural cues. *Music Perception*, *17*(1), 43–64. https://doi.org/10.2307/40285811

Balkwill, L.-L., Thompson, W. F., & Matsunaga, R. (2004). Recognition of emotion in Japanese, Western, and Hindustani music by Japanese listeners. *Japanese Psychological Research*, *46*(4), 337–349. https://doi.org/10.1111/j.1468-5584.2004.00265.x

Bänziger, T., Patel, S., & Scherer, K. R. (2013). The role of perceived voice and speech characteristics in vocal emotion communication. *Journal of Nonverbal Behavior*, *38*(1), 31–52. https://doi.org/10.1007/s10919-013-0165-x

Barrett, L. F. (2006). Solving the emotion paradox: Categorization and the experience of emotion. *Personality and Social Psychology Review*, *10*(1), 20–46. https://doi.org/10.1207/s15327957pspr1001_2

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. https://doi.org/10.1177/030573569302100102

Bowling, D. L., Sundararajan, J., Han, S., & Purves, D. (2012). Expression of emotion in Eastern and Western music mirrors vocalization. *PLoS ONE*, *7*(3), e31942. https://doi.org/10.1371/journal.pone.0031942

Bowman, C., & Yamauchi, T. (2016). Perceiving categorical emotion in sound: The role of timbre. *Psychomusicology: Music, Mind, and Brain*, *26*(1), 15–25. https://doi.org/10.1037/pmu0000105

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2016). *Time series analysis: Forecasting and control* (5th ed.). John Wiley & Sons, Inc.

Cannam, C., Landone, C., & Sandler, M. (2010). Sonic visualiser: An open source application for viewing, analysing, and annotating music audio files. *Proceedings of the ACM Multimedia 2010 International Conference*, 1467–1468. https://dl.acm.org/doi/pdf/10.1145/1873951.1874248

Cespedes-Guevara, J. (2021). A constructionist approach to emotional experiences with music. https://doi.org/10.31234/osf.io/sfzm2

Cespedes-Guevara, J., & Eerola, T. (2018). Music communicates affects, not basic emotions – A constructionist account of attribution of emotional meanings to music. *Frontiers in Psychology*, *9*, 215. https://doi.org/10.3389/fpsyg.2018.00215

Chan, M. C. (2003). *Chinese orchestral music is better because of you-an appreciation of Chinese orchestral music* (Vol. 1). Joint Publishing H.K. Pte Ltd.

Cowen, A. S., Fang, X., Sauter, D., & Keltner, D. (2020). What music makes us feel: At least 13 dimensions organize subjective experiences associated with music across different cultures. *Proceedings of the National Academy of Sciences*, *117*(4), 1924–1934. https://doi.org/10.1073/pnas.1910704117

Cross, I. (2009). The evolutionary nature of musical meaning. *Musicae Scientiae*, *13*(2_suppl), 179–200. https://doi.org/10.1177/1029864909013002091

Dance, F. E. (1970). The "concept" of communication. *Journal of Communication*, *20*(2), 201–210.

Davies, S. (2019). *Musical meaning and expression*. Cornell University Press. https://doi.org/10.7591/9781501733987

Dean, R. T., & Bailes, F. (2010). Time series analysis as a method to examine acoustical influences on real-time perception of music. *Empirical Musicology Review*, *5*(4), 152–175. https://doi.org/10.18061/1811/48550

Dean, R. T., & Bailes, F. (2011). Modelling perception of structure and affect in music: Spectral centroid and Wishart's *Red Bird*. *Empirical Musicology Review*, *6*(2). https://doi.org/10.18061/1811/51217

Dean, R. T., & Bailes, F. (2014). Influences of structure and agency on the perception of musical change. *Psychomusicology: Music, Mind, and Brain*, *24*(1), 103–108. https://doi.org/10.1037/pmu0000034

Dilbeck, K. E. (2017). Factor analysis: Varimax rotation. In M. Allen (Ed.), *The SAGE encyclopedia of communication research methods* (pp. 532–533). SAGE Publications, Inc. http://dx.doi.org/10.4135/9781483381411.n191

Duke, R. A., & Colprit, E. J. (2001). Summarizing listener perceptions over time. *Journal of Research in Music Education*, *49*(4), 330–342. https://doi.org/10.2307/3345616

Durlauf, S. N., & Blume, L. (2008). Impulse response function (2nd ed.). http://catdir.loc.gov/catdir/toc/ecip085/2007047205.html

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. https://doi.org/10.1525/mp.2012.30.1.49

Eerola, T., Lartillot, O., & Toiviainen, P. (2009). Prediction of multidimensional emotional ratings in music from audio using multivariate regression models. *Proceedings of the International Conference on Music Information Retrieval*, 621–626.

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. https://doi.org/10.1177/0305735610362821

Egermann, H., Fernando, N., Chuen, L., & McAdams, S. (2015). Music induces universal emotion-related psychophysiological responses: Comparing Canadian listeners to Congolese Pygmies. *Frontiers in Psychology*, *5*, 1341. https://doi.org/10.3389/fpsyg.2014.01341

Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion*, *6*(3-4), 169–200. https://doi.org/10.1080/02699939208411068

Elkin, L. A., Kay, M., Higgins, J. J., & Wobbrock, J. O. (2021). An aligned rank transform procedure for multifactor contrast tests. *UIST '21: The 34th Annual*

ACM Symposium on User Interface Software and Technology, 754–768. https://doi.org/10.1145/3472749.3474784

Evans, P., & Schubert, E. (2008). Relationships between expressed and felt emotions in music. *Musicae Scientiae*, *12*(1), 75–99. https://doi.org/10.1177/102986490801200105

Farraj, J. (2018). Maqam world. http://www.maqamworld.com/en/index.php

Farraj, J., & Shumays, S. A. (2019). *Inside Arabic music: Arabic maqam performance and theory in the 20th century*. Oxford University Press.

Friedrich, S., & Pauly, M. (2018). MATS: Inference for potentially singular and heteroscedastic MANOVA. *Journal of Multivariate Analysis*, *165*, 166–179. https://doi.org/10.1016/j.jmva.2017.12.008

Gabrielsson, A. (2001). Emotion perceived and emotion felt: Same or different? *Musicae Scientiae*, *5*(1_suppl), 123–147. https://doi.org/10.1177/10298649020050s105

Gabrielsson, A., & Juslin, P. N. (2003). Emotional expression in music. In R. Davidson & K. R. Scherer (Eds.), *Handbook of affective sciences* (pp. 503–534). Oxford University Press.

Gabrielsson, A., & Lindström, E. (2010). The role of structure in the musical expression of emotions. In P. N. Juslin & J. A. Sloboda (Eds.), *Handbook of music and emotion: Theory, research, applications* (pp. 368–402). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199230143.003.0014

Goodchild, M., Wild, J., & McAdams, S. (2019). Exploring emotional responses to orchestral gestures. *Musicae Scientiae*, *23*(1), 25–49. https://doi.org/10.1177/1029864917704033

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

Hailstone, J. C., Omar, R., Henley, S. M., Frost, C., Kenward, M. G., & Warren, J. D. (2009). It's not what you play, it's how you play it: Timbre affects perception of emotion in music. *Quarterly Journal of Experimental Psychology*, *62*(11), 2141–2155. https://doi.org/10.1080/17470210902765957

Hanslick, E. (1854). *The beautiful in music* (G. Cohen, Trans.). Liberal Arts Press.

Harrison, P., & Pearce, M. T. (2020). Simultaneous consonance in music perception and composition. *Psychological Review*, *127*(2), 216. https://doi.org/10.1037/rev0000169

Hatten, R. S. (2004). *Interpreting musical gestures, topics, and tropes: Mozart, Beethoven, Schubert*. Indiana University Press.

Heng, L., & McAdams, S. (2022). *Timbre's function in the perception of affective intentions: Differences between musical traditions* [Manuscript submitted for publication; chapter 2 of this dissertation].

Heng, L., Wei, C., & McAdams, S. (2023). *Continuous response in music listening: Training in different musical traditions influence perception of affective intentions* [Manuscript in preparation; chapter 3 of this dissertation].

Huron, D. (2001). Is music an evolutionary adaptation? *Annals of the New York Academy of Sciences*, *930*(1), 43–61. https://doi.org/10.1111/j.1749-6632.2001.tb05724.x

Huron, D., Anderson, N., & Shanahan, D. (2014). "You can't play a sad song on the banjo:" Acoustic factors in the judgment of instrument capacity to convey sadness. *Empirical Musicology Review*, *9*(1), 29–41. https://doi.org/10.18061/emr.v9i1.4085

International Organization for Standardization. (2004). Acoustics — Reference zero for the calibration of audiometric equipment — Part 8: Reference equivalent threshold sound pressure levels for pure tones and circumaural earphones [ISO Standard No. 389-8]. https://www.iso.org/obp/ui/#iso:std:iso:389:-8:ed-1:v1:en

Juslin, P. N. (2000). Cue utilization in communication of emotion in music performance: Relating performance to perception. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(6), 1797–1812. https://doi.org/10.1037/0096-1523.26.6.1797

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. https://doi.org/10.1016/j.plrev.2013.05.008

Juslin, P. N. (2019). *Musical emotions explained: Unlocking the secrets of musical affect*. Oxford University Press. https://doi.org/10.1093%2Foso%2F9780198753421.001.0001

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. https://doi.org/10.1037/0033-2909.129.5.770

Juslin, P. N., & Laukka, P. (2004). Expression, perception, and induction of musical emotions: A review and a questionnaire study of everyday listening. *Journal of New Music Research*, *33*(3), 217–238. https://doi.org/10.1080/0929821042000317813

Juslin, P. N., & Timmers, R. (2010). Expression and communication of emotion in music performance. In P. N. Juslin & J. A. Sloboda (Eds.), *Handbook of music and emotion: Theory, research, applications* (pp. 453–489). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199230143.003.0017

Kazazis, S. (2020). *Psychophysical scaling of timbre-related audio descriptors.* (Publication No. 28383840). [Doctoral dissertation, McGill University] ProQuest Dissertations Publishing.

Kazazis, S., Depalle, P., & McAdams, S. (2021). The Timbre Toolbox version R2021a, user's manual. https://github.com/MPCL-McGill/TimbreToolbox-R2021a

Kivy, P. (1980). *The corded shell: Reflections on musical expression* (Vol. 39). Princeton University Press.

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

Koelsch, S. (2011). Towards a neural basis of processing musical semantics. *Physics of Life Reviews*, *8*(2), 89–105. https://doi.org/10.1016/j.plrev.2011.04.004

Konečni, V. J. (2008). Does music induce emotion? A theoretical and methodological analysis. *Psychology of Aesthetics, Creativity, and the Arts*, *2*(2), 115–129. https://doi.org/10.1037/1931-3896.2.2.115

Konietschke, F., Bathke, A. C., Harrar, S. W., & Pauly, M. (2015). Parametric and nonparametric bootstrap methods for general MANOVA. *Journal of Multivariate Analysis*, *140*, 291–301. https://doi.org/10.1016/j.jmva.2015.05.001

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). Excerpta Medica.

