

Affective Qualities of Sustained Instrumental Blends

Yifan Huang



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 of the degree of
Master of Arts.

© 2023 Yifan Huang

Abstract

Musical sounds can be combined into timbral blends with perceptual properties that result from the overall acoustic features of the mixture. We examine the affective qualities of blended sounds. Previous studies have found that instrumental blends can have a range of distinct timbral characteristics that are different from those of the constituent sounds, which makes the perceived affects of an instrumental blend unknown and requires further research. In our experiment, 40 participants listened to 45 blended unison pairs created from 10 sustained instruments at pitch D#4. They were asked to rate the perceived affect along three dimensions (valence, tension arousal, and energy arousal). They also rated the degree of blend for each blended pair in a separate block and completed a musical sophistication questionnaire at the end of the experiment.

Our findings showed that blending not only creates new sounds with distinct timbral properties but also evokes people's perception of affects that are different from those of the constituent sounds, and which are related to different audio descriptors than those of the constituent sounds. Blended sounds also span a broader range of the emotion space than the constituent sounds, so it may be helpful for musicians to use blends to express more varied emotions. A small set of acoustic features was useful to explain the affects of both the blended and individual sounds, and we found the variance over time of the acoustic features may play a more important role in the perception of the blended sounds than of individual sounds. In the three-dimensional affect space, the composite sounds are not simply in between the constituent sounds but in a triangular configuration. Some blends are within the emotional scope of the constituents, whereas others are beyond that scope. Also, some constituents dominate in their influence on the affect of the blend. We did not find a direct relationship between the degree of blend and perceived affect, but "good blenders" may tend to "soften" the sound and might thus lower the perception of arousal. Although blend in musical contexts is not always limited to the instrumental unison dyads used in the present study, this study does attest to the importance of timbral and orchestration features in conveying affect.

Résumé

Les sons musicaux peuvent être combinés pour former des mélanges timbrales dont les propriétés perceptives résultent des caractéristiques acoustiques globales du mélange. Nous examinons les qualités affectives des sons mélangés. Des études antérieures ont montré que les mélanges instrumentaux peuvent avoir une gamme de caractéristiques timbrales distinctes qui sont différentes de celles des sons constitutifs, ce qui rend l'émotion perçue d'un mélange instrumental inconnue et nécessite des recherches plus approfondies. Dans notre expérience, 40 participants ont écouté 45 paires d'unissons mélangés créés à partir de 10 instruments soutenus à la hauteur D#4. On leur a demandé d'évaluer l'affect perçu selon trois dimensions (valence, tension et énergie), ainsi que le degré de mélange pour chaque paire mélangée dans des blocs distincts. Ils ont également répondu à un questionnaire sur leur sophistication musicale à la fin de l'expérience.

Nos résultats ont montré que le mélange ne crée pas seulement de nouveaux sons avec des propriétés timbrales distinctes, mais évoque également la perception d'affects qui sont différents de ceux des sons constitutifs, qui sont par ailleurs liés à des descripteurs acoustiques différents de ceux des sons constitutifs. Les sons mélangés couvrent également une gamme plus large de l'espace émotionnel que les sons constitutifs, de sorte qu'il peut être utile pour les musiciens d'utiliser des mélanges pour exprimer des émotions plus variées. Un petit ensemble de caractéristiques acoustiques a permis d'expliquer les effets des sons mélangés et des sons individuels, et nous avons constaté que la variation au cours du temps des caractéristiques acoustiques peut jouer un rôle plus important dans la perception des sons mélangés que dans celle des sons individuels. Dans l'espace émotionnel tri-dimensionnel, les sons composites ne se situent pas simplement à un point intermédiaire entre les sons constitutifs, mais dans une configuration triangulaire. Certains mélanges se situent dans le champ émotionnel des constituants, tandis que d'autres se situent au-delà de ce champ. De même, certains constituants dominent dans leur influence sur l'affect du mélange. Nous n'avons pas trouvé de relation directe entre le degré de mélange et l'affect perçu, mais les « bons mélangeurs » peuvent avoir tendance à « adoucir » le son et donc à diminuer la perception de l'activation. Bien que le mélange dans les contextes musicaux ne soit pas toujours limité aux dyades instrumentales à l'unisson utilisées dans la présente étude, celle-ci atteste de l'importance des caractéristiques timbrales et orchestrales dans la transmission de l'affect.

Acknowledgements

I never thought that I could have such a precious two years and meet so many wonderful people here. I want to express my greatest appreciation to Professor Stephen McAdams, who has been very patient in guiding my thesis for the past two years, even though I was a novice in psychoacoustics at the beginning. I still remember the days when I sent an email to Professor McAdams to ask whether I could be his student and how happy I was when I got confirmed. As an international student living abroad for the first time, I have to admit that these two years were not easy for me. Professor McAdams always encourages and trusts me, making me more confident in myself. He is also very thoughtful, and I still remember I was so grateful that when I was so very nervous that I didn't understand the questions very well and couldn't find a correct English word to explain during my presentation. He always helped me to explain my thoughts to the audience. He also took care of me, not only for research but also for life, and he always reaches out to me when I need help. He is not only my academic mentor but also my life mentor. I have learned a lot from his conduct and approach to life. It is him who makes me feel like MPCL is my home while being abroad, and the people at MPCL have become my family.

Yes, MPCL is the place where every day, I am amazed at how wonderful everyone is! First, I want to thank Bennett Smith, who has been dedicated to the programming of my experiments, helped me set them up in MPCL, and helped me to prepare everything I needed to conduct experiments in China. He is always full of wisdom to solve the problems. His optimism and humor are also infecting me. Erica is my experimental mentor and my friend. She was very patient in teaching me all the details of how to conduct the experiments, for which I was so grateful. She was also very thoughtful in helping me integrate into MPCL when I was very shy at the beginning. I also want to express my great thanks to all our lab members and friends (Corinne, Linglan, Lena, Ben, Joshua, Andrés, Jade, André, Yuval, Félix, Iza), who are all always really, really kind and give me so much support and great suggestions not only in academics but also in life. There is so much that I could learn from everyone, and I am eager to see everyone being a great star in the future.

I am also grateful to all the professors and friends in the Music Technology area who broadened my horizon in this field and helped me a lot. I remember all the courses and discussions and all the talks in our kitchen. I would also like to thank my roommate, my best friend, Eto, and

my family and friends in China, who have been supportive and loving me. I will always honour the memory of my mentor Professor Shengchun Zhou of Morningside Scholars, who passed away last year but shed light on my life forever.

Everything I did, I owe to everyone here who helped me. I am so happy and so lucky that I can continue to be here for my future Ph.D. studies. I still remember that I wrote in my personal statement when applying for McGill University that “Music is my light.” Now I want to change that to: All of you are my light.

CONTENTS

Abstract	i
Résumé	ii
Acknowledgements	iii
Contents	v
List of Figures	ix
List of Tables	xi
1 Introduction	1
1.1 Music and Emotion.....	1
1.1.1 Locus of Emotion.....	2
1.1.2 Emotion Models.....	2
1.1.3 Emotion studies on Timbre.....	3
1.2 Instrumental Blends.....	5
1.2.1 Perceptual Studies on Instrumental Blends	5
1.2.2 Blend Space	6
1.3 Current Study	7
1.3.1 Motivation and Objectives	7
1.3.2 Research Questions and Hypothesis	8
1.3.3 Thesis Overview	8
2 Methodology	11
2.1 Experiments	11

2.1.1	Participants	11
2.1.2	Stimuli	11
2.1.3	Experimental design	14
2.1.4	Questionnaire	17
2.1.5	Apparatus.....	17
2.2	Acoustic description	17
2.2.1	Audio Representations.....	17
2.2.2	Hierarchical Cluster Analysis.....	18
3	Results and Analysis	23
3.1	Affective Analysis	24
3.1.1	Affective Analysis for Blended Pairs	24
3.1.2	Comparative Geometric Analysis.....	26
3.2	Acoustic Analysis	30
3.2.1	Regression Analysis	30
3.3	Analysis of the Role of Degree of Blend.....	37
3.3.1	Correlation analysis	37
3.3.2	Multidimensional Scaling of Blend Space	37
3.3.3	Geometric Analysis for Blend Space	40
3.3.4	Acoustic Analysis for Blend Space.....	41
3.3.5	Social Network Analysis.....	44
4	Discussion	47
4.1	Timbral blend and emotion	47
4.2	Relation of emotions of constituent sounds to composite sounds.....	48
4.3	Acoustics and emotion	48

4.4	The degree of blend.....	49
4.5	Emotion and the degree of blend.....	50
4.6	Musicians and nonmusicians	51
5	Conclusion	53
5.1	General conclusion.....	53
5.2	Future study	54
Appendix A:	Loudness-Matched Levels	55
Appendix B:	Onset Synchronization	57
Appendix C:	Participant Validation	59
Appendix D:	Assumption Check.....	61
Bibliography	63

LIST OF FIGURES

Figure 2.1 Valence-Energy Arousal plot of constituent sounds in the McAdams et al. (2017) emotion space.....	12
Figure 2.2 Valence-Tension Arousal plot of constituent sounds in the McAdams et al. (2017) emotion space.....	13
Figure 2.3 Emotion rating interface	15
Figure 2.4 Blend rating interface.	16
Figure 2.5 Hierarchical cluster dendrogram of individual (left) and blended sounds (right).	19
Figure 3.1 Scatter plot of mean energy arousal and valence for both composite sounds and constituent sounds. The values of the affective qualities of constituent sounds are taken from McAdams et al.'s (2017) data.....	27
Figure 3.2 Scatter plot of tension arousal and valence for both composite sounds and constituent sounds. The values of the affective qualities of constituent sounds are taken from McAdams et al.'s (2017) data.	28
Figure 3.3 Geometric configuration types.....	28
Figure 3.4 Valence-Energy arousal plot for Trumpet-Trombone and Oboe-Tuba blended pairs.	29
Figure 3.5 Graphic representation of dominant difference and intermediateness.....	30
Figure 3.6 Trace plot of cross-validation curve for lasso regression of valence for blended sounds.	32
Figure 3.7 Trace plot of cross-validation curve for lasso regression of tension arousal for blended sounds.....	32
Figure 3.8 Trace plot of cross-validation curve for lasso regression of energy arousal for blended sounds.....	33
Figure 3.9 Trace plot of cross-validation curve for lasso regression of valence for individual sounds.....	34

Figure 3.10 Trace plot of cross-validation curve for lasso regression of tension for individual sounds.....	35
Figure 3.11 Trace plot of cross-validation curve for lasso regression of energy for individual sounds.....	35
Figure 3.12 Trace plot of cross-validation curve for lasso regression of dominant difference.	36
Figure 3.13 AIC for different dimensionalities in Identity MDS.	38
Figure 3.14 R^2 for different MDS dimensionalities of Identity, INDSCAL, and IDIOSCAL algorithms.	39
Figure 3.15 MDS blend space.....	39
Figure 3.16 The degree of blend for each pair.	40
Figure 3.17 The sum of blend distances.....	41
Figure 3.18 Blend space with colour indicating the instrumental family along Dimension 1.....	42
Figure 3.19 Trace plot of cross-validation curve for lasso regression of first dimension of blend space.....	42
Figure 3.20 Trace plot of cross-validation curve for lasso regression of the second dimension. .	43
Figure 3.21 Social network representation of blend relationship.	45
Figure D.1 Q-Q plot for normality check of perceived valence for blended sounds.....	61
Figure D.2 Q-Q plot for normality check of perceived energy for blended sounds.....	61
Figure D.3 Q-Q plot for normality check of perceived tension for blended sounds.....	61

LIST OF TABLES

Table 2.1	Audio descriptors for composite sounds and constituent sounds.....	20
Table 3.1	Statistics for Cronbach's α and average inter-participant correlation.....	24
Table 3.2	Pearson correlation matrix for valence, tension, energy, and blend.	25
Table 3.3	ANCOVA result for valence, tension, and energy.	26
Table 3.4	Linear regression coefficients and adjusted R^2 for blended sounds.....	33
Table 3.5	Linear regression coefficients and adjusted R^2 for individual sounds.....	34
Table 3.6	Linear regression coefficient and adjusted R^2 for dominant difference.....	36
Table 3.7	Linear regression coefficients of the first dimension of blend space.....	43
Table 3.8	Linear regression coefficients of the second dimension of blend space.....	43
Table A.1	Level adjustment applied to each constituent sound in loudness matching to the bassoon sound.	55
Table A.2	Sound pressure level for final stimuli.....	56
Table B.1	Time offsets between constituents of each blend pair (values are the number of milliseconds that a row's stimulus should be delayed to be in perceived in synchrony with a column's stimulus), e.g., EH precedes Vc by 21.05 ms, whereas flute is delayed relative to Vc by 10.20 ms.....	57
Table C.1	Cronbach's α table for individual participant reliability statistics. The second column "Cronbach's α if item dropped" shows how the overall Cronbach's α would change if one participant is removed from the dataset. The third column "Item-rest correlation" shows the correlation between each participant and the rest of the participants in the scale.....	59
Table D.1	Homogeneity of variance of perceived valence for blended sounds.	62
Table D.2	Homogeneity of variance of perceived energy for blended sounds.....	62
Table D.3	Homogeneity of variance of perceived tension for blended sounds.	62

1 INTRODUCTION

It is widely acknowledged that the emotional impacts are one of the greatest motivating factors for people to listen to music and engage in musical activities (Juslin & Laukka, 2004; Krumhansl, 2002; Sloboda & O'Neill, 2001). It is always a topic of interest for music listeners, performers, composers, and researchers. The relationship between emotional effects and music has been widely studied, especially for global structural music cues such as melody, harmony, tempo, and so on (Cespedes-Guevara & Eerola, 2018). Timbre, as a multidimensional music attribute, has recently caught the attention of researchers and has been studied in terms of its emotional impact over the years. Some acoustic correlates of timbre have already been used to interpret the listener's perceived emotion. At the same time, musical sounds can be combined into timbral blends with perceptual properties that result from the overall acoustic features of the mixture. Blend or the fusion of concurrent sounds is very common in music today, and composers also try to use these techniques to convey emotion. However, it is hard to find a study investigating the emotion of blends. The present study investigates the perceived affect of sustained instrumental blends. This section will report the background on emotion studies in music, especially on timbre, and the background of studies on blends. Finally, the motivation, research questions, and structure of the thesis will be illustrated.

1.1 Music and Emotion

Research has investigated the relationship between emotion and music. Many musical cues, such as pitch and tempo, have been studied in terms of their role in emotion perception (Cespedes-Guevara & Eerola, 2018). In this thesis, we prefer to use the term "affect" as it includes emotions, moods and also all evaluative or valenced (positive/negative) states (Juslin & Sloboda, 2010). Emotions are considered to be more intense and shorter-term, whereas moods are less intense and longer-term. The term "affect" used in this thesis is to take into account the fact that music can not only cause changes in emotion but may also lead to some fluctuations in mood, and no matter

whether an emotion, mood, or other evaluative states are involved in music listening, they all belong to affective response. At the same time, some related studies continue to use the terms “emotion” and “affect” interchangeably (e.g., McAdams et al., 2017), so these two terms are also used alternately in this thesis.

1.1.1 Locus of Emotion

The notion of “locus of emotion” was proposed by Evans and Schubert (2008) and indicates whether the study addresses felt (induced) emotions or perceived (recognized) emotions. A review by Eerola and Vuoskoski (2013) pointed out that the fundamental question of music and emotion studies—“How does music evoke emotions in listeners?”—can be broken down into separate questions. Among the questions, felt emotion would be mainly related to “What are the putative emotions induced by music?” In contrast, perceived emotion is related to “What are the emotions conveyed by music?” The reason we distinguish these two concepts is that sometimes listeners may not feel the same emotion as the emotion the music conveys (e.g., Vuoskoski & Eerola, 2017), and thus they need to be treated differently in research. In a word, perceived emotion refers to the ability of a listener to identify an emotion that has been conveyed without necessarily experiencing that emotion (Juslin & Västfjäll, 2008). In the present study, we focus on perceived emotion.

1.1.2 Emotion Models

A review by Eerola and Vuoskoski (2013) concluded from 251 studies that the theoretical models of emotion could be generally divided into four classes: discrete, dimensional, miscellaneous, and music-specific. The discrete model is related to the theory of basic emotions, which holds that all emotions can be derived from a finite set of basic emotions, including fear, anger, disgust, sadness, and happiness. They found that categories are easy to discriminate and to be explained and recognized. However, the choice of emotion categories may be limited, and it may be hard to compare the results across different studies because researchers don’t always use same set of categories. Miscellaneous models for emotion in music consist of a diverse group of emotion concepts, such as intensity, preference, similarity, tension, or any other concept that is closely associated with emotions in general. Researchers try to characterize the aspects that are ignored by individual models, but it is hard to find concepts that can provide a deep enough understanding of emotions in music. Music-specific models only employ the emotions and underlying factors

that are directly relevant for music, such as Zentner et al. (2008). It is encouraging that some models share many factors (feeling moved, nostalgic, relaxed, and enchanted) and provide uniquely aesthetic emotions, but there is no evidence that music-specific emotions would be easier to perceive or express than the common discrete emotions.

Dimensional models represent emotions as a mixture of core dimensions. The commonly used dimensional models have two or three dimensions. The most representative study adopting a two-dimensional model is Russell's circumplex model (Russell, 1980). This model includes two core bipolar dimensions: valence (describes the evaluation of the emotion, from pleasure to displeasure) and arousal (describes the activity of the emotion, from arousal to sleep), which are orthogonal and continuous in the affect space. However, Schimmack and Grob (2000) found that the two-dimensional model does not fit the data, and they suggested that the poor fit of two-dimensional model is due to the arousal dimension being poorly defined in the pleasure–arousal model. Therefore, they proposed a three-dimensional (3D) affect model including valence, tension arousal, and energy arousal (Schimmack & Grob, 2000). Tension arousal measures the affective state from tension to relaxation, and energy arousal measures it from awake to sleepiness. The 3D model provides a more thorough model to capture affects, as the tension arousal and energy arousal do not need to collapse into a single dimension. For example, excitement and astonishment could have positive valence and high energy arousal but excitement has low tension arousal whereas astonishment has high tension arousal, so it is hard to represent in a two-dimensional modal with only one arousal dimension. Also, a 3D model provides the possibility of collapsing to a two-dimensional models if there is a high correlation between two of the dimensions (Eerola & Vuoskoski, 2011). At the same time, the present study was designed to be compared to the results of McAdams et al. (2017), who also used a 3D model. Therefore, to include more subtle and broader emotions and be more flexible in analysis and comparison, the present study chose to adopt a 3D affect model to investigate the effect of blends.