Krumhansl, C. L. (1998). Topic in music: An empirical study of memorability, openness, and emotion in Mozart's String Quintet in C Major and Beethoven's String Quartet in A Minor. *Music Perception*, *16*(1), 119–134. https://doi.org/10.2307/40285781

Krumhansl, C. L., & Kessler, E. J. (1982). Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychological Review*, *89*(4), 334–368. https://doi.org/https://doi.org/10.1037/0033-295x.89.4.334

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

Lartillot, O. (2022). MIRtoolbox. MATLAB Central File Exchange. https://www.mathworks.com/matlabcentral/fileexchange/24583-mirtoolbox

Laukka, P., Eerola, T., Thingujam, N. S., Yamasaki, T., & Beller, G. (2013). Universal and culture-specific factors in the recognition and performance of musical affect expressions. *Emotion*, *13*(3), 434–449. https://doi.org/10.1037/a0031388

Lennie, T. M., & Eerola, T. (2022). The CODA model: A review and skeptical extension of the constructionist model of emotional episodes induced by music. *Frontiers in Psychology*, *13*, 822264. https://doi.org/10.3389/fpsyg.2022.822264

Levitin, D. J., Nuzzo, R. L., Vines, B. W., & Ramsay, J. (2007). Introduction to functional data analysis. *Canadian Psychology/Psychologie Canadienne*, *48*(3), 135–155.

Lütkepohl, H. (2005). *New introduction to multiple time series analysis.* Springer Science & Business Media.

Margulis, E. H. (2017). An exploratory study of narrative experiences of music. *Music Perception*, *35*(2), 235–248. https://doi.org/10.1525/mp.2017.35.2.235

Margulis, E. H., Miller, N., Mitchell, N., Orsini Windholz, M., Williams, J., & McAuley, J. D. (2022). Intersubjectivity and shared dynamic structure in narrative imaginings to music. *Music Theory Online*, *28*(4). https://www.mtosmt.org/issues/mto.22.28.4/mto.22.28.4.margulis.php

Margulis, E. H., Williams, J., Simchy-Gross, R., & McAuley, J. D. (2022). When did that happen? The dynamic unfolding of perceived musical narrative. *Cognition*, *226*, 105180. https://doi.org/10.1016/j.cognition.2022.105180

Margulis, E. H., Wong, P. C. M., Simchy-Gross, R., & McAuley, J. D. (2019). What the music said: Narrative listening across cultures. *Palgrave Communications*, *5*(1). https://doi.org/10.1057/s41599-019-0363-1

Margulis, E. H., Wong, P. C. M., Turnbull, C., Kubit, B. M., & McAuley, J. D. (2022). Narratives imagined in response to instrumental music reveal culture-bounded intersubjectivity. *Proceedings of the National Academy of Sciences*, *119*(4), e2110406119. https://doi.org/10.1073/pnas.2110406119

Martin, F. N., Champlin, C. A., et al. (2000). Reconsidering the limits of normal hearing. *Journal of the American Academy of Audiology*, *11*(2), 64–66. https://doi.org/10.1055/s-0042-1748011

McAdams, S. (1989). Psychological constraints on form-bearing dimensions in music. *Contemporary Music Review*, *4*(1), 181–198. https://doi.org/10.1080/07494468900640281

McAdams, S. (2019). The perceptual representation of timbre. In K. Siedenburg, C. Saitis, S. McAdams, A. Popper, & R. Fay (Eds.), *Timbre: Acoustics, perception, and cognition* (pp. 23–57). Springer. https://doi.org/10.1007/978-3-030-14832-4_2

McAdams, S., Douglas, C., & Vempala, N. N. (2017). Perception and modeling of affective qualities of musical instrument sounds across pitch registers. *Frontiers in Psychology*, *8, 153*. https://doi.org/10.3389/fpsyg.2017.00153

McAdams, S., Vines, B. W., Vieillard, S., Smith, B. K., & Reynolds, R. (2004). Influences of large-scale form on continuous ratings in response to a contemporary piece in a live concert setting. *Music Perception*, *22*(2), 297–350. https://doi.org/10.1525/mp.2004.22.2.297

McAdams, S., Winsberg, S., Donnadieu, S., Soete, G. D., & Krimphoff, J. (1995). Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes. *Psychological Research*, *58*(3), 177–192. https://doi.org/10.1007/bf00419633

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

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

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

Meyer, L. B. (1994). *Music, the arts, and ideas*. The University of Chicago Press. https://doi.org/10.7208/chicago/9780226521442.001.0001

Micallef Grimaud, A., & Eerola, T. (2022). Emotional expression through musical cues: A comparison of production and perception approaches. *PLoS ONE*, *17*(12), e0279605. https://doi.org/10.1371/journal.pone.0279605

Nagel, F., Kopiez, R., Grewe, O., & Altenmüller, E. (2007). EMuJoy: Software for continuous measurement of perceived emotions in music. *Behavior Research Methods*, *39*(2), 283–290. https://doi.org/10.3758/bf03193159

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

Paraskeva, S., & McAdams, S. (1997). Influence of timbre, presence/absence of tonal hierarchy and musical training on the perception of musical tension and relaxation schemas. *Proceedings of the International Computer Music Conference*.

Patel, A. D. (2010). *Music, language, and the brain*. Oxford University Press. https://doi.org/10.1093%2Facprof%3Aoso%2F9780195123753.001.0001

Peeters, G., Giordano, B. L., Susini, P., Misdariis, N., & McAdams, S. (2011). The Timbre Toolbox: Extracting audio descriptors from musical signals. *The Journal of the Acoustical Society of America*, *130*(5), 2902–2916. https://doi.org/10.1121/1.3642604

Pierce, C. S. (2014). On the nature of signs. In J. Hooper (Ed.), *Peirce on signs: Writings on semiotic by Charles Sanders Peirce* (pp. 141–143). The University of North Carolina Press.

R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Ramsay, J., Hooker, G., & Graves, S. (2009). Introduction to functional data analysis. In *Functional data analysis with R and MATLAB* (pp. 1–19). Springer.

Ratner, L. G. (1980). *Classic music: Expression, form, and style*. Schirmer Books.

Reisenzein, R. (1994). Pleasure-arousal theory and the intensity of emotions. *Journal of Personality and Social Psychology*, *67*(3), 525–539. https://doi.org/10.1037/0022-3514.67.3.525

Rickard, N. S. (2004). Intense emotional responses to music: A test of the physiological arousal hypothesis. *Psychology of Music*, *32*(4), 371–388. https://doi.org/10.1177/0305735604046096

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*(6), 1161–1178. https://doi.org/10.1037/h0077714

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, *110*(1), 145–172. https://doi.org/10.1037/0033-295x.110.1.145

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. https://doi.org/10.1080/0929821042000317822

Scherer, K. R., Clark-Polner, E., & Mortillaro, M. (2011). In the eye of the beholder? Universality and cultural specificity in the expression and perception of emotion. *International Journal of Psychology*, *46*(6), 401–435. https://doi.org/10.1080/00207594.2011.626049

Scherer, K. R., & Oshinsky, J. S. (1977). Cue utilization in emotion attribution from auditory stimuli. *Motivation and Emotion*, *1*(4), 331–346. https://doi.org/10.1007/bf00992539

Schiavio, A., van der Schyff, D., Cespedes-Guevara, J., & Reybrouck, M. (2017). Enacting musical emotions. sense-making, dynamic systems, and the embodied mind. *Phenomenology and the Cognitive Sciences*, *16*(5), 785–809. https://doi.org/10.1007/s11097-016-9477-8

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. https://doi.org/10.1002/1099-0984(200007/08)14:4<325::aid-per380>3.0.co;2-i

Schubert, E. (1999). Measuring emotion continuously: Validity and reliability of the two-dimensional emotion-space. *Australian Journal of Psychology*, *51*(3), 154–165. https://doi.org/10.1080/00049539908255353

Schubert, E. (2004). Modeling perceived emotion with continuous musical features. *Music Perception*, *21*(4), 561–585. https://doi.org/10.1525/mp.2004.21.4.561

Schubert, E. (2009). The fundamental function of music. *Musicae Scientiae*, *13*(Special issue), 63–81. https://doi.org/10.1177/1029864909013002051

Schubert, E. (2010). Continuous self-report methods. In P. N. Juslin & J. A. Sloboda (Eds.), *Handbook of music and emotion: Theory, research, applications* (pp. 223–254). Oxford University Press.

Schubert, E. (2013). Reliability issues regarding the beginning, middle and end of continuous emotion ratings to music. *Psychology of Music*, *41*(3), 350–371. https://doi.org/10.1177/0305735611430079

Schutz, M., Huron, D., Keeton, K., & Loewer, G. (2008). The happy xylophone: Acoustics affordances restrict an emotional palate. *Empirical Musicology Review*, *3*(3), 126–135. https://doi.org/10.18061/1811/34103

Sethares, W. A. (1993). Local consonance and the relationship between timbre and scale. *The Journal of the Acoustical Society of America*, *94*(3), 1218–1228. https://doi.org/10.1121/1.408175

Sethares, W. A. (1998). Consonance-based spectral mappings. *Computer Music Journal*, *22*(1), 56–72. https://doi.org/10.2307/3681045

Shaver, P., Schwartz, J., Kirson, D., & O'Connor, C. (1987). Emotion knowledge: Further exploration of a prototype approach. *Journal of Personality and Social Psychology*, *52*(6), 1061–1086. https://doi.org/10.1037/0022-3514.52.6.1061

Shepherd, J., & Wicke, P. (1997). *Music and cultural theory*. Polity Press.

Siedenburg, K. (2017). Instruments unheard of: On the role of familiarity and sound source categories in timbre perception. In T. Bovermann, A. de Campo, H. Egermann, S.-I. Hardjowirogo, & S. Weinzierl (Eds.), *Musical instruments in the 21st century*. Springer.

Siedenburg, K., Jones-Mollerup, K., & McAdams, S. (2016). Acoustic and categorical dissimilarity of musical timbre: Evidence from asymmetries between acoustic and chimeric sounds. *Frontiers in Psychology*, *6*, 1977. https://doi.org/10.3389/fpsyg.2015.01977

Siedenburg, K., & McAdams, S. (2017). Four distinctions for the auditory "wastebasket" of timbre. *Frontiers in Psychology*, *8*, 1747. https://doi.org/10.3389/fpsyg.2017.01747

Sloboda, J. A. (1985). *The musical mind: The cognitive psychology of music*. Oxford University Press.

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

Soden, K., Saks, J., & McAdams, S. (2019). *Timbral, temporal, and expressive shaping of musical material and their respective roles in musical communication of* *affect* [Paper presented at the Auditory Perception, Cognition, and Action Meeting, Montréal, QC].

Spencer, H. (2021). The origin and function of music. In *Essays: Scientific, political and speculative* (pp. 400–451). Routledge. https://doi.org/10.4324%2F9781003191834-15

Spitzer, J., & Zaslaw, N. (2004). *The birth of the orchestra: History of an institution, 1650-1815*. Oxford University Press.

Stilp, C. E., Rogers, T. T., & Kluender, K. R. (2010). Rapid efficient coding of correlated complex acoustic properties. *Proceedings of the National Academy of Sciences*, *107*(50), 21914–21919. https://doi.org/10.1073/pnas.1009020107

Susino, M., & Schubert, E. (2017). Cross-cultural anger communication in music: Towards a stereotype theory of emotion in music. *Musicae Scientiae*, *21*(1), 60–74. https://doi.org/https://doi.org/10.1177/1029864916637641

Thompson, W. F., & Balkwill, L.-L. (2010). Cross-cultural similarities and differences. In P. N. Juslin & J. A. Sloboda (Eds.), *Handbook of music and emotion: Theory, research, applications* (pp. 755–788). Oxford University Press.

Thompson, W. F., Bullot, N. J., & Margulis, E. H. (2022). The psychological basis of music appreciation: Structure, self, source. *Psychological Review*, *130*(1), 260–284. https://doi.org/10.1037/rev0000364

Thoresen, L. (2015). *Emergent musical forms: Aural explorations*. University of Western Ontario.

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. https://doi.org/10.1037/0278-7393.30.5.1131

Tomlinson, G. (1984). The web of culture: A context for musicology. *19th-Century Music*, *7*(3), 350–362. https://doi.org/10.2307/746387

Vines, B. W., Nuzzo, R. L., & Levitin, D. J. (2005). Analyzing temporal dynamics in music: Differential calculus, physics, and functional data analysis techniques. *Music Perception*, *23*(2), 137–152. https://doi.org/10.1525/mp.2005.23.2.137

Vuoskoski, J. K., & Eerola, T. (2011). Measuring music-induced emotion. *Musicae Scientiae*, *15*(2), 159–173. https://doi.org/10.1177/1029864911403367

Wallmark, Z., & Kendall, R. A. (2018). Describing sound. In *The Oxford handbook of timbre* (pp. 578–608). Oxford University Press. https://doi.org/10.1093%2Foxfordhb%2F9780190637224.013.14

Wang, C. (2021). Aleppo. *Resonances.* [Recorded by Taipei Chinese Orchestra].

Wang, X., Wei, Y., Heng, L., & McAdams, S. (2021). A cross-cultural analysis of the influence of timbre on affect perception in Western classical music and Chinese music traditions. *Frontiers in Psychology*, *12*, 732865. https://doi.org/10.3389/fpsyg.2021.732865

Wang, X., Wei, Y., & Yang, D. (2021). Cross-cultural analysis of the correlation between musical elements and emotion. *Cognitive Computation and Systems*, *4*(2), 116–129. https://doi.org/10.1049/ccs2.12032

Warrenburg, L. A. (2020). Comparing musical and psychological emotion theories. *Psychomusicology: Music, Mind, and Brain*, *30*(1), 1–19. https://doi.org/10.1037/pmu0000247

West, G. (Ed.). (1920). *Film folio no. 1.* The Boston Music Company.

Wobbrock, J. O., Findlater, L., Gergle, D., & Higgins, J. J. (2011). The aligned rank transform for nonparametric factorial analyses using only ANOVA procedures. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 143–146. https://doi.org/10.1145/1978942.1978963

Wold, S., Sjöström, M., & Eriksson, L. (2001). PLS-regression: A basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, *58*(2), 109–130. https://doi.org/10.1016/s0169-7439(01)00155-1

Zacharakis, A., Pastiadis, K., & Reiss, J. D. (2012). An interlanguage study of musical timbre semantic dimensions and their acoustic correlates. *Music Perception*, *31*(4), 339–358. https://doi.org/10.1525/mp.2014.31.4.339

Zhang, J., & Xie, L. (2017). Analysis of timbre perceptual discrimination for Chinese traditional musical instruments. *2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, 1–4.

Appendix A

Table A1

Four-way ANOVA on Rank Transformed Data for Note Accuracy

	df	*df*.res	*F*	*p*	partial η^2	signif.
Listener Group (ListGrp)	2	87	62.24	< .001	.59	***
Affective Intention (AffInt)	3	2001	158.31	< .001	.19	***
Instrument Culture (InstrCult)	1	2001	12.80	< .001	.006	***
Instrument Category (InstrCat)	2	2001	6.01	.0025	.006	**
ListGrp:AffInt	6	2001	44.96	< .001	.12	***
ListGrp:InstrCult	2	2001	1.05	.35	.001	
AffInt:InstrCult	3	2001	45.31	< .001	.06	***
ListGrp:InstrCat	4	2001	0.60	.66	.001	
AffInt:InstrCat	6	2001	33.59	< .001	.09	***
InstrCult:InstrCat	2	2001	7.14	< .001	.007	***
ListGrp:AffInt:InstrCult	6	2001	8.58	< .001	.02	***
ListGrp:AffInt:InstrCat	12	2001	7.28	< .001	.04	***
ListGrp:InstrCult:InstrCat	4	2001	0.38	.82	.0008	
AffInt:InstrCult:InstrCat	6	2001	18.28	< .001	.05	***
ListGrp:AffInt:InstrCult:InstrCat	12	2001	1.59	.089	.009	

Note. Signif. codes: *** .001 **.01 *.05

Table A2

Four-way ANOVA for Measure Accuracy

	df	*df*.res	*F*	*p*	partial η^2	signif.
Listener Group (ListGrp)	2	87	82.64	< .001	.66	***
Affective Intention (AffInt)	3	2001	185.23	< .001	.22	***
Instrument Culture (InstrCult)	1	2001	3.54	.06	.002	
Instrument Category (InstrCat)	2	2001	18.02	< .001	.02	***
ListGrp:AffInt	6	2001	33.05	< .001	.09	***
ListGrp:InstrCult	2	2001	0.09	.91	.00009	
AffInt:InstrCult	3	2001	38.72	< .001	.05	***
ListGrp:InstrCat	4	2001	1.76	.13	.004	
AffInt:InstrCat	6	2001	20.44	< .001	.06	***
InstrCult:InstrCat	2	2001	7.77	< .001	.008	***
ListGrp:AffInt:InstrCult	6	2001	1.64	.13	.005	
ListGrp:AffInt:InstrCat	12	2001	4.91	< .001	.03	***
ListGrp:InstrCult:InstrCat	4	2001	0.48	.75	.001	
AffInt:InstrCult:InstrCat	6	2001	30.82	< .001	.08	***
ListGrp:AffInt:InstrCult:InstrCat	12	2001	8.15	< .001	.05	***

Note. Signif. codes: *** .001 **.01 *.05

Table A3

4-way ANOVA for Phrase Accuracy

	df	*df*.res	*F*	*p*	partial η^2	signif.
Listener Group (ListGrp)	2	87	50.45	< .001	.54	***
Affective Intention (AffInt)	3	2001	92.85	< .001	.12	***
Instrument Culture (InstrCult)	1	2001	0.22	.64	.0001	
Instrument Category (InstrCat)	2	2001	21.44	< .001	.02	***
ListGrp:AffInt	6	2001	11.20	< .001	.03	***
ListGrp:InstrCult	2	2001	3.65	.03	.004	*
AffInt:InstrCult	3	2001	19.50	< .001	.03	***
ListGrp:InstrCat	4	2001	1.02	.40	.002	
AffInt:InstrCat	6	2001	14.39	< .001	.04	***
InstrCult:InstrCat	2	2001	2.99	.05	.003	
ListGrp:AffInt:InstrCult	6	2001	3.42	.002	.01	**
ListGrp:AffInt:InstrCat	12	2001	3.86	< .001	.02	***
ListGrp:InstrCult:InstrCat	4	2001	4.93	< .001	.01	***
AffInt:InstrCult:InstrCat	6	2001	18.66	< .001	.05	***
ListGrp:AffInt:InstrCult:InstrCat	12	2001	3.06	< .001	.02	***

Note. Signif. codes: *** .001 **.01 *.05

Table A4

Post Hoc Comparisons for Instrument Category Across All Listener Groups and Contexts

Instrument Category	*df*	*t* **ratio**	*p* **value**
Low-arousal/negative-valence (L-)			
Bow,Measure - Pluck,Measure	1506	5.02	< .001
Bow,Measure - Wind,Measure	1506	4.70	< .001
Bow,Note - Pluck,Note	1506	5.27	< .001
Bow,Note - Wind,Note	1506	6.65	< .001
Bow,Phrase - Pluck,Phrase	1506	6.99	< .001
Bow,Phrase - Wind,Phrase	1506	4.63	< .001
Pluck,Measure - Wind,Measure	1506	–0.32	1.0000

Continued on next page

Table A4 – continued from previous page

Instrument Category	*df*	*t* **ratio**	*p* **value**
Pluck,Note - Wind,Note	1506	1.37	1.0000
Pluck,Phrase - Wind,Phrase	1506	–2.36	.1831
High-arousal/negative-valence (H-)			
Bow,Measure - Pluck,Measure	1506	–1.14	1.0000
Bow,Measure - Wind,Measure	1506	–6.03	< .001
Bow,Note - Pluck,Note	1506	2.70	.1327
Bow,Note - Wind,Note	1506	–1.39	1.0000
Bow,Phrase - Pluck,Phrase	1506	2.42	.2689
Bow,Phrase - Wind,Phrase	1506	–3.28	.0232
Pluck,Measure - Wind,Measure	1506	–4.89	< .001
Pluck,Note - Wind,Note	1506	–4.09	.001
Pluck,Phrase - Wind,Phrase	1506	–5.70	< .001
High-arousal/positive-valence (H+)			
Bow,Measure - Pluck,Measure	1506	–1.45	.58
Bow,Measure - Wind,Measure	1506	–4.96	< .001
Bow,Note - Pluck,Note	1506	–1.26	.63
Bow,Note - Wind,Note	1506	–4.79	< .001
Bow,Phrase - Pluck,Phrase	1506	–2.56	.06
Bow,Phrase - Wind,Phrase	1506	–8.59	< .001
Pluck,Measure - Wind,Measure	1506	–3.50	.004
Pluck,Note - Wind,Note	1506	–3.54	.004
Pluck,Phrase - Wind,Phrase	1506	–6.03	< .001
Low-arousal/positive-valence (L+)			
Bow,Measure - Pluck,Measure	1506	–3.80	.004
Bow,Measure - Wind,Measure	1506	–2.90	.07
Bow,Note - Pluck,Note	1506	–5.63	< .001
Bow,Note - Wind,Note	1506	–4.05	.001
Bow,Phrase - Pluck,Phrase	1506	–6.05	< .001
Bow,Phrase - Wind,Phrase	1506	–5.63	< .001
Pluck,Measure - Wind,Measure	1506	0.91	1.0000
Pluck,Note - Wind,Note	1506	1.57	1.0000
Pluck,Phrase - Wind,Phrase	1506	0.42	1.0000

Table A5

Post Hoc Comparisons for Listener Groups Across All Instruments for Note Context

Affective intention: low-arousal/negative-valence (L-)	*df*	*t* ratio	*p* value
CHM - WM	87	2.69	.03
CHM - NM	87	1.84	.14
WM - NM	87	–0.85	.40
Affective intention: high-arousal/negative-valence (H-)			
CHM - WM	87	4.77	< .001
CHM - NM	87	8.20	< .001
WM - NM	87	3.44	< .001
Affective intention: high-arousal/positive-valence (H+)			
CHM - WM	87	0.38	.70
CHM - NM	87	1.86	.20
WM - NM	87	1.48	.29
Affective intention: low-arousal/positive-valence (L+)			
CHM - WM	87	–1.39	.50
CHM - NM	87	–1.26	.50
WM - NM	87	.13	.90

Table A6

Post Hoc Comparisons for Listener Groups Across All Instruments for Measure Context

Affective intention: low-arousal/negative-valence (L-)	*df*	*t* ratio	*p* value
CHM - WM	87	6.79	< .001
CHM - NM	87	4.88	< .001
WM - NM	87	–1.91	.06
Affective intention: high-arousal/negative-valence (H-)			
CHM - WM	87	3.59	< .001
CHM - NM	87	7.22	< .001
WM - NM	87	3.64	< .001
Affective intention: high-arousal/positive-valence (H+)			
CHM - WM	87	5.22	< .001
CHM - NM	87	5.13	< .001
WM - NM	87	–0.09	.93
Affective intention: low-arousal/positive-valence (L+)			
CHM - WM	87	–0.24	.83
CHM - NM	87	0.86	.83
WM - NM	87	1.10	.83