1.1.3 Emotion studies on Timbre

According to Holmes (2011), it has been reported that performers and composers utilize timbre to convey their intended emotional expression to their audience. Timbre, a multidimensional acoustic attribute comprised of spectral, temporal, and spectrotemporal acoustic factors (McAdams et al.,

1995), has recently captured the interest of emotion researchers. The fact that people can perceive emotion from timbre is supported by a listener's ability to perceive emotion from extremely short sound samples. According to previous research conducted by Peretz et al. (1998) and Filipic et al. (2010), it has been observed that as little as 250 ms of a musical excerpt holds enough information to perceive emotion in a consistent manner across listeners. Bigand et al. (2005) also found that even a single note provides listeners with enough cues to form an emotional judgment. Moreover, musical expertise does not have a significant effect on the recognition of the music and the emotional evaluations based on minimal acoustic cues.

Based on this, many perceptual studies on the emotion of timbre have been conducted over the years. Juslin and Laukka (2004) found that timbre is related to certain discrete emotions; in general, bright sounds are associated with happiness, dull sounds with sadness, sharp sounds with anger, and soft sounds with both fear and tenderness. Eerola et al. (2012) systematically investigated the affect qualities of 110 isolated instrumental sounds with different timbres. They conducted three experiments with two sets of sounds (the second set was selected from the first one, and the acoustic features were manipulated) to investigate the role of timbre in the perception of affective dimensions in music. The rating scales included valence (pleasant/unpleasant), energy arousal (awake/tired), tension arousal (tense/relaxed), and preference (like/dislike). They confirmed that listeners were able to rate isolated instrument samples along the affect dimensions and found the affect structure in the experiments to be best represented by two dimensions (valence and energy arousal) due to high correlation between energy arousal and tension arousal. They found that valence was significantly correlated with the ratio of high-frequency to low-frequency (HF-LF) energy (brightness), spectral regularity (the degree of uniformity of the successive peaks of the spectrum), and sub-band no.6 flux (the frequency of fluctuation of the energy in the 800–1600 Hz frequency band) across the original instrumental sounds and the manipulated sounds. The energy was significantly correlated with the Ratio of HF-LF energy, spectral skewness (asymmetry of spectrum around spectral centroid), and temporal envelope centroid (percussive–sustained). According to the results, the acoustic features selected, as well as the affect ratings for sounds in both experiments, were mainly stable across the dimensions. They confirmed that a small set of acoustic features can be used to predict the listeners' ratings.

McAdams et al. (2017) investigated individual sounds in a 3D affective model of valence (displeasure-pleasure), energy arousal (tired-awake), and tension arousal (tension-relaxation)

across a broad range of pitch registers playing a D# pitch class. They found that changes in pitch are accompanied by significant changes in timbral properties and their corresponding perceived emotions.

However, to our knowledge most emotion studies on timbre have been done on individual sounds, and it is difficult to find an emotion study on instrumental blends. At the same time, the instrument selection process by composers and arrangers is careful and aimed at producing specific characteristics and emotional nuances within the musical framework (Schutz et al., 2008). It is time to pay attention to the perceived emotion of instrumental blends.

1.2 Instrumental Blends

1.2.1 Perceptual Studies on Instrumental Blends

Generally speaking, blend is very common in many non-solo musical pieces, and it has also been a deep problem for composers and researchers. It is always a fascinating topic because blending may create new timbres. Sandell (1991) concluded that a blend can augment existing timbres, soften timbres, invent timbres, and be used for timbral imitation. At the same time, Auditory Scene Analysis (Bregman, 1990) has also provided psychological support for the perception of concurrent sounds, as concurrent grouping principles result in the fusion of acoustic information into auditory events with emergent properties. McAdams (2019, p. 218) indicated that timbre emerges from this perceptual fusion into a single auditory event, which may be conceived as resulting in “virtual” sound source created from the blending of separate instrumental sounds. As for this new “virtual” sound, Sandell (1995) also indicated that instrumental blends could have a range of timbres that are distinct from those of their constituent sounds.

To evaluate blend, Sandell (1991) concluded from treatises and empirical studies that blend spans a continuum and is not an all-or-nothing effect, which means that a group of concurrent sounds could blend very well or poorly. Some instrumental pairs tend to blend very well, such as the trumpet and clarinet pair. They are more likely to fuse into one sound that may create an emergent timbre. Some instrumental pairs, like the violin and trumpet, may blend poorly, and listeners can probably recognize both sounds separately with equal clarity. To achieve a better blend, Bregman (1990) mentioned that onset synchrony can override other cues that promote segregation, thereby improving fusion. Sandell (1991) concluded from orchestration manuals that authors always emphasize the synchronized onsets for obtaining blend as well. McAdams (1984)

mentioned that concurrent grouping is affected by perceived coherent cues, such as onset synchrony, harmonicity, and coherent frequency and amplitude behaviour. Several of these cues are related to the Gestalt principle of common fate, in which the sounds that change in similar ways are likely to be perceived as coming from the same source (Bregman, 1990). The present study also used these coherent cues to create blend stimuli in the onset synchronization procedure.

The acoustic features of constituent and composite sounds can also influence how well they blend, in other words, the degree of blend. Sandell (1995) investigated the acoustic correlates of the degree of blend. He used synthesized instrumental tones at Eb4 from Grey's (1975) study to create 120 blended unisons. The results suggested that as the overall spectral centroid of composite sounds, the difference in attack time between constituent sounds, and the dissimilarity of temporal loudness increased, the blend worsened. Notably, he designed a rating scale with "oneness" and "twoness" at the end points, which also inspired our experiment design. Tardieu and McAdams (2012) reported that a better degree of the blend is related to lower spectral centroids and slower attacks for combinations of pitched impulsive and sustained sounds.

Based on these studies, my thesis will begin with relatively well-blended, sustained unison dyads and investigate the correlation between the degree of blend and perceived emotion. We wanted to test whether the degree of blend influences the perceived affect and what acoustic features underlying timbre perception are related to blend and emotion perception.

1.2.2 Blend Space

1.2.2.1 Multidimensional scaling (MDS)

Multidimensional scaling (MDS) is a useful technique to visualize and analyze the similarities or dissimilarities among sounds with different timbres. It models the dissimilarity ratings as Euclidean distances, and the output is a space with a small number of perceptual dimensions shared among listeners. Researchers have adopted this method to create timbre spaces to analyze the perceptual dimensions of timbre (such as Grey, 1977; Lakatos, 2000; McAdams et al., 1995). As with timbre space, MDS is also very useful for visualizing and analyzing blends. Kendall and Carterette (1991) asked participants to rate the similarity of blend dyads and adopted the MDS to create a 3D blend timbre space. They interpreted the three dimensions as nasality, richness, and complexity. Sandell (1991) used MDS to analyze blend ratings. He took the degree of blend as a measure of proximity and used MDS to analyze and create a two-dimensional blend space to

interpret the degree of blend among all the pairs, where the closer the instrument is, the better they blend. He found the first dimension to be correlated with perceptual attack time and pitch variability of the tones, and the second dimension to be correlated with spectral centroid and acoustic dissonance. Kendall and Carterette (1993) used a similar approach as Sandell (1991) to rate the degree of blend and create a blend space for unison pairs using MDS, which they found to be nearly identical to the blend space created by a musicologist based on a mental image of timbre. The present study also adopted this blend-rating method to investigate the blend space.

1.2.2.2 Social Network Analysis

Social Network Analysis (SNA) is a commonly used tool in social science to understand a community's social structure. Usually, a node represents an individual in the network, and the edge represents the relationship between individuals. To the best of our knowledge, there is no previous study using SNA to analyze the degree of the blend. We propose that blend is also a small “social relationship” between instrumental sounds, so we adopted SNA in the present study to uncover relations among the sounds in terms of the degree of blend as a supplement to MDS.

1.3 Current Study

1.3.1 Motivation and Objectives

As shown above, previous studies have shed light on the perceived emotion of isolated sounds. However, many timbral effects need to be studied to elucidate aspects of orchestration practice. We want to help answer the following questions: when composers combine different sounds or synthesize new sounds, what kind of sound could help convey the intended emotion? When the performer and conductor try to adjust the performance together with each other, what kind of resulting concurrent sound could help to express the emotion? Also, how do listeners perceive the emotion from the timbral effects? This study starts with relatively basic sustained timbral blends to try to uncover the emotion they convey.

Also, instrumental blend is probably the most fundamental building block for many orchestration techniques and sound effects. So, the more profound question is: can we extend our study of the emotional effect conveyed by unison blended sounds to more complicated orchestration techniques and more general sounds (such as vocal blend, synthesized sounds, and daily life sounds used in music)? Do the acoustic features provide us with an answer? Also, as we

have many thorough studies on the emotion of the timbres of individual sounds, can we use those timbres to predict the emotion of composite sounds? This study makes a first step toward answering this question.

1.3.2 Research Questions and Hypothesis

In this study, the main research question is to investigate the perceived affects of sustained instrumental blends. This question can be broken down into these sub-questions:

- Do different blends result in different perceived emotions?
- What is the relationship between the emotion of constituent sounds and composite sounds?
Can the constituent sounds be used to predict the emotion of composite sounds?
- What audio descriptors are correlated to listeners' perceived emotion of blended sounds?
- Will the degree of blend influence the perceived emotion?
- Will the musical sophistication of listeners influence the perceived emotion of timbre?

As shown in the previous study, blend creates a new “virtual sound source” (McAdams, 2019); therefore, we hypothesize that different blends are able to result in different perceived emotions. We also hypothesize that it is possible to use constituent sounds to predict the emotion of composite sounds. As in the previous study, a small group of audio descriptors can interpret a listener's perceived emotion, so we hypothesize this will also be the case for blended sounds. As the degree of blend is an essential perceptual feature of blend, we hypothesize that it might be a helpful feature in interpreting the perceived emotion. Moreover, for musical sophistication, we assume that will be a covariate that potentially influences the perceived emotion, with participants having higher musical sophistication being more sensitive to emotion perception and thus perceiving more extreme measures of valence and arousal.

1.3.3 Thesis Overview

Chapter 2, “Methodology,” describes the design of the experiment, stimuli, apparatus, and information about participants. Chapter 3, “Results and Analysis,” presents the experimental results and analyses, including ANCOVAs, lasso and linear regressions, geometric analysis, multidimensional scaling, and social network representation analysis. Chapter 4, “Discussion,” interprets the data in light of the hypotheses and discusses some observations based on the results.

The final chapter, "Conclusion," will provide a general conclusion on the current study and discuss some ideas for future studies.

2 METHODOLOGY

Experiments were conducted to test the hypothesis posed in 1.3.2. Participants rated the perceived affective qualities and the degree of blend for each stimulus in separate blocks. Subsequently, a questionnaire was administered to evaluate the musical sophistication of each participant. Acoustic descriptors were also extracted from all the sounds used in the experiment for further data analysis.

2.1 Experiments

2.1.1 Participants

Forty participants consisted of 23 females, 16 males, and 1 who preferred not to disclose their gender (age $M = 24.03$, $SD = 5.82$). We recruited the participants from either a mailing list at McGill University or a web-based advertisement in the general Montreal community. Before the experiment, participants passed a pure-tone audiometric test with octave-spaced frequencies from 125 Hz to 8000 Hz at a hearing threshold of 20 dB HL (ISO 389–8, 2004; Martin & Champlin, 2000). All participants passed the audiometric screening and were compensated for their participation. This study was certified for ethical compliance by the McGill University Research Ethics Board II.

2.1.2 Stimuli

Forty-five stimuli were used in the experiments. The blend stimuli were created from all different pairs of ten selected sustained individual sounds.

2.1.2.1 Constituent sustained sounds

The individual sound samples were carefully selected based on the research results in McAdams et al. (2017). As the pitch was shown to influence the emotion of sounds in their study, and our research is mainly concerned with blend and not pitch, we selected the sounds with the same pitch. To have as many instruments as possible and reasonable differences in the emotion space, we used all the sustained sounds at pitch D#4. Ten sustained sounds were selected: tuba (abbreviated as “Tu” in tables and figures in this thesis; same for the other instruments), tenor trombone (Tb), horn (Hn), trumpet (Tp), bassoon (Bn), Bb clarinet (Cl), oboe (Ob), English horn (EH), alto flute (Fl), and cello (Vc). The mean affective qualities of these sounds in the 3D emotion space from the experimental results across 40 participants of McAdams et al. (2017) are shown in Figure 2.1 (Valence vs. Energy Arousal) and Figure 2.2 (Valence vs. Tension Arousal).

The sounds were taken from the Vienna Symphonic Library (<https://vsl.co.at>). Audio signals were sampled at 44.1 kHz with 16-bit amplitude resolution. As in the study of McAdams et al. (2017), the stimuli were edited to have a consistent duration of 500 ms with a raised-cosine ramp applied to fade them out over the final 50 ms. The dynamic level was forte.

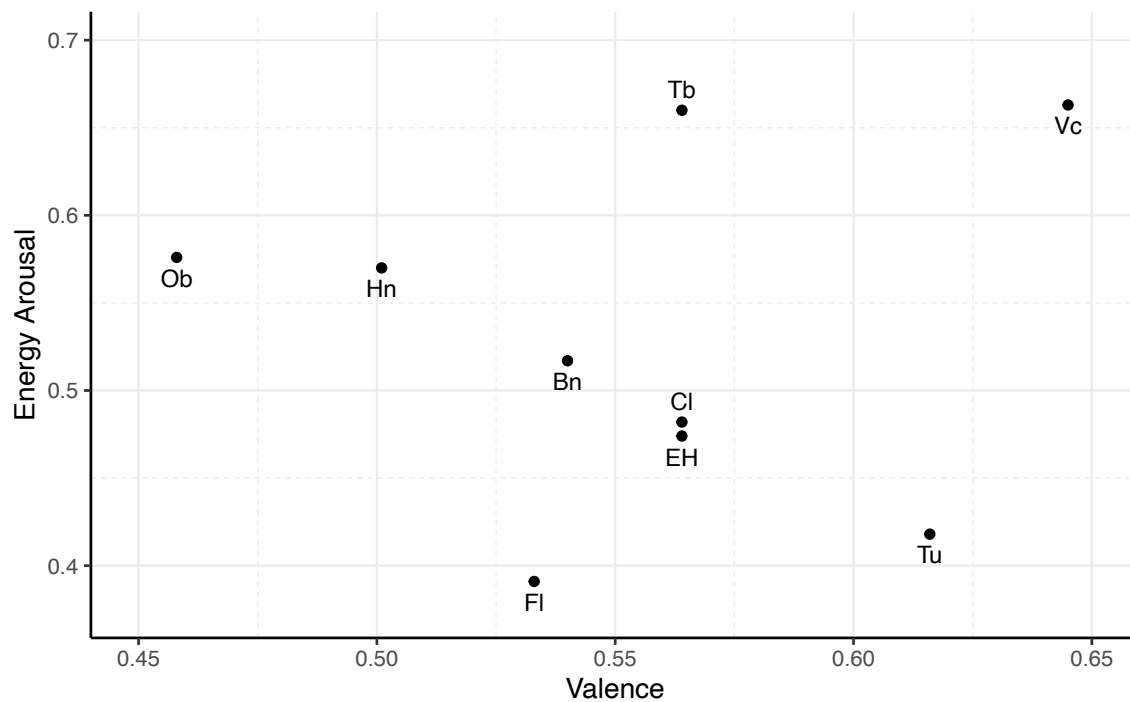


Figure 2.1 Valence-Energy Arousal plot of constituent sounds in the McAdams et al. (2017) emotion space.

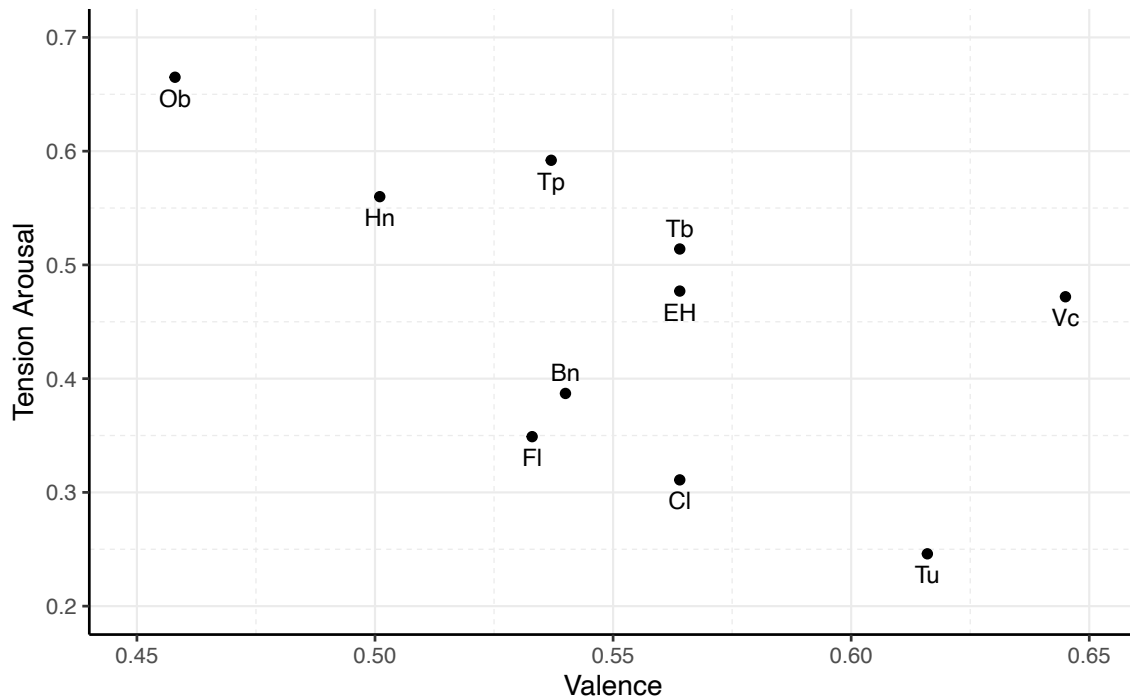


Figure 2.2 Valence-Tension Arousal plot of constituent sounds in the McAdams et al. (2017) emotion space.