Table A7

Post Hoc Comparisons for Listener Groups Across All Instruments for Phrase Context

Affective intention: low-arousal/negative-valence (L-)	*df*	*t* ratio	*p* value
CHM - WM	87	4.79	< .001
CHM - NM	87	3.32	.003
WM - NM	87	–1.47	.14
Affective intention: high-arousal/negative-valence (H-)			
CHM - WM	87	4.27	< .001
CHM - NM	87	5.99	< .001
WM - NM	87	1.73	.09
Affective intention: high-arousal/positive-valence (H+)			
CHM - WM	87	2.98	.007
CHM - NM	87	4.55	< .001
WM - NM	87	1.56	.12
Affective intention: low-arousal/positive-valence (L+)			
CHM - WM	87	–0.59	.55
CHM - NM	87	1.41	.32
WM - NM	87	2.01	.14

Table A8

Post Hoc Comparisons for Affective Intentions Across All Listener Groups and Instruments for Each Context

Note	*df*	*t* ratio	*p* value
(H-) - (H+)	2001	14.04	< .001
(H-) - (L-)	2001	6.64	< .001
(H-) - (L+)	2001	20.49	< .001
(H+) - (L-)	2001	–7.41	< .001
(H+) - (L+)	2001	6.45	< .001
(L-) - (L+)	2001	13.86	< .001
Measure			
(H-) - (H+)	2001	5.97	< .001
(H-) - (L-)	2001	–0.90	0.37
(H-) - (L+)	2001	19.94	< .001
(H+) - (L-)	2001	–6.88	< .001
(H+) - (L+)	2001	13.97	< .001
(L-) - (L+)	2001	20.85	< .001
Phrase			
(H-) - (H+)	2001	–6.26	< .001
(H-) - (L-)	2001	–0.44	0.66
(H-) - (L+)	2001	10.14	< .001
(H+) - (L-)	2001	5.82	< .001
(H+) - (L+)	2001	16.40	< .001
(L-) - (L+)	2001	10.58	< .001

Table A9

Post Hoc ANOVAs for Listener Group × Instrument for Note Context

Affective Intention	*df*	*df*.**res**	*F*	*p*	**signif.**
$(L-)$ Valence					
ListGrp	2		12.51	< .001	***
Instr	5		10.94	< .001	***
ListGrp:Instr	10	522	6.73	< .001	***
$(L-)$ Arousal					
ListGrp	2		10.07	< .001	***
Instr	5		96.19	< .001	***
ListGrp:Instr	10	522	3.62	< .001	***
$(H-)$ Valence					
ListGrp	2		34.42	< .001	***
Instr	5		13.14	< .001	***
ListGrp:Instr	10	522	8.68	< .001	***
$(H-)$ Arousal					
ListGrp	2		96.19	< .001	***
Instr	5		50.81	< .001	***
ListGrp:Instr	10	522	1.03	0.42	
$(H+)$ Valence					
ListGrp	2		3.59	0.03	*
Instr	5		11.53	< .001	***
ListGrp:Instr	10	522	4.59	< .001	***
$(H+)$ Arousal					
ListGrp	2		27.35	< .001	***
Instr	5		38.68	< .001	***
ListGrp:Instr	10	522	0.61	0.81	
$(L+)$ Valence					
ListGrp	2		3.57	0.03	*
Instr	5		13.35	< .001	***
ListGrp:Instr	10	522	5.18	< .001	***

Continued on next page

Table A9 – continued from previous page

Affective Intention	*df*	*df*.**res**	*F*	*p*	**signif.**
($L+$) Arousal					
ListGrp	2		10.02	$< .001$	***
Instr	5		12.78	$< .001$	***
ListGrp:Instr	10	522	3.08	$< .001$	***

Note. Signif. codes: *** .001 **.01 *.05

Table A10

Post Hoc ANOVAs for Listener Group × Instrument for Measure Context

Affective Intention	*df*	*df*.**res**	*F*	*p*	**signif.**
($L-$) Valence					
ListGrp	2		52.64	$< .001$	***
Instr	5		18.28	$< .001$	***
ListGrp:Instr	10	522	5.47	$< .001$	***
($L-$) Arousal					
ListGrp	2		92.57	$< .001$	***
Instr	5		66.60	$< .001$	***
ListGrp:Instr	10	522	3.23	$< .001$	***
($H-$) Valence					
ListGrp	2		52.87	$< .001$	***
Instr	5		15.25	$< .001$	***
ListGrp:Instr	10	522	11.65	$< .001$	***
($H-$) Arousal					
ListGrp	2		95.13	$< .001$	***
Instr	5		32.25	$< .001$	***
ListGrp:Instr	10	522	4.26	$< .001$	***
($H+$) Valence					
ListGrp	2		38.41	$< .001$	***
Instr	5		21.02	$< .001$	***
ListGrp:Instr	10	522	5.89	$< .001$	***

Continued on next page

Table A10 – continued from previous page

Affective Intention	*df*	*df***.res**	*F*	*p*	**signif.**
(*H*+) Arousal					
ListGrp	2		35.12	< .001	***
Instr	5		29.83	< .001	***
ListGrp:Instr	10	522	1.47	0.15	
(*L*+) Valence					
ListGrp	2		19.64	< .001	***
Instr	5		19.09	< .001	***
ListGrp:Instr	10	522	4.52	< .001	***
(*L*+) Arousal					
ListGrp	2		80.90	< .001	***
Instr	5		32.66	< .001	***
ListGrp:Instr	10	522	3.01	0.001	**

Note. Signif. codes: *** .001 **.01 *.05

Table A11

Post Hoc ANOVAs for Listener Group × Instrument for Phrase Context

Affective Intention	*df*	*df***.res**	*F*	*p*	**signif.**
(*L*−) Valence					
ListGrp	2		12.25	< .001	***
Instr	5		18.78	< .001	***
ListGrp:Instr	10	522	3.19	< .001	***
(*L*−) Arousal					
ListGrp	2		40.60	< .001	***
Instr	5		42.92	< .001	***
ListGrp:Instr	10	522	1.55	0.12	
(*H*−) Valence					
ListGrp	2		36.93	< .001	***
Instr	5		16.76	< .001	***
ListGrp:Instr	10	522	8.44	< .001	***

Continued on next page

Table A11 – continued from previous page

Affective Intention	*df*	*df*.**res**	*F*	*p*	**signif.**
($H-$) Arousal					
ListGrp	2		50.76	$< .001$	***
Instr	5		13.59	$< .001$	***
ListGrp:Instr	10	522	4.32	$< .001$	***
($H+$) Valence					
ListGrp	2		28.94	$< .001$	***
Instr	5		16.12	$< .001$	***
ListGrp:Instr	10	522	3.35	$< .001$	***
($H+$) Arousal					
ListGrp	2		16.57	$< .001$	***
Instr	5		18.50	$< .001$	***
ListGrp:Instr	10	522	1.23	0.27	
($L+$) Valence					
ListGrp	2		8.33	$< .001$	***
Instr	5		21.20	$< .001$	***
ListGrp:Instr	10	522	5.05	$< .001$	***
($L+$) Arousal					
ListGrp	2		40.29	$< .001$	***
Instr	5		30.68	$< .001$	***
ListGrp:Instr	10	522	2.93	0.001	**

Note. Signif. codes: *** .001 **.01 *.05

Appendix B

Figure B1

Mean and Standard Deviation of Emotional Intensity Ratings by CHM

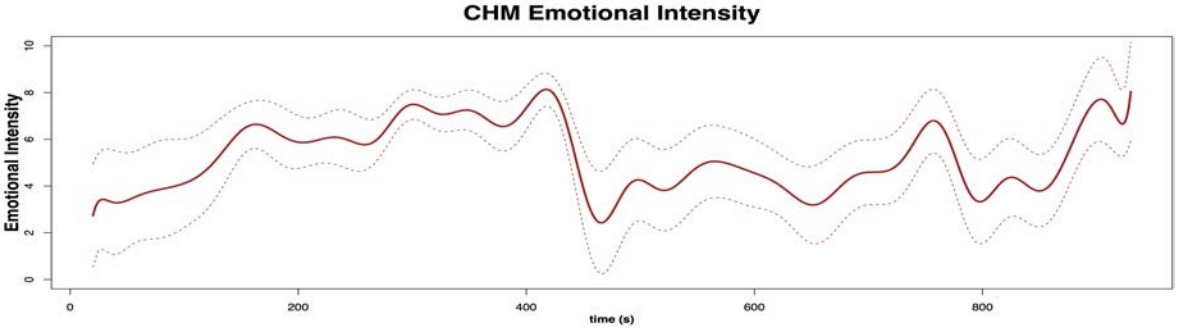


Figure B2

Mean and Standard Deviation of Emotional Intensity Ratings by WM

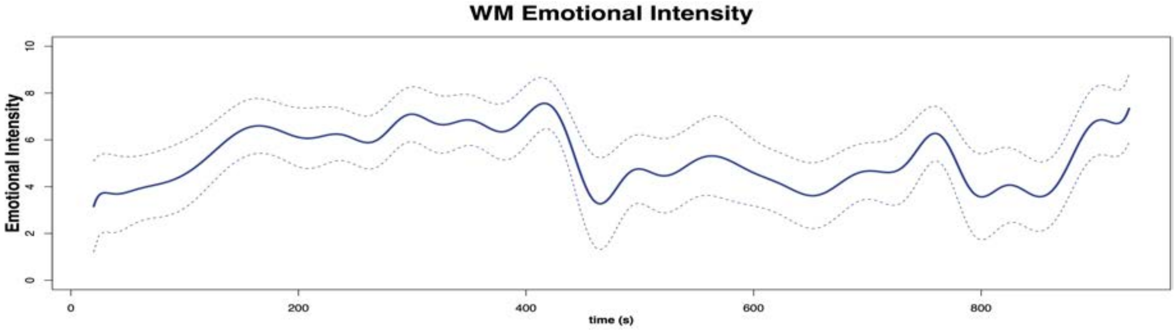


Figure B3

Mean and Standard Deviation of Emotional Intensity Ratings by NM

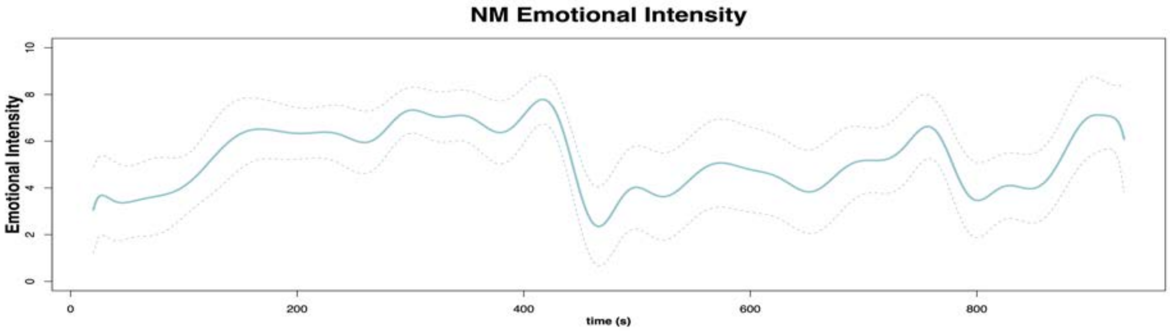


Figure B4

Mean and Standard Deviation of Arousal Ratings by CHM

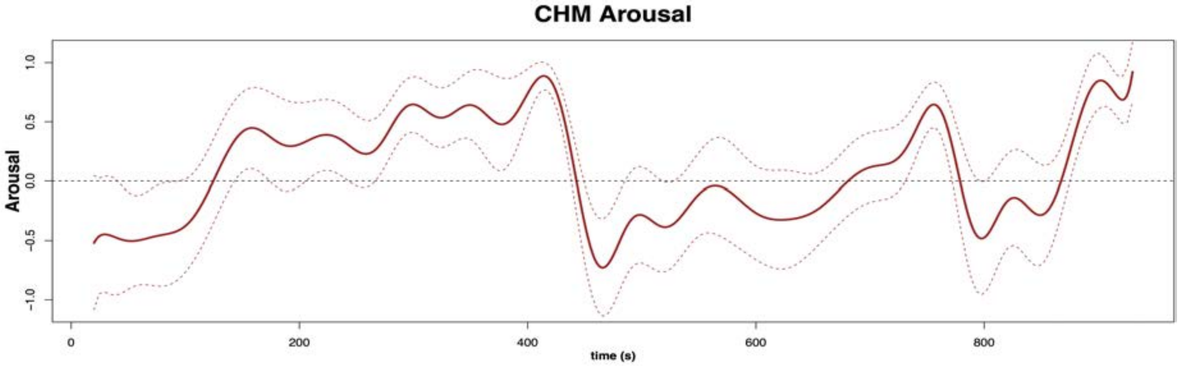


Figure B5

Mean and Standard Deviation of Arousal Ratings by WM

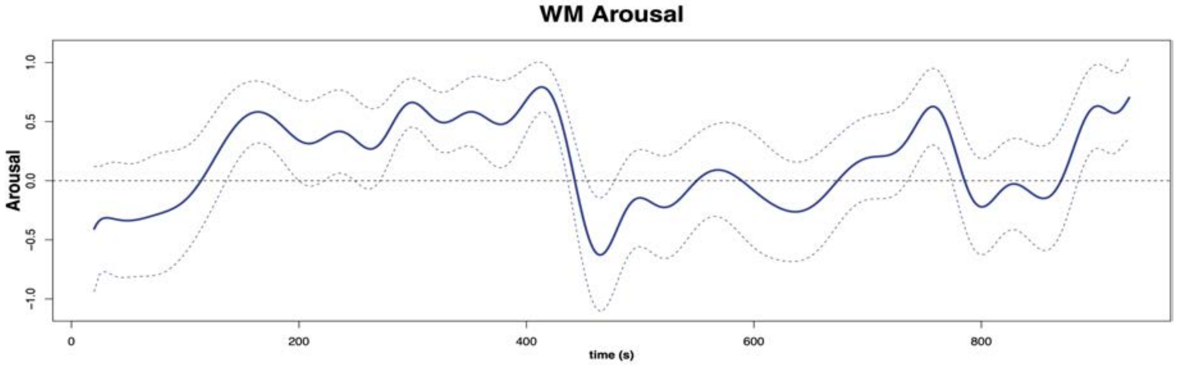


Figure B6

Mean and Standard Deviation of Arousal Ratings by NM

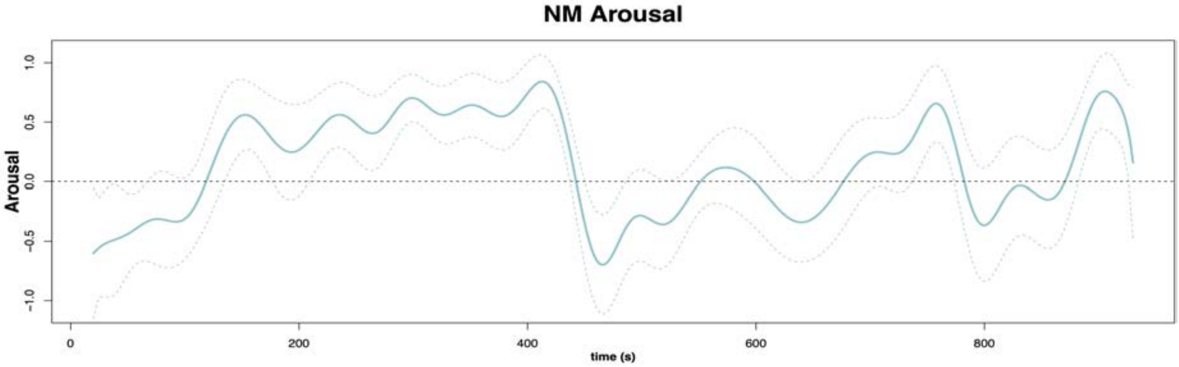


Figure B7

Mean and Standard Deviation of Valence Ratings by CHM

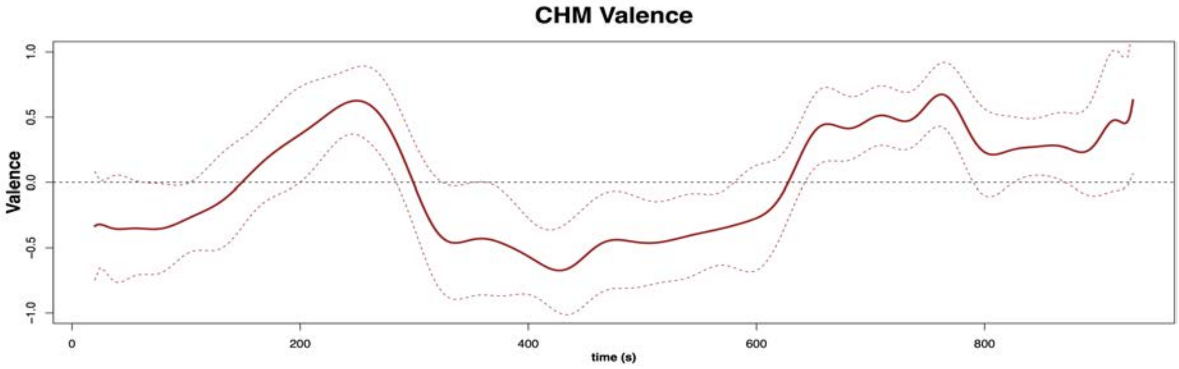


Figure B8

Mean and Standard Deviation of Valence Ratings by WM

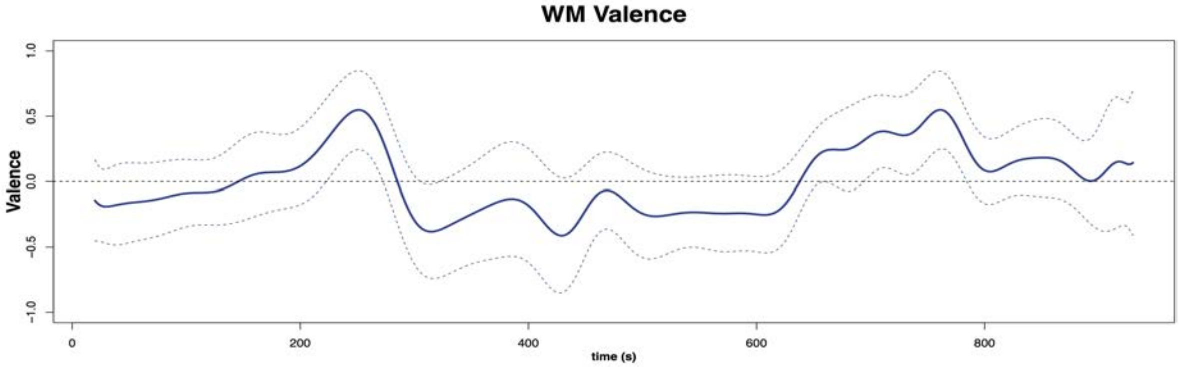


Figure B9

Mean and Standard Deviation of Valence Ratings by NM

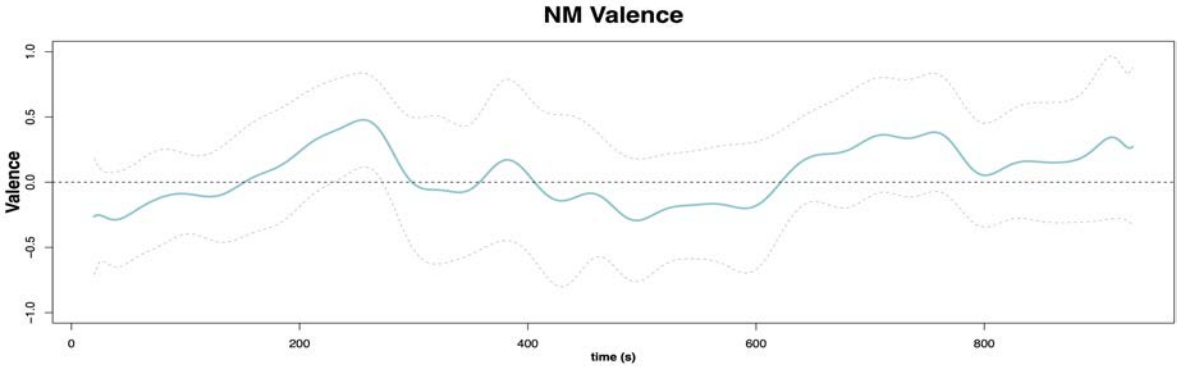
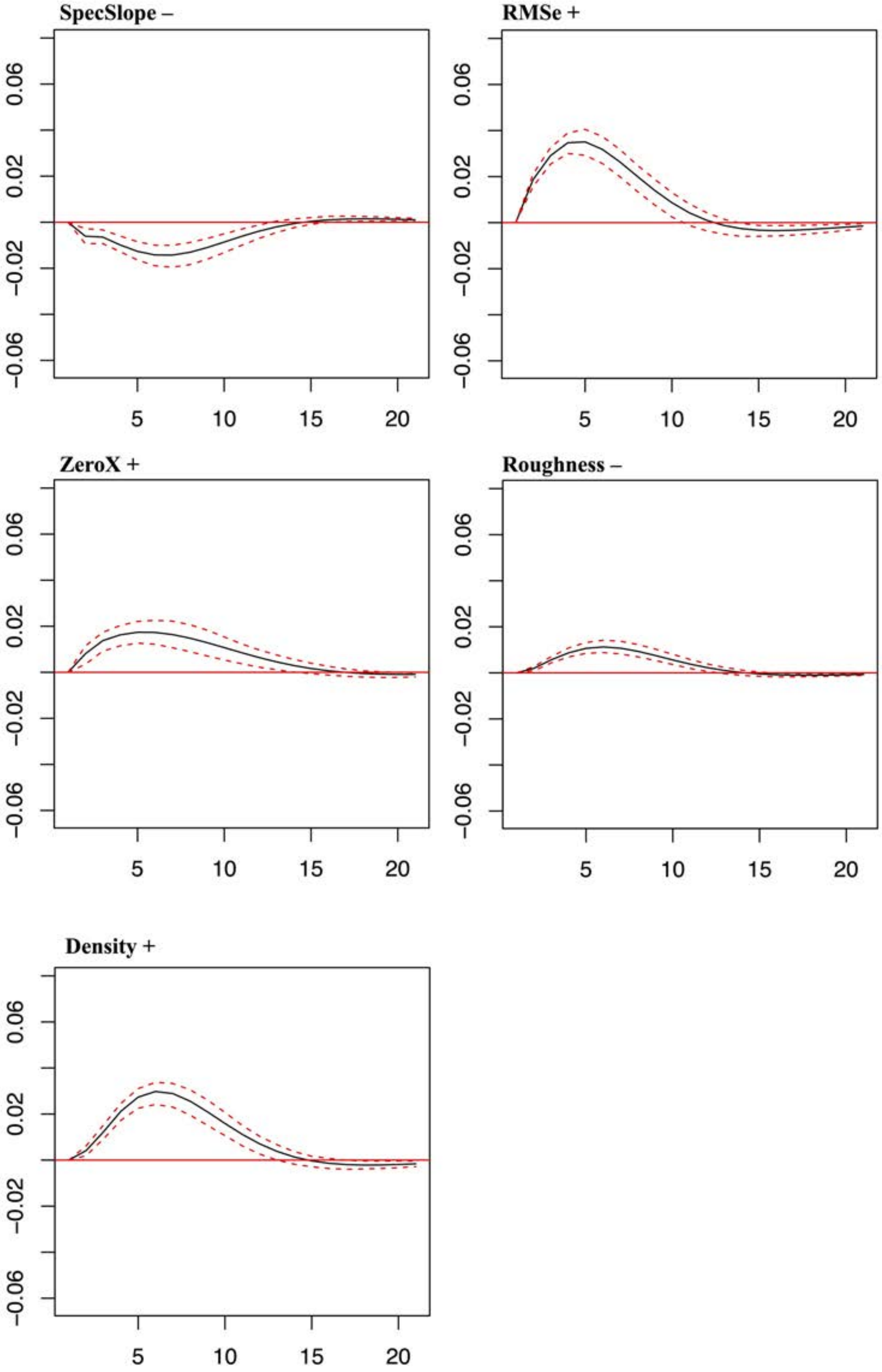