2.1.2.2 Blend pairs

Loudness Equalization: As this study is mainly concerned with the emotion of unison blended sounds, other musical parameters that may affect the affective qualities should be controlled as much as possible. As we have already controlled the pitch, duration, and dynamics, another important factor is loudness. To equalize the perceived loudness, eight music researchers participated in a loudness-matching experiment in which they had to adjust the level of each individual sound to match that of the standard sound. The bassoon was selected as the standard sound, because we all agreed that it was the most comfortable sound that could be listened to over and over again in the loudness-matching process. The median adjusted sound level across these listeners was used to equalize the 10 individual sounds. (The values of loudness equalization applied to individual sounds are documented in Appendix A. Loudness-Matched Level.)

Onset Synchronization: As our study focuses mainly on the emotion of timbral blends, other factors influencing blend should be controlled. Besides the timbre itself, onset synchronization has proven to be a significant factor in perceptual fusion in auditory scene analysis (McAdams, 2019). Physically synchronizing onsets, however, does not guarantee perceptual attack synchronization due to differences in perceptual attack times between instruments (Gordon, 1987). Therefore, after

the loudness equalization, eight music researchers were invited to adjust the temporal offset of loudness-compensated unison dyads until they were perceived as maximally synchronized. The median offset time was used to generate the final 45 pairwise experimental stimuli. (The values of onset synchronization applied to generate the 45 stimuli are documented in Appendix B. Onset Synchronization)

We generated two versions of the stimuli: a dichotic version with each instrument in a different speaker and a monophonic version with a mix of both sounds coming from both speakers. We choose to use dichotic version in the main experiment rather than the mono version, because we found that with the dichotic version, it was much clearer to identify the degree of blend according to the feedback from our pilot study participants. It also avoided phase-interference effects present in the mono mix that distorted the timbres.

2.1.3 Experimental design

2.1.3.1 Procedure

The experiment consisted of two parts: affect ratings and blend ratings, followed by a questionnaire on musical sophistication (Gold-MSI; Müllensiefen et al., 2014). For each rating part, the experiment was a one-way repeated-measures design with a covariate. The repeated-measures factor was the 45 instrument pairs, and the covariate was the music sophistication score. The perceived Affect Rating and Blend Rating experiments were conducted sequentially with a short break in between.

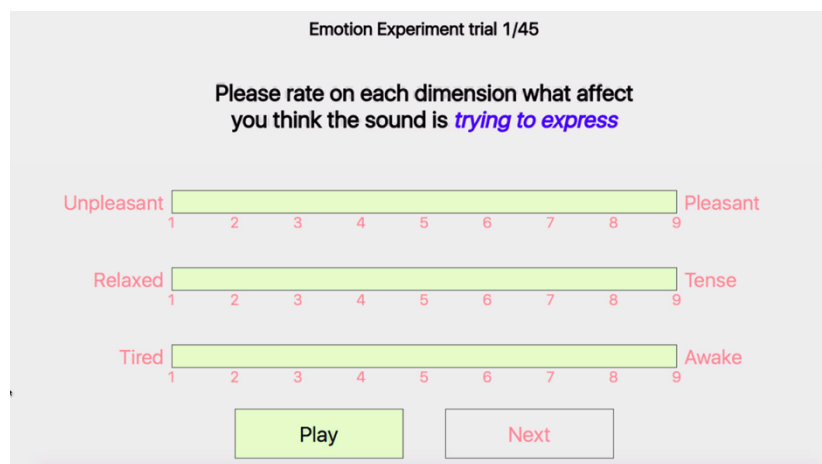
2.1.3.1.1 Experiment 1: Perceived Affect Ratings

Terminology was defined for the participants before beginning the experiment. To help participants differentiate perceived emotion from felt emotion, we defined perceived emotions as the emotions they think the music or sound is trying to convey or communicate to listeners. And we also mentioned that perceived emotion may or may not be the same as the emotions they are currently feeling in response to the sound. The inconsistency between perceived emotion and felt emotion was also illustrated to help participants understand the terminology. We then described the three dimensions of emotions as follows: valence describes the range of pleasantness on a rating scale from unpleasant on the left to pleasant on the right; tension arousal describes the degree of tension an emotion might have on a scale from relaxed to tense; and energy arousal describes the

amount of energy an emotion might have on a scale from tired to awake. We also provided concrete examples, such as happiness has a positive valence, but sadness has a negative valence, to help participants understand the terminology. To compare these results with the affective qualities of individual instrument timbres in McAdams et al. (2017), we used the same affective dimensions and scale endpoint labels.

Participants were instructed on the procedure of the experiment verbally and through written instructions. Following the instructions, they completed six practice trials to familiarize themselves with the interface, which followed the same format and paradigm as the experimental blocks. The practice stimuli were selected according to the results of a pilot study to cover the whole range of each dimension of emotion with the aim of helping participants establish an idea of the possible range of emotions. The experimenter was with the participants during the practice session, and they were allowed to pose questions of clarification regarding the procedure before beginning the experimental session.

The experimental session had one block of 45 trials. The 45 blended pair stimuli were played in random order for each participant. In each trial, the participant was instructed to play the stimulus by clicking on “Play”, and then they were asked to rate the perceived emotion along the three dimensions on sliders with a continuous scale from 1 to 9 (interface shown in Figure 2.3). Ratings were scaled to 0–1 for the following analyses. The order of the three dimensions was also random for each participant but consistent across trials for a given participant. Participants were allowed to play the stimuli a maximum of twice to refresh their memories and determine their ratings.



Emotion Experiment trial 1/45

Please rate on each dimension what affect you think the sound is *trying to express*

Unpleasant 1 2 3 4 5 6 7 8 9 Pleasant

Relaxed 1 2 3 4 5 6 7 8 9 Tense

Tired 1 2 3 4 5 6 7 8 9 Awake

Play Next

Figure 2.3 Emotion rating interface

2.1.3.1.2 Experiment 2: Blend Rating

Following a brief break, participants were given instructions for this experiment. To eliminate the ambiguity of the definition of blend, we defined it as a perceptual phenomenon characterizing the fusion of different sounds into a perceptual unity. Also, we mentioned that a strongly blended sound is perceived as coming from a single source and cannot be separated into more than one source. The more easily the two constituent sounds can be perceived as coming from separate sound sources, the lower the degree of blend. According to participants frequently asked questions, we also mentioned that the stimulus consists of two individual sounds, and “more than one” can be perceived as anywhere from one to two. After the definition, participants were shown two examples of instrumental pairs that were perceived as strongly blended (bassoon and tuba) and relatively less blended (oboe and tuba pair with oboe delayed by 60 ms), based on the results of the pilot study.

Similarly to the experiment on perceived affect rating, participants were instructed on the procedure of the experiment verbally and through written instructions. Before the experimental session, participants completed six practice trials with the experimenter to become familiar with the interface. The stimuli were selected according to the results of a pilot study, which tried to cover the whole range of degrees of blend.

The experimental session had one block of 45 trials with stimuli presented in random order for each participant. In each trial, the participants clicked “Play” to hear the stimulus, and then they were asked to rate how well the two sounds blended on a slider with a continuous scale from 0 to 8 (interface shown in Figure 2.4). Ratings were scaled to 0–1 for the following analyses. The stimuli could only be played once in each trial.



Figure 2.4 Blend rating interface.

2.1.4 Questionnaire

Based on our hypothesis that music sophistication may have a potential effect on perceived affect, we included music sophistication as the covariate. Therefore, after the two rating experiments, participants were instructed to fill out a Goldsmiths Musical Sophistication Index (Gold-MSI) questionnaire (Müllensiefen et al., 2014). The Gold-MSI is a self-reported inventory for individual differences in music sophistication that measures the ability to engage with music.

(See the questionnaire: <https://exp.music.mcgill.ca/questionnaire/blendemotion.php>)

2.1.5 Apparatus

The experiment was conducted in the Music Perception and Cognition Lab at McGill University. The experimental session was run with the PsiExp computer environment (Smith, 1995). Sounds stored on a Mac Pro 5 computer running OS 10.6.8 (Apple Computer, Inc., Cupertino, CA) were amplified through a Grace Design m904 monitor (Grace Digital Audio, San Diego, CA) and presented over Dynaudio BM6a loudspeakers (Dynaudio International GmbH, Rosengarten, Germany) arranged at about 60°, facing the listener at a distance of 1.5 m. Participants were seated in an IAC model 120act-3 double-walled audiometric booth (IAC Acoustics, Bronx, NY). The amplification level of the monitor was chosen in advance by the experimenter after pilot sessions to ensure a comfortable level (Maximum = 81.3 dB, $M = 76.3$ dB, $SD = 2.1$ dB) for listening to all stimuli in the experiment and remained fixed for all participants.

2.2 Acoustic description

2.2.1 Audio Representations

Considering the perceptual multidimensionality of timbre, it is necessary for us to analyze multiple acoustic features to determine the features potentially influencing participants' perception of emotion and the degree of blend. We used the Timbre Toolbox (Peeters et al., 2011, recently updated by Kazazis et al., 2022). The Timbre Toolbox works in three steps. In the first step, the audio files are analyzed to estimate the temporal and spectral parameters for each input audio representation. The audio representations we used here are Temporal Energy Envelope representation (TEErep), which uses the power amplitude envelope of the audio signal, as well as the raw waveform; the Harmonic representation (HARMrep), which relies on partial tracking for

estimating harmonic and inharmonic components; the Audio Signal representation (ASrep), which is computed directly from the raw audio signal; and the Equivalent Rectangular Bandwidth representation (ERBrep), which depends on a Short-Term Fourier Transform (STFT) with a frequency scale transformed to a physiological scale related to the distribution of frequencies along the basilar membrane in the inner ear as modelled by a scale derived from the Equivalent Rectangular Bandwidth representation (Moore & Glasberg, 1983).

The second step is to extract the audio descriptors from each representation. The descriptors derived from TEErep are scalar values that capture the global temporal features of a sound. Each scalar is capable of summarizing one specific aspect of the temporal characteristics of a sound and was used directly for further analysis. The descriptors derived from HARMrep, ERBrep, and ASrep are time series. The local features extracted from each window at each hop constitute the time series. In terms of computing, for TEErep, the waveform is first segmented using a window length of 200 ms and a hop size of 100 ms. For HARMrep, the window length is 2048 samples, and the hop size is 512 samples. For ERBrep and ASrep, the window length is 20 ms and a hop size is 10 ms. All other parameters are used from default settings in Timbre Toolbox.

The third step in the Timbre Toolbox is to summarize the time series descriptors according to summary statistics. In our study, the median value and the interquartile range (IQR) value were used to summarize the descriptors from HARMrep, ERBrep, and ASrep, which represent robust versions of central tendency and variability, respectively.

2.2.2 Hierarchical Cluster Analysis

Peeters et al. (2011) showed that many acoustic features were highly correlated with each other and clustered together for a wide range of musical instrument sounds. To avoid using very highly co-linear descriptors in the following regressions, hierarchical cluster analyses were used to select the representative acoustic features. As the acoustic features may differ between the composite and constituent sounds, we conducted a hierarchical cluster analysis for each set. The correlation matrixes of the acoustic features were transformed into distance matrixes to calculate the clusters.

As shown in Figure 2.5, the red line indicates the height at 0.8 (Pearson's $r = 0.2$) where the correlation is small enough to choose the representative audio descriptors in each cluster. At this height, we have 13 clusters of acoustic features for blended sounds and 9 clusters for individual

sounds. The final chosen acoustic features are shown in Table 2.1. In order to compare the two sets of descriptors, we tried to choose similar audio descriptors and also considered the features of the sounds and the commonly used descriptors in relevant studies (such as McAdams et al., 2017; Tardieu & McAdams, 2012).

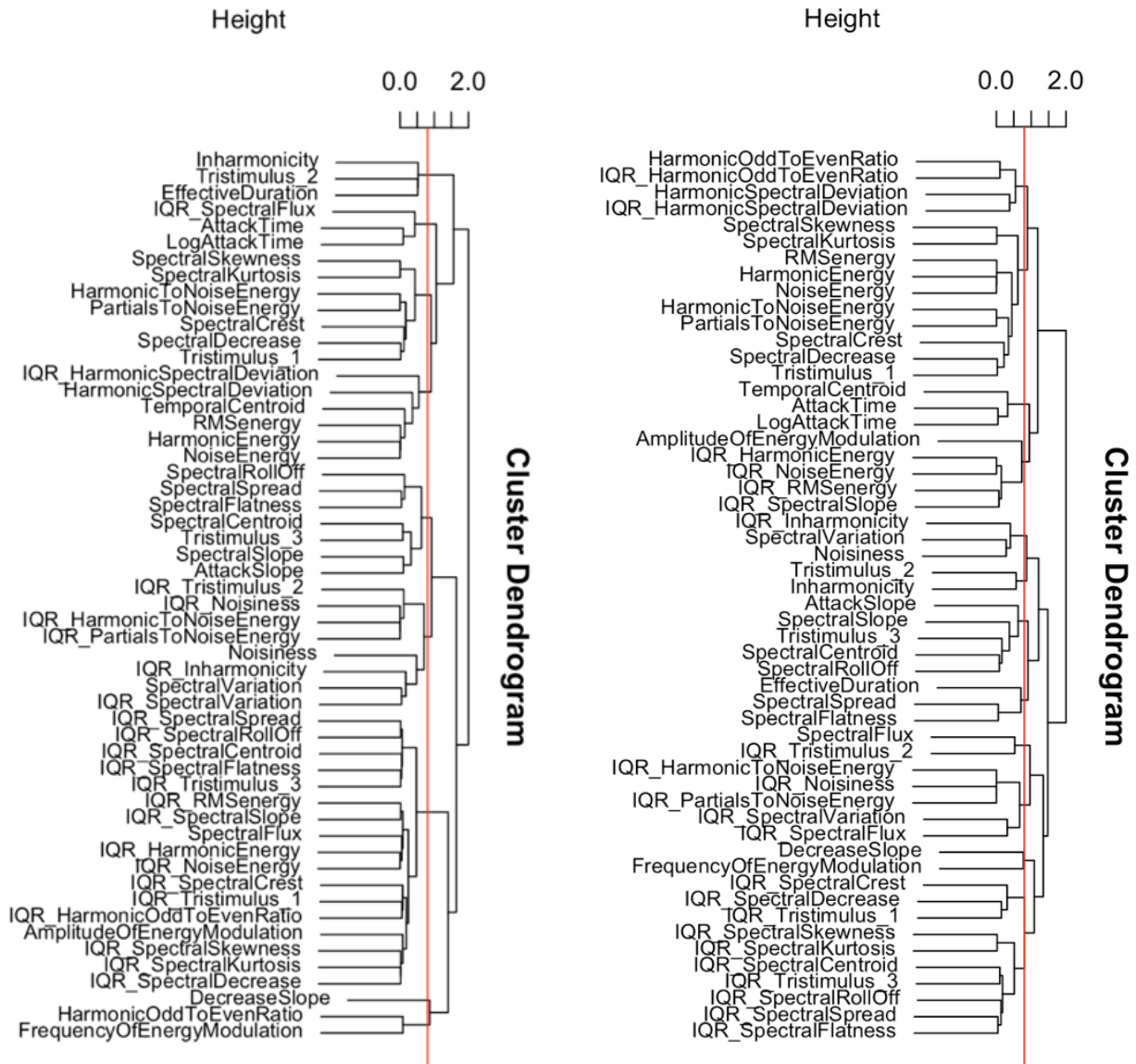


Figure 2.5 Hierarchical cluster dendrogram of individual (left) and blended sounds (right).

Table 2.1 Audio descriptors for composite sounds and constituent sounds.

Composite Sound Descriptors	Constituent Sound Descriptors
Spectral Centroid (Median, IQR)	Spectral Centroid (Median)
Spectral Spread (Median)	
Spectral Skewness (Median)	Spectral Skewness (Median)
Spectral Variation (Median, IQR)	Spectral Variation (Median)
Spectral Flux (Median)	Spectral Flux (Median)
Spectral crest (IQR)	
RMS Energy (IQR)	
Harmonic Odd-to-Even Ratio (Median)	Harmonic Odd-to-Even Ratio (Median)
Tristimulus 2 (Median)	Tristimulus_2 (Median)
	Temporal Centroid
Log Attack Time	Log Attack Time
Decrease Slope	Decrease Slope

For the composite sounds, we chose spectral centroid, spectral spread, spectral skewness, spectral variation, spectral flux, spectral crest, spectral variation, harmonic odd-to-even ratio, tristimulus 2, RMS energy, log attack time, and decrease slope. For the constituent sounds, we chose spectral centroid, spectral skewness, spectral variation, spectral flux, temporal centroid, harmonic odd-to-even ratio, tristimulus 2, log attack time, and decrease slope.

- Spectral descriptors (from ERBrep).** *Spectral centroid* refers to the spectral centre of gravity, which is related to auditory brightness (low centroid values indicate a dark sound, and high values a bright sound). It also increases in the presence of noise, and it tends to fluctuate to a great extent during the transient regions of sound events. *Spectral spread* measures the standard deviation of the spectrum around the spectral centroid, where high values indicate a rich spectrum. *Spectral crest* is a measure of the peakiness of the spectrum, where low values indicate a flat spectrum and high values indicate a peaky spectrum consisting of strong sinusoidal components. *Spectral skewness* is a measure of the asymmetry of the spectrum around the spectral centroid, where zero skewness indicates a symmetric distribution, negative skewness indicates more energy at lower frequencies, and positive skewness indicates more energy at higher frequencies. *Spectral flux* represents the amount of variation in the spectrum over time. Similar to spectral flux, *spectral variation* is another measure of spectrum variability over time, where low values indicate low amplitude waveform segments or a stationary spectrum, which can be noisy, and high

values indicate strong spectral changes. The difference between the spectral flux and spectral variation is that spectral flux is calculated as the Euclidean distance between two spectra of consecutive frames, whereas spectral variation is calculated as the cosine distance between two spectra of consecutive frames, which is more focused on the dissimilarity between spectra.