Figure B10

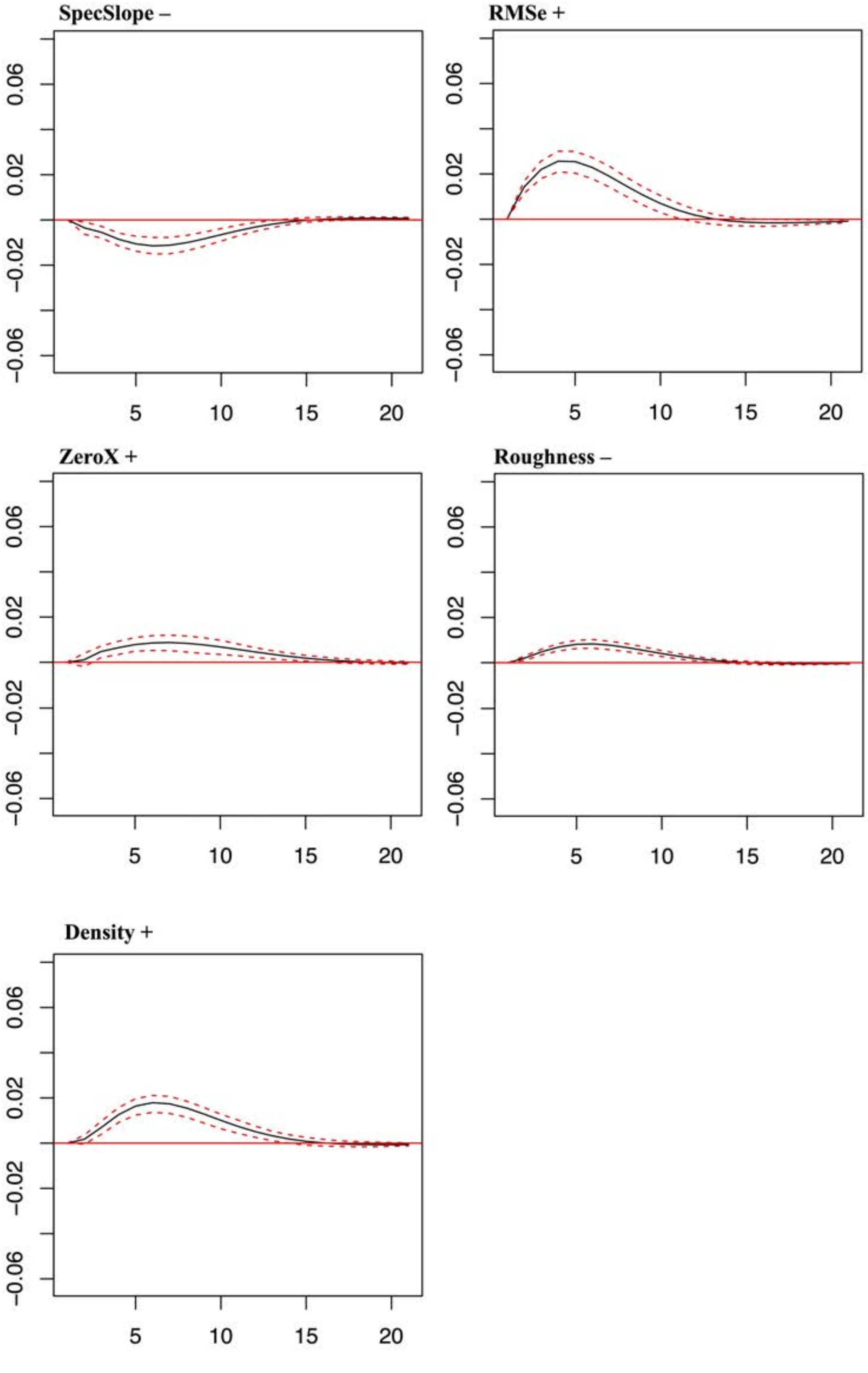
Impulse Response Functions for CHM Emotional Intensity Responses



Note. Only features with 95% C.I of impulse response functions breaching 0 are listed.

Figure B11

Impulse Response Functions for WM Emotional Intensity Responses



SpecSlope –

RMSe +

ZeroX +

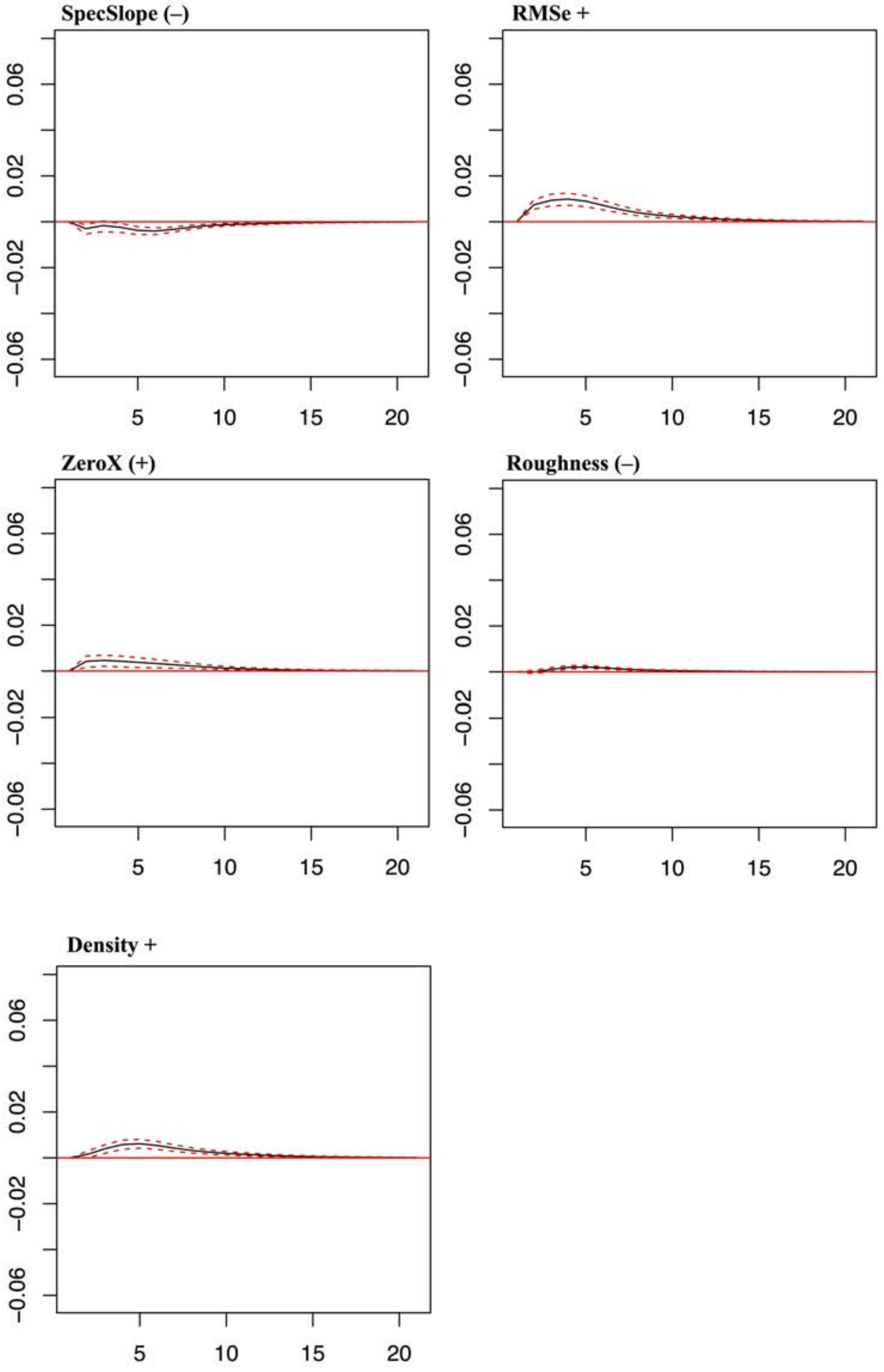
Roughness –

Density +

Note. Only features with 95% C.I of impulse response functions breaching 0 are listed.

Figure B12

Impulse Response Functions for NM Emotional Intensity Responses



Note. Only features with 95% C.I of impulse response functions breaching 0 are listed.

Figure B13

Impulse Response Functions for CHM Arousal Responses

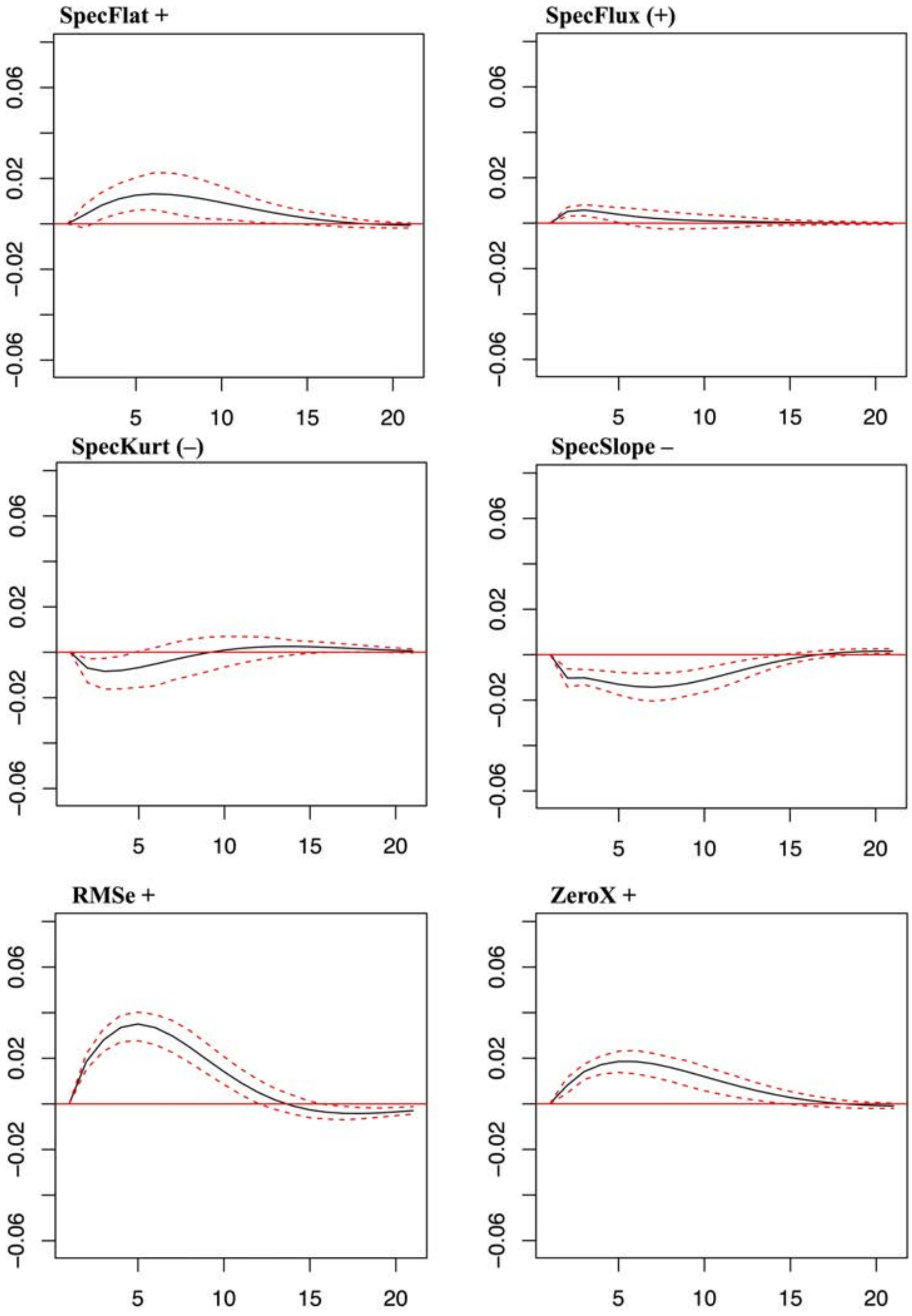
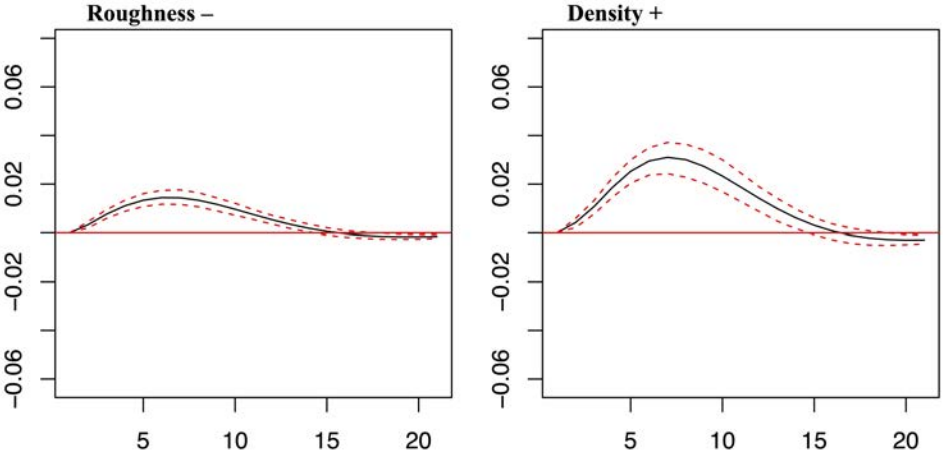


Figure B13 (cont.)

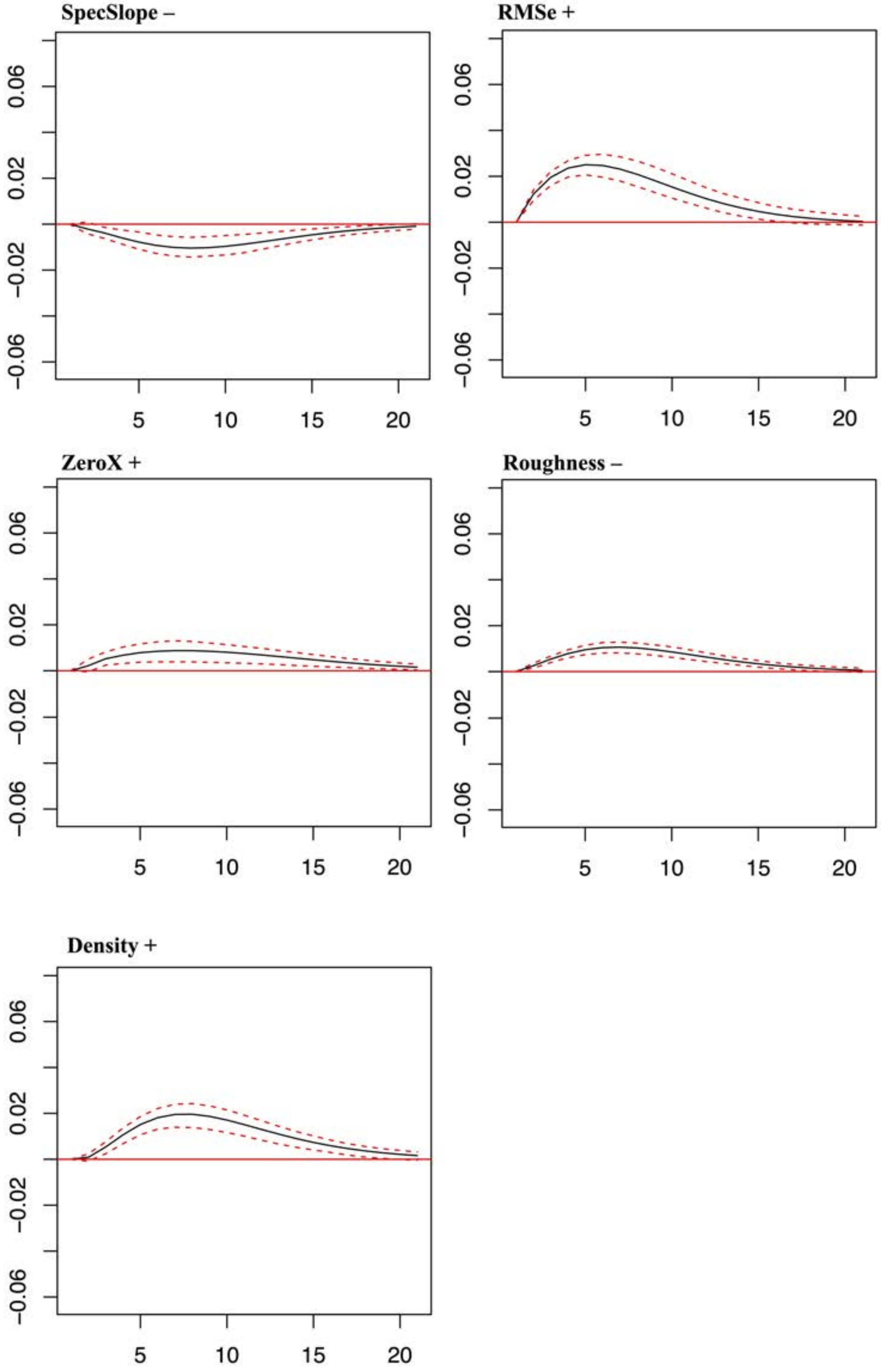
Impulse Response Functions for CHM Arousal Responses



Note. Only features with 95% C.I of impulse response functions breaching 0 are listed.

Figure B14

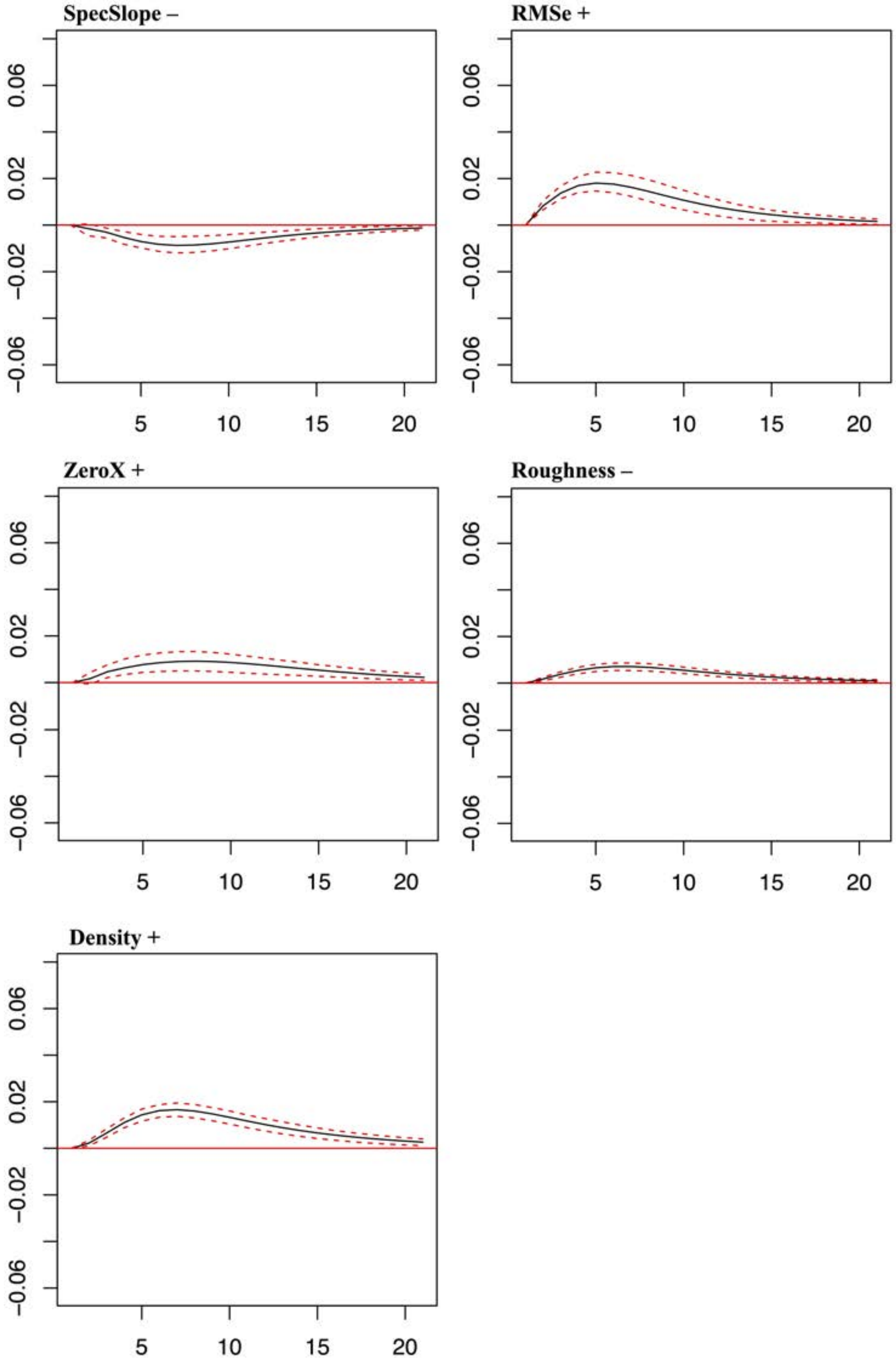
Impulse Response Functions for WM Arousal Responses



Note. Only features with 95% C.I of impulse response functions breaching 0 are listed.