- **Harmonic descriptors (from HARMrep).** The *harmonic odd-to-even ratio* is the ratio of energies of the odd to the even harmonics, where a high ratio indicates more energy in the odd harmonics (e.g., the clarinet) and often results in “hollow” sounds, and a lower ratio indicates a smoother spectrum and a “fuller” sound. The second tristimulus value (*tristimulus 2*) is defined as the relative amplitude of harmonics 2–4 compared to that of all frequency components.
- **Audio signal descriptors (from ASrep).** *RMS energy* is computed as the root mean square of the frame energy of the signal.
- **Temporal energy envelope descriptors (from TEErep).** *Attack time* is defined as the time it takes the waveform to reach its maximum level from a defined threshold level (0 dBFS). *Decrease slope* measures the rate of decrease of the signal energy during the sustained part of the sound. *Temporal centroid* represents the centre of gravity of the energy envelope over the duration of the sound.

Notably, the results of hierarchical cluster analyses both for the ten constituent sounds and the 45 composite pairs show different acoustic clusters at the same cluster height of 0.8, which indicates that the structure of the acoustic features of the composite sounds is different from that of the constituent sounds. As shown in Table 2.1, more IQR (Interquartile Range) values appeared in clusters that are independent of the median values for composite sounds compared to constituent sounds, which suggests that the composite sounds have more unique features on both the variance of the spectral properties (spectral centroid, spectral crest, spectral variation) and one temporal property (RMS energy).

3 RESULTS AND ANALYSIS

Emotion ratings of blended pairs were evaluated through descriptive analysis, correlation analysis, and ANCOVA analysis. Descriptive analysis was used to show the overall distribution of the Emotion ratings. Correlation analysis was used to show the relationships among valence, energy, and tension. ANCOVA was used to investigate the influence of different combinations of sustained instruments on perceived affect with the covariate of musical sophistication. After analyzing the emotion ratings of composite sounds, a comparative analysis of perceived emotion between the composite sounds and constituent sounds was conducted. Geometric analysis was used in the affective space to uncover the relationship of the perceived affect between the composite sounds and corresponding constituent sounds.

Acoustic analyses were then conducted to determine whether a small set of audio descriptors could be found to interpret the acoustic origins of participants' perceived emotions and the differences between the perceived emotions of composite and constituent sounds.

The degree-of-blend ratings were then analyzed using correlation analysis, multidimensional scaling, and a social network representation. The correlation analysis was used to find the linear relationship between the perceived emotion and the degree of blend. Multidimensional scaling and the social network representation were used to discover participants' perceived space of the degree of blend. Additional acoustic analyses were also used to help interpret each dimension of the degree-of-blend space.

3.1 Affective Analysis

3.1.1 Affective Analysis for Blended Pairs

3.1.1.1 Validity Check

Before we calculated the statistics, Cronbach's α and average inter-participant correlation were computed to check the validation of data. The Cronbach's alpha is excellent, and the average inter-participant correlation is within the ideal range (Piedmont, 2014) for 45 ratings across 40 participants, as shown in Table 3.1. Also, a jackknife analysis of the Cronbach alphas with each participant removed is also excellent (see Appendix C: Participant Validation).

Table 3.1 Statistics for Cronbach's α and average inter-participant correlation.

Estimate	Cronbach's α	Average inter-participant correlation
Point estimate	0.924	0.251
95% CI lower bound	0.907	0.210
95% CI upper bound	0.938	0.292

3.1.1.2 Correlation Analysis among Dimensions of Valence, Energy, and Tension

Correlation analysis was conducted to examine the overall relationship among the three dimensions. All ratings were averaged across 40 participants for each blend stimulus. As shown in Table 3.2, among the three dimensions of affect ratings of the blended pairs, energy arousal and tension arousal were highly positively correlated (95%). However, we continued using three dimensions in the following analyses in order to compare with the McAdams et al. (2017) study.

Table 3.2 Pearson correlation matrix for valence, tension, energy, and blend.

		Valence	Tension	Energy	Blend
Valence	<i>r</i>	—			
	<i>p</i>	—			
Tension	<i>r</i>	.304*	—		
	<i>p</i>	.042	—		
Energy	<i>r</i>	.501***	.953***	—	
	<i>p</i>	< .001	< .001	—	
Blend	<i>r</i>	-.441**	-.352*	-.398**	—
	<i>p</i>	.002	.018	.007	—

* $p < .05$, ** $p < .01$, *** $p < .001$

3.1.1.3 ANCOVA

To investigate the influence of different combinations of sustained instruments on perceived affect with the covariate of musical sophistication, a one-way ANCOVA was conducted for each affect dimension (Valence, Tension, and Energy). Ratings were range-normalized from 0 to 1 for each participant before analysis. An assumption check was conducted before the test. According to the Q-Q plots, which compare the sample distribution with a normal distribution (Appendix D: Assumption Check), all the ratings of the three dimensions can be considered normally distributed. According to Levene's test, the homogeneity of variance assumption of valence ratings ($p = 0.360$) and tension ratings ($p = 0.573$) was satisfied, but the assumption for energy ratings ($p = 0.013$) was slightly violated. Considering that we had done participant-wise correlation before to check for outliers and had a very large sample size (1800 ratings for each dimension), we selected to continue analysis with all three dimensions.

For both ANCOVAs, the dependent variable was the perceived emotion rating, the independent variable was the 45 blend stimuli, and the covariate was each participant's General Musical Sophistication score.

As shown in Table 3.3, our results indicated that the musical sophistication score had no significant unique effect on perceived valence, $F(1, 1754) < 1$, and tension, $F(1, 1754) < 1$, beyond that of the instrumental blend pairs. The music sophistication score had a significant unique effect

on perceived energy; however, the effect size is very small, $F(1, 1754) = 4.11$, $p = .043$, $\omega^2 = .001$. Therefore, we did not further consider the musical sophistication score in our analyses.

Different combinations of blended pairs have significantly different perceived emotions, confirming that blended sounds have distinct timbral properties that give rise to perceived affect [valence: $F(44,1754) = 1.72$, $p = .002$, $\omega^2 = .017$; tension: $F(44,1754) = 13.20$, $p < .001$, $\omega^2 = .230$; energy: $F(44,1754) = 20.37$, $p < .001$, $\omega^2 = .321$]. Note that the effect size for valence is an order of magnitude smaller than that for the arousal dimensions.

Table 3.3 ANCOVA result for valence, tension, and energy.

Cases	Sum of Squares	df	Mean Square	F	p	ω^2
Valence						
stimuli	3.139	44	0.071	1.72	.002	.017
rawSophistication	0.001	1	0.001	0.04	.852	0
Residuals	72.652	1754	0.041			
Tension						
stimuli	23.444	44	0.533	13.20	< .001	.23
rawSophistication	0.015	1	0.015	0.37	.543	0
Residuals	70.794	1754	0.04			
Energy						
stimuli	35.388	44	0.804	20.37	< .001	.321
rawSophistication	0.162	1	0.162	4.11	.043	.001
Residuals	69.253	1754	0.039			

3.1.2 Comparative Geometric Analysis

3.1.2.1 Emotion Space

In order to investigate whether we could use the affective qualities of constituent sounds to predict those of the composite sounds, a geometric analysis was used (Caetano et al., 2022). As shown in Figure 3.1 and Figure 3.2, we found that, in general, the blended sounds occupied more of the emotion space than the individual sounds, especially in energy arousal: blended sounds could thus

have higher or lower energy arousal than the individual constituent instruments. The blended sounds could also have higher tension arousal ratings. For valence, participants tended to give the same compact ratings of blended sounds as the individual sounds. Note that the range of variation of average valence ratings is about 0.2, whereas that of the arousal dimensions is about 0.45-0.5, indicating greater variation of the ratings of blend stimuli on these latter dimensions.

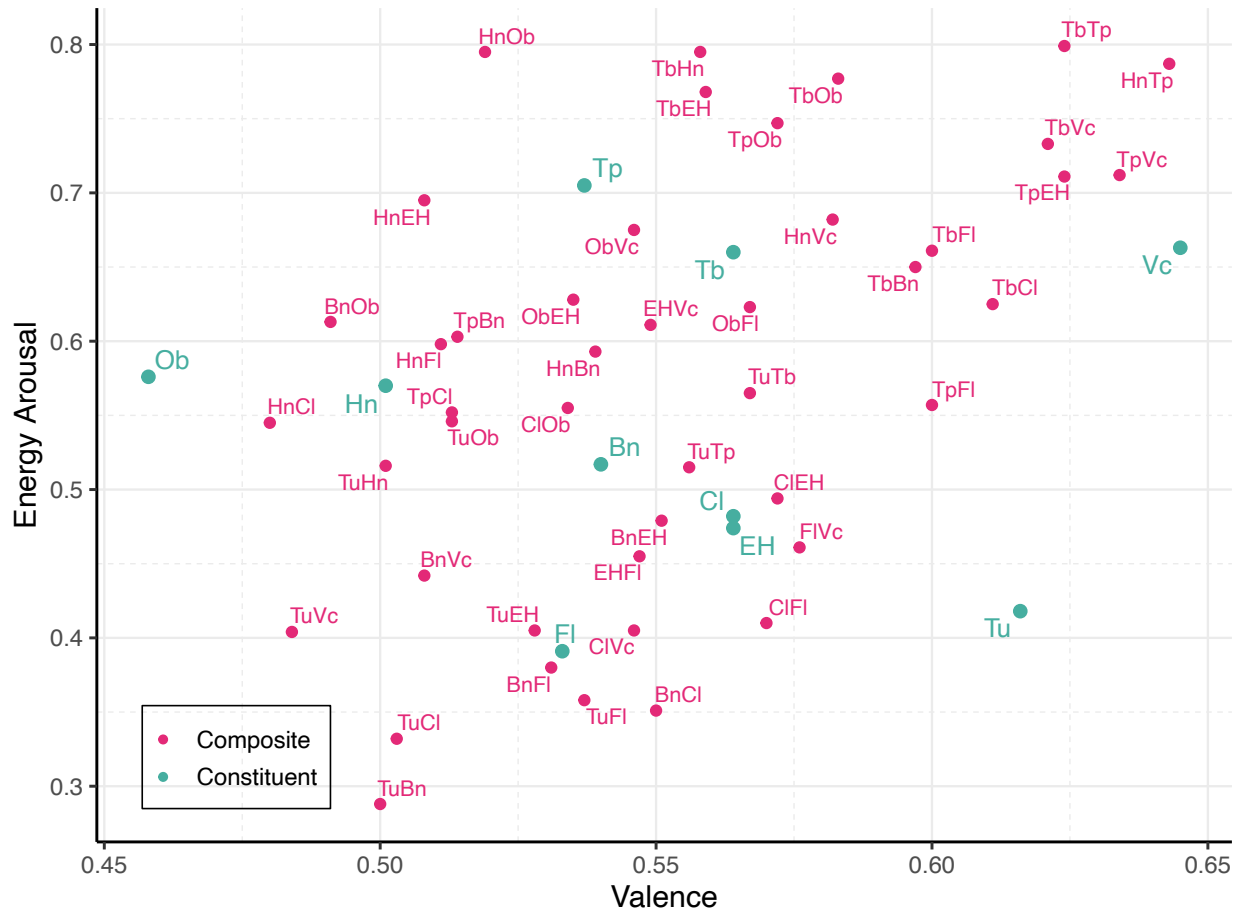


Figure 3.1 Scatter plot of mean energy arousal and valence for both composite sounds and constituent sounds. The values of the affective qualities of constituent sounds are taken from McAdams et al.'s (2017) data.

We took the mean value of each perceptual quality of both constituent sounds and composite sounds in the 3D emotion space and calculated the distances δ : $s1 = \delta(S1, B)$, $s2 = \delta(S2, B)$, $b = \delta(S1, S2)$. All 45 instrumental combinations reveal a triangular relationship, which means that the emotional quality of a blended sound does not simply lie on a line between the constituent sounds in the emotion space. There are also some interesting patterns.

Internal and Beyond: As shown in Figure 3.4, the position of the affect of the composite sound can be somewhere in between the two constituent sounds, like the Tuba–Oboe pair, although it can also be beyond the range of constituent sounds, as with the Trumpet–Trombone pair. In the first case, the values of valence, energy arousal, and tension arousal of composite sound are all less than the maximum and greater than the minimum of the corresponding values of constituent sounds. In the second case, the composite sound would be rated higher or lower in either one of the affective dimensions than both constituents. So energy and valence coordinates of TuOb are between the corresponding ranges of coordinates of Ob and Tu, whereas the coordinates of TbTp are outside the range of coordinates of Tp and Tb on both dimensions.

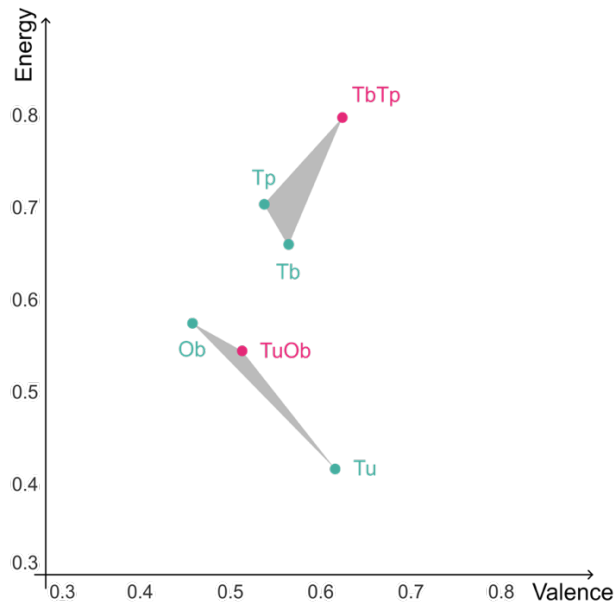


Figure 3.4 Valence-Energy arousal plot for Trumpet-Trombone and Oboe-Tuba blended pairs.

Intermediateness and Dominance: Furthermore, the composite sound is perceptually intermediate between the two constituent sounds in none of the pairs, and the affect of some of the blended sounds are dominated by the one of their constituent sounds. As shown in Figure 3.5, calculating the difference d between the distances from the blend position to both constituent

sound positions reveals that some instruments show dominance in the perceived emotion space. For example, in the Oboe–Tuba pair, the Oboe “drags” the blend position closer to it and away from the Tuba. If d is zero, in other words, s_1 equals to s_2 , it means that the blended sound is perceptually intermediate between the two constituent sounds. However, there are no sounds showing this feature in this study, although some are quite close s_1 - s_2 equality.

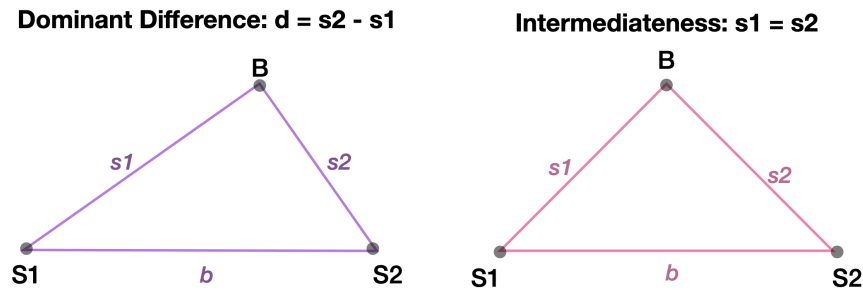


Figure 3.5 Graphic representation of dominant difference and intermediateness.

As shown in the results of geometric analysis, the composite sounds are not simply in between the constituent sounds but in a triangular configuration. It appears difficult to develop a generalizable formula to predict the emotion coordinates of blended sounds across all blended pairs from the emotion coordinates of the individual sounds, but there are still some patterns showing that some blends are situated within the range of the constituents’ coordinates, whereas others are beyond that range. Also, some constituents dominate in their influence on the affect of the blend.

3.2 Acoustic Analysis

Acoustic analyses were used to interpret the timbral cues underlying people’s perception of the emotional qualities of the sounds. We first selected audio descriptors using hierarchical cluster analysis for both composite and constituent sounds, as shown in 2.2.2. In this part, lasso regression was used to reduce the pool of descriptors, and multiple linear regression was used to analyze the relationship between the selected audio descriptors and the perceived emotion ratings.

3.2.1 Regression Analysis

We performed lasso regressions for variable selection (Tibshirani, 2011) among the 13 audio descriptors for composite sounds and 9 for individual sounds, we first use lasso regression to reduce the dimensionality to reduce overfitting and increase interpretability. Then multiple linear

regression was applied to the selected variables to determine the relationship between the audio descriptors and the perceived affects. R was used to conduct the analyses (R Core Team, 2022), where `glmnet` package (Friedman et al., 2010) was used to conduct cross validation.

3.2.1.1 Lasso Regression and Multiple Linear Regression Analyses

Lasso stands for “least absolute shrinkage and selection operator” (Tibshirani, 1996). It performs both variable selection and regularization to enhance prediction accuracy and interpretability with multiple independent variables. Regularization is a technique to avoid overfitting and helps to reduce the number of parameters and simplify the model. Lambda is the parameter that controls the amount of regularization in lasso regression. Usually, cross-validation is used to select the optimal value of lambda. Cross-validation is a technique to evaluate the prediction performance of a model by splitting the data into multiple subsets and using some subsets for testing after modelling the remaining items in the dataset. For example, as shown in Figure 3.6, we can use the trace plot of the cross-validation curve to select lambda: on the x-axis at the bottom are the logarithm values of lambda and the values at the top are the corresponding numbers of parameters; the y-axis shows the mean cross-validation error and the vertical bars on each MSE measure show plus and minus one standard deviation. Mean-Squared error (MSE) is adopted in the R `glmnet` package and was used in the present study to measure the cross-validation error. The two dashed lines indicate two methods to select the optimal value of lambda—one-standard-error criterion (1se) and minimum criterion (min). The 1se method chooses the largest lambda such that the MSE is within one standard error of the minimum, which means a more conservative and simpler model. The min method chooses the lambda that gives the minimum MSE, which means the model that fits the data better but is more complex. In the present study, min was adopted if it did not select too many variables that decrease the interpretability (i.e., more than four variables), otherwise, 1se was adopted to get a simpler and more compact model.

3.2.1.1.1 Blended sounds

The regularization parameter lambda was determined using 10-fold cross-validation out of the initial set of 13 predictor variables for 45 blended sounds in predicting each affective dimension. Considering that min selected too many parameters, we used 1se to select the simpler model, as shown in Figure 3.6, Figure 3.7, and Figure 3.8 for valence, tension arousal and energy arousal, respectively.

For valence, lasso regression selected a subset of two predictors from the initial set. The selected predictors were the median and IQR of the Spectral Centroid. For tension arousal, the three selected predictors were the medians of Spectral Centroid, Spectral Spread, and Spectral Skewness. For the energy arousal, five predictors were selected: Spectral Centroid (Median and IQR), Spectral Spread (Med), Spectral Skewness (Med), and Log Attack Time. Multiple linear regressions were conducted to investigate further the relationship between the perceived affect and the acoustic predictors selected by lasso regression. All predictors identified by lasso regression were included in the multiple linear regression model. As shown in Table 3.4, blended sounds with higher spectral centroid and spectral spread tend to have more positive energy arousal and tension arousal. Energy arousal also positively related to shorter log attack time. Blended sounds with a larger median and IQR of spectral centroid tend to have more positive valence.

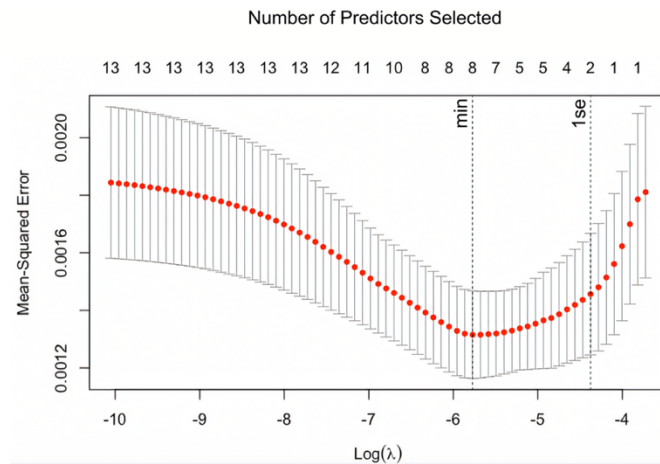


Figure 3.6 Trace plot of cross-validation curve for lasso regression of valence for blended sounds.

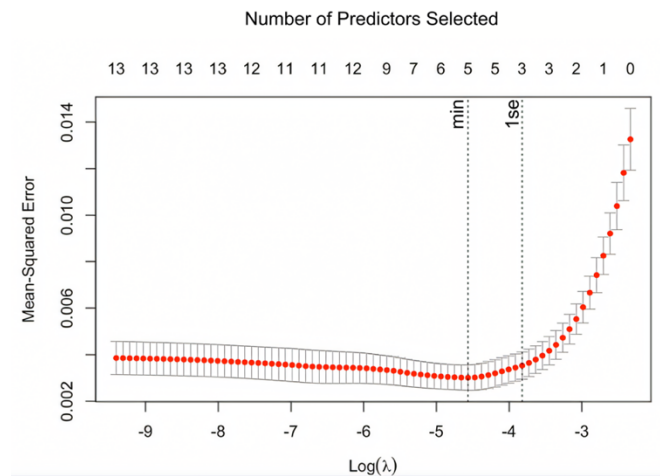


Figure 3.7 Trace plot of cross-validation curve for lasso regression of tension arousal for blended sounds.

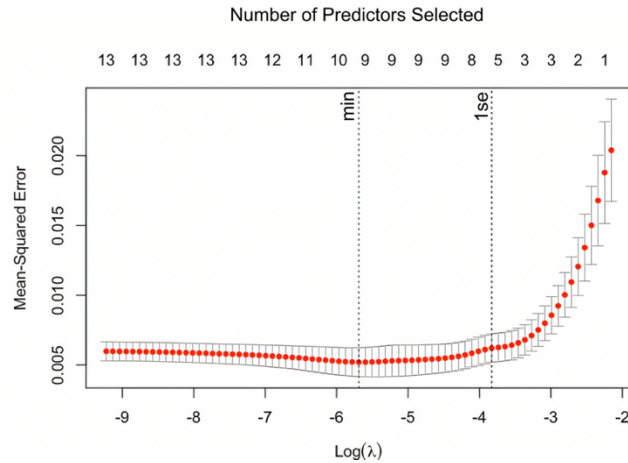


Figure 3.8 Trace plot of cross-validation curve for lasso regression of energy arousal for blended sounds.

Table 3.4 Linear regression coefficients and adjusted R² for blended sounds.

	Valence	Tension	Energy
Adjusted R ²	.382	.783	.789
SpectCent	0.053*	0.308***	0.339***
SpectVar			
OddToEvenRatio			
SpectSpread		0.169**	0.158*
SpectSkew		-0.029	-0.089
LogAttTime			-0.104*
SpectCent (IQR)	0.093***		0.086
SpectCrest (IQR)	0.053		
*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$			

3.2.1.1.2 Individual sounds

Considering the relatively small sample size (10 sounds) of the individual sounds, the regularization parameter was determined using 5-fold cross-validation out of the initial set of 9 predictor variables. We used the min criterion to select predictors for valence because the 1se method selected no parameter, as shown in Figure 3.9. Considering that the 1se criterion did not increase the MSE much more than the min for tension and energy, we used 1se to select the predictors for these dependent variables, as shown in Figure 3.10 and Figure 3.11.

For valence, lasso regression selected the medians of Spectral Centroid, Spectral Flux, and Tristimulus_2, as well as Log Attack Time. For tension arousal, the two selected predictors were the medians of Spectral Centroid and Harmonic Odd-to-Even Ratio. For energy arousal, only the

median of Spectral Centroid was selected. The multiple linear regression model included all predictors identified by lasso regression. The results are shown in Table 3.5. Individual sounds with a higher spectral centroid tend to have higher tension and energy arousal. Sounds with a lower spectral centroid, higher spectral flux, and sharper log attack time have more positive valence.

Comparing the regression results of blended and individual sounds, we found that the arousal resulting from both individual instrument timbres and blended timbres had similar acoustic features, although different properties underlie valence perception in the two sound sets. Notably, the IQR values played a more important role in blended sounds than in individual sounds.

Table 3.5 Linear regression coefficients and adjusted R^2 for individual sounds.

	Valence	Tension	Energy
Adjusted R^2	.820	.871	.634
SpectCent	-0.071*	0.364***	0.177
SpectFlux	0.098*		
OddToEvenRatio	0.044		
Tristim2	-0.069		
LogAttTime	-0.119*		
TempCent			-0.140

*** $p < 0.001$, * $p < 0.05$

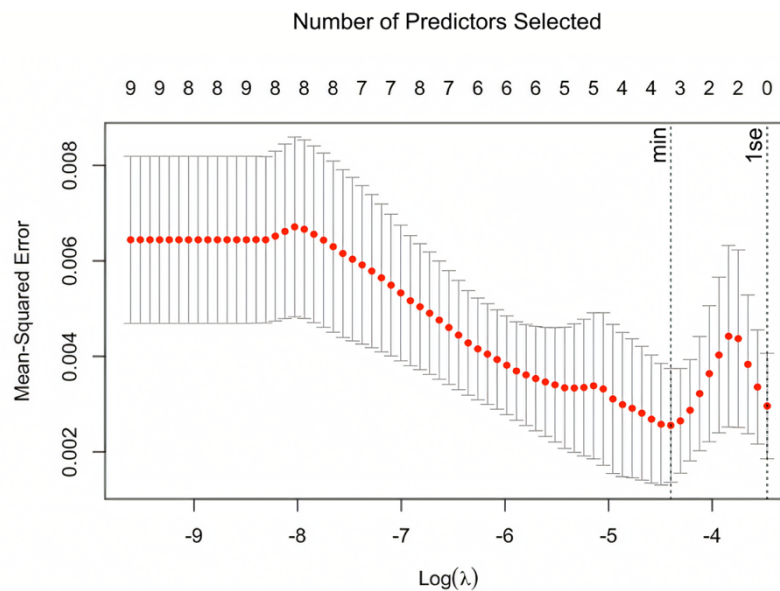


Figure 3.9 Trace plot of cross-validation curve for lasso regression of valence for individual sounds.

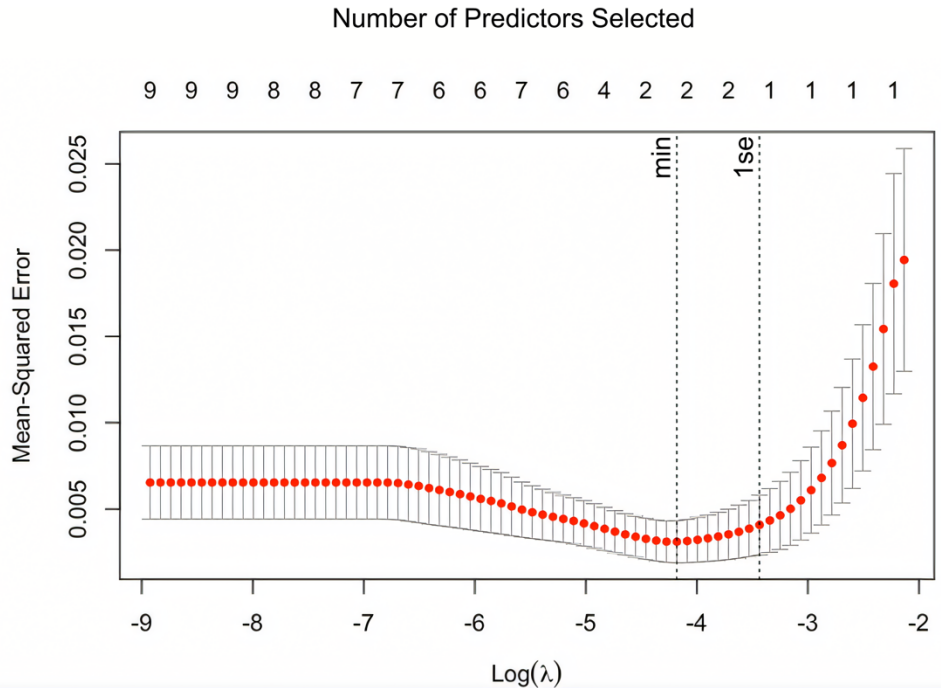


Figure 3.10 Trace plot of cross-validation curve for lasso regression of tension for individual sounds.

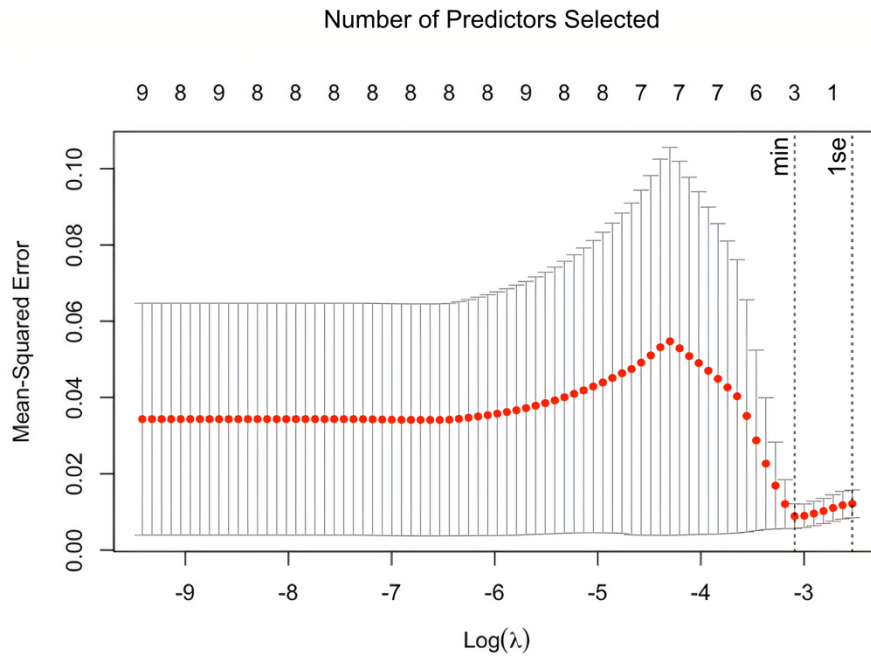


Figure 3.11 Trace plot of cross-validation curve for lasso regression of energy for individual sounds.

3.2.1.1.3 Dominance analysis

We also conducted lasso and multiple linear regression analyses for the dominant difference. The dominant difference d is the difference between the distances from the composite

sound to each of the two constituent sounds in the 3D emotion space, as shown in the previous section (Figure 3.5). In the regression for the difference d , we use d as the dependent variable and the difference of acoustic features between s_1 and s_2 to predict d . The min criterion was used in lasso regression for selecting predictors since min and 1se both selected one predictor, as shown in Figure 3.13 and Table 3.7. Only the difference between spectral centroids of constituent sounds was selected for the linear regression. The result shows that the difference between spectral centroids is negatively correlated with the difference d , which means that the composite sound's perceived affect was closer to that of the constituent sound with a higher spectral centroid (adjusted $R^2 = .54$): brighter timbres appear to be more dominant in the perceived emotion of blended sounds.

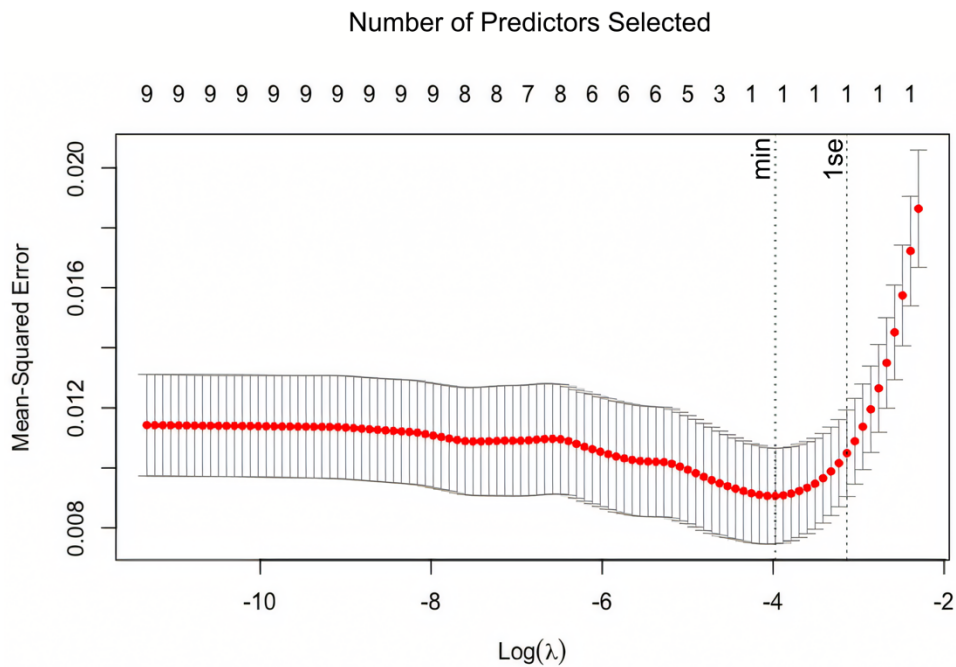


Figure 3.12 Trace plot of cross-validation curve for lasso regression of dominant difference.

Table 3.6 Linear regression coefficient and adjusted R^2 for dominant difference.

	Estimate coefficient
Adjusted R^2	.54
Spectral Centroid	-0.370***
*** $p < 0.001$	

In conclusion, using small sets of acoustic features is helpful to understand listeners' perceived affects. Notably, audio feature variability may play a more important role in blended sounds than individual sounds. At the same time, acoustic features are also helpful to explain the dominance of individual constituents in the affect of a blended sound.

3.3 Analysis of the Role of Degree of Blend

We conducted a correlation analysis to determine the relationship between the perceived degree of blend and the perceived emotion, followed by multidimensional scaling and social network representation analyses to reveal the perceived relations among sounds in terms of their degree of blend.

3.3.1 Correlation analysis

The correlation analysis tested whether the degree of blend predicts the perceived affect. As shown in Table 3.2, the correlations between the degree of blend and each of the emotion dimensions, although significantly different from 0, were all weak and negative (for valence, $r = -0.44$; for tension, $r = -0.35$; for energy, $r = -0.40$). The linear regressions all show low explained variance (adjusted $R^2 < 17\%$), suggesting that the degree of blend is not a strong predictor for perceived affects. Therefore, we did not further consider degree of blend as a factor in the interpretation of perceived emotion in this study. We will focus on the blend space in the following analyses.

3.3.2 Multidimensional Scaling of Blend Space

Multidimensional scaling (MDS) was used to find the latent dimensions of the perceived space of degree of blend. Given that we had blend ratings for all sound pairs, we flipped all the blend ratings to get dissimilarity values as the input for the MDS algorithm. SMACOF (Scaling by Majorizing a Complicated Function) was used as the main algorithm (de Leeuw & Mair, 2009). A jackknife strategy was also used to reduce bias, which systematically leaves out one or more observations from the dataset, estimating the statistic of interest each time, and then calculating the bias and variance based on these repeated estimates (Elliott et al., 2013). R was used to conduct MDS algorithm (R Core Team, 2022) with the SMACOF package (de Leeuw & Mair, 2009 ; Mair et al., 2022) and maximum dimensions were set to seven for this experiment.

In our analysis, we compared three MDS algorithms, including Identity MDS, INDSCAL (Individual Differences Scaling), and IDIOSCAL (Individual Differences in Orientation Scaling). Among the three algorithms, Identity is the simplest model, which assumes that all listeners use the same perceptual dimensions in their ratings. INDSCAL is a more general model, which allows individual listeners to have different perceptual weights for different dimensions. IDIOSCAL is the most general model, which allows individuals to not only have different perceptual weights but also to have different relative orientations for different dimensions.

All three algorithms have the lowest Akaike information criterion (AIC)(Akaike, 1973) (evaluating how well a model fits the data it was generated from) for a two-dimensional model, as shown in Figure 3.13. At the same time, INDSCAL and IDIOSCAL did not show much improvement in R^2 beyond the two-dimensional model, as shown in Figure 3.14. Therefore, we chose a two-dimensional Identity MDS model as the final derived space, shown in Figure 3.15. This derived space could be understood as a blend space, in which instruments perceived to blend better are closer in the space. For instance, the Tuba–Bassoon pair has the highest degree of blend in Figure 3.16 and also has the closest distance in MDS blend space; the English Horn–Cello pair has the lowest degree of blend and also has the farthest distance in the blend space.

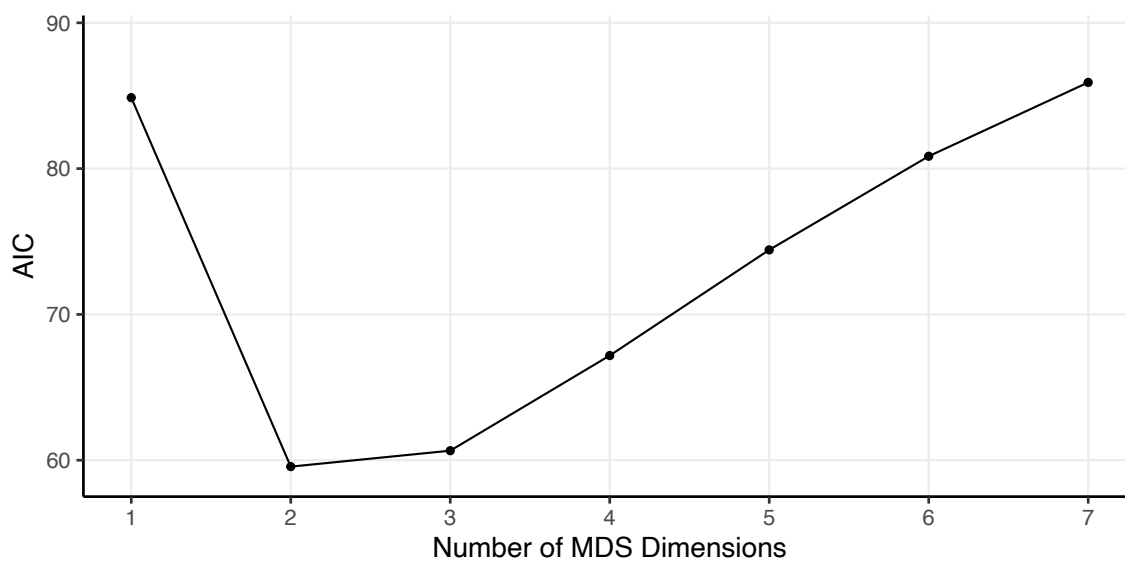


Figure 3.13 AIC for different dimensionalities in Identity MDS.

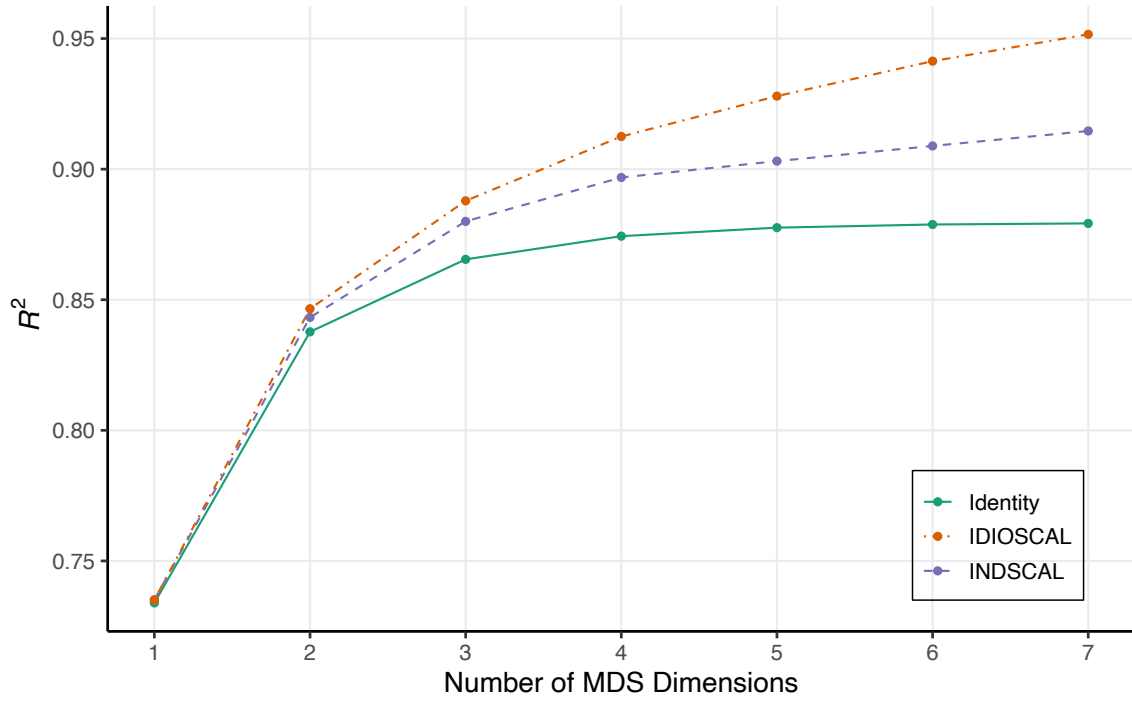


Figure 3.14 R^2 for different MDS dimensionalities of Identity, INDSCAL, and IDIOSCAL algorithms.

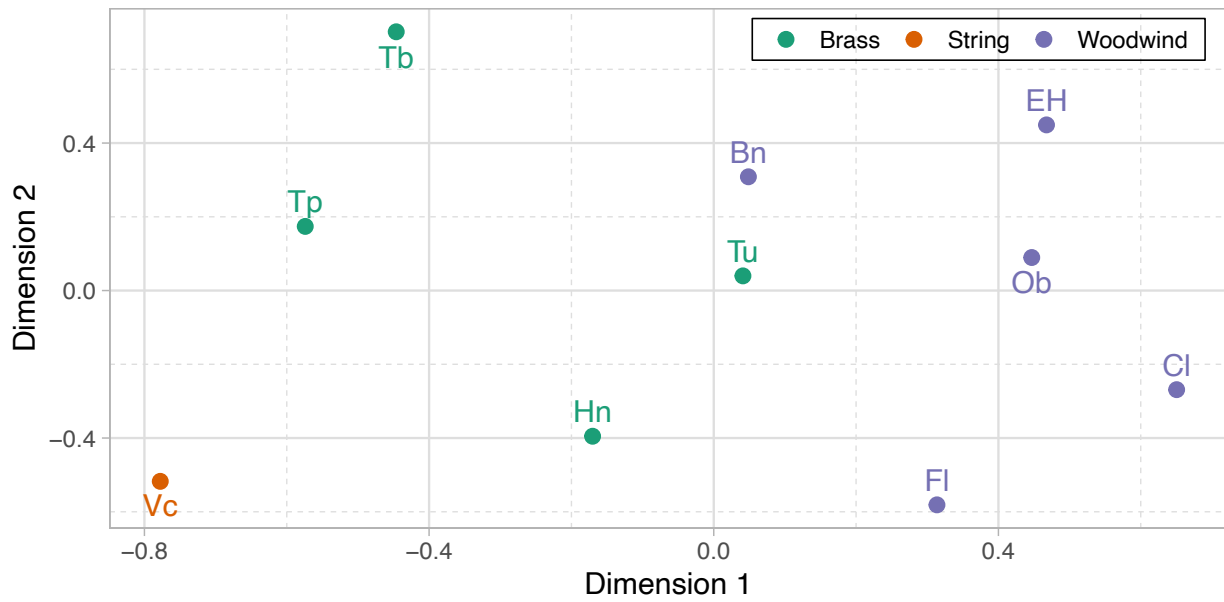


Figure 3.15 MDS blend space.

3.3.3 Geometric Analysis for Blend Space

To investigate the relative success of blending of instruments with all other instruments, we designed a geometric analysis that calculates the sum of the distances between each instrument and all other instruments. The blend distance sum is shown in Figure 3.17. It is clear to see that flute is the best blender, followed by the bassoon and tuba. Cello is the poorest blender. Notably, it was the only string instrument among the constituent sounds.

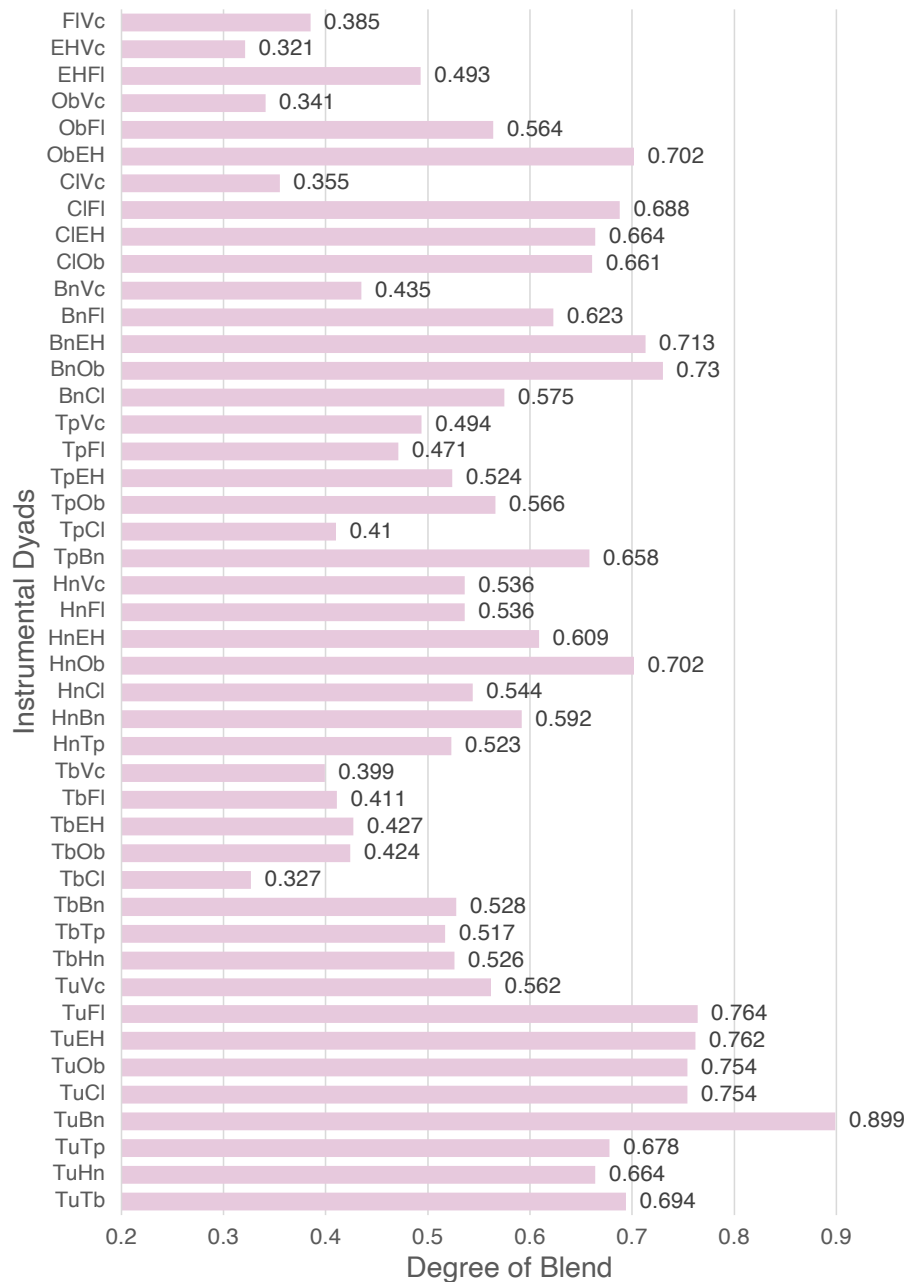


Figure 3.16 The degree of blend for each pair.

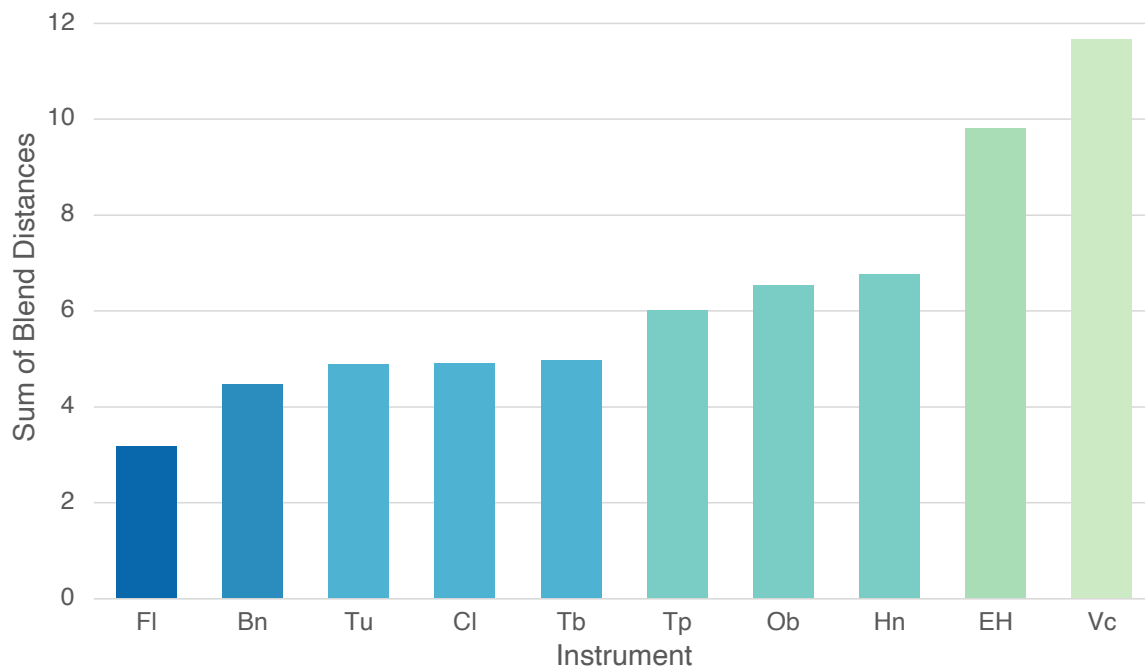


Figure 3.17 The sum of blend distances.

3.3.4 Acoustic Analysis for Blend Space

As with the perceived emotion space, acoustic analysis was also helpful in interpreting the dimensions of the blend space (Figure 3.18). Lasso regression was first conducted for both dimensions. The regularization parameter was determined using 5-fold cross-validation from the same set of 9 predictor variables as the individual sounds. We used the minimum criterion to select predictors, as shown in Figure 3.20.

For the first dimension, the medians of Spectral Variation, Spectral Flux, Temporal Centroid, and Tristimulus_2 were selected. According to the linear regression results in Table 3.7, Temporal Centroid and Spectral Flux were significant, and the adjusted R^2 was high (89%). It is evident in the blend-space plot that the first dimension of the blend space divided the instrument families, as shown with the three background colours in Figure 3.18 (orange–string; green–brass; purple–woodwind). Therefore, it makes sense that both temporal and spectral descriptors are significant in interpreting the dimension. The medians of Spectral Flux and Temporal Centroid, as well as Log Attack Time, were selected for the second dimension. Temporal Centroid significantly predicted the dimension (adjusted $R^2 = 80\%$), as shown in Table 3.8. This result

indicates that the second dimension of the blend space is more likely related to a temporal descriptor: the more similar the family of instruments (predicted by both temporal centroid and spectral flux) and the closer the temporal centroid of the instruments, the greater the degree of blend.

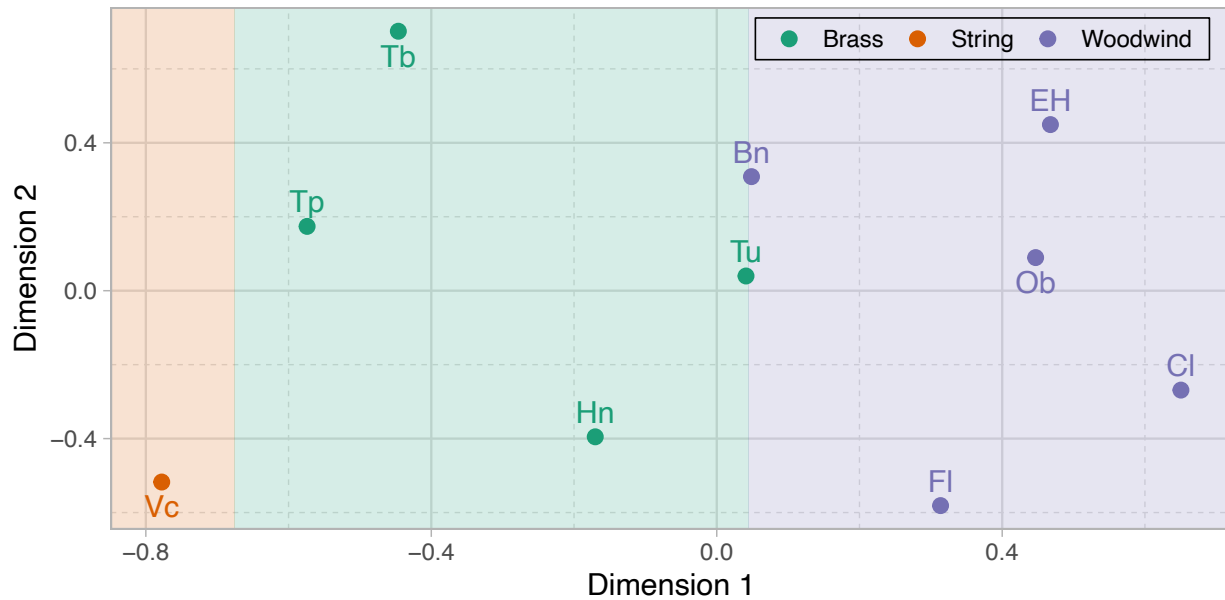


Figure 3.18 Blend space with colour indicating the instrumental family along Dimension 1.

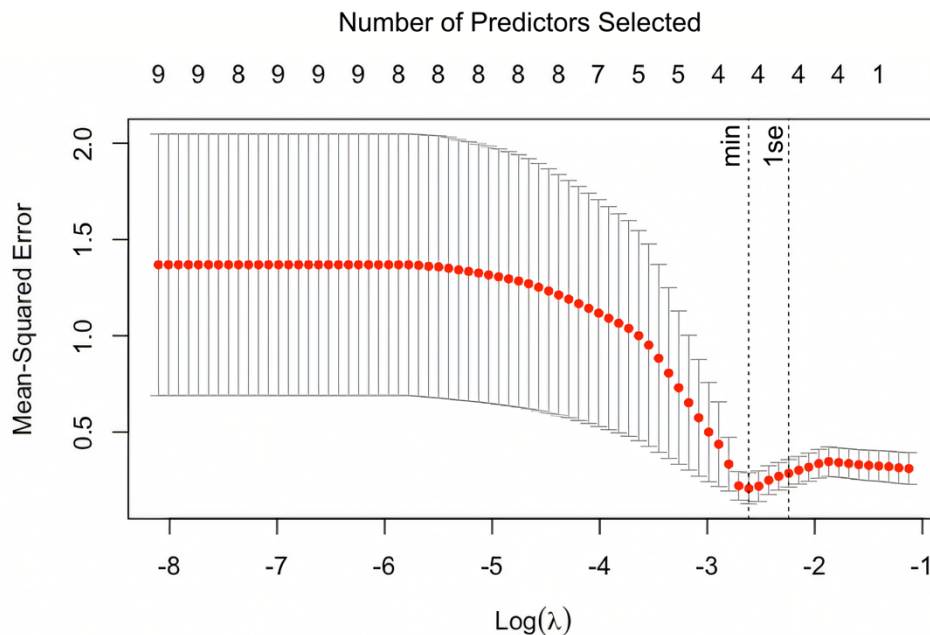


Figure 3.19 Trace plot of cross-validation curve for lasso regression of first dimension of blend space.

Table 3.7 Linear regression coefficients of the first dimension of blend space.

	Estimate	Std. Error	<i>t</i>	Pr(> <i>t</i>)
(Intercept)	0.187	0.152	1.227	0.27
SpectralVariation	-0.070	0.184	-0.382	0.72
SpectralFlux	-0.778	0.162	-4.796	0.005**
TemporalCentroid	0.825	0.203	4.069	0.010**
Tristimulus_2	0.173	0.129	1.340	0.24

*** $p < 0.001$, ** $p < 0.01$

Adjusted R^2 : 0.89

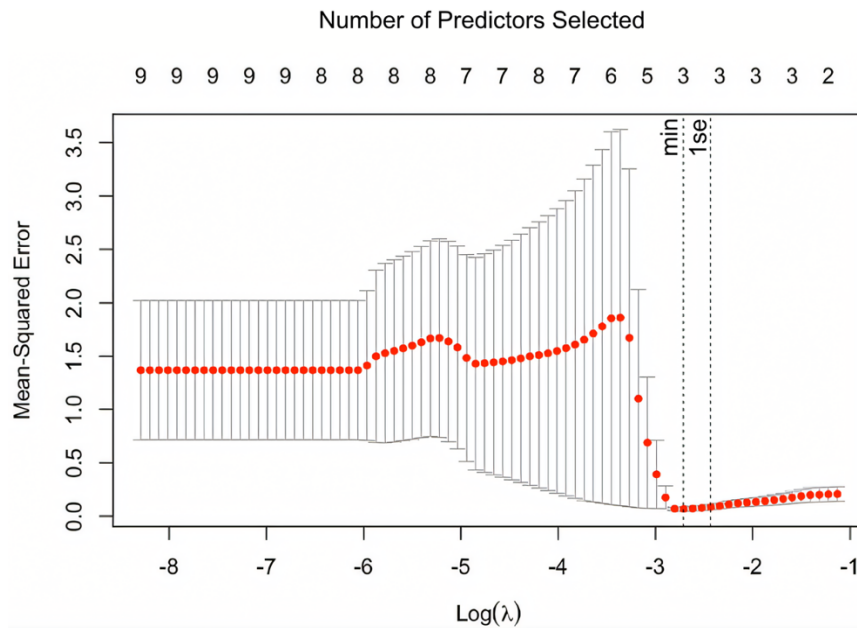


Figure 3.20 Trace plot of cross-validation curve for lasso regression of the second dimension.

Table 3.8 Linear regression coefficients of the second dimension of blend space.

	Estimate	Std. Error	<i>t</i>	Pr(> <i>t</i>)
(Intercept)	1.087	0.122	8.923	0.0001***
SpectralFlux	-0.374	0.186	-2.015	0.09
TemporalCentroid	-0.524	0.210	-2.489	0.047*
LogAttackTime	-0.409	0.248	-1.648	0.18

*** $p < 0.001$, * $p < 0.05$

Adjusted R^2 : 0.80

3.3.5 Social Network Analysis

Social Network Analysis (SNA) (Wasserman & Faust, 1994) is a useful tool in sociology to investigate and visualize social structures based on measures of proximity or strength of relation between individuals. To the best of our knowledge, no research has yet used this tool to analyze blend data. Blend can also be understood as a proximity relationship between each pair of instruments. The degree of blend can also be understood as the strength of the relationship. We took each instrument as a node, the blended relationship as an edge, and the degree of the blend as the weight of the edge to analyze the relationship, which was also indicated in the thickness of edge (the higher the degree of blend, the thicker the edge). The ForceAtlas2 algorithm (Jacomy et al., 2014) was used, which is helpful to spatialize small world and scale-free networks, where Gephi 0.9.7 (Bastian et al., 2009) software was used to visualization the network. ForceAtlas2 is a force layout, which has the specificity of placing each node depending on its relation to the other nodes. In this way, nodes that are more connected or similar are placed closer together, whereas nodes that are less connected or different are farther apart. As we are using the degree of blend representing the similarity, the nodes blends well if they are placed closer to each other. Therefore, it is very useful for showing the “community” of timbres, where the timbres will cluster together if they blend well.

As shown in Figure 3.21, the resulting network is a 2D representation of the network, and nodes are connected to each other because we used a complete set of pair-wise data. In this network, it is easy to see that instruments were much easier to blend within the same family as they clustered together. Cello was the only one that did not blend very well with other instruments but was also the only one that belonged to the string family. However, some of the instruments are also blend well across families, such as the Tuba and Bassoon, which are clustered very closely.

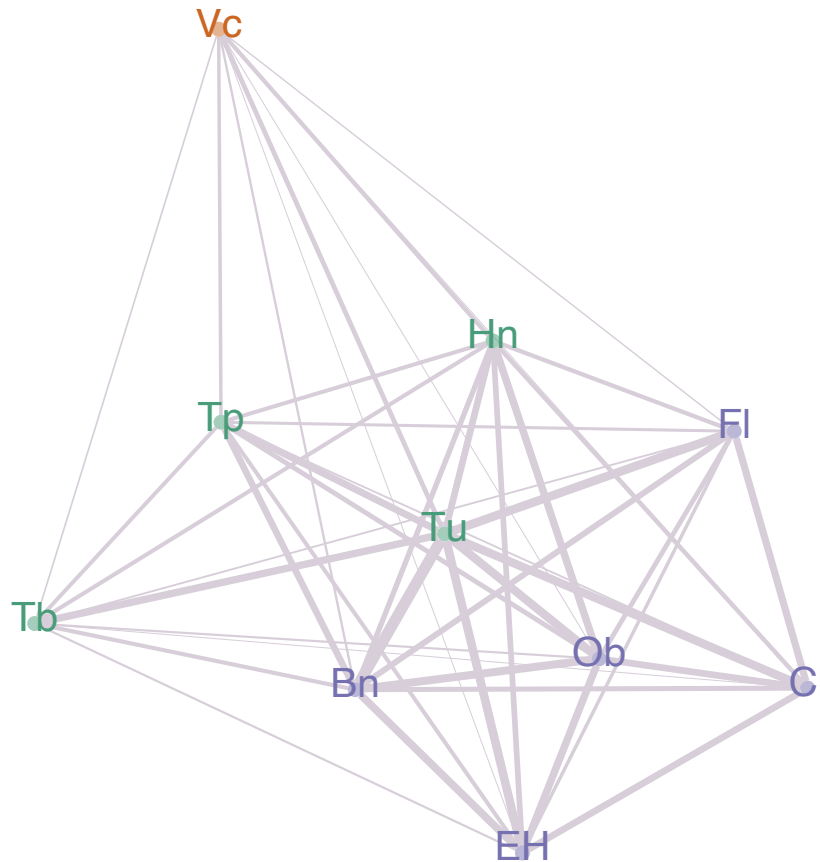


Figure 3.21 Social network representation of blend relationship.

In conclusion, the degree of blend shows weak correlation with all affective qualities, but the role of blend in perceived affect still needs to be studied in the future. As for the degree of blend, MDS and SNA are useful to analyze the blend space. A small set of spectral and temporal acoustic features can well explain the dimensions of perceptual blend space.

4 DISCUSSION

4.1 Timbral blend and emotion

In previous studies (Eerola et al., 2012, McAdams et al., 2017), it was evident that single instrumental sounds can provide enough information to convey a wide range of emotions. At the same time, researchers have asserted that new timbres can be created from the emergence of instrumental blends (McAdams, 2019). Sandell (1995) concluded from a review of orchestration treatises that blend can “invent timbres.” Therefore, in the present study, we hypothesized that these new timbres created by blends can also provide enough acoustic information to be perceived as having a wide range of emotions.

The result of ANCOVA analyses on valence, tension arousal, and energy arousal indicated that different combinations of blended pairs have significantly different perceived emotions after controlling for the participants’ musical sophistication. Furthermore, according to the result of geometric analysis, the blends also occupy an expanded range of emotions compared to the constituent sounds alone. These results suggest that the new timbres created by blend can thus have a wider range of emotions.

However, one limitation in the geometric analysis is that the experimental context of the present study and McAdams et al. (2017) is different. Participants were exposed to different sets of stimuli in the two studies, so the ratings of the two experiments need to be standardized. In particular, the wider range of sounds in McAdams et al. (2017) may have resulted in a more compact structure of the subset of constituent samples used in the present study. A supplemental experiment will be conducted in the near future to standardize the experimental context of affect ratings on individual sounds and blended sounds. All ten individual sounds will be included along with some of the blended stimuli samples, which will be randomly selected from the stimulus set used in this experiment. The selection strategy may refer to Eerola et al.’s study (2012) in selecting

18 sounds from 110 sounds according to the affect space of 110 sounds, which used an even grid (3 * 3, defined by 33.3% and 66.7% percentiles in data) overlaid on the emotion space with two sounds being sampled from the original sounds in each cell of the grid. This strategy will be adopted to make sure the blended stimuli covering the whole affective space. All geometric analysis will be reconducted using the standardized data.

4.2 Relation of emotions of constituent sounds to composite sounds

According to the geometric analysis, all pairs of constituent sounds and their corresponding composite sounds have a triangular relationship, and there was no intermediateness, which means that the emotion of a blend does not simply lie on a line between the emotions of the constituent sounds. It is therefore difficult to predict the position in emotion space of the blended sound from the positions of the constituent sounds. It is interesting to observe that many of the pairs reveal certain patterns: the affect of a blend may be internal to or beyond the emotional scope of the constituents and may also be dominated by one constituent over the other. We found that the dominance of a constituent can be well explained by the spectral centroids of the constituent sounds with the sound having a brighter timbre tending to pull the blend closer to it in the emotion space. However, we have not found a way to predict when the perceived emotion of a blended sound will be within or beyond the range of coordinates in emotion space of the constituent sounds. In Sandell's (1991) study, he analyzed the degree of blend using the difference and the sum of the audio descriptors of constituent sounds, which maybe also useful in analyzing the relations in triangles in the future.

4.3 Acoustics and emotion

In general, according to our results, a small number of acoustic features (less than eight) is useful to interpret listeners' perceived emotion. The valence of blended sounds is positively correlated with spectral centroid; in other words, brighter sounds may result in a more positive valence. Valence is also positively correlated with the IQR of spectral centroid, which means that blended sounds with more changes in brightness may also have a more positive valence. Similar to valence, tension and energy arousal are positively correlated with spectral centroid, and they are also

positively correlated with spectral spread, which means that brighter blended sounds with a richer spectrum may have higher arousal. Also, energy arousal is significantly negatively correlated with the log attack time, which means the blended sounds with a sharper attack may have higher energy arousal. In the present study, 38% of the variance in valence, 78% of the variance in tension arousal, and 79% of the variance in energy arousal were explained by audio descriptors. So valence remains somewhat more elusive in terms of its acoustic underpinnings.

Compared to the emotional qualities of ten individual sounds selected from McAdams et al. (2017), tension arousal shows a similar positive correlation with spectral centroid. Although energy arousal did not have a significant predictor, the spectral centroid was still selected by lasso regression, and the coefficient had similar trends as the energy arousal of blended sounds. However, valence has a different relation to spectral centroid. It is also negatively correlated with the log attack time and positively correlated with the tristimulus 2, similar to Eerola et al. (2012). However, when compared to the Principal Components (PC) result in McAdams et al. (2017) for all 137 instrumental sounds, the acoustics for valence show consistency with our results, which were positively correlated with spectral centroid median and IQR in PC1. Therefore, we found that arousal shows more consistency, and the acoustic features usually show a good ability to explain the variance across different studies. Although the audio descriptors can moderately explain the variance in valence, it is still hard to find the overall consistency across the studies, which may need more musical context to evaluate. Eerola et al. (2012) also found that valence was hard to predict, and arousal was much more robustly explained by audio descriptors.

Differences between the acoustic features of blended and individual sounds still need to be discussed. According to the Hierarchical Cluster Analysis, the temporal variability (the IQRs of acoustic features) of audio descriptors shows more uniqueness in blended sounds than in individual sounds. Similarly, the IQR of spectral centroid is a highly significant predictor of valence. This shows that the variance of the acoustic features may play an important role in the emotional perception of the blend, which is not that obvious in individual sounds.

4.4 The degree of blend

In the present study, the similarity of constituent sounds in terms of spectral flux and temporal centroid had the most significant influence on the degree of blend (as the sounds are closer in

blend space with similar spectral flux and temporal centroid), which means the similar variance of the spectrum and the similar behaviour of the amplitude envelope could achieve a better degree of blend. This result is also consistent with the Gestalt principle of common fate according to which sounds that change in a similar manner are likely to be perceived as originating from the same source (Bregman, 1990). Compared to the acoustic result from Sandell (1991), although we are using different audio descriptors, we both found that the similarity of both spectral and temporal behaviour achieves a better blend.

As for the visualization of the blend space, the social network analysis shows a similar set of relations as MDS structure in the present study, which could also be a choice to analyze blend. Furthermore, it is very simple to use and has a lot of settings and parameters to finetune the visualization. It is also very helpful to see the structural information in the data, such as clusters, which means a group of sounds blend well with each other in this context. However, given the small size of our dataset, this technique may be more useful for a larger stimulus set.

4.5 Emotion and the degree of blend

One of our hypotheses was that the degree of blend could be used to interpret listeners' perceived affects. However, the correlations between the degree of blend and all three dimensions of affect are small. There are, nonetheless, some interesting patterns that the degree of blend could also help to interpret. In Figure 3.17, the flute, bassoon, and tuba show the best degree of a blend among all the instruments; it is interesting to see that the energy of many of the pairs with these three instruments becomes lower than the energy of at least one of the constituent sounds as shown in Figure 3.16, such as the FLEH, TuTp, and BnVc pairs. Also, all pairs with low energy and tension arousal (<0.4 ; normalized from 0 to 1) include flute, tuba, or bassoon. This may be because good blenders such as flute, tuba, or bassoon "soften" the other sound and thus be perceived as conveying a lower degree of arousal. Interestingly, Sandell (1991) found that some orchestration treatises mentioned good blenders "softening" other timbres, and he found that the most frequent agent for softening other instruments is the flute. Rogers (1951), for example, stated that, "The high brass loses brilliance when doubled in unison by woodwinds. Its tone becomes thicker but less incisive. Some of the flashing edges are lost." (p. 105)

Therefore, although we could not find a direct linear relationship between the degree of blend and the perceived affects, it is still early to say that there is no relationship. More studies can be done to investigate their relations in the future.

4.6 Musicians and nonmusicians

Although we did not specifically recruit participants in musician and nonmusician groups, and the results did not show a significant effect of musical sophistication on emotion, we still observed some differences during the experiment. Participants' ratings were not very different in the emotion rating part, but participants who reported having a musical background tended to finish the blend rating part faster than the nonmusicians according to observations during the experiment. We also found that it is usually more difficult for nonmusician participants to understand the concept of blend, perhaps because we used relatively well-blended stimuli so the differences in blend may have been subtle. However, after the practice phase and based on informal reports after the experiment, nonmusicians were able to judge the degree of blend as well as musicians. According to this observation, we think it may be because the concept of the blend is used more often in the music field. So nonmusicians may find it hard to understand, but the ratings of blend show consistency among the participants.

5 CONCLUSION

5.1 General conclusion

This study investigated listeners' perceived affect of sustained instrumental blends. Forty-five sustained instrumental blended pairs were investigated. The experiment was a self-report experiment with two separate blocks of emotion ratings and blend ratings. According to our analysis, a blend creates a new timbre with emergent acoustic features and conveys different emotions. Blended sounds may also span a broader range of the emotion space than the constituent sounds, so it may be helpful for musicians to use blends to express more varied emotions. A compact set of acoustic features was useful to explain the emotional qualities of both the blended and individual sounds, and we found that the variability over time of the acoustic features often plays a more important role in the perception of the blended sounds than of individual sounds.

In the emotion space, the composite sounds are not simply in between the constituent sounds but in a triangle configuration: some blends are within the emotional scope of the constituents, whereas others are beyond that scope. Also, some constituents dominate in their influence on the affect of the blend. It is still hard to simply use the emotion of constituent sounds to predict the emotion of the blended sounds; more patterns and the relationship between the acoustic features may help in the future. We did not find a direct relationship between the degree of blend and perceived affect, but "good blenders" tend to "soften" the timbre and thus might lower the perception of arousal. A deeper relationship could be investigated in the future. Although blend in a musical context is not always limited to instrumental unison dyads like those used in the present study, this study does attest to the importance of timbral and orchestration features in conveying affect.

5.2 Future study

A supplemental experiment will be conducted in the near future to standardize the experimental context of affect ratings on individual sounds and blended sounds. All ten individual sounds will be included, and some blended stimuli will be randomly selected from the stimulus set used in this experiment. The selection strategy may refer to Eerola et al.'s study (2012) as mentioned in section 4.1.

Broader future research could examine the affect conveyed by timbre in more complicated orchestration techniques and include more musical context. This could start from commonly used orchestration combinations, like string quartet, wind quintet (e.g., Kendall & Carterette, 1991), etc. Some broader contexts of timbre, like contemporary sound effects, synthesized sounds, vocals and instruments, and cross-cultural instrumental blends, could also be investigated in the future.

Appendix A: LOUDNESS-MATCHED LEVELS

Table A.1 Level adjustment applied to each constituent sound in loudness matching to the bassoon sound.

Instrument	Loudness Adjustment (dB)
Tu	2.27
Tb	3.37
Hn	-1.15
Tp	0.82
Bn (standard)	0.00
Cl	0.42
Ob	-0.52
EH	-0.21
Fl	0.06
Vc	-1.71

Table A.2 Sound pressure level for final stimuli.

Blend Pairs	Sound Level (dB)	Blend Pairs	Sound Level (dB)	Blend Pairs	Sound Level (dB)
TuTb	74.0	TpBn	75.2	ObEH	78.7
TuHn	77.7	TpCl	77.5	ObFl	77.0
TuTp	76.1	TpOb	75.4	ObVc	77.4
TuBn	80.5	TpEH	78.3	EHTu	77.1
TuCl	79.2	TpFl	75.5	EHTb	76.6
TuOb	76.9	TpVc	74.4	EHHn	76.5
TuEH	76.7	BnTu	81.3	EHTp	76.7
TuFl	75.8	BnTb	75.3	EHBn	77.1
TuVc	74.1	BnHn	79.1	EHCl	78.1
TbTu	74.2	BnTp	74.6	EHOb	78.4
TbHn	76.4	BnCl	79.0	EHFl	77.2
TbTp	73.5	BnOb	77.9	EHVc	77.8
TbBn	75.2	BnEH	76.9	FlTu	76.3
TbCl	75.0	BnFl	78.6	FlTb	73.4
TbOb	76.9	BnVc	73.9	FlHn	76.3
TbEH	75.5	ClTu	77.3	FlTp	75.4
TbFl	72.3	ClTb	73.6	FlBn	80.2
TbVc	71.0	ClHn	78.5	FlCl	79.1
HnTu	76.8	ClTp	76.5	FlOb	75.8
HnTb	74.3	ClBn	78.7	FlEH	74.2
HnTp	74.8	ClOb	75.4	FlVc	72.5
HnBn	79.1	ClEH	77.8	VcTu	73.4
HnCl	79.5	ClFl	76.5	VcTb	69.9
HnOb	78.4	ClVc	76.2	VcHn	75.6
HnEH	76.8	ObTu	76.9	VcTp	74.8
HnFl	77.2	ObTb	76.2	VcBn	71.9
HnVc	76.6	ObHn	77.3	VcCl	76.3
TpTu	75.2	ObTp	79.9	VcOb	76.7
TpTb	73.2	ObBn	77.4	VcEH	75.8
TpHn	75.4	ObCl	76.9	VcFl	73.1

Appendix B: ONSET SYNCHRONIZATION

Table B.1 Time offsets between constituents of each blend pair (values are the number of milliseconds that a row's stimulus should be delayed to be in synchrony with a column's stimulus), e.g., EH precedes Vc by 21.05 ms, whereas flute is delayed relative to Vc by 10.20 ms.

Median Offset (ms)	Tu	Tb	Hn	Tp	Bn	Cl	Ob	EH	Fl	Vc
Tu	-	-	-	-	-	-	-	-	-	-
Tb	14.55	-	-	-	-	-	-	-	-	-
Hn	23.20	-12.35	-	-	-	-	-	-	-	-
Tp	21.80	-2.90	2.20	-	-	-	-	-	-	-
Bn	8.70	-16.70	-6.55	-8.70	-	-	-	-	-	-
Cl	-16.70	-44.25	-26.10	-47.15	-24.65	-	-	-	-	-
Ob	23.25	0.75	6.55	-8.70	7.25	26.10	-	-	-	-
EH	23.20	4.35	4.35	-2.20	7.30	23.25	-4.40	-	-	-
Fl	-7.95	-13.10	-18.15	-23.20	-5.10	7.25	-13.05	-29.00	-	-
Vc	26.10	-12.35	8.70	-13.10	1.45	50.05	-13.05	-21.05	10.20	-

Appendix C: PARTICIPANT VALIDATION

Table C.1 Cronbach's α table for individual participant reliability statistics. The second column "Cronbach's α if item dropped" shows how the overall Cronbach's α would change if one participant is removed from the dataset. The third column "Item-rest correlation" shows the correlation between each participant and the rest of the participants in the scale.

Participant ID	Cronbach's α if item dropped	Item-rest correlation
1	0.921	0.550
2	0.921	0.510
3	0.922	0.483
4	0.922	0.477
5	0.922	0.552
6	0.922	0.467
7	0.922	0.433
9	0.919	0.676
10	0.921	0.644
11	0.922	0.479
12	0.920	0.570
13	0.921	0.577
14	0.922	0.519
16	0.921	0.552
17	0.922	0.450
18	0.922	0.478
19	0.923	0.363
20	0.924	0.191
21	0.922	0.434
22	0.922	0.448
23	0.922	0.475
24	0.922	0.572
25	0.922	0.408
26	0.922	0.415
27	0.922	0.484
28	0.924	0.277
29	0.923	0.361
30	0.920	0.610
31	0.922	0.487

Participant ID	Cronbach's α if item dropped	Item-rest correlation
32	0.920	0.601
33	0.923	0.387
34	0.921	0.563
35	0.924	0.299
37	0.922	0.429
40	0.922	0.523
41	0.923	0.345
42	0.921	0.538
43	0.920	0.622
44	0.921	0.561
45	0.922	0.462

Appendix D: ASSUMPTION CHECK

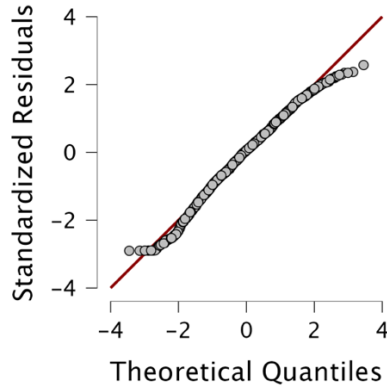


Figure D.1 Q-Q plot for normality check of perceived valence for blended sounds.

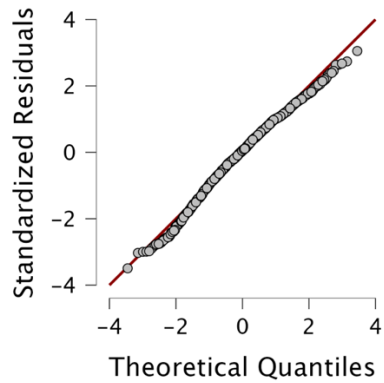


Figure D.2 Q-Q plot for normality check of perceived energy for blended sounds.

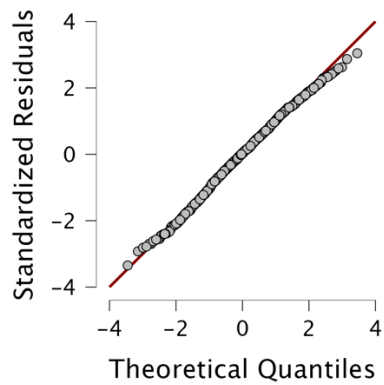


Figure D.3 Q-Q plot for normality check of perceived tension for blended sounds.

Table D.1 Homogeneity of variance of perceived valence for blended sounds.

Test for Equality of Variances (Levene's)			
F	df1	df2	p
1.064	44	1755	0.360

Table D.2 Homogeneity of variance of perceived energy for blended sounds.

Test for Equality of Variances (Levene's)			
F	df1	df2	p
1.540	44	1755	0.013

Table D.3 Homogeneity of variance of perceived tension for blended sounds.

Test for Equality of Variances (Levene's)			
F	df1	df2	p
0.946	44	1755	0.573

BIBLIOGRAPHY

- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & B. F. Csaki (Eds.), *2nd International Symposium on Information Theory* (pp. 267–281). Akadémiai Kiadó: Budapest.
- Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: An open source software for exploring and manipulating networks. *Proceedings of the Third International AAAI Conference on Web and Social Media*, 3(1), 361–362. <https://doi.org/10.1609/icwsm.v3i1.13937>
- Bigand, E., Vieillard, S., Madurell, F., Marozeau, J., & Dacquet, A. (2005). Multidimensional scaling of emotional responses to music: The effect of musical expertise and of the duration of the excerpts. *Cognition & Emotion*, 19(8), 1113–1139. <https://doi.org/10.1080/02699930500204250>
- Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. The MIT Press. <https://doi.org/10.7551/mitpress/1486.001.0001>
- Caetano, M., Depalle, P., & McAdams, S. (2022, August 6). Multidimensional scaling (MDS) of simplified musical instrument sounds morphed with additive synthesis. *Society for Music Perception and Cognition*. Portland, OR.
- 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>
- de Leeuw, J., & Mair, P. (2009). Multidimensional scaling using majorization: SMACOF in R. *Journal of Statistical Software*, 31(3), 1–30. <https://doi.org/10.18637/jss.v031.i03>
- 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>
- Eerola, T., & Vuoskoski, J. K. (2013). A review of music and emotion studies: Approaches, emotion models, and stimuli. *Music Perception*, 30(3), 307–340. <https://doi.org/10.1525/mp.2012.30.3.307>
- 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>
- Elliott, T. M., Hamilton, L. S., & Theunissen, F. E. (2013). Acoustic structure of the five perceptual dimensions of timbre in orchestral instrument tones. *The Journal of the Acoustical Society of America*, 133(1), 389–404. <https://doi.org/10.1121/1.4770244>

- 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>
- Filipic, S., Tillmann, B., & Bigand, E. (2010). Judging familiarity and emotion from very brief musical excerpts. *Psychonomic Bulletin & Review*, 17(3), 335–341. <https://doi.org/10.3758/PBR.17.3.335>
- Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33(1), 1–22. <https://doi.org/10.18637/jss.v033.i01>
- Gordon, J. W. (1987). The perceptual attack time of musical tones. *The Journal of the Acoustical Society of America*, 82(1), 88–105. <https://doi.org/10.1121/1.395441>
- Grey, J. M. (1975). *An exploration of musical timbre*. [Doctoral dissertation, Stanford University].
- Grey, J. M. (1977). Multidimensional perceptual scaling of musical timbres. *The Journal of the Acoustical Society of America*, 61(5), 1270–1277.
- Holmes, P. A. (2011). An exploration of musical communication through expressive use of timbre: The performer's perspective. *Psychology of Music*, 40(3), 301–323. <https://doi.org/10.1177/0305735610388898>
- ISO389–8. (2004). *Acoustics – Reference zero for the calibration of audiometric equipment – Part 8: Reference equivalent threshold sound pressure levels for pure tones and circumaural earphones* (Tech. Rep.). Geneva, Switzerland: International Organization for Standardization.
- Jacomy, M., Venturini, T., Heymann, S., & Bastian, M. (2014). ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PLoS ONE*, 9(6), 1–12. <https://doi.org/10.1371/journal.pone.0098679>
- 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., & Sloboda, J. A. (2010). *Handbook of music and emotion: Theory, research, applications*. Oxford University Press.
- Juslin, P. N., & Västfjäll, D. (2008). Emotional responses to music: The need to consider underlying mechanisms. *Behavioral and Brain Sciences*, 31(5), 559–575. <https://doi.org/10.1017/S0140525X08005293>
- Kazakis, S., Depalle, P., & McAdams, S. (2022). *The Timbre Toolbox User's Manual*. <https://github.com/MPCL-McGill/TimbreToolbox-R2021a>
- Kendall, R. A., & Carterette, E. C. (1991). Perceptual scaling of simultaneous wind instrument timbres. *Music Perception*, 8(4), 369–404.
- Kendall, R. A., & Carterette, E. C. (1993). Identification and blend of timbres as a basis for orchestration. *Contemporary Music Review*, 9, 51–67. <https://doi.org/10.1080/07494469300640341>

- Krumhansl, C. L. (2002). Music: A link between cognition and emotion. *Current Directions in Psychological Science*, 11(2), 45–50. <https://doi.org/10.1111/1467-8721.00165>
- Lakatos, S. (2000). A common perceptual space for harmonic and percussive timbres. *Perception & Psychophysics*, 62(7), 1426–1439.
- Martin, F. N., & Champlin, C. A. (2000). Reconsidering the limits of normal hearing. *The Journal of the American Academy of Audiology*, 11(2), 64–66.
- Mair, P., Groenen, P. J. F., & de Leeuw, J. (2022). More on multidimensional scaling and unfolding in R: Smacof Version 2. *Journal of Statistical Software*, 102(10), 1–47. <https://doi.org/10.18637/jss.v102.i10>
- McAdams, S. (1984). The auditory image: A metaphor for musical and psychological research on auditory organization. In W. R. Crozier & A. J. Chapman (Eds.), *Cognitive processes in the perception of art* (pp. 289–323). North Holland. [https://doi.org/10.1016/S0166-4115\(08\)62356-0](https://doi.org/10.1016/S0166-4115(08)62356-0)
- McAdams, S. (2019). Timbre as a structuring force in music. In K. Siedenburg, C. Saitis, S. McAdams, A. N. Popper, & R. R. Fay (Eds.), *Timbre: Acoustics, perception, and cognition* (pp. 211–243). Springer International Publishing. https://doi.org/10.1007/978-3-030-14832-4_8
- 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., Winsberg, S., Donnadieu, S., De Soete, G., & Krimphoff, J. (1995). Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes. *Psychological Research*, 58, 177–192. <https://doi.org/10.1007/BF00419633>
- Moore, B. C. J., & Glasberg, B. R. (1983). Suggested formulae for calculating auditory-filter bandwidths and excitation patterns. *The Journal of the Acoustical Society of America*, 74, 750–753. <https://doi.org/10.1121/1.389861>
- Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). The musicality of non-musicians: An index for assessing musical sophistication in the general population. *PLoS ONE*, 9(2), e89642. <https://doi.org/10.1371/journal.pone.0089642>
- 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>
- Peretz, I., Gagnon, L., & Bouchard, B. (1998). Music and emotion: Perceptual determinants, immediacy, and isolation after brain damage. *Cognition*, 68, 111–141.
- Piedmont, R. L. (2014). Inter-item correlations. In A. C. Michalos (Ed.), *Encyclopedia of quality of life and well-being research* (pp. 3303–3304). Springer Netherlands. https://doi.org/10.1007/978-94-007-0753-5_1493
- R Core Team. (2022). *R: A language and environment for statistical computing*. <https://www.r-project.org/>

- Rogers, B. (1951). *The art of orchestration: Principles of tone color in modern scoring*. Appleton-Century-Crofts.
- Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6), 1161–1178.
- Sandell, G. J. (1991). *Concurrent timbres in orchestration: A perceptual study of factors determining “blend.”* [Doctoral dissertation, Northwestern University].
- Sandell, G. J. (1995). Roles for spectral centroid and other factors in determining “blended” instrument pairings in orchestration. *Music Perception*, 13(2), 209–246.
<https://doi.org/10.2307/40285694>
- 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](https://doi.org/10.1002/1099-0984(200007/08)14:4<325::AID-PER380>3.0.CO;2-I)
- 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>
- Sloboda, J. A., & O’Neill, S. A. (2001). Emotions in everyday listening to music. In P. N. Juslin & J. A. Sloboda (Eds.), *Music and emotion: Theory and research* (pp. 415–430). Oxford University Press.
- Smith, B. K. (1995). PsiExp: An environment for psychoacoustic experimentation using the IRCAM musical workstation. In D. L. Wessel (Ed.), *Society for Music Perception and Cognition Conference, Berkeley, CA* (pp. 83–84).
- Tardieu, D., & McAdams, S. (2012). Perception of dyads of impulsive and sustained instrument sounds. *Music Perception*, 30(2), 117–128. <https://doi.org/10.1525/MP.2012.30.2.117>
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267–288.
<https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>
- Tibshirani, R. (2011). Regression shrinkage and selection via the lasso: A retrospective. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(3), 273–282.
<https://doi.org/10.1111/j.1467-9868.2011.00771.x>
- Vuoskoski, J. K., & Eerola, T. (2017). The pleasure evoked by sad music is mediated by feelings of being moved. *Frontiers in Psychology*, 8, 439. <https://doi.org/10.3389/fpsyg.2017.00439>
- Wasserman, S., & Faust, K. (1994). *Social Network Analysis: Methods and applications (structural analysis in the social sciences)*. Cambridge University Press.
<https://doi.org/https://doi.org/10.1017/CBO9780511815478>
- Zentner, M., Grandjean, D., & Scherer, K. R. (2008). Emotions evoked by the sound of music: Characterization, classification, and measurement. *Emotion*, 8(4), 494–521.
<https://api.semanticscholar.org/CorpusID:447039>