Figure B15

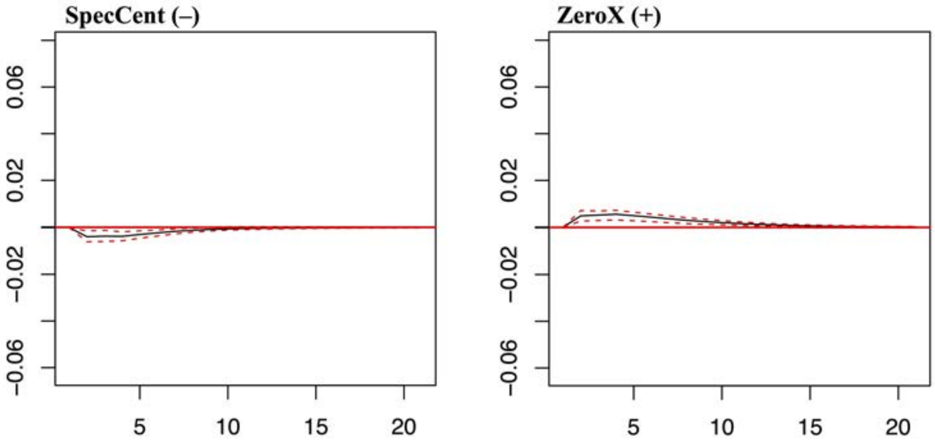
Impulse Response Functions for NM Arousal Responses



Note. Only features with 95% C.I of impulse response functions breaching 0 are listed.

Figure B16

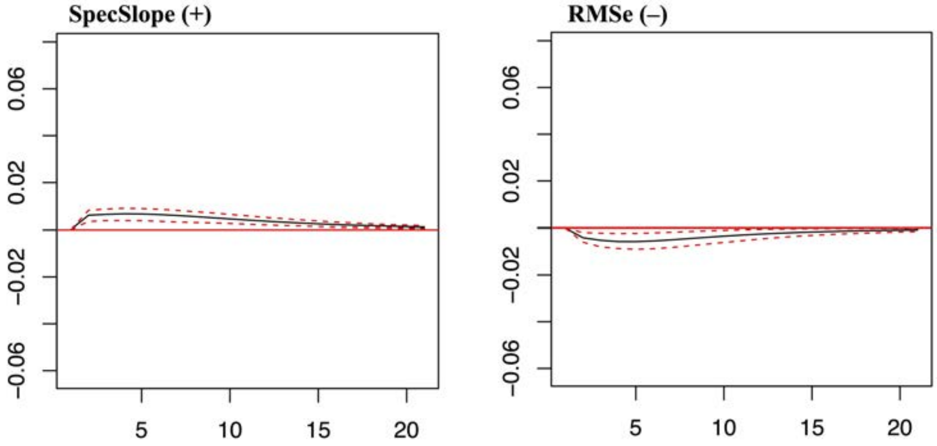
Impulse Response Functions for CHM Valence Responses



Note. Only features with 95% C.I of impulse response functions breaching 0 are listed.

Figure B17

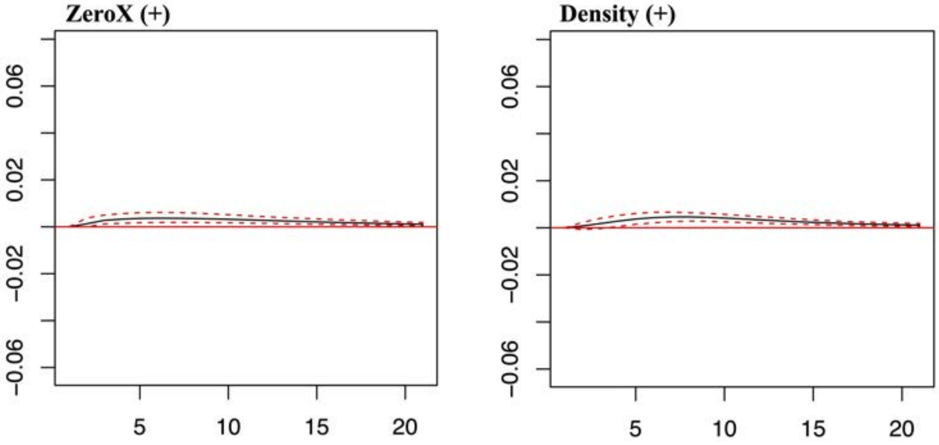
Impulse Response Functions for WM Valence Responses



Note. Only features with 95% C.I of impulse response functions breaching 0 are listed.

Figure B18

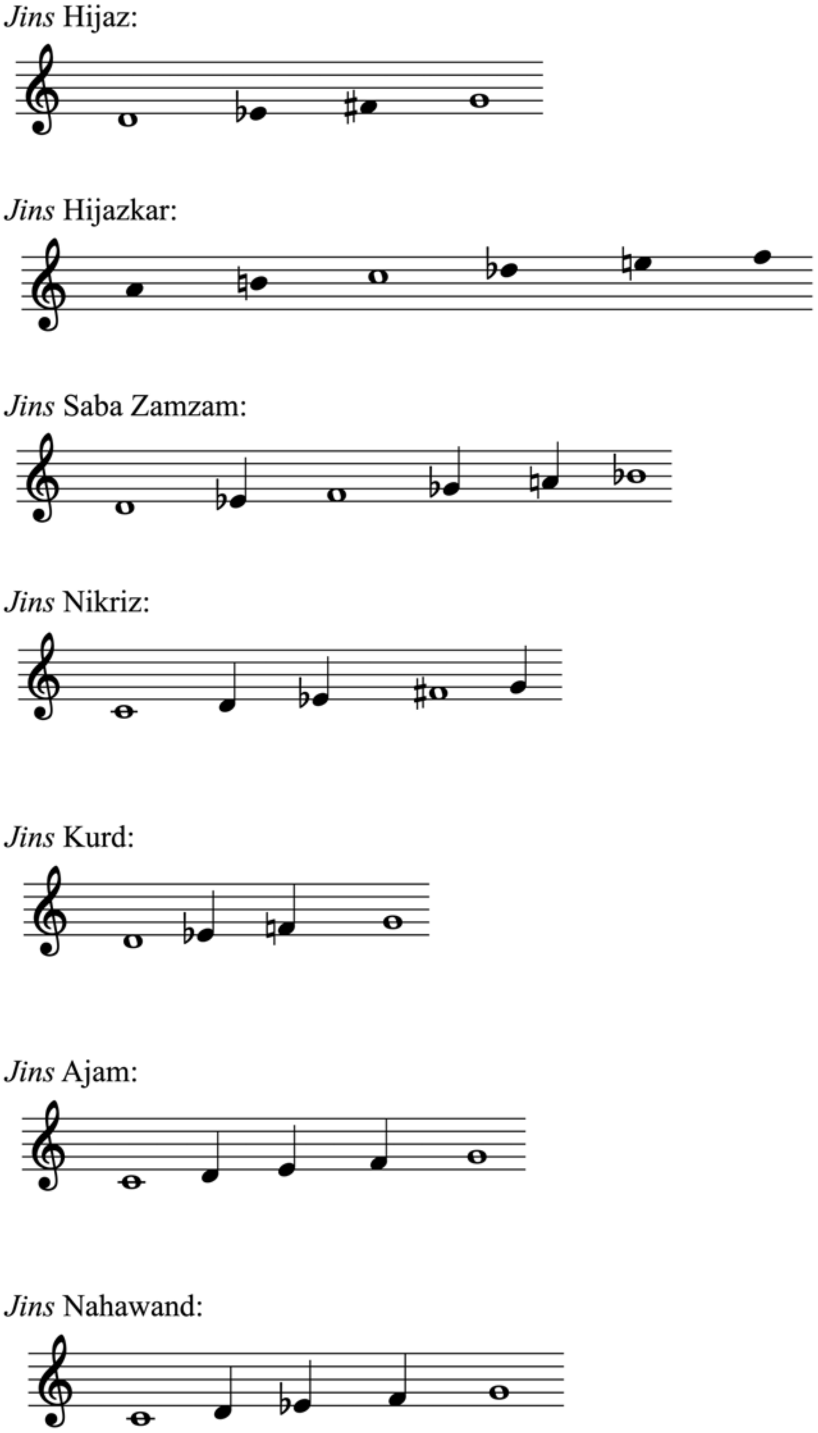
Impulse Response Functions for NM Valence Responses



Note. Only features with 95% C.I of impulse response functions breaching 0 are listed.

Figure B19

Ajnās Mentioned in Text



Jins Hijaz:

Jins Hijazkar:

Jins Saba Zamzam:

Jins Nikriz:

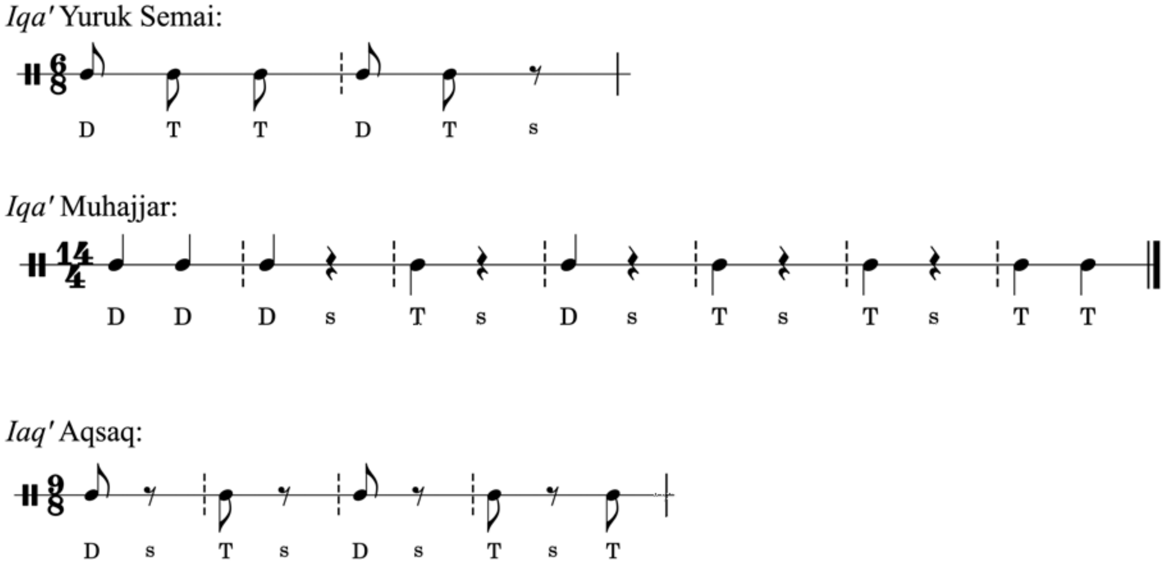
Jins Kurd:

Jins Ajam:

Jins Nahawand:

Figure B20

Iqā'āt Mentioned in Text



Iqa' Yuruk Semai:

Iqa' Muhajjar:

Iaq' Aqsaq: