Follow the Bouncing Ball: Music, Motion, and Emotion

A Thesis

Submitted to the Faculty

in partial fulfillment of the requirements for the degree of

Master of Arts

in

DIGITAL MUSICS

by

Beau Sievers

DARTMOUTH COLLEGE

Hanover, New Hampshire

May 30, 2010

(chair) Michael Casey

Thalia Wheatley

Larry Polansky

Brian W. Pogue, Ph.D.
Dean of Graduate Studies

Abstract

We suggest musical emotion operates mimetically with respect to emotion-signifying movement. In service of this hypothesis, we present an experiment in which subjects use a computer program to create representations of several emotions in either music or animated motion, and the dynamic properties of these representations are quantitatively compared. We show the representations created by subjects are significantly more similar within emotion groups than within task groups, providing strong evidence in support of our hypothesis. Along the way, we propose a schema for classifying cross-modal mappings, and argue that cross-modal mapping in general, and music-motion-emotion mappings in particular, are good candidates for universal human predispositions.

Acknowledgements

Thanks to my parents, William and Daria, for their unconditional support of everything I do, no matter how obscure, and to Joanne Cheung, for being endlessly encouraging and exceedingly patient. To my advisors: Thalia Wheatley, Larry Polansky, and Michael Casey; without their inspiration and assistance, none of this work would have been possible. Special lucky-bonus thanks to Newton Armstrong, who provided invaluable council. To Jon Appleton, founder of the Dartmouth Digital Musics program, for his help and advice. To Daniel Leopold, James Hughes, and Swaroop Guntupalli for helping with the behavioral and fMRI experiments. To Rebecca Fawcett, for keeping everything in order. To everybody who talked with me about the work, much of which is inspired by conversations with Walter Sinnott-Armstrong, Theodore Levin, Peter Tse, Yune-Sang Lee, David Dunn, Christine Looser, Chris Peck, Guy Madison, Ioana Chitoran, Spencer Topel, Patrick Barter, Michael Chinen, Kristina Wolfe, Paul Osetinsky, Dave Yoss, Nicolai Buhr, and Brendan Landis. To my professors at Dartmouth: Kui Dong, Charles Dodge, Doug Perkins, Aden Evens, Amy Allen, Jody Diamond, Fred Haas, Don Glasgo, Bill Kelley, Scot Drysdale, and Afra Zomorodian. And to Bronwen Evans, for awakening my interests in cognitive neuroscience and philosophy of mind.

Contents

List of Tables

List of Figures

1 Introduction

We perceive the world all at once. Making sense of our environment requires the integration of multiple simultaneous perceptual streams. Tracking relationships between these streams is how we understand what is happening around us. This is a matter of great evolutionary importance: we have good reason to be alert if what we are looking at jumps and makes a sound, or if we hear something at or beyond the limits of our field of vision. Correspondences such as these are not rationally identified – solemnly contemplating the roar leads into the lion's mouth – but are a function of involuntary perceptual processes. Co-occurrence, for example, is one of several natural indicators of correspondence: when the dynamic contours of a sound and a movement are synchronized, we tend to perceive the two as a unified whole, even if the movement is not the source of the sound. This tendency is flexible and promiscuous: we often hear sound as signifying movement even if there is no real movement at all. The capacity to cross-modally map sound to implied movement helped our ancestors survive in the wilderness; a more common use is the enjoyment of a well-made film. We argue that this capacity is what allows us to hear music as signifying emotion. Following Meyer's (1956) hypothesis that "music can be heard as imitating the dynamics of behavior", we suggest musical emotion operates mimetically with respect to emotion-signifying movement.

The present work has two goals. First, we hope to provide a preliminary theoretical framework for understanding cross-modal mapping, and to point toward possible programs for future research. To this end, we undertake an extensive review of the relevant literature. This review serves to draw connections between projects which have something to say about cross-modal

mapping, but have remained more-or-less isolated. In doing so, we veer far afield of emotion recognition in its most basic sense, exploring such subjects as synesthesia, infant language learning strategies, and the history of universalist versus relativist approaches to musicological practice. In the course of this review, we make two material offerings. We propose a schema for classifying cross-modal mappings, and we provide evidence that cross-modal mapping in general, and music-motion-emotion mappings in particular, are good candidates for universal human predispositions. Our second goal is to provide empirical evidence that musical emotion is imitative of emotion-signifying movement. To this end, we present an experiment in which subjects use a computer program to create representations of several emotions in both music and animated motion, and the dynamic properties of these representations are quantitatively compared.

2 Background

2.1 What is emotion?

Any study concerning emotion is obliged to provide at least a cursory, working explanation of what "emotion" might be. The question "What is emotion?" is ancient, and most or all proposed definitions seem inadequate – either too loose for inclusion in scientific discourse, or so rigid as to sever every link to ordinary language. Taxonomizing and cataloging the history of this problem is beyond the scope of this work, as is any serious attempt to solve it. Rather than heroically proposing some new gymnastic definitional criteria, we dodge the problem by splitting it in two.

Directly following the work of Andrea Scarantino (2005), we see the question "What is emotion?" as concealing two related but fundamentally

separate problems. Many (if not most) attempts to define "emotion" have been stymied by a tendency to conflate these two problems; our strategy will be to focus on one and delegate the other to ambitious philosophers. The first problem is finding a definition of emotion suitable for scientific use. Implied here is the process of examining the inexact, scientifically unsuitable, ordinary use of the term "emotion" and proceeding to explicate, a la Carnap (1950), new related theoretical constructs. These constructs would be much narrower in meaning than those implied by ordinary language, but may outline categories constituting what Quine (1969) calls *natural kinds*, or groups of things about which scientific generalizations are possible. Scarantino (2005) refers to the production of these constructs as the Explicating Emotion Project.

The second problem is figuring out how the ordinary language term "emotion" is used. Rather than constructing scientific categories with necessary and sufficient conditions for inclusion, analysis of ordinary language suggests using a family resemblance model (Wittgenstein, 1953). That is, something can be considered a member of the emotion family if it has a certain number of "emotional" traits, but it is possible for family members to possess non-overlapping sets of traits and thus be fundamentally dissimilar. An account of emotion in these terms would be completely compatible with (and indeed derived from) ordinary language, but resistant to scientific generalization. Emotions in this sense may not form a natural kind (Griffiths, 2004). Scarantino calls the task of explaining emotional family resemblance the Folk-Emotion Project. In its understanding of emotion as a general concept, the present work takes the Folk-Emotion Project as its basis.

Investigation of how "anger" – or any other emotional concept – manifests can proceed perfectly well without a precise, scientific definition. It is sufficient for most purposes to ensure that all exemplars of the invoked concept

are empirically certified as authentic. That is, if subjects in an experiment view or listen to a stimulus and label it as "angry", and this labeling is shown to be statistically significant, we should trust them. In other words, at least with respect to natural language and everyday human experience, "emotion" is whatever people say it is.

The details of an experimental methodology may imply a certain theoretical stance toward emotion. For example, asking subjects to rank the emotion in a piece of music on a scale from "positive" to "negative" assumes that emotion varies on some dimension which corresponds to those terms. Our methodology assumes only that emotions can be grouped into categories based on similarity. We hope this assumption is relatively theory-neutral; that is, it may mesh fruitfully with dimensional, social, or other theories of emotion, so long as those theories admit emotions may be compared to other emotions on the basis of similarity or dissimilarity. That this is true with respect to everyday emotional thinking in Western subjects has been confirmed by Shaver et al. (1987). Shaver asked subjects to sort cards printed with emotion words into categories, creating an emotional distance matrix which was subjected to cluster analysis. This analysis found emotions fit into five broad categories: love, joy/surprise, anger, sadness, and fear.

The present study focuses on recognition, and not induction, of emotion. Emotion induction is a complicated process not cleanly reducible to a small set of investigable factors. It is easy to conceive, for example, of a happy-seeming piece of music which makes a listener feel sad by way of an ironic contextual presentation, such as in a film when a character has just died. Further, the idea that emotional induction during music listening can occur at all is somewhat controversial. Some philosophers of music, such as Kivy (1989) and Meyer (1956), believe listeners habitually confuse induction and recognition. Listeners

may say they are feeling emotions sympathetically with some music, when in fact they are only recognizing those emotions as being expressed. Determining the fact of the matter is difficult, in no small part because the problem itself calls into question the judgment of the experimental subjects. A number of strategies have been devised to work around this stumbling block, including observing cognitive changes during listening (Martin & Metha 1997; Balch et al, 1999), measuring physiological arousal (Bartlett, 1996; Krumhansl 1997), and taking detailed subject reports (Kenealy 1988; Zentner et al, 2000; Sloboda & O'Neill, 2001). A review of over 100 studies by Juslin and Laukka (2004) concludes there is sufficient evidence to claim that music induces emotions, but also notes that the range of emotions induced by music seems to be very different from the range of emotions which can be recognized in it.

The reduced range of musically induced emotion is demonstrated by Sloboda and O'Neill (2001). Using an experimental paradigm inspired by Csikszentmihalyi and Lefevre (1989), the authors equipped subjects with pagers and paged them at random once within every 2-hour interval during the day. When paged, subjects were instructed to write down in a log book information about their most recent experience of music listening. Subjects reported music either made them feel "more positive, more alert, and more focused in the present", or caused them to become nostalgic, thinking about "things and people not present". The difference in range between these reports and everyday thinking about musically induced emotion is good evidence that induction and recognition are decoupled.

Conveniently, emotion recognition in music is much easier to study and substantiate than induction. If a subject claims to recognize an emotion in music, there is little reason to distrust them, so subject reports become much more

useful. Further, while verifiable reports of emotional induction are ambiguous and sparse, reports of emotion recognition are specific and plentiful.

2.2 Emotion Recognition

A survey of the literature on recognition of emotions in various modalities follows, with special emphasis on results which suggest recognition in different modalities may be governed by similar processes.

2.2.1 Emotion recognition in music

That emotion may be recognized in music is uncontroversial. More interesting is to ask which emotions may be recognized, and which are off-limits. Certainly some music is happy, and other music is sad, but is there jealous music? Guilty music? Further, what properties allow us to distinguish happy from sad music? Are these judgments unreliable and subjective, or is there a broad consensus about what music is happy and what music is sad?

Juslin and Laukka's (2004) meta-study concluded that music listeners' judgments regarding emotion are "systemic and reliable, and can thus be predicted with reasonable accuracy." Such judgments were only marginally affected by the level of musical training, age, and gender of the listener. That is, there is broad consensus regarding the recognition of emotional content in music. Further, the emotion recognition process is nearly immediate. Peretz et al. (1998) found subjects were able to accurately judge the emotional tone of musical excerpts in well under 3 seconds. However, subjects were less likely to agree on differences within emotional categories, indicating limits on the precision with which music can represent emotion. Juslin (2005) lists three possible reasons for this imprecision. First, it may be that music's ability to

communicate emotions is “heavily dependent on its similarity to other forms of non-verbal communication” and is thus similarly limited. Second, there is a great deal of redundancy built in to the structure of musical features which communicate emotion, placing a limit on expressive specificity. Third, music is not designed solely to convey emotion, and so other considerations may often take precedence.

Juslin and Laukka (2004) also compiled a list of the emotional states most frequently recognized by listeners along with their attendant musical features. Their approach implies a parameterization of music based on the terminology of Western music theory. Their results are summarized in table 1. [1] While this analysis is limited by its reliance on the vocabulary of Western music theory, it does indicate that emotion recognition is reliably accompanied by certain sets of musical features. That is, the presence of these musical features seems to be a sufficient condition for recognition of the emotion.

Juslin (2000) suggests a probabilistic model for quantifying the contributions of various musical features to both the production and recognition of emotion in music. This model is based on Brunswik's (1956) lens model and Hursch's (1964) lens model equation (LME). Thirty listeners judged the emotions expressed in performances of three short musical pieces by three professional guitarists. Multiple regression analysis was used to determine the relationship between musical features and performer intention, as well as the relationship between musical features and listener interpretation. The LME was then applied to quantify how closely the expressive codes of the performers matched the interpretive codes of the listeners. This approach yielded results similar to those summarized above: “Anger was associated with fast tempo, high

[1] While “timing variability” is associated with rubato, Juslin & Laukka are unclear about how timing variability is distinguished from tempo variability.

Anger	Fast tempo, small tempo variability, minor mode, atonality, dissonance, high sound level, small loudness variability, high pitch, small pitch variability, ascending pitch, major 7th and augmented 4th intervals, raised singer's formant, staccato articulation, moderate articulation variability, complex rhythm, sudden rhythmic changes (e.g., syncopations), sharp timbre, spectral noise, fast tone attacks/decays, small timing variability, accents on tonally unstable notes, sharp contrasts between "long" and "short" notes, accelerando, medium-fast vibrato rate, large vibrato extent, micro-structural irregularity
Happiness	Fast tempo, small tempo variability, major mode, simple and consonant harmony, medium-high sound level, small sound level variability, high pitch, much pitch variability, wide pitch range, ascending pitch, perfect 4th and 5th intervals, rising micro intonation, raised singer's formant, staccato articulation, large articulation variability, smooth and fluent rhythm, bright timbre, fast tone attacks, small timing varibility, sharp contrasts between "long" and "short" notes, medium-fast vibrato rate, medium vibrato extent, micro-structural regularity
Tenderness	Slow tempo, major mode, consonance, medium-low sound level, small sound level variability, low pitch, fairly narrow pitch range, lowered singer's formant, legato articulation, small articulation variability, slow tone attacks, soft timbre, moderate timing variability, soft contrasts between long and short notes, accents on tonally stable notes, medium fast vibrato, small vibrato extent, micro-structural regularity
Sadness	Slow tempo, minor mode, dissonance, low sound level, moderate sound level variability, low pitch, narrow pitch range, descending pitch, "flat" (or falling) intonation, small intervals (e.g., minor 2nd), lowered singer's formant, legato articulation, small articulation variability, dull timbre, slow tone attacks, large timing variability (e.g., rubato), soft contrasts between "long" and "short" notes, pauses, slow vibrato, small vibrato extent, ritardando, micro-structural irregularity
Fear	Fast tempo, large tempo variability, minor mode, dissonance, low sound level, large sound level variability, rapid changes in sound level, high pitch, ascending pitch, wide pitch range, large pitch contrasts, staccato articulation, large articulation variability, jerky rhythms, soft timbre, very large timing variability, pauses, soft tone attacks, fast vibrato rate, small vibrato extent, micro-structural irregularity

Table 1: *Summary of results from Juslin & Laukka (2004)*

sound level, a lot of HF [high frequency] energy in the spectrum, legato articulation, and small articulation variability; sadness was associated with slow tempo, low sound level, little HF energy in the spectrum, legato articulation, and small articulation variability; happiness was associated with fast tempo, high sound level, intermediate amount of HF energy in the spectrum, staccato articulation, and much articulation variability; fear was associated with slow tempo, very low sound level, little HF energy in the spectrum, staccato articulation, and large articulation variability." (Juslin, 2000) Expressive codes used differed from performer to performer, but despite this, each performer was intelligible to all of the listeners. Presumably because melody is difficult to parameterize, and performers are scarce, the melodies used were fixed throughout the experiment, and melody-related parameters were not included in the regression analysis. This limitation prevented the study from assessing the effects of changes in contour, step size, consonance, etc.

With respect to emotional experience in day-to-day life, it is important to note that the emotions identified by the above studies (happiness, sadness, anger, fear, and tenderness) are a very small subset of states typically considered emotional. Shaver et al. (1987) asked 100 subjects to rate 213 possible emotion names on a scale of 1 to 4, with 1 meaning "I would definitely not call this an emotion" and 4 meaning "I would definitely call this an emotion". A number of high ranking candidates are notably absent from most studies on music and emotion, including jealousy (3.81), grief (3.65), guilt (3.53), embarrassment (3.49), shame (3.43), and disgust (3.42). What emotions can and cannot be expressed in music may be valuable information with respect to how, exactly, the recognition of musical emotions takes place.

2.2.2 Emotion recognition in speech

A complete review of historical approaches to emotional communication in speech is beyond the scope of the present work. I offer a brief survey of recent empirical work, drawing heavily from Scherer (2003).

Scherer (2003) notes that "the basis of any functionally valid communication of emotion via vocal expression is that different types of emotion are actually characterized by unique patterns or configurations of acoustic cues. [...] Without such distinguishable acoustic patterns for different emotions, the nature of the underlying speaker state could not be communicated reliably". Additionally, changes in acoustic parameters are linked to physiological changes, i.e. *experiencing* a given emotion changes the way people speak. The following quote is representative of how this idea is typically framed in the literature: "For instance, many of us have experienced talking in an unwittingly loud voice when feeling gleeful, speaking in an uncharacteristically high-pitched voice when greeting a sexually desirable person, or talking with marked vocal tremor while giving a public speech." (Bachorowski, 1999) It is important to note that this is not always true. Deliberate, expressive modulation of acoustic parameters not related to some authentically experienced emotional state is certainly possible, as in the cases of acting and deception. (Ekman et al., 1976; Anolli et al., 1997) The composition of music offers an analogous circumstance; a work may express or convey emotions the composer did not feel during composition, nor the performer during exhibition.

Studies reviewed by Scherer (2003) of how emotion is encoded into speech examined recordings of people in emotionally trying situations, subjects under the influence of emotion-altering psychoactive drugs, subjects who had undergone a battery of laboratory procedures designed to induce emotion, and actors simulating emotional states. Findings across these studies were consistent,

showing a stable set of acoustic features associated with each emotion studied. For example, anger was associated with increases in intensity, speed, the mean value of formant 0 (F0), and descending sentence contours; fear with increases in intensity, speed, indeterminate F0 range, and indeterminate contour; sadness with decreased intensity, speed, F0 mean and range, and falling contours; joy with increased intensity, speed, and F0 mean and range. It is significant that all of these acoustic parameters have equivalents in music.

2.2.3 Emotion recognition in human movement

Like music and speech, bodily movement can act as a conduit for the expression of emotion. The standard practice for capture and analysis of human movement in isolation is the point-light model, where small lights are attached to the joints of actors who are filmed moving in a room (Johansson, 1973). Viewing the lights, and not the body attached to them, provides a way to observe the movement of a person isolated from anything which could complicate interpretation, such as facial expression. While viewing point-light movement, people are able to identify such attributes as gender (Kozlowski and Cutting, 1977), vulnerability (Gunns et al., 2002), and emotion (Atkinson et al., 2004; Makeig, 2001; Pollick et al., 2001). Emotion may be perceived in point-light motion even when emotional expression is not the primary goal of the actor. (Pollick et al., 2001) Unlike research into music and speech, point-light motion studies tend not to address the relationship between parameterization of the stimuli and subjects' categorical judgments. This is probably because complex point-light motion does not have any intuitive or obvious parameterization, leading most researchers to favor machine learning analyses. These analyses yield dimensions on which automated classification of emotional movement is possible, but which do not necessarily have any relationship to human perception. Below, we focus on

studies which either take human perception as a starting point, or reframe the results of machine learning analyses in terms of perceptually valid categories.

Castellano et al. (2007) used automated classification techniques to analyze video of human movement with respect to emotion. They parameterized movement in terms of a number of properties related to amplitude variation and spectral centroid (a weighted average level of energy), and parameterized emotion along the dimensions of valence and arousal. While moderately successful, their model confused negative emotions with positive emotions having similar arousal characteristics (e.g. angry and happy), and confused positive emotions with other positive emotions with opposite arousal characteristics (e.g. happy and peaceful).

Amaya et al. (1996) created a system for modifying captured neutral human movement such that it expressed various emotions. Their system focused on amplitude variation, corresponding with Pollick's (2001) activation dimension, but they suggest that other emotional changes in the movement (perhaps variation in Pollick's pleasantness dimension) may be related to phase relations between joints. Badler et al. (1999) suggest a parameterization based on Laban Movement Analysis (Laban, 1960) – possibly a very interesting direction – but do not quantify their model or demonstrate how it might be used.

Pollick et al. (2001), using a dimensional model of emotion based on "activation" and "pleasantness", found high-activity emotions were associated with greater velocity, acceleration, and "jerk" in the expressive movement. Bernhardt and Robinson (2007) suggest a similar model. Interestingly, Pollick et al. (2001) found their results held with respect to the activation dimension even when the point-light displays were scrambled so they were no longer consistent with human movement. This suggests certain dynamic features present in point-

light motion are sufficient for emotion recognition even without a visible, moving human body.

2.3 Emotion as Dynamic Contour

In the modalities discussed above, each recognizable emotion is associated with a set of unique features. In music, for example, happiness is associated with fast tempo, major mode, ascending pitch, and so on. Remarkably, there are similarities in parametric variation *across modalities* for each emotion. Sad music, for example, has a slow tempo, low and descending pitch, and is dissonant (Juslin and Laukka, 2003). Sad speech has a slow articulation rate, low fundamental frequency, descending pitch, and is dissonant (Scherer, 2003). Sad door-knocking movement is "slow and slack" (Bernhardt and Robinson, 2007) and associated with low velocity, acceleration, and jerk (Pollick et al., 2001). Tempo, articulation rate, and slowness are all variations on the theme of speed, and sadness is slow whether in music, speech, or movement. This suggests the signifying power of a given parameter isn't limited to one medium, but can cut across media and perceptual modalities. Indeed, at least for speech and music, this tendency toward cross-modal parametric similarity has been confirmed by Juslin and Laukka (2003), a meta-study of 104 papers which concludes that "music performance involves mainly the same emotion-specific patterns of acoustic cues as does vocal expression".

Moving forward, we review a number of studies which consider in more detail mappings which act across perceptual modalities. We use the term cross-modal mapping to refer to any reliable association of activity in one modality to activity or implied activity in another. We suggest cross-modal mappings may be classified using the following three-level schema. First, a mapping may be either perceptual or cognitive. A cognitive mapping requires conscious effort to

understand; an example is looking at a chart and evaluating its meaning. In contrast, perceptual mappings are automatic, obligatory and unconscious: when we hear a fast, intense series of footsteps, we know immediately that there is a person running, without engaging in any active "reading" process. Second, mappings may be either intuitive or learned. Intuitive mappings are present from birth and require no acculturation or study. A good candidate for an intuitive mapping between movement and music in infants is described by Phillips-Silver and Trainer (2007) (discussed in detail in the following section). Learned mappings require study or acculturation; a familiar example of a learned mapping is the relationship between words in a language and their meanings. Provisionally, we suggest a third level. Mappings may be either *isomorphic*, *analogical*, or arbitrary. A mapping is isomorphic when the source and the result parameters occupy the same representational space. For example, slow music mapping to slow movement. We refer to this kind of mapping as *cross-modal parametric isomorphism*. Conversely, a mapping is analogical when the source and result parameters occupy different representational spaces which are mapped to one another, usually in terms of intensity or magnitude. For example, visual brightness mapping to pitch height. We call this type of mapping *inter-parametric analogy*. Finally, phenomena mapped arbitrarily may have no representational similarity, and are typically associated by rote memorization. Although the three levels of this schema are conceptually independent, there are correlations between levels; for example, arbitrary mappings are always learned.

We suggest that to recognize emotion in music, speech, or movement is, in part, to compare the dynamic contour of a stimulus via cross-modal mapping to an internal dynamic model of emotion. The cross-modal mappings implicated in this process may have any of the six properties we suggest in our classification schema; they may be perceptual or cognitive, intuitive or learned, and

isomorphic or analogical. In particular, we would like to highlight the possible importance of automatic perceptual and intuitive mappings in the understanding of emotion, especially insofar as that understanding appears to be a human universal (this is discussed at length in section 2.6).

2.4 Cross-modal connections

How does cross-modal mapping come to be possible? What are its uses, tendencies, and limits? The following section is a survey of research which, rather than focusing on a particular medium or modality, directly investigates cross-modal mapping itself.

Film and television viewing is a commonplace scenario where the cross-modal influence of music plays an important role. Cohen (1993) found that subjects' judgments of the affect of a bouncing ball were modified by music in a roughly additive manner. That is, when subjects viewed a happily bouncing ball, playing happy music made it appear more happy, and playing sad music made it appear less happy. Cohen (1993) also found that when the music contained ascending and descending major triads, this increased perceived happiness, but when minor triads were played, subjects' judgments were indeterminate. Cohen (1993) also describes two further experiments which used more realistic musical and visual stimuli and achieved similar results.

Eitan and Granot (2006) asked 95 college students to, while listening to music, visualize internally and then describe using a forced-choice questionnaire the movement of an imaginary animated character. The musical pieces used were short melodies, designed in pairs such that one piece featured an increase in some musical parameter and the other featured a decrease, with other parameters held constant. Parameters varied included dynamics, pitch direction, pitch intervals, attack rate or inter-onset interval, motivic pace, and articulation. They found

changes in the music affected imagined motion "significantly and diversely" as subjects employed a variety of music-to-motion mapping strategies. The most statistically significant results include increases in volume mapping to an approach motion or an increase in speed, decreasing volume mapping to descending motion or movement away from the subject, ascending pitch to ascending motion, descending pitch to descending motion, and decreasing inter-onset intervals (accelerandi) mapping increases in speed and vice-versa. Eitan and Granot note that a number of the mappings used by subjects were analogical in nature, such as ascending pitch "height" mapping to ascending position in imaginary space. Further, some cross-modal mappings were directionally asymmetrical, with musical parameter changes having a greater effect on imagined motion in one direction than the other; e.g. falling pitch mapped very strongly to descending motion, but the relationship of rising pitch to ascending motion was weaker. All results were largely invariant with respect to the level of subjects' musical training, although subjects with training tended to apply mapping strategies more consistently.

An earlier study by Eitan and Granot (2003) also found inter-parametric analogies were important for determinations of stimulus similarity in music. Subjects judged musical phrases with similar parametric contours as similar to one another, even if those contours were applied to different parameters in each musical phrase. For example, phrases with accelerating tempo were judged similar to phrases with increasing pitch. This demonstrates subjects are able to perceive parametric contours as distinct from musical phrases as gestalts.

Phillips-Silver and Trainer (2007) demonstrated the effect of bodily movement on rhythm perception, showing that, as they put it, "*how we move* will influence *what we hear*". They played rhythmically ambiguous musical phrases to infant subjects aged 7-months while gently bouncing the subjects in a rhythm

that implied either a march or a waltz. In a listening test, where the musical phrases were rhythmically disambiguated by the addition of accents in either duple or triple meter, the subjects listened significantly longer to the rhythm that matched their movement. While their results didn't depend on visual stimulation, in their introduction they draw an interesting analogy between rhythmically and visually ambiguous figures. They compare their march-waltz phrases, which can be heard either in duple or triple meter, to Rubin's (1915) face-vase image, which can be seen as either a vase or two faces looking at one another.

Saenz and Koch (2008) presented evidence of perceptual (i.e. automatic and impenetrable) mappings in visual-audio synesthetes, a class of subjects who involuntarily experience sound sensations alongside visual changes such as flashing lights and fast movements. Their experiment exploited a well-known cross-modal asymmetry: normal people are pretty good at identifying auditory rhythms and evaluating their similarity, and pretty bad at accomplishing the same with visual rhythms. To confirm reports of visual-audio synesthesia, Saenz and Koch played pairs of short rhythms to two groups of subjects: one normal group, and the other a group of synesthetes who claimed to hear visual flashes as auditory beeps. Half of the rhythm pairs were presented as audio, and the other half as visual flashes. Subjects were asked to evaluate whether the two rhythms in each pair were the same or different. Normal subjects performed well on the auditory task, but not the visual. The synesthetes, who claimed to hear visual flashes as auditory beeps, performed well in both domains. Interestingly, over the course of the experiment, the synesthetes reported the synesthetic sounds they heard along with the visuals changed to match the real sounds played during the auditory tests.

Synesthetic connections such as the above are not mere flukes or irregularities, unpredictable variations from a cross-modally segregated norm, but in fact are widespread and thought to undergird the perceptual processes of normal individuals. In their excellent review of recent research on synesthesia, Spector and Maurer (2009) suggest two possible developmental causes of synesthetic perception. According to the cross-activation theory, synesthesia is the result of incomplete pruning of synaptic connections between adjacent brain areas. The disinhibited feedback theory suggests synesthesia is caused by reentrant feedback from higher cortical areas failing to inhibit the effects of connections between primary sensory cortical areas. Spector and Maurer state that both of these theories predict synesthesia would be ubiquitous among normal individuals in early childhood and would persist to some extent in normal adults.

Groups of synesthetes (e.g. visual-audio, word-color, etc) tend to exhibit the same cross-modal mappings. Some of these mappings are based upon inter-parametric analogy. For example, synesthetic adults who perceive auditory pitch as visual color tend to map pitch height to brightness, with higher pitches resulting in brighter colors (Spector and Maurer, 2009). This cross-modal interaction is sufficiently strong that their ability to discriminate pitch is affected by the luminosity of the hearing environment. In pitch identification tests with a light used as a distractor, synesthetic subjects responded slower and less accurately if the distractor was opposed to the synesthetic percept of the pitch; i.e. shining a light at a synesthete with colored hearing interferes with their ability to hear correctly. (Marks, 1987, as cited by Spector and Maurer, 2009) The mapping of higher pitch to greater brightness also holds to some extent in normal adults and toddlers, suggesting that there may be a "natural

mapping" (what we would classify as an intuitive, perceptual mapping) between pitch height and brightness.

Sound and shape are also associated. Spector and Maurer (2009) note that "sharp visual shapes go with words that produce a small, constricted movement of the tongue and mouth (e.g., spike, point)." This is exemplified by an experiment where children and adults were asked to match a nonsense word (e.g. takete, kiki, maluma, bauba) to a 2-dimensional shape. Words like "takete" and "kiki" were reliably matched with jagged, spiky shapes, while words such as "maluma" and "bauba" were matched to more rounded, bulbous shapes (Köhler, 1929). Spector and Maurer (2009) elaborated on these experiments, testing for sound-shape mappings in toddlers using a wide variety of sounds and shapes, finding the association between non-rounded vowels and jagged shapes, and rounded vowels and rounded shapes was consistent. They also determined the effect occurred early enough in development to influence the learning of language. Further, the association of rounded sounds and rounded forms, and sharp sounds and angular forms seems to hold across cultures. The takete/maluma experiment was performed on 14 year-old children who spoke no English, but Swahili and the Bantu dialect of Kitongwe, with similar results to English speaking subjects (Davis, 1969). Further, there appear to be shape correspondences between real, non-nonsense words which hold across language barriers. Koriat and Levy (1979) show that Hebrew speaking adults could match Chinese characters with their corresponding Hebrew word with better-than-chance accuracy, and Berlin (1994) showed that English speakers were able to accurately sort Huambison words based on whether they referred to a bird or a fish. Ramachandran and Hubbard (2001) speculate that these phenomena "arise from connections between contiguous cortical areas mediating decoding of the visual percept of the nonsense shape (round or angular), the appearance of the

speaker's lips (open and round or wide and narrow), and the feeling of saying the same words oneself", invoking the idea of a "natural mapping." Spector and Maurer (2009) continue this train of thought, taking the cross-language results to suggest that this natural sound-shape mapping has significant influence on cross-cultural evolution of language.[2]

Spector and Maurer (2009), in a section entitled "A Common Code for Magnitude", note that a great many synesthetic effects can be explained in terms of cross-modal parametric isomorphism or parametric analogy. They suggest a natural predisposition to map magnitude cross-modally exists from birth, and posit a likely evolutionary explanation: it would leave more energy for the learning and working out of arbitrary mappings not related to magnitude, which tend to be "individually meaningful". Their example of an individually meaningful mapping is from the timbre of Mom's voice to her face.

2.5 Infants

Trehub (2001) shows infants are capable of and predisposed toward the recognition of melodic and rhythmic contours as distinct from melodic gestalts. Infants are able to recognize melodies after transposition, and rhythms after phrases have been sped up or slowed down, so long as the relative lengths of the notes remain the same. Trehub identifies melodic contour as the most salient musical feature for infant listeners, and references studies (Fernald, 1991;

---

[2] Spector and Maurer (2009) suggest that in the case of sound-shape mappings like those used in the kiki/bauba experiment, the mapping is the result of physical sympathy, where the sharpness in the language is analogous to tightness or sharpness in the mouth. Another parsimonious interpretation is that this relationship is not one of analogy, but perceptual isomorphism: the auditory spikiness may be a perceptually salient dimension isomorphic to visual spikiness. Where the word "takete" might map to an image with a small number of spikes, "takakekatakekteke" might map to an image with a great deal of spikes. This kind of spikiness may be related to the rate and amplitude contour of variation in the spectral centroid of the sound over time.

Fernald et al., 1987; Papoušek, 1992; and Lewis, 1951) which suggest melodic contour may also be the most salient feature of mothers' speech to prelinguistic infants. When mothers speak to infants, they slip into *motherese* or infant-directed speech, which is marked by increased pitch and dramatically exaggerated melodic contour. The melodic exaggeration of infant-directed speech occurs in all cultures (Trehub, 2000). Trehub (2001) notes that infant-directed singing is distinguished from typical singing styles by increased pitch (although not quite as high as infant-directed speech), slow tempo, and slurred articulation of words. The functions of infant-directed speech and singing seem to be to capture attention, moderate arousal, and develop the emotional bond between mother and infant.

The contour-affect relationship implicated in infant-directed speech is of a different kind than the mappings discussed above: there is no evidence that infants listening to infant-directed singing or motherese recognize them as representing anything. Because pre-linguistic infants are unable to explicitly relate a narrative of their experiences, it would be difficult to design an experiment which would provide satisfactory evidence that representation was at play. Nevertheless, that these modes of communication evoke affective states suggests the fundamental relationship between contour and affect exists prior to acculturation.[3]

Saffran et al. (1999 and 2008) examine the language learning process in infants. They note that "languages exemplify exactly those structures that humans are best able to learn", suggesting "at least some aspects of structure may emerge from constraints imposed by learning itself". In addition to being a possible bedrock on which more explicitly representational contour-emotion

[3] Treheb (2000) and Dissanayake (2000) suggest a number of evolutionary functions this relationship might serve.

mappings may develop, we conjecture that contour-affect associations may be beneficial to the language learning process, and therefore may in turn broadly affect the development of language on a global scale. This would go some lengths toward explaining some of the inter-language effects described in the previous section. This is also probably a two-way relationship: listening to ambient adult-directed speaking influences the contour of affective expressions on the part of the infant. Mampe et al. (2009) shows some preliminary evidence that the contour of newborns' crying melodies are shaped by the contour of their native language.

Representation of emotion as cross-modal dynamic contour recalls Stern's (1985) concept of *vitality affects* or *vitality contours* – basically, feelings represented by abstract 'forms' – which he suggests are important in early communication between mother and child. Stern's own work leaves vitality contours vaguely defined and empirically impenetrable (Køppe et al., 2008). The present work may to some degree assist in the explication of vitality contours as investigable phenomena.

2.6 Cross-cultural emotion

2.6.1 Moving past naïve universalism and relativism

What if this entire discourse is polluted by the cultural contingency of language and thought? It may be that the notion of happiness is represented entirely differently in one culture than another; indeed, all representation of emotion in music may be partially or wholly destabilized when transported across cultural borders. If this is so, how can it make sense to refer to emotion in music without some geographic or cultural qualification, e.g. Western music, Hindustani music, etc.? Evidence from studies of synesthesia suggests there are certain linguistic-

formal associations which hold across cultures – perhaps there are musical-emotional associations which hold as well. In this section, I discuss problems with what I refer to as the naïve universalist and naïve relativist positions on cross-cultural emotion in music, present evidence from the literature that there are cross-culturally valid musical-emotional associations, and discuss how these associations neither exclude genuine difference between cultures, nor imply any universal musical systems (i.e. no "universal language" of music).

Music, as understood by what I will call the naïve universalist position, may be superficially different in one culture or another, but has some common core which all people experience the same way. The explanation for this shared experience is the shared structure of the human body: everybody's body works more or less the same way; we all have ears and a brain. Universal musical experiences are the result of an interaction between the physical properties of a sound and the structures of our sense organs. Authors taking this approach tend to conflate feelings with perceptual experiences with physical phenomena. If we experience music as having a feeling (e.g. sadness), which seems to correlate with a percept (e.g. dissonance), and we have got universal sameness on our mind, then the natural thing to do is treat either the percept, the feeling, or both as physical properties of the sound.

Plato's treatment of music in The Republic is a paradigmatic example of this approach (Plato, 1992). Plato associates each musical mode with an emotional state, and bans from the Republic those modes which evoke emotions undesirable from the perspective of government. Implied in this approach are two assumptions: 1) that certain collections of pitches have certain evocative effects, and 2) that some of these effects are good for culture, and some bad. Plato is undertaking an engineering project where music (or control over musical dissonance) is a means and culture is the end. His assumption that music's

emotional effects are independent of acculturation implies the naïve universalist position: emotional responses to music are determined by sonic content, not cultural context. Significantly, this same assumption implies a theory of music perception in which dissonance, and thereby emotional feeling, is thought of as a physical property of sound and not a subjective percept.

Many later theorists have tried to make Plato's implication explicit. Leibniz famously quipped that “music is the pleasure the human mind experiences from counting without being aware that it is counting”, and suggested in (Leibniz, 1714) that the perception of dissonance was related to the subconscious calculation of frequency ratios. While the exact physical correlate of dissonance varies from author to author, basically similar positions are held by Euler (1739), Stumpf (1890), Helmholtz (1912), etc. A brief summary of these authors' theories of dissonance perception can be found in (Lundin, 1947). All of these theories of sound perception are related to cultural universalism via their treatment of dissonance. If a cue-based model of emotion perception is assumed, where cues like dissonance are considered physical properties and not subjective percepts, then it seems reasonable to expect musical emotion to be basically invariant across cultures.

The naïve universalist approach is attacked by Lundin (1947) on the basis that its account of dissonance is incoherent, and by ethnomusicologists such as Meriam (1964) and Blacking (1965) on the basis that it fails to account for diversity and difference in the music of non-Western cultures. Taking dissonance as its focus, Lundin's strategy is to drive a wedge between percepts and physical phenomena, suggesting our perceptual experiences are culturally contingent: the way we hear is affected by the culture in which we live. He challenges the physical correlates of dissonance suggested by Euler, Helmholtz, and Stumpf: dissonance can't simply be the subconscious calculation of frequency ratios,

where larger frequency ratios mean greater dissonance, for example, because some out-of-tune intervals, e.g. 99:201 Hz, are still perceived as consonant. There must be some process, probably conditioned by experience, which allows us to perceive 99:201 Hz as close enough for rock 'n' roll. According to Lundin, then, dissonance is not a physical phenomenon in itself, but a percept resulting from a "discriminative reaction" or subconscious judgment made on the basis of cultural experience.

Lundin's wedge between percepts and physical properties is not unlike the division Scarantino establishes between Folk Emotion and Explicated Emotion. According to Lundin, dissonance (like emotion) is whatever people say it is, and people in different cultural contexts may well say different things. The contrasting approaches of Leibniz et al. are attempts to explicate a theoretical construct called "dissonance" which could be scientifically useful, but may not line up with ordinary language use. Again assuming a cue-based model of emotion perception, but where cues such as dissonance are culturally conditioned judgments, one would expect to see dramatic variations in how different cultures express emotion in music.

Naïve relativism so expressed disallows any crosstalk between folk and explicated understandings of musical features, and any interaction between the physical or biological and the cultural. This approach is exemplified by Meriam (1964) and much of Blacking's early work on the music of the Venda, a people living in the Transvaal in northern South Africa (e.g. Blacking, 1965). For these authors, culture is the supreme and perhaps single factor constitutive of musical form: "Every piece of music has its own inherent logic, as the creation of an individual reared in a particular cultural background, and in terms of this there is ultimately only one explanation of its structure and meaning." (Blacking, 1973 as quoted by Agawu, 1997) For insights into the structure of music, Blacking

looks toward relationships with dance, language, and social organization rather than any sort of biological predisposition. The relativist approach is skeptical of any claim that music from one cultural context is similar to music from another. Superficial similarities may be entirely coincidental. While, to Western ears, it may seem that some culturally and geographically separate groups utilize the same musical material, the cultural contexts and organizational principles at play may be entirely different. Only deep anthropological study of culture can shed light on how music is heard.

While offering obvious benefits (not the least of which is avoiding the traps of naïve universalism suggested above), the drawbacks of this position are substantial. If musical cultures are fundamentally incommensurable with one another, what are we to make of certain striking similarities? Without making cross-cultural comparisons, how are we to undertake analysis of musical cultures which offer no internal analytic vocabulary? As Lundin (1947) pokes holes in the universalist position by undermining too-tightly explicated definitions of dissonance, Kolinski (1967) challenges the naïve relativist view to account for some striking empirical observations. Taking the universality of the human vocal apparatus as a starting point, he asks "1) what causes the singer to select certain tones out of this pitch continuum and to organize them into coherent structures; and 2) why similar patterns of tonal construction can be found in widely separated areas and in strongly contrasting cultures". He goes on to suggest octave equivalence (that is, the recognition of pitches with frequencies related by an approximately 2:1 ratio as being members of a pitch class) as a musical universal, as well as the presence of fifths and other small frequency ratios, and categorical discrimination of pitches. Evidence of the universality of these features is offered in Kolinski (1967), Trehub (2000), McDermott & Hauser (2005), and Nettl (1956, 1983).

Kolinski's (1967) approach to this evidence suggests a softening of both the relativist and universalist positions such that both nature and culture are allowed influence. This view is afforded by the identification of very simple perceptual universals, such as octave equivalence, as phenomena on which complex notions such as dissonance or musical scales must supervene. After being faced with evidence of universality, Blacking (1995) offered the following instructive approach: "I suggest that an accurate and comprehensive description of a composer's cognitive system will provide the most fundamental and powerful explanation of the patterns that the music takes. 'Cognitive system' includes, of course, all cerebral activity involved in motor coordination, feelings, and cultural experiences, as well as the composer's social, intellectual, and musical activities. Even if we regard them solely as 'sonic objects,' the notes of a piece of music are the products of cognitive processes." The solution is not to clarify a transcendental boundary between biology and culture, but to acknowledge that they form a coupled system, with each having an influence on the other.

We follow this moderated relativist approach in our reliance on folk terminology and our grounding of terms such as "dissonance" in perceptual studies instead of theory. As we understand them, culturally conditioned ideas such as dissonance do not refer to a single physical phenomenon, but a package of loosely linked properties such as frequency ratio characteristics, loudness, context of presentation, timbre, and so on. Learning what "dissonance" means is associating this set of properties with their proper name as situated within a cultural context. Some of these properties are themselves thorny, densely packed cultural terms – "timbre", for example, has no obvious physical correlate and myriad uses. Other properties may be basic to the human perceptual apparatus. For example, categorical pitch perception, while not being a sufficient condition

for a cultural understanding of dissonance, is probably universal, and ideas like dissonance probably depend upon it. Once the idea is learned, the collection of properties is perceived *as* the identifying term, e.g. a tone at a high volume, with a harsh timbre, a high-numbered frequency ratio, and in a certain cultural context is perceived *as* a dissonance. At the same time, activation of the idea may also trigger reflection upon the interaction between the cultural context and the set of distinguishable properties available to the sensory apparatus, which may in turn inform or update what the idea means. This feedback loop between reflection and perception allows for the presence of cross-cultural universals, but packed into culturally relative terms in different combinations and degrees. It also allows for intercultural difference, as well as the slippage of meaning over time. This approach has additional implications for our plans to test our research cross-culturally, outlined in section 3.7.

2.6.2 Evidence for cross-cultural music-emotion mappings

Ekman's classic studies (Ekman et al., 1969; 1971) demonstrated consistent emotion recognition in facial expressions in numerous literate and preliterate cultures, some of which had minimal contact with Westerners prior to the experiment. This finding was the first positive evidence that emotions are construed and expressed similarly across cultural boundaries. Scherer (2003) extended this line of study beyond facial expression, finding that vocal expressions of emotion are also recognized with better than chance accuracy across cultures. This and other findings from Scherer et al. (2001) are interpreted by the authors "as evidence for the existence of universal inference rules from vocal characteristics to specific emotions across cultures". (Scherer, 2003) Sauter et al. (2010) added to these findings, showing that nonverbal emotional vocalizations were bidirectionally recognizable between Western participants in

their study and culturally isolated Namibian villagers. The similarities and correspondences between musical and linguistic expressions of emotion addressed above suggest a concerted study of cross-cultural musical expression. Some preliminary steps are summarized below.

Balkwill and Thompson (1999) played selections of Kyrgyz, Hindustani, and Navajo music to Western subjects in order to compare their assessment of the music's emotional content with the music's cultural-emotional association. For an initial pilot experiment, they used a model of emotion limited to separate ratings of joy and sadness. They found the Western listeners assigned higher joy scores to music considered joyful in all three traditions, and higher sad scores to music traditionally considered sad. They also found that the joy rankings were positively correlated with the tempo of the music, while the sadness rankings were negatively correlated. These results were followed by a more in-depth study of emotion recognition by Westerners in Hindustani music, where the emotional palette was expanded to include anger and peacefulness, and the tempo, melodic complexity, rhythmic complexity, pitch range, and timbre of the music were analyzed to determine how each of these musical parameters contributed to emotion recognition. They found that Western listeners correctly rated the emotions of the ragas in every case except for peace, which was confused with sadness. They found joy was correlated with tempo and melodic complexity, sadness was correlated with melodic complexity, but negatively correlated with tempo, anger was correlated with sharp timbre, and peace was negatively correlated with rhythmic complexity.

Fritz et al. (2009) present the most compelling evidence for cross-cultural validity music-emotion mappings. Their subject population consisted of twenty-one members of the Mafa ethnic group in Northern Cameroon who, prior to the study, had never been exposed to Western music. Subjects were played musical

examples meant to convey happiness, sadness, and fear, ranking each example by placing a slider on a continuum between a cartoon happy face and a cartoon grimace. The Mafa subjects were able to correctly assess the emotional content of each musical example, although compared to a German control group, the Mafa responses were less extreme. After the recordings were digitally altered such that formerly consonant harmonies became dissonant, the ratings of the Mafa group became considerably more negative. While this study is emotionally narrow and limited to only two cultures, the results are compelling enough to lend weight to the notion that dynamic emotional signs are understood similarly regardless of acculturation.

If musical signs for emotion are understood cross culturally, then those signs cannot be pointing toward concepts which are entirely culturally contingent. There must be some universally occurring thing to which emotional musical signs refer. Ekman (1999) suggests emotions are "distinctive universal signals" for "inform[ing] conspecifics, without choice or consideration, about what is occurring: inside the person (plans, memories, physiological changes), what most likely occurred before to bring about that expression (antecedents), and what is most likely to occur next (immediate consequences, regulatory attempts, coping)". All kinds of actions are packed into emotions: those which lead up to the emotional experience, those which are a part of or coincide with its occurrence, and those to which it is an antecedent. Following this observation, it seems likely that cross-cultural emotions are accompanied by predictable patterns of behavior. Examples include fighting or yelling when angry, receding or crying when sad, or moving slowly and becoming still when peaceful. We would like to suggest that emotional signs in music bear an iconic relationship to these and similar activities, including both physical actions and modes of speaking affected by emotional state. There is some inconclusive

evidence that emotions and actions may be associated in a way which would support this semiotic relationship. In addition to the studies of movement and gesture summarized above, Ekman (1999) summarizes studies which indicate emotions may be reliably accompanied by certain physiological changes which could predispose subjects to certain activities. None of the studies summarized offer evidence as to whether these predispositions are innate or the result of acculturation, so this is still an open question. However, if evidence of cross-cultural validity of emotional signs in music is shown to be conclusive, that would strongly suggest that emotional predispositions to behavioral action are cross-cultural as well.

2.7 Where do we go from here?

Evidence from the literature shows consistent mappings between music, motion, and emotion which appear to be determined by cross-modal parametric isomorphisms and inter-parametric analogies. However, this evidence for cross-modal mappings is mostly implicit: most of the studies surveyed focus on examples of emotion-signifying stimuli in a single modality which are either created prior to the experiment or by actors. Subjects typically assess the emotionality of the stimulus, and then the experimenters parameterize and analyze those stimuli with respect to subjects' judgments. The primary problem with this approach is the segregation of different modalities. Subjects only assess stimuli in a single modality at a time, and researchers typically analyze each modality separately, so relationships between modalities are rarely described in detail.[4] In the case of music, this problem is compounded by a willingness to take music-theoretical terminology (especially “major” and “minor”) as basic aspects

[4] Eitan and Granot (2003) and (2006) are notable exceptions.

of musical experience, obscuring the possibility that lower-level parameters could be at play. Finally, the use of human actors for producing stimuli imposes severe limits on the number of stimuli generated and the extent to which those stimuli reflect the full breadth of expressive possibility.

3 A behavioral experiment

Our experiment avoids the methodological issues described above by inverting the standard create stimuli, then get subject assessments, then analyze process. We developed a novel experimental paradigm where subjects are presented with a computer program which allows them to manipulate slider bars corresponding to parameters in a statistical model generating dynamic contours. The output of this model is fed to computer programs which simultaneously create stimuli in two different modalities – music and motion – in real time, with similar dynamic contours. Two groups of subjects, one for each modality, use the model as an authoring tool, creating stimuli which they think best express a set of emotions. Afterward, the results from the two groups are compared. If the same statistical properties of dynamic contour are similarly implicated in emotion recognition in both music and motion, then effect of class (emotion) on slider position should be than the effect of modality (music, motion). In addition to avoiding the shortcomings described above by providing precisely (in fact, programmatically) explicated definitions of terminology, this approach also results in the production of a generative model for creating numerous statistically and emotionally similar stimuli.

3.1 A notable methodological precedent

In addition to the literature described above, there is one study with which the present work shares a certain kinship. Its nearest methodological neighbor is Scherer and Oshinsky (1977), which is the first study of musical emotion to use stimuli generated based on regular sampling of a musical parametric space. Scherer and Oshinsky used a Moog synthesizer to create eight-tone melodies based on the division of musical space into various parameters, each of which was further divided into levels of intensity. Because the melodies were produced manually, they were limited in terms of the resolution of their divisions, which in turn limited the number of stimuli produced (164 in total, which were narrowed down to a smaller group for testing). While this is quite a large selection relative to other contemporaneous studies, the present work expands this further by automating the melody generation process and handing control of the parameter settings to the subjects, enabling fine-grained exploration of a very large parametric space.

3.2 Parameterization and emotion choices

Based on a review of the literature outlined above, we selected five emotions on which to focus our research. We chose emotions likely to be recognizable in both music and simple movement. These emotions were represented in our study by five-word clusters, following Hevner (1935). Each cluster is topped by a single word we decided was the clearest and simplest expression of the emotion-group. These top words are: “happy”, “sad”, “angry”, “scared”, and “peaceful”.

We then selected a group of parameters implying a model we thought could represent simple music and biological motion. This model was the basis of the stimulus-generating computer program described in the following section. By

"simple" music and biological motion, we mean to indicate our goal was not realism, but mere recognizability; the musical output of our program is not going to sit alongside Mozart in the canon, but should be recognizable as "happy music", "sad music", etc. Likewise, the program does not generate realistic human movement, but instead bounces a ball around, the motion of which should be recognizable as "happy", "sad", and so on.[5] A combination of evidence from the literature and our intuition suggested the parameters of tempo or inter-onset interval (measured in beats per minute), jitter (standard deviation from the mean tempo), musical consonance/visual spikiness, tendency to make big movements vs. small movements, and tendency to move upward or downward. Each of these parameters are isomorphic in both music and motion, with the exception of consonance/spikiness. For consonance/spikiness we implemented a mapping from a simple model of musical consonance to the visual spikiness of the moving figure.

3.3 The model

The program for the behavioral experiment was written in Max/MSP (Puckette, 1991; Zicarelli, 1998), JavaScript, and Processing (Reas & Fry, 2006).

Subjects were presented an interface with slider-bars corresponding to the five dimensions of our statistical model: tempo, jitter, or scale choice (also referred to as dissonance or consonance), step size, and step direction. We selected these parameters based on intuition and experience, augmented by a review of the literature, identifying each parameter as either crucial for emotional expression or as low-level ground upon which higher-level ideas might

[5] That is to say, if this model is an explication of music, motion, or emotional dynamics, it a humble, pragmatic one. Its aim is to balance recognizability with ease of use, not to be the most accurate model in town.

depend (dissonance, for example, depends upon scale choice). The five sliders controlled parametric values fed to an algorithm which probabilistically moved the position of a marker around a discrete number-line in real time. We will refer to the movements of this marker as a path. The position of the marker at each step in the generated path was mapped to either music or animated movement.

The number-line traversal algorithm can be split into two parts. The first part, called the metronome, controlled the timing of trigger messages sent to the second part, called the path generator, which kept track of and controlled movement on the number line. The tempo and jitter parameters were fed to the metronome, and the scale choice (also referred to as consonance), step size (also referred to as bigsmall), and step direction (also referred to as updown) parameters were fed to the path generator. When the subject pressed the space bar on the computer keyboard, the metronome turned on, sent sixteen trigger messages to the path generator (variably timed as described below), and then turned off. The beginnings and endings of paths correspond to the ons and offs of the metronome.

Tempo was measured in beats-per-minute (bpm), and constrained to a minimum of 30bpm and a maximum of 400bpm. Jitter was expressed as a coefficient of the tempo with a range between 0 and 0.99. When jitter was set to 0, the metronome would send out a stream of events at evenly spaced intervals as specified by the tempo slider. If the jitter slider were above zero, then specific per-event delay values were calculated nondeterministically as follows. Immediately prior to each event, a uniformly random value was chosen between 0 and the current value of the jitter slider. That value was multiplied by the period in milliseconds as specified by the tempo slider, and then the next event was delayed by a number of milliseconds equal to the result. The effect was that

as the value of the jitter slider increased, the timing of event onsets became less predictable while the mean event density remained the same.

The path generator can be conceived of as a "black box" with a memory slot which could store one number and which responded to a small set of messages: reset; select next number; and output next number. Whenever the path generator was sent the reset message, a new starting position was picked and stored in the memory slot (the exact value of the starting position was constrained by the value of the scale choice slider as explained below). Whenever the path generator was sent the select next number message, it picked a new number according to the constraints specified by the slider bars – first, the size of the interval was selected, then the direction (up or down), then a specific number according to the position of the scale choice slider. The output next number message caused the path generator to output the next number to the music and motion generators, described in section 3.4.

When selecting a new number, the path generator first chose a step size, or the distance between the previous number (stored in the memory slot) and the next. This value was calculated nondeterministically based on the position of the step size slider. The step size slider had a minimum value of 0 and a maximum value of 1. When choosing a step size, a uniformly random number between 0 and 1 was generated. This number was then used as the x value in the following equation, where a = the value of the step size slider:

$$r = \begin{cases} \frac{1-a}{-a} \cdot x + 1 & x \leq a \\ \frac{-a}{1-a} \cdot (x-1) & x > a \end{cases}$$

The result r was multiplied by 4 and then rounded up to the nearest integer to give the step size of the event. As the value of the step size slider increased, the

likelihood of a small step size decreased, and vice versa. If the slider was in the minimum position, all the steps would be as small as possible. If it was in the maximum position, all the steps would be as large as possible. If it was in the middle position, there would be an equal likelihood of all possible step sizes. Other positions skew the distribution one way or the the other, where higher values resulted in a larger average step size. Note that these step size units did not correspond directly to the units of the number line; they were flexibly mapped to the number line as directed by the user's scale selection, described below.

After the step size was chosen, the path generator determined the direction of the next step: up or down. As with step size, the step direction was calculated nondeterministically based on the position of the step direction slider. The step direction slider had a minimum value of 0 and a maximum value of 1. When choosing step direction, a uniformly random number between 0 and 1 was generated. If that number was less than or equal to the value of the step direction slider, then the next step would be downward; otherwise the next step would be upward.

Finally, the number was mapped on to one of 38 unique scales. As the notion of a scale is drawn from Western music theory, this decision requires some elaboration. In Western music theory, a collection of pitches played simultaneously or in sequence may be heard as consonant or dissonant. The perception of a given musical note as consonant or dissonant is not a function of its absolute pitch value, but of the collection of intervals between all pitches comprising the current chord or phrase. The relationship between interval size and dissonance is non-linear. For example, an interval of 7 half steps, or a perfect fifth, is considered quite consonant, whereas an interval of 6 half steps, or a tritone, is considered quite dissonant. Intervallic distance, consonance/

dissonance, and equivalency are closely related. If a collection of pitch classes X (a pitch class set, or PC set) has the same set of intervallic relationships as another PC set Y, those two PC sets will have the same degree of consonance and are transpositionally identical (and in certain conditions equivalent).

Absolute pitches also posses this property of transpositional equivalency. When the frequency of a note is doubled, it is perceived as belonging to the same pitch class. For example, the A key closest to the middle of a piano has a fundamental frequency of 440Hz, while the A an octave higher has a fundamental frequency of 880Hz; both are heard as an A. Western music divides the octave into twelve pitch classes, called the chromatic scale, from which all other scales are derived. Because we wanted to investigate musical dissonance and possible functional analogs in the modality of motion, our number-line scales were designed to be analogous to musical scales, where a number-line scale is a 5-member subset of the *chromatic* set [0,1,2,3,4,5,6,7,8,9,10,11]. There are 768 subsets of the chromatic set, many of which are (in the domain of music) transpositionally or inversionally equivalent. Our scale list was created by generating the prime forms (Forte, 1973) of these 768 subsets, and then removing duplicates, yielding 38 unique scales.[6] These scales were ordered by their aggregate dyadic consonance as defined by Huron (1994).

The choice of a definition of consonance determined exclusively by pitch-class relationships may seem at odds with our reflection-perception feedback model of conceptual understanding outlined in section 2.6. However, while motivated by Western music theory, Huron's (1994) aggregate dyadic consonance is a perceptual measure. It is derived from the results of three separate studies (Malmberg, 1918; Kameoka and Kuriyagawa, 1969; Hutchinson

[6] There are 38 prime 5-member PC sets, but only 35 unique interval vectors, so two of the entries on our slider bar were redundant.

and Knopoff, 1979; all as cited by Krumhansl, 1990) in which subjects were surveyed as to the relative dissonance of various combinations of notes played on a piano. We think that this, a measure based solely on these subjective judgments, in a context which closely matches that of our experiment, is the best we can do to balance the need for explication imposed by computer modeling with our need to rely on a folk notion of dissonance. This approach is not without its limits: Huron's metric is only applicable to listeners acculturated to Western music, and does not take into account the effects of melody or pitch order, loudness, pitch register, or any musical parameters other than interval class. While this limits to some extent the generalizability of our results, and the applicability of the experiment to other cultural contexts, we believe it is sufficient for the present work.

The algorithm for generating a specific path across the number line was as follows. The number line consisted of the integers from 0 to 127 inclusive. When the algorithm began, three variables were stored. First, a starting-point offset between 0 and 11 was selected uniformly at random, then an octave bias variable was set to 5, and a scale position variable was set to 0. The current scale class was determined by using the scale position variable as an index to the array of scale elements specified by the position of the scale slider. For example, if the current selected scale was [0, 3, 4, 7, 10] and the current scale position variable was 2, then the current scale class would be 4 (indices start from 0). The current position on the number line was given by multiplying the octave bias by 12, adding the starting-point offset, and then adding the current scale class value. For example, if the octave bias was 5, the starting-point offset was 4, and the scale class value was 7, then the current position on the number line would be 71.

When the select next number message was received, an interval and note direction were selected as described above. If the note direction was upward, then the new scale position value was given by the following:

(current scale position + new interval value) % 5

If the note direction was downward, then the new scale position value was given by:

5 + (current scale position - new interval value)

Either of these conditions may imply a modular "wrapping around" the set of possible values (0 to 4). If this is the case, then the current octave variable is either incremented by 1 in the case of an upward interval, or decremented by 1 in the case of a downward interval. If a step in the path would move the position on the number line outside of the allowed range, 12 would be either added to or subtracted from the new position. This to some extent distorted the contour of paths with very large step sizes which had an extreme tendency toward either upward or downward movement.

3.4 The mappings

The subjects were divided into two groups. For the first group, number-line values were mapped to musical notes, and for the second group, number-line values were mapped to animated movement.

3.4.1 Music

Our mapping from movement across a number-line to Western music was straightforward, as its most significant modality-specific features were taken care of by the very design of the number-line algorithm. The division of pitches into pitch-classes and scales is accounted for by the scale-class and scale selection system used by the algorithm, as is the modulo 12 equivalency of pitch-classes. Each number was mapped to a specific pitch which was sounded as the algorithm selects the number. The number 60 was mapped to middle-C, or C4. Movement of a distance of 1 on the number line corresponded to a pitch change of a half-step, with higher numbers being higher in pitch. For example, 40 maps to E2, 0 maps to A0, and 127 maps to G9. This matches the mapping described by the MIDI Manufacturers Association (1996). Notes were triggered via MIDI and played on the grand piano instrument included with Apple GarageBand.

A piano timbre was picked because of the instrument's ubiquity in Western music and relative emotional neutrality. Unlike the violin, guitar, clarinet, etc., the piano appears in almost every genre of Western music, and is routinely used to express the full spectrum of musical emotions. The violin or cello, for example, could for some listeners carry a connotation of sadness. Further, the piano does not necessarily carry any extra-musical connotations – unlike, for example, a pure sine tone, which is often used to signify the future or advanced technology. This is not to suggest the piano provides a truly neutral timbre, or that it cannot be used to point in an extra-musical direction, but simply to say that no emotional information or extra-musical context may be reliably inferred from its use.

3.4.2 Motion

Mapping from movement across a number-line to animated movement was less straightforward. Our animated character was a red ellipsoid egg with cubic "eyes". It sat atop a rectangular dark grey "floor" on a light grey background. An ellipsoid was chosen because it can be seen as rotating around a center. The addition of eyes was intended to engage cognitive processes related to the perception of biological motion. We wanted our subjects to perceive the egg as having its own subjectivity; that it could be capable of communicating or experiencing happiness, sadness, etc. The movement of our character (henceforth referred to as "the Egg") was limited to bouncing up and down, rotating forward and backward, and modulating the spikiness of its surface. Technical details follow.

The Egg was rendered using OpenGL (Rost, 2004) and programmed using Processing (Reas & Fry, 2006). The Egg was drawn as a red 3-dimensional sphere composed of a limited number of triangular faces which were transformed into an ellipsoid by scaling its y-axis by a factor of 1.3. The Egg was positioned such that it appeared to be resting on a rectangular floor beneath it. Its base appeared to flatten where it made contact with the floor. The total visible height of the Egg when it is above the floor was 176 pixels; this was reduced to 168 pixels when the Egg was making contact with the floor. Its eyes were small white cubes located about 23% downward from the top of the ellipsoid. The Egg and the floor are rotated about the y-axis such that it appeared the Egg was looking somewhere to the left of the viewer.

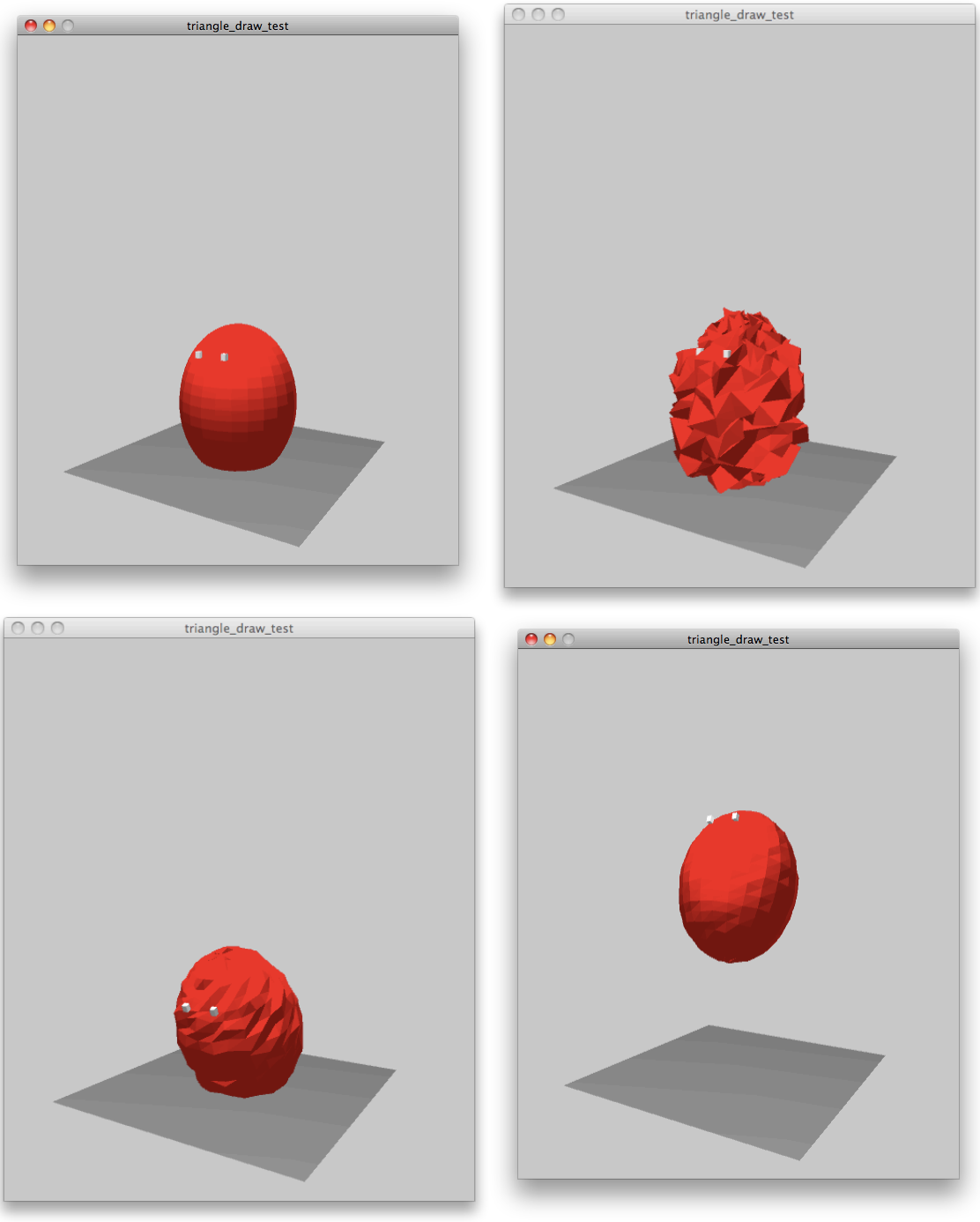

Figure 1. *Aspects of the Egg*

Every time the current position on the number line changed, the Egg bounces. A bounce is the translation of the Egg to a position somewhere above its resting position and back down again. Bounce duration was equal to 93% of the current period of the metronome. The 7% reduction was intended to create a perceptible "landing" between each bounce. Bounce height was determined by the difference between the current position on the number line and the previous position. A difference of 1 resulted in a bounce height of 20 pixels. Each additional addition of 1 to the difference increased the bounce height by 13.33 pixels, e.g. a difference of 5 would result in a bounce height of 73.33 pixels. The Egg reached its translational apex when the bounce was 50% complete. The arc of the bounce followed the first half of a sine curve, i.e. at any point during the bounce, the current vertical translation of the Egg relative to its original position was given by the formula:

$$sin(\pi \cdot p) \cdot h$$

Where p is a decimal value between 0 and 1 representing the percentage of the bounce completed and h is the total height of the bounce.

The Egg would rotate, leaning forward or backward, depending on the current number line value. High values caused the Egg to lean backward, such that it appeared to look upward, and low values caused the Egg to lean forward or look down. When the current value of the number line was 60, the Egg's angle of rotation was 0 degrees. An increase of 1 on the number line decreased the Egg's angle of rotation by 1 degree; conversely, a decrease of 1 on the number line increased the Egg's angle of rotation by 1 degree. For example, if the current number-line value were 20, the Egg's angle of rotation would be 40 degrees. If

the current number-line value were 90, the Egg's angle of rotation would be -30 degrees.

The Egg could also be more or less “spiky”. The amplitude of the spikes, or perturbations of the Egg's surface, were analogically mapped to musical dissonance. The visual effect was achieved by adding noise to the x, y, and z coordinates of each vertex in the set of triangles comprising the Egg. Whenever a new position on the number-line was chosen, the aggregate dyadic consonance (Huron, 1994) of the interval formed by the new position and the previous position was calculated. The maximum aggregate dyadic consonance was 0.8, the minimum was -1.428. The results were scaled such that when the consonance value was 0.8, the spikiness value was 0, and when the consonance value was -1.428, the spikiness value was 0.2. Changes in consonance of 0.01 resulted in a change of 0.008977 to the spikiness value. For each vertex on the Egg's surface, spikiness offsets for each of the three axes were calculated. Each spikiness offset was a number chosen uniformly at random between -1 and 1, which was then multiplied by the Egg's original spherical radius times the current spikiness value.

3.5 Experimental Method

Subjects were divided into two groups, the Motion group and the Music group. The same program was used for both groups of subjects, except that the Motion group only saw the motion output, whereas the Music group only heard the music. The program was explained to the subjects as follows: whenever the space bar was pressed, a musical phrase would begin to play or the ball would begin to bounce. While the music was playing or ball was bouncing, a visual indicator would appear on the screen, and additional presses of the space bar would have no effect. Moving the slider bars immediately caused the music or the way the

ball bounced to change. Subjects were given an opportunity to play with the slider bars in an open-ended way for as long as they liked. When they were ready, they were instructed to press a button on the screen which displayed the "emotional targets" (explained below) and began the experimental task.

Five emotional targets were displayed on the screen. Each target consisted of a group of five emotional words, a save button, and a load button. The targets appear in figure 2, a screenshot of the user interface.

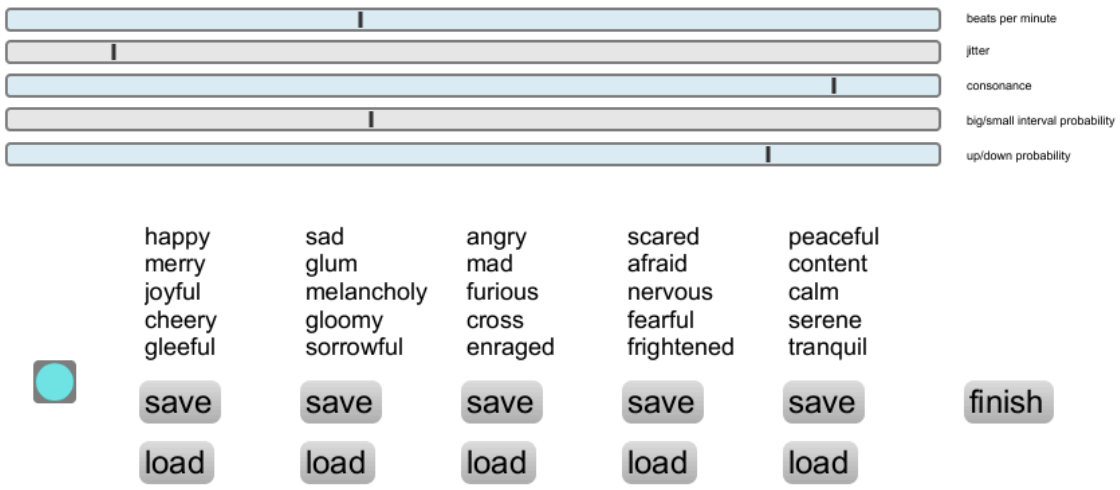

Figure 2. *Screenshot of the user interface*

Pressing the save button for a group stored the slider bar positions in memory. Pressing the load button for a group restored the slider bars to the position saved for that group. Subjects were instructed to save slider bar settings corresponding to each emotional group. No group order was mandated, that is, subjects were free to work on hitting each emotional target in whatever order they chose, and they were instructed to load and revise settings as changes occurred to them. Each saved slider bar setting was meant to make sense with respect to the other saved settings, i.e. the subject was told that their "happy" settings should make

sense relative to their "sad" settings, and so forth. The difference between recognized and evoked emotion was emphasized: subjects were told that their task was to make the emotions recognizable to an observer but that they should not worry about trying to make the clips emotionally evocative. There was no time limit imposed on the experiment; when subjects saved slider settings for all five emotional targets and were satisfied with the results, they pressed a button which finished the experiment.

3.6 Results

3.6.1 Multi-way ANOVA/GLM

Unless otherwise noted, Mauchly's Test of Sphericity was significant for all within subject effects. To compensate, all results listed below have Greenhouse-Geisser correction applied. Emotion had the largest effect on slider position ($F(2.97, 142.44) = 185.56$, $p < 0.001$). The partial Eta2 was .79, which means that Emotion by itself accounted for 79% of the overall (effect+error) variance. This main effect of emotion was qualified by an Emotion x Feature interaction indicating that different emotions required different configurations of dynamic features ($F(4.81, 230.73) = 112.90$, $p < 0.001$; partial Eta2 = .70). Importantly, while there was a significant main effect of Modality ($F(1,48) = 4.66$, $p < .05$) this effect was small (partial Eta2 = .09) and did not interact with Emotion (Emotion x Modality: $F(2.97, 142.44) = .97$, $p > .4$; partial Eta2 = .02). The three way interaction between Feature, Emotion, and Modality was significant, albeit modest ($F(4.81, 230.73) = 4.50$, $p < 0.001$; partial Eta2 = .09).

The three-way interaction can be read as a measure of variance per parameter explained by the combination of emotion and modality. That is, a rough measure of the extent to which the statistical codes for emotion in our

model differ between music and motion. This measure combines those differences which are a function of the human perceptual system with differences caused by limitations and inaccuracies in our model. Its modest size suggests, unsurprisingly, that the domains of music and motion are to an extent fundamentally different, but also that they are sufficiently similar, and our models sufficiently accurate, for the purposes of our experiment.

3.6.2 Analyses by emotion class

The following sections describe the data for each emotion in detail. Means with standard deviations are provided for each slider bar and task combination, inter-parametric correlations with magnitudes > 0.3 are discussed, and linear discriminant analysis (LDA) is used to assess which parameters best distinguish data points in each emotion class from all out-of-class data points. To describe the results of LDA, we provide the proportion of the linear combination of predictor variables which describe the rotation of the discriminant for each parameter. This is a relatively abstract measure; suffice it to say that high values indicate the given parameter is important for discriminating the current emotion from the others. Parameters represented in the model as values between 0 and 1 are scaled to between 0 and 100. The possible ranges of each parameter are as follows: BPM, 30-400; jitter, 0-99; consonance, 0-37; bigsmall, 0-100; updown, 0-100.

3.6.2.1 Angry

	bpm	jitter	consonance	bigsmall	updown
Mean all	331.00	53.70	8.00	67.92	76.94
SD all	73.02	34.53	11.44	33.24	18.62
Mean Music	340.04	42.72	11.84	65.44	75.36
SD Music	81.10	35.83	12.83	29.07	20.24
Mean Motion	321.96	64.68	4.16	70.40	78.52
SD Motion	64.33	29.99	8.46	37.39	17.12

Table 2: *Angry: means and standard deviations*

Music and motion	Consonance, updown	r = -0.39, p<0.005, 95% CI -0.6 to -0.13
	BPM, updown	r = -0.32, p<0.025, 95% CI -0.55 to -0.04
Music	Consonance, updown	r = -0.39, p<0.05, 95% CI -0.68 to 0
	BPM, updown	r = -0.33, p<0.1, 95% CI -0.64 to 0.07
	Jitter, bigsmall	r = -0.43, p<0.03, 95% CI -0.7 to -0.04
Motion	Consonance, updown	r = -0.39, p<0.052, 95% CI -0.68 to 0
	Consonance, bigsmall	r = -0.45, p<0.023, 95% CI -0.72 to -0.07

Table 3: *Angry: correlations with magnitude greater than 0.3*

	music and motion	music	motion
bpm	0.022	0.036	0.009
jitter	0.006	0.020	0.002
consonance	0.577	0.265	0.779
bigsmall	0.040	0.045	0.031
updown	0.355	0.634	0.179

Table 4: *Angry: LDA results*

Music and motion settings for angry were very similar. Anger is quick, jittery, dissonant, takes large steps, and tends to move rather steeply downward. Although BPM values for motion had a lower standard deviation than music, a majority of music subjects pushed the BPM slider to its highest possible value. The minimum BPM value for angry music was 136, whereas the minimum for angry music was 269. Angry motion also tended to be more jittery, with a mean value of 61.13, vs. 33.6 for music.

In both angry music and angry motion, consonance and updown were negatively correlated, meaning that as subjects steepened the downward trajectory of the path, they also decreased the consonance. In music alone, BPM and updown as well as jitter and bigsmall were negatively correlated, meaning that as subjects steepened the downward trajectory of the path, they also slowed down the tempo, and as they chose to increase the step size, they also chose to decrease the jitter. In motion alone, consonance and bigsmall were negatively correlated, meaning that as subjects decreased the consonance (and so increased the spikiness of the Egg), they also decreased the step size. The jitter-bigsmall correlation in music and the consonance-bigsmall correlation in motion are of similar size (-0.43 and -0.45) and in the same direction; also, within the dataset as a whole (both tasks for all emotions), jitter and consonance are negatively

correlated (a discussion of the whole dataset correlations is below). This leads us to hypothesize that motion subjects' spikiness choices and music subjects' consonance choices are similarly motivated.

Angry data points are best discriminated from other data points by assessing their position on the consonance-updown plane. Consonance is the most important parameter (0.577), followed by updown (0.355). LDA results for angry motion alone are similar but exaggerated. For angry music alone the result is different: consonance and updown are still the most important dimensions for distinguishing angry from the other emotions, updown is more important (0.634) than consonance (0.265).

3.6.2.2 Happy

	bpm	jitter	consonance	bigsmall	updown
Mean all	280.12	33.24	32.00	49.36	35.56
SD all	87.55	32.56	8.40	34.43	16.93
Mean Mu	321.84	43.08	31.24	32.16	31.92
SD Mu	59.36	32.78	8.27	30.12	15.55
Mean Mo	238.40	23.40	32.76	66.56	39.20
SD Mo	92.19	29.80	8.64	29.94	17.77

Table 5: *Happy: means and standard deviations*

Music and motion	BPM, jitter	r = 0.51, p<0.0001, 95% CI 0.27 to 0.69
	BPM, bigsmall	r = -0.33, p<0.02, 95% CI -0.56 to -0.06
Music	BPM, bigsmall	r = -0.36, p<0.08, 95% CI -0.66 to 0.04
	Updown, bigsmall	r = 0.4, p< 0.043, 95% CI 0.02 to 0.69
Motion	BPM, jitter	r = 0.62, p<0.002, 95% CI 0.29 to 0.81

Table 6: *Happy: correlations with magnitude greater than 0.3*

	music and motion	music	motion
bpm	0.007	0.013	0.003
jitter	0.000	0.002	0.001
consonance	0.925	0.801	0.913
bigsmall	0.008	0.020	0.071
updown	0.060	0.164	0.011

Table 7: *Happy: LDA results*

Happy is fast, slightly jittery, consonant, has medium step size, and tends moderately upward. There are a number of significant differences between happy music and happy motion. Happy music tends to be faster. The BPM value for happy motion has a large standard deviation and a relatively low minimum value of 98; it's possible that motion may appear happy so long as it is above a certain threshold. Happy music tends to be more jittery than motion. 8 of 15 happy music subjects chose jitter values of 26 or below, while the rest chose values between 48 and 95. It may be that the high-jitter subjects were trying to create more elaborate rhythms.

In happy music and motion considered together, BPM and bigsmall are negatively correlated, meaning that as speed increases, step size decreases. This suggests that subjects want to achieve an increase in speed which does not also dramatically increase the path's angle of upward movement. Perhaps because this angle of upward movement is probably less perceptible in the motion domain than the music domain, this correlation is stronger in music ($r = -0.36$, vs. $r = -0.13$). This hypothesis is strengthened by the correlation in music of updown and bigsmall ($r = 0.4$), demonstrating that as step size increases, the ratio of upward to downward movements decreases, moderating the angle of upward movement in the same way as the BPM-bigsmall correlation. Across both tasks, BPM and jitter are relatively strongly correlated, suggesting that as tempo increases, more jitter is necessary to achieve the same effect. When only considering happy music and motion, this correlation is much stronger in motion ($r = 0.62$) than in music ($r = 0.22$). This may be because of differences in accuracy in visual vs. auditory rhythm processing (Saenz and Koch, 2008). This correlation isn't unique to happiness: it holds across all emotions for both tasks ($r = 0.45$, $p<5.33e-13$, 95% CI 0.35 to 0.54), for just the music task ($r = 0.39$, $p< 5.69e-06$, 95% CI 0.23 to 0.53), and for just the motion task ($r = 0.52$, $p<6.801e-10$, 95% CI 0.37 to 0.63).

Consonance is by far the most important dimension for distinguishing happy music and motion data points from the other emotions. For music and motion together, the second most important dimension is updown (0.06), and updown is still more important for music on its own (0.164). For motion alone, however, bigsmall is more important (0.071) than updown (0.011).

3.6.2.3 Peaceful

	bpm	jitter	consonance	bigsmall	updown
Mean O	69.48	11.30	32.34	23.66	38.34
SD O	41.90	20.19	9.04	24.41	20.07
Mean Mu	77.16	12.68	29.76	24.44	41.32
SD Mu	41.51	16.48	9.70	22.83	14.26
Mean Mo	61.80	9.92	34.92	22.88	35.36
SD Mo	41.70	23.59	7.68	26.34	24.51

Table 8: *Peaceful: means and standard deviations*

Music and motion	Bigsmall, consonance	r = -0.31, p<0.026, 95% CI -0.54 to -0.04
	Bigsmall, jitter	r = 0.33, p<0.018, 95% CI 0.06 to 0.56
	Consonance, jitter	r = -0.49, p<0.00032, 95% CI -0.67 to -0.24
Music	Consonance, jitter	r = -0.32, p<0.13, 95% CI -0.63 to 0.09
Motion	Bigsmall, consonance	r = -0.5, p<0.012, 95% CI -0.74 to -0.12
	Bigsmall, jitter	r = 0.56, p<0.0036, 95% CI 0.21 to 0.78
	Consonance, jitter	r = -0.68, p<0.00017, 95% -0.84 to -0.39

Table 9: *Peaceful: correlations with magnitude greater than 0.3*

	music and motion	music	motion
bpm	0.047	0.043	0.049
jitter	0.020	0.075	0.007
consonance	0.379	0.573	0.317
bigsmall	0.030	0.178	0.001
updown	0.523	0.132	0.626

Table 10: *Peaceful: LDA results*

Peacefulness has a slow or slow-medium tempo, very low jitter, is quite consonant, takes small steps, and tends upward. Peaceful music tends to be faster than peaceful motion, with a mean BPM of 73.6, within the average normal human heart rate range of 75 +- 2 (Mancia et al., 1983). Across both tasks, consonance and jitter are negatively correlated, bigsmall and jitter are positively correlated, and bigsmall and consonance are negatively correlated, suggesting that dissonance, jitter and large step size are similar insofar as they work to disrupt the peace. The correlation between bigsmall and jitter in peaceful music alone is very weak ($r = 0.04$); rather than suggesting a fundamental difference between the tasks, this may because the standard deviation of jitter values in peaceful music is so dramatically small, i.e. jitter so effectively disrupts peacefulness that in the music task subjects eschewed it almost completely. This effect may be more extreme in music because of increased rhythmic acuity in audition vs. vision (Saenz and Koch, 2008).

Peaceful music and motion together are best distinguished from the other emotions by position on the consonance-updown plane, although updown (0.523) is more important than consonance (0.378). Peaceful motion is similar. For peaceful music on its own, consonance is the most important dimension (0.573), followed by bigsmall (0.178) and updown (0.132); position on the

consonance-updown plane alone is not enough to distinguish peaceful music from other kinds of music, three dimensions are necessary.

3.6.2.4 Sad

	bpm	jitter	consonance	bigsmall	updown
Mean O	53.74	19.44	22.66	26.28	78.30
SD O	26.80	25.43	15.22	31.40	26.84
Mean Mu	60.80	19.56	17.32	47.52	64.76
SD Mu	29.42	23.51	14.99	30.87	26.36
Mean Mo	46.68	19.32	28.00	5.04	91.84
SD Mo	22.29	27.70	13.75	10.93	19.85

Table 11: *Sad: means and standard deviations*

Music and motion	Updown, bigsmall	r = -0.35, p<0.013, 95% CI -0.57 to -0.08
	Updown, consonance	r = 0.38, p<0.007, 95% CI 0.11 to 0.59
	Jitter, consonance	r = -0.42, p<0.0022, 95% CI -0.62 to -0.16
Music	Jitter, updown	r = -0.36, p<0.079, 95% CI -0.66 to 0.04
	Jitter, consonance	r = -0.47, p<0.018, 95% CI -0.73 to -0.09
Motion	Updown, bigsmall	r = -0.63, p<0.00075, 95% CI -0.82 to -0.31
	Updown, consonance	r = 0.42, p<0.036, 95% CI 0.033 to 0.7
	Jitter, consonance	r = -0.44, p<0.029, 95% CI -0.71 to -0.05
	BPM, consonance	r = 0.31, p<0.13, 95% CI -0.09 to 0.63

Table 12: *Sad: correlations with magnitude greater than 0.3*

	music and motion	music	motion
bpm	0.071	0.119	0.028
jitter	0.009	0.006	0.021
consonance	0.175	0.546	0.014
bigsmall	0.037	0.007	0.108
updown	0.708	0.323	0.829

Table 13: *Sad: LDA results*

Sadness is slow, has low jitter, is moderately dissonant, takes small steps, and moves decisively downward. Sad motion has a smaller step size (mean: 1.26; sd: 2.91) than sad music (mean: 48.4; sd: 31.56), and sad music tends downward at a much slower rate than sad motion. Sad motion tends to be quite a bit more consonant than sad music, with a majority of subjects placing the slider at the most consonant position. This indicates our analogical mapping of consonance to spikiness is inappropriate for sadness in the context of our model. This makes intuitive sense; spikes seem angry or active, and sadness is sedate and slow moving. We suggest an improvement of the analogical mapping would provide a more natural result, e.g. instead of simply mapping note-to-note consonance to spike length, it could also map to spike sharpness/dullness, so intermediate levels of dissonance would produce angles and bumps but not "spikes" per se. It may also be possible that our mapping is correct, in which case this result could demonstrate a fundamental difference between the two domains.

Across both tasks, and in each task individually, jitter and consonance are negatively correlated, suggesting that increasing jitter and decreasing consonance serve a similar function for sadness. Across both tasks, updown and consonance are correlated, meaning that the steeper the downward trajectory the more consonance is applied. Updown and bigsmall are negatively correlated,

meaning that as the downward trajectory becomes steeper, the step size becomes smaller. Similarly, in the music task, jitter and updown are negatively correlated. In each of these three correlations, the setting in one parameter seems to moderate the effects of the other. In the motion task, BPM and consonance are weakly correlated.

Sad music and motion together are best distinguished from the other emotions based on consonance-updown plane position, with updown as the most important dimension (0.708). Sad music alone is also based distinguished on the consonance-updown plane, but consonance is the most important dimension (0.545), followed by updown (0.323), and then BPM (0.119). Sad motion alone is best distinguished from the other emotion based on updown (0.829) and bigsmall (0.108).

3.6.2.5 Scared

	bpm	jitter	consonance	bigsmall	updown
Mean O	289.68	58.22	10.08	52.28	51.12
SD O	108.83	36.30	11.32	37.17	31.85
Mean Mu	293.92	57.52	10.16	62.72	54.56
SD Mu	114.16	38.59	12.11	31.63	30.52
Mean Mo	285.44	58.92	10.00	41.84	47.68
SD Mo	105.40	34.64	10.73	39.92	33.40

Table 14: *Scared: means and standard deviations*

Music and motion	None	
Music	None	
Motion	Updown, jitter	r = 0.34, p<0.1, 95% CI -0.06 to 0.65

Table 15: *Scared: correlations with magnitude greater than 0.3*

	music and motion	music	motion
bpm	0.003	0.002	0.006
jitter	0.042	0.089	0.019
consonance	0.888	0.802	0.748
bigsmall	0.006	0.074	0.110
updown	0.061	0.034	0.116

Table 16: *Scared: LDA results*

Scared is fast, quite jittery, quite dissonant, and tends neither upward nor downward. Scared music tends to have a moderately large step size, whereas scared motion has a medium-small step size. This difference seems like a fundamental difference between the two modalities; while both modalities seem to depend on moment-to-moment unpredictability (equal probability of upward and downward motion), scared movement seems tentative, as if walking through a haunted house, whereas scared music is more active, as if being chased by some active threat. The data for scared are almost entirely uncorrelated, with the exception of updown and jitter in scared motion.

In every case, consonance is the most important dimension for distinguishing sadness from the other emotions. In music and motion together, updown (0.06) and jitter (0.04) are also important; in music alone, jitter (0.088),

bigsmall (0.074) and updown (0.034) are important, and in motion alone bigsmall (0.11) and updown (0.116) are important.

3.6.3 Whole dataset analyses

Music and motion	BPM, jitter	r = 0.45, p<5.33e-13, 95% CI 0.35 to 0.54
	BPM, bigsmall	r = 0.35, p<1.363e-08, 95% CI 0.23 to 0.45
	BPM, consonance	r = -0.32, p<2.04e-07, 95% CI -0.43 to -0.2
	Consonance, jitter	r = -0.44, p<1.79e-13, 95% CI -0.54 to -0.34
	Consonance, bigsmall	r = -0.33, p<1.16e-07, 95% CI -0.43 to -0.21
Music	BPM, jitter	r = 0.39, p< 5.69e-06, 95% CI 0.23 to 0.53
Motion	BPM, jitter	r = 0.52, p<6.801e-10, 95% CI 0.37 to 0.63
	BPM, bigsmall	r = 0.5, p<1.67e-09, 95% CI 0.36 to 0.63
	BPM, consonance	r = -0.48, p<1.21e-08, 95% CI -0.6 to -0.33
	Consonance, jitter	r = -0.58, p<8.074e-13, 95% CI -0.69 to -0.45
	Consonance, bigsmall	r = -0.35, p<4.89e-05, 95% CI -0.5 to -0.19

Table 15: *Whole dataset: correlations with magnitude greater than 0.3*

Across the whole dataset, there are five moderately correlated parameter pairs. As subjects increase the tempo, they tend to decrease consonance, increase jitter, and increase step size. As subjects increase consonance, they tend to decrease jitter and decrease step size. When the data are limited to the music group alone, there is only one significant correlation with a magnitude greater than 0.3: as tempo is increased, jitter is increased. The motion group is correlated like the

combined dataset, except the sizes of the correlations are slightly larger. There are more significant inter-parametric correlations within motion than music.

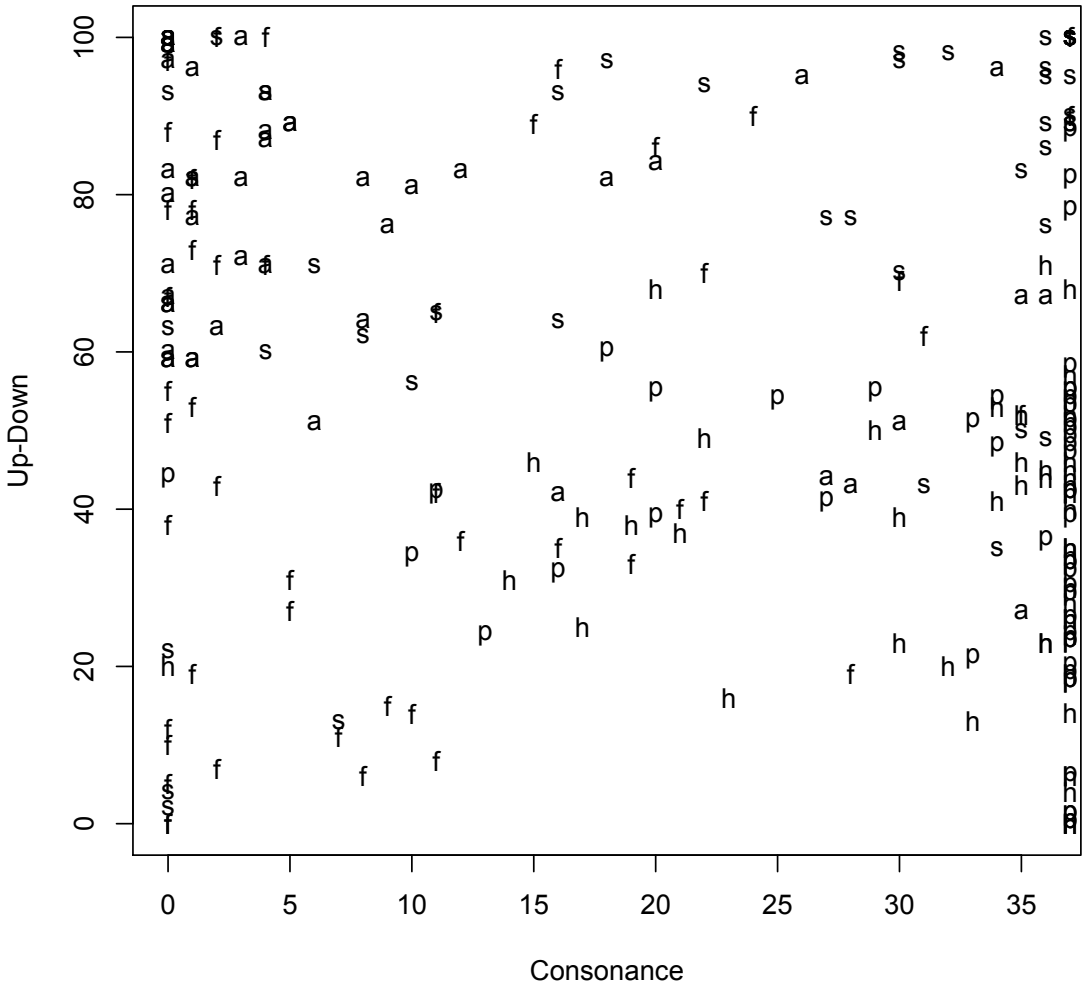

Figure 3: *All points on the consonance-updown plane.*
a = angry, h = happy, p = peaceful, s = sad, f = scared.

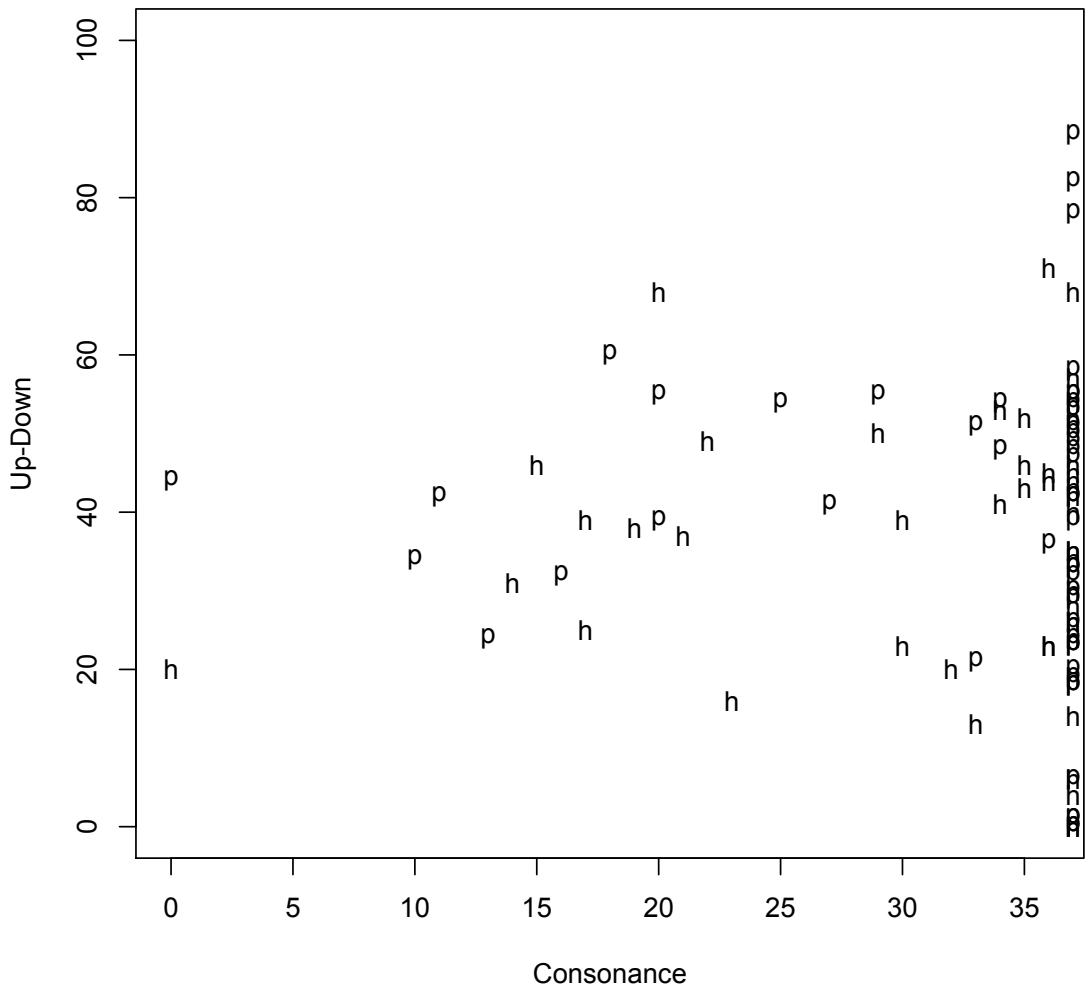

Figure 4: *Happy and peaceful data points on the consonance-updown plane.*
h = happy, p = peaceful

When the dataset is considered as a whole, each emotion is best discriminated from the others on the consonance-updown plane. This is not to suggest the other dimensions are unimportant, or that examining position on the consonance-updown plane is sufficient to accurately determine the emotion class of a data point. As the figure 4 shows, despite the results of LDA, discriminating

between happy and peaceful is impossible in terms of consonance and updown. LDA assumes the data can be modeled by gaussian distributions. This is more-or-less true for angry, sad, and scared, but not true for happy and peaceful, within which the positions of the consonance slider bar are clumped up around the maximum. To determine the parameters which best distinguish happy and peaceful, we looked at the within-class covariance of each parameter with itself, for both emotions.

	Happy	Peaceful
BPM	7664.31	1755.93
Jitter	1060.06	407.64
Consonance	70.61	81.70
Bigsmall	1185.34	595.66
Updown	286.54	402.84

Table 18: *Within class covariance for happy and peaceful*

Relatively large within-class covariance indicates the data points clump together in that dimension, suggesting its importance for between-class discrimination. For both happy and peaceful, within class covariance is highest for BPM and bigsmall. And, indeed, plotting happy and peaceful together on the bigsmall-BPM plane (figure 5) makes it possible to cleanly discriminate between the two emotions.

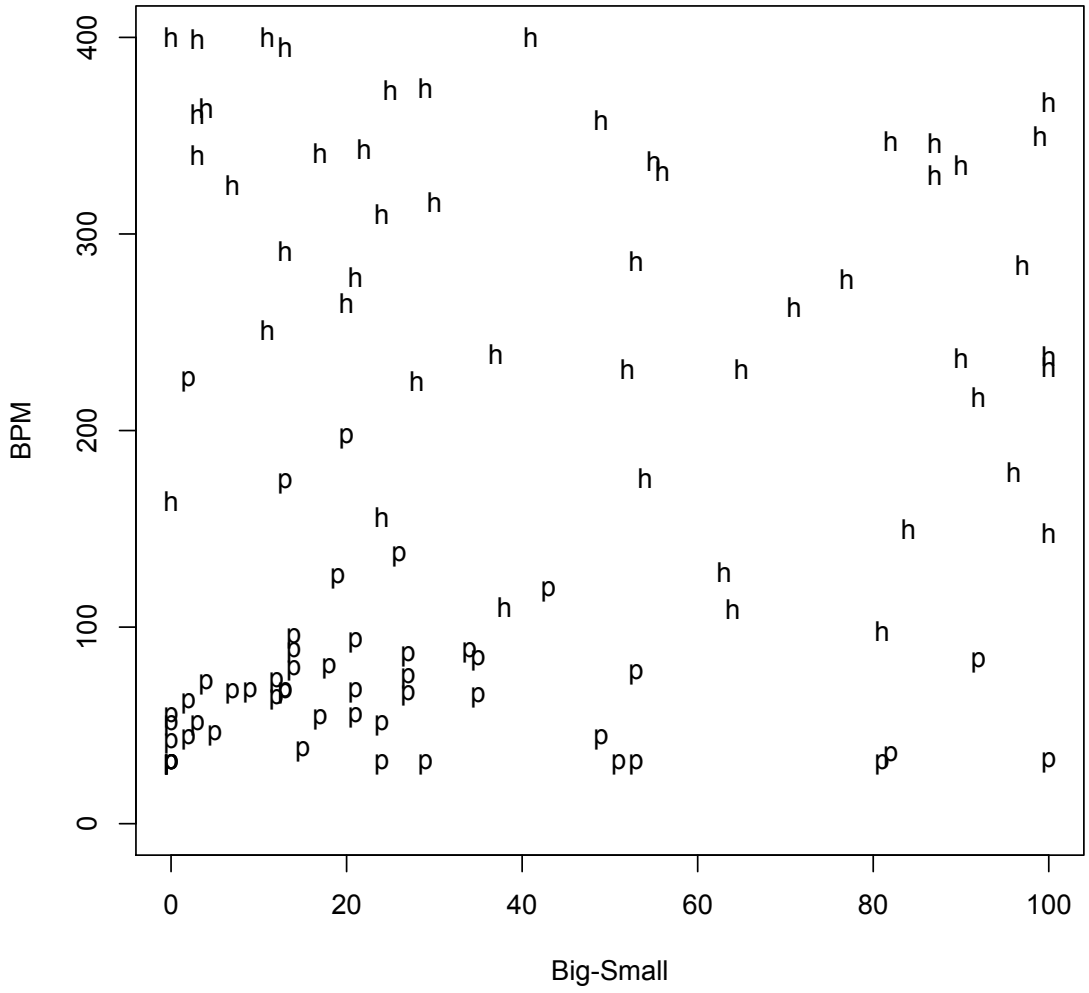

Figure 5: *Happy and peaceful data points on the bigsmall-BPM plane*
h = happy, p = peaceful

As seen in table 19, consonance was found to be the parameter most important for discriminating between music and emotion data points. This indicates that, in relative terms, music subjects used consonance differently than motion subjects used spikiness. Examining the data in absolute terms, consonance values

across modalities are quite similar. A high LDA importance value may be one way of distinguishing between analogical and isomorphic mappings.

BPM	0.013
Jitter	0.122
Consonance	0.621
Bigsmall	0.013
Updown	0.231

Table 19: *Music versus motion LDA*

3.6.4 Similarity analysis/hierarchical clustering

The distance matrix in figure 6 was created by taking the Euclidean distance from every data point to every other data point. The data were not regularized. Broad structural features of the data are readily apparent: angry, happy, and scared are similar to one another, and peaceful and sad are similar to one another. Because BPM has the largest range of all of the parameters, but is not always the most important for distinguishing between emotions, the distances between emotions at different BPM levels seem exaggerated. The regularized distance matrix in figure 7 provides more detail.

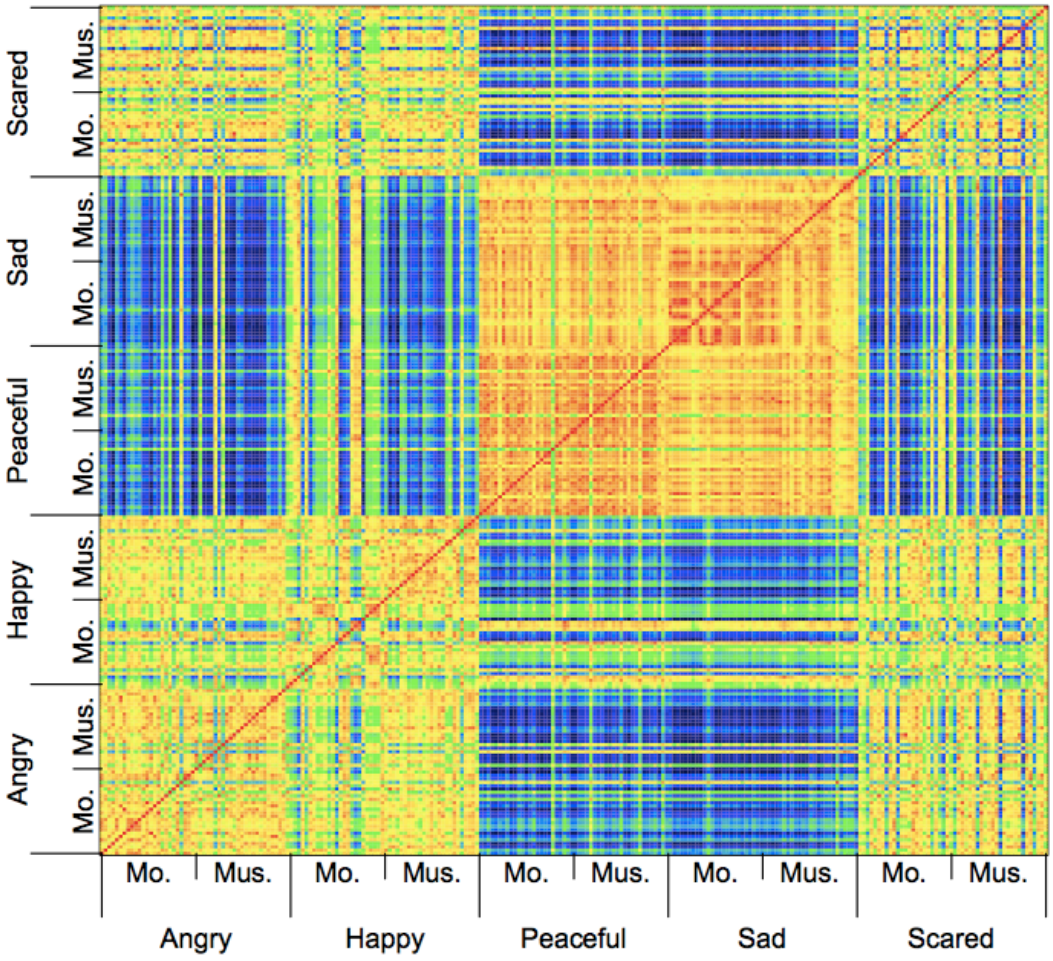

Figure 6: *Raw distance matrix*
Blue = distant, red= close

Where, in the raw distance matrix (figure 6), happy, angry, and scared formed one almost indistinguishable block, and peaceful and sad formed another block, in the normalized matrix each emotion is distinct. We can see that happy and peaceful are more similar than happy and sad. Peaceful reveals itself as the most self-similar emotion, while scared is the least self-similar.

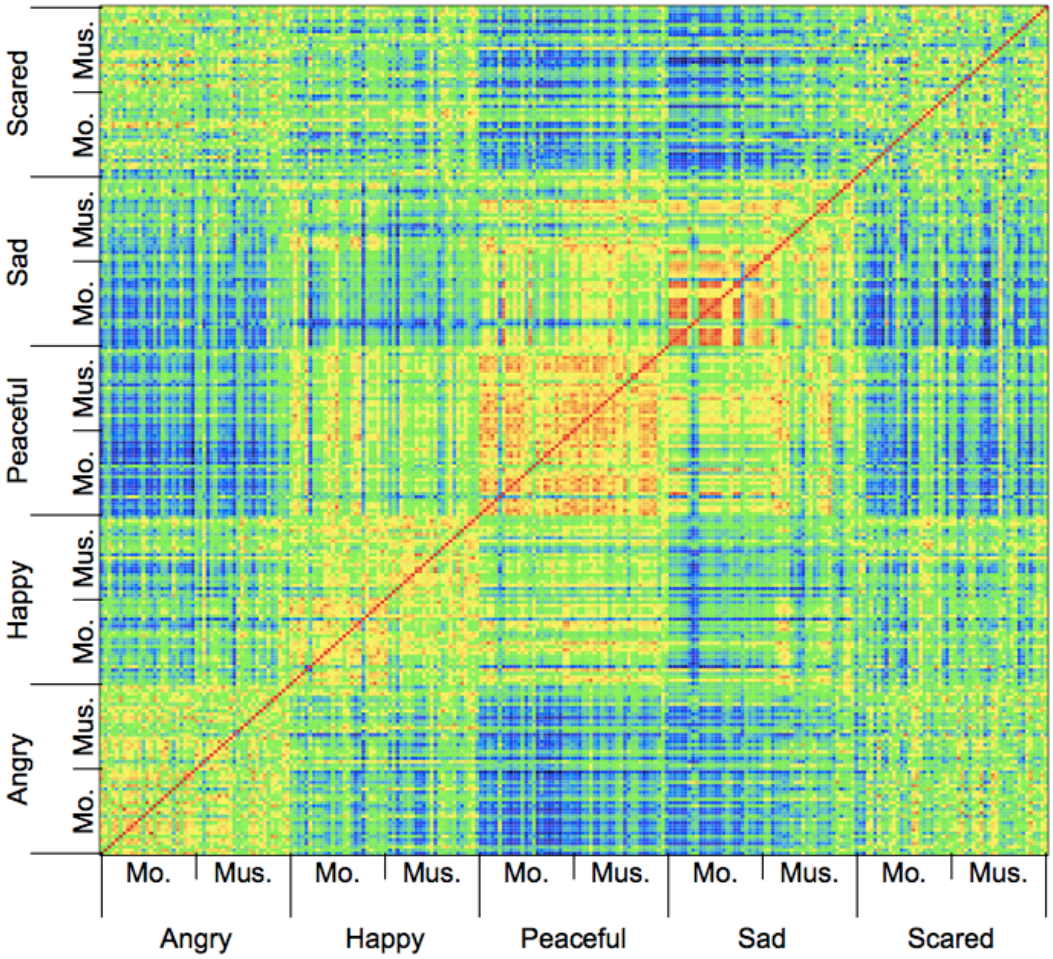

Figure 7: *Regularized distance matrix*
Blue = distant, red= close

The dendrogram in figure 8 was generated from the regularized distance matrix, using the average distance between points as the affinity function. The clustering was performed by R's (R Development Core Team, 2008) hclust function, which, at each merge, places the tighter subtree at the top. Across the dendrogram, music and motion are mixed together, providing further confirmation that the

dynamics of emotion function similarly across these modalities. Of the two first-level subtrees, the top subtree contains mostly happy, sad, and peaceful data points, while the bottom subtree contains mostly angry and scared. Most of the sad data points are contained within their own fourth-level subtree, whose closest (also fourth-level) neighbor is a subtree comprising most of the peaceful data points. The bottommost first-level subtree, containing mostly angry and scared data points, is more confused. Angry data points tend to clump together in large groups, and there are a few small groups of similar scared data points, especially in the bottommost third-level subtree. Scared, however, is confused with angry in the bottommost first-level subtree, but occasionally sneaks into the topmost first-level subtree as well. Notably, the first-level of the tree does not divide the data points by either valence or arousal.

Please see the fold-out insert (p. 69) for figure 8, *Dendrogram from regularized distance matrix*. In Figure 8, the numbers on each leaf record the raw slider bar positions of each parameter in the following order: BPM, jitter, consonance, big-small, up-down.


0.0 1.0 2.0 3.0
happy motion 284 0 37 97 35
happy music 277 4 37 77 43
happy music 330 6 37 87 46
happy music 238 3 37 100 57
happy motion 237 0 29 90 50
sad music 104 20 27 94 77
sad music 93 20 30 71 70
happy music 148 26 37 100 34
happy music 179 24 35 96 46
happy music 217 30 35 92 52
sad music 82 2 35 100 50
peaceful music 31 7 34 100 54
happy music 98 2 37 81 44
happy music 150 0 37 84 52
sad motion 37 1 22 1 94
sad motion 63 30 36 19 89
sad motion 47 28 37 3 100
sad motion 126 5 32 2 98
sad motion 30 0 37 0 100
sad motion 30 0 37 0 100
sad motion 30 0 37 0 100
sad motion 30 6 37 0 100
sad motion 36 5 36 0 100
sad motion 52 0 37 4 95
sad motion 56 0 37 2 100
sad motion 79 0 37 0 100
sad motion 70 0 37 0 100
sad motion 30 0 37 0 89
peaceful motion 43 0 37 2 88
sad music 33 0 35 0 83
peaceful music 41 0 37 0 78
sad motion 73 7 37 1 90
sad motion 44 7 37 11 89
peaceful motion 63 0 37 12 82
sad music 30 2 36 47 76
sad music 32 9 36 53 95
sad music 69 16 36 46 86
scared music 55 32 22 29 70
sad music 53 15 28 25 77
sad motion 44 52 30 37 98
sad music 49 66 36 4 96
sad motion 30 46 30 0 97
sad music 49 47 16 1 93
scared motion 102 83 15 9 89
peaceful motion 30 64 37 0 53
peaceful music 67 65 37 13 23
peaceful motion 30 12 37 51 0
happy motion 109 0 37 64 6
happy music 176 2 36 54 23
happy music 128 5 37 63 35
sad music 39 8 34 66 35
peaceful motion 30 0 37 53 29
peaceful motion 76 1 37 53 26
peaceful music 64 2 37 35 19
peaceful music 87 8 36 34 36
peaceful motion 118 0 37 43 30
happy music 278 0 37 21 40
happy motion 164 0 37 0 54
peaceful motion 173 0 37 13 39
peaceful motion 37 0 37 15 6
peaceful motion 45 2 37 5 1
peaceful motion 61 0 37 2 1
peaceful motion 50 0 37 0 33
peaceful motion 30 0 37 0 20
peaceful motion 50 0 37 3 18
peaceful music 54 9 33 21 21
peaceful music 92 0 37 21 32
peaceful motion 78 0 37 14 33
peaceful music 53 4 37 17 24
peaceful motion 72 0 37 12 18
peaceful motion 94 0 37 14 23
sad music 46 9 36 49 49
peaceful music 43 9 37 49 49
peaceful music 30 2 37 29 42
peaceful motion 30 0 37 24 58
happy motion 110 4 37 38 51
peaceful motion 83 0 37 35 55
peaceful motion 74 0 37 27 54
peaceful motion 30 0 37 0 55
peaceful motion 30 0 37 0 51
peaceful motion 54 0 37 0 42
peaceful motion 71 0 37 4 47
peaceful music 67 0 37 9 48
peaceful music 66 18 29 7 55
peaceful music 66 24 37 13 50
peaceful music 65 18 25 27 54
peaceful music 85 11 27 27 41
peaceful music 79 4 33 18 51
sad music 79 1 31 24 43
peaceful music 136 53 20 26 55
sad music 92 45 11 37 65
peaceful music 196 5 18 20 60
happy motion 156 0 20 24 68
peaceful music 50 26 20 24 39
peaceful music 67 26 16 21 32
sad music 82 22 10 24 56
peaceful music 125 13 11 19 42
peaceful music 87 23 10 14 34
sad motion 38 14 16 0 64
sad music 46 0 6 9 71
sad music 40 5 8 34 62
sad music 35 3 4 34 60
happy music 332 51 17 56 39
happy music 337 65 19 55 38
happy music 367 71 34 100 41
happy music 346 52 30 87 39
happy music 335 60 36 90 44
angry music 275 70 28 72 43
scared music 297 82 28 81 19
scared motion 292 86 24 100 90
scared music 393 88 20 78 86
happy music 358 66 36 49 71
angry music 398 84 36 100 67
happy music 347 78 37 82 68
happy music 400 89 37 11 0
happy music 399 99 37 3 4
happy music 239 52 30 37 23
happy music 343 68 33 22 13
happy music 291 60 37 13 28
happy music 265 77 36 20 23
happy music 374 48 35 29 43
happy music 316 31 36 30 45
happy music 231 59 34 52 53
happy music 225 44 37 28 45
angry music 341 80 30 36 51
scared music 337 99 35 23 52
angry music 361 94 35 22 67
happy music 395 95 37 13 30
angry motion 400 99 27 10 44
happy music 310 68 22 24 49
scared motion 328 85 19 11 44
scared music 319 98 11 37 65
happy music 373 93 15 25 46
angry music 400 99 16 43 42
happy music 286 1 37 53 0
happy music 263 23 37 71 14
happy music 231 22 37 65 26
scared motion 332 0 22 57 41
angry music 400 13 35 64 27
happy music 400 0 37 41 42
scared motion 400 0 31 15 62
scared music 367 12 30 6 69
peaceful motion 225 0 37 2 0
happy music 341 9 32 17 20
happy music 400 0 37 0 24
happy music 340 0 37 3 19
happy music 361 0 37 3 26
happy music 325 26 23 7 16
happy music 364 20 21 4 37
scared motion 291 0 21 3 40
angry music 380 54 6 10 51
scared music 390 7 0 1 55
scared motion 400 52 0 0 0
scared motion 268 0 8 0 6
happy music 251 53 14 11 31
scared motion 232 44 10 3 14
scared music 154 52 9 6 15
sad motion 30 67 0 0 100
sad motion 30 99 0 0 100
scared motion 105 99 0 0 88
sad music 35 0 0 7 93
sad motion 30 3 2 1 100
scared motion 95 29 0 29 97
sad music 154 54 4 17 93
angry music 152 52 0 2 99
scared motion 399 95 37 1 90
scared motion 400 99 4 0 97
angry motion 400 99 0 1 99
scared motion 400 99 0 8 78
scared motion 374 92 2 6 87
angry music 351 67 12 9 83
angry music 322 66 1 4 96
angry music 343 64 1 1 77
angry music 343 77 1 20 82
angry motion 400 76 0 42 59
angry music 385 82 0 55 67
scared music 284 82 1 53 78
angry motion 327 92 4 55 87
angry music 387 88 4 90 93
angry motion 400 99 3 95 100
angry motion 300 82 0 87 99
angry motion 333 83 5 81 89
angry motion 400 80 0 100 59
angry motion 330 77 2 89 63
angry motion 308 61 0 100 60
scared motion 322 66 1 92 53
angry motion 292 99 0 100 71
scared music 343 98 1 90 82
scared music 354 99 0 78 67
scared music 386 88 4 74 71
scared music 346 84 2 80 71
scared music 352 88 3 86 72
angry music 355 88 4 91 71
sad music 35 55 0 80 99
angry music 199 92 0 78 83
angry music 159 58 0 87 100
angry music 192 48 5 84 89
angry motion 277 58 0 100 100
angry motion 269 67 0 100 80
sad music 72 0 18 96 97
sad music 65 0 0 81 63
scared music 97 27 1 90 73
sad music 57 18 1 100 82
angry music 218 5 0 53 97
angry music 136 7 10 87 81
angry music 233 0 8 71 82
angry music 389 2 3 71 82
angry music 400 6 8 67 64
angry music 400 11 9 66 76
angry motion 282 0 0 85 66
angry motion 400 21 0 92 66
scared motion 339 14 0 100 51
angry music 361 25 1 100 59
angry music 343 18 1 95 59
angry music 394 20 18 100 82
scared music 371 1 2 94 100
angry music 387 4 0 100 100
angry motion 291 34 4 98 88
angry motion 400 35 0 100 100
angry motion 400 35 0 100 100
angry music 325 13 34 87 96
scared music 400 0 37 91 100
angry music 306 5 26 38 95
scared music 376 17 16 47 96
angry music 354 8 20 72 84
happy music 232 0 17 100 25
scared music 265 4 0 96 10
scared music 400 7 0 93 0
scared motion 348 96 0 51 5
scared motion 368 64 12 100 36
scared motion 330 59 2 98 7
scared motion 352 89 0 100 12
happy motion 350 76 0 99 20
scared motion 369 42 1 77 19
scared motion 395 46 11 65 8
scared music 211 39 5 78 31
scared music 292 59 5 62 27
scared motion 215 69 7 69 11
peaceful music 34 10 13 82 24
sad music 33 36 7 46 13
sad motion 34 0 0 40 4
sad music 58 67 0 13 2
sad music 56 82 0 35 22
scared music 63 99 0 30 38
peaceful music 82 66 34 92 48
peaceful motion 30 83 0 81 44
scared music 118 85 11 61 42
scared motion 132 96 2 78 43
scared music 123 85 16 97 35
scared motion 155 63 19 67 33

3.7 Discussion and directions for further research

The results confirm our hypothesis that emotionally expressive contours are very similar in simple music and movement. For the most part, our mean parameter values line up with the musical feature-emotion associations identified by Juslin and Laukka (2004). Further, our results are good evidence in favor of the theory that recognition of emotion in music operates mimetically with respect to motion, and that cross-modal mapping is implicated in the recognition process.

Our approach also reveals a crucial shortcoming of the typical approach to emotion-feature association: it is important to examine any putative feature – especially the use of major and minor scales – very carefully to see whether it might be a subset of some broader concept. Our subjects did not simply push the dissonance slider to one side or the other, dividing the notion of dissonance into the major scale and everything else. Nor did they leave the slider predominantly on the consonant side, limiting the results to the major and minor scales. A broad range of consonance values were selected, with consistency within categories, showing that major and minor scales alone are not sufficient to account for the role of musical dissonance in emotion perception. In general, the organization of empirical research around music theoretical terminology such as "major" and "minor" reifies Western cultural practice. This reification both limits the applicability of experimental results and institutionalizes a vocabulary which cannot speak coherently about the musics of non-Western cultures.

Subjects in an in-progress follow-up fMRI experiment (discussed below) confirmed that the stimuli generated in the behavioral experiment are recognizable as conveying the intended emotions. Because subjects in both the music and motion groups used the slider bars in similar ways, our mappings from music to motion must have some perceptual salience. Each mapping can be read as an implicit hypothesis about cross-modal expression of emotion; e.g. mapping

dissonance to spikiness can be read as a claim that, for the task at hand, dissonance and spikiness are functionally similar. The three mapping hypotheses confirmed by our experiment are: 1) dissonance is like spikiness, 2) pitch height is like angle of orientation (i.e. looking up, looking down, leaning back, leaning forward), and 3) pitch interval size is like bounce height.

This raises some important issues. First, what kinds of mappings would *not* work? Presumably if the directionality of any of these three mappings were inverted in one of the modalities, music and motion subjects would use the sliders dramatically differently. An interesting change to the experiment would be to systematically adjust the scaling values for each mapping, or to allow subjects to adjust the scaling values themselves. This would allow us to measure exactly how movement in one modality corresponds to movement in another. Second, we don't have a good sense of how interactions between the parameters and the mappings affected the results. Again, because the slider positions were similar across modalities, we may assume that the inter-parametric interactions were also similar, but we would like to be able to measure these interactions precisely. This recommends a future experiment in which each of five groups of subjects is only given one parameter to adjust, while all the other parameters are held constant. This procedure would then be repeated combinatorially with groups of two, three, and four active parameters at a time.

A straightforward direction for further work is improvement of the underlying statistical model. In particular, the status of the parameter of dissonance is uneasy. Although our parameterization is based on perceptual studies, all of those studies focused on decontextualized, note-to-note pairwise dissonance. This is a poor approximation of how dissonance functions in real music. More preferable would be a model of dissonance based on cross-cultural perceptual studies which take musical context into account, and were freed from

the constraints of Western equal-temperament. Other possible improvements to the model include expanding its notion of contour, which emerges in our study from the interaction of upward/downward movement and big/small step size probabilities. Our model allows for the production of rough trajectories, but fails to model contour as it occurs in a typical musical context. Contours which could be described as departing from and returning to a center, or contours which decisively change velocity and direction are impossible with our model. These sorts of contours would be achievable with the introduction of the ability to describe how the slider bar settings should change over time, perhaps with hand-drawn parameter envelopes or the use of low-frequency oscillators as parameter modulation signals. Still more interesting would be a perceptually grounded model of contour which drew from cross-cultural research on its perception in various modalities and possible methods of parameterization. Also, in our model, contour and range are coupled; future models may uncover interesting results by decoupling the starting point of a path (or mean note frequency) from its contour.

Also promising are improvements and extensions to the stimulus generators themselves. Simple music and motion are a good start but with some effort more detailed, realistic stimuli could be created. In the case of music, more human sounding articulation and sound could be achieved, characteristic rhythmic patterns included, and so on. In the case of motion, the shape and movement of the ball could be made more realistic, smoothing its triangular surfaces and incorporating gravity, elasticity, weight, and other similar parameters into the rendering process. Improving the rendering of the ball might well improve the analogical mapping from dissonance to spikiness: if the spikes associated with smaller dissonance levels were rounded (low energy roughness) instead of pointy (high energy roughness), slider positions for sad motion might

move closer to those for sad music. Beyond simple improvements to our present approach, stimuli could be generated in a number of other modalities. Instead of a bouncing ball, the statistical model could be mapped to a walking human, the gait of a dog, the movement of a car, and so on. Any of these proposed changes may in turn suggest additional modifications to the underlying model.

Following Juslin (2000), an experiment which asks subjects to judge the emotionality of stimuli generated by other subjects and assessing the results using Hursch's (1964) lens model equation could offer further confirmation of the perceptual validity of the mappings used. The results from this, the present study, and others like it could be used as analytical tools. A number of attempts have been made to develop systems for automatic emotion recognition in speech (Ververidis et al., 2004), movement (Bernhardt and Robinson, 2007), and music (Liu et al., 2003). These approaches could be augmented by the inclusion of data from generative experiments. The data could also be used to assist sonification designers in reliably conveying emotion (Childs, 2003). To these ends, it will be necessary to be explicit about what sorts of musical behaviors correspond with what movements and emotions, exactly. As Alexander (1977) developed a catalog of architectural patterns, one could construct a catalog of patterned relations of music, movement, and emotion. One should note that music-movement correspondence is in no way limited to emotional signification, biological motion or human movement. In this view, cataloging patterns within the music-motion-emotion nexus could be seen as a filling-in of Smalley's (1995) notion of the indicative field. Iyer (2002) also points in this direction.

We are presently engaged in an fMRI experiment to explore the neural correlates of cross-modal emotion recognition. We hypothesize cross-modally mappable contours share neural representations regardless of presentation modality, i.e. that there is an area of the brain which responds to e.g. happy

music and happy motion in the same way. Our experiment presents subjects with a battery of music and motion stimuli based on the results of the behavioral experiment described above. A preliminary general linear model analysis of the data has yielded encouraging results, and a detailed multi-voxel pattern analysis (Norman et al., 2006) is forthcoming.

3.7.1 Testing for cross-cultural validity

We believe the evidence presented above for cross-modal perception of emotion from studies of feature-emotion association, visualization, synesthesia, infant-directed speech and music (as well as in the present study) are compelling, and that dynamic contour sits at the core of a certain kind of emotion perception. We suggest that there exists a set of emotions which may be represented by characteristic patterns of dynamic information, in any modality which will bear those patterns. That is, these emotions are expressed by the presence of certain features (e.g. slowness, smoothness, upwardness, bigness, consonance, etc.) in a gesture, event, or utterance in any modality where those features may obtain (by parametric isomorphism, analogy, or otherwise). Further, we would also like to tentatively suggest that there exists a subset of these emotions which may be represented without allusion to any specifically cultural context, i.e. that they are recognizable by their similarity to universal human behavioral predispositions.

This is an empirical claim which can be tested by generalizing the present research to a wider variety of modalities, and performing the experiment described in wider variety of cultural contexts. This is a challenging proposal, and we'd like to suggest some modifications to the present work which might be necessary to follow through with it.

Although we've taken pains to avoid overuse of specifically Western music-theoretical concepts, our model of musical dissonance has some characteristically Western features which could be a barrier to executing the experiment in other cultural contexts. First, our use of equal temperament, a characteristically Western system, would make our stimuli sound quite unusual to someone with no exposure to Western music. While it's likely that categorical pitch perception would smooth over some of the differences, the cross-cultural interaction of tuning systems and dissonance perception has not been studied in a way which would allow us to make any presumptions. Secondly, our choice of a piano sound, while more-or-less neutral within a Western context would have an entirely different set of connotations to someone who grew up under e.g. colonial British rule, opening up the experiment to exactly the kind of allusion we need to avoid.

Further, these kinds of problems will persist no matter what models, sounds, and modes of presentation are engineered. While we claim there is something universal in musical contour, any expression of that contour will necessarily be specific and culturally contingent. Rather than attempting to remove all traces of culture from the model in the present study, it would be more fruitful to aim for relative neutrality *within* the context of the host culture in which the experiment were being performed. For example, running the experiment in Cambodia might require the use of various other intonation systems, and a switch from piano to a wooden flute sound. This also opens up the possibility of running the experiment using sounds and temperaments contrary to the normal practices of the host culture, which would certainly yield interesting results. Rather than sweeping cultural differences aside, we need to acknowledge and accept them as an area of inquiry: what we are studying is not just how people are the same, but also how they're different. A challenge

associated with this relativistic approach would be finding a method of recording and analyzing data sufficiently general to account for each included culture.

Related to the question of cross-cultural validity is the possibility of using stimulus generation systems to test how certain populations conceive of the dynamics of different emotions. Once the bounds of normal behavior (either within a culture or universally) are discovered, it becomes possible to identify and study subjects who are outside of those bounds. Of particular interest are psychopaths, patients suffering from clinical depression or autism spectrum disorder, and others whose approach to emotion is abnormal.

Our schema for classifying cross modal mappings (presented in section 2.3) could be expanded and improved. Specifically, the possibility that analogical mappings could in fact be second-order parametric isomorphisms seems worth investigating; a study of the neural correlates of analogical vs. parametrically isomorphic mappings could provide useful evidence. We hypothesize that most analogical mappings depend upon what Spector and Maurer call "a common code for magnitude", while most parametrically isomorphic mappings are based on shared neural responses to similar stimuli in different modalities; e.g. increases in speed in both the visual and auditory domains may elicit a higher neural firing rate.

3.7.2 Implications of the experimental paradigm

This question of how emotion is encoded in parameters common to music and movement is a specific instance of a more general problem, which is the empirical investigation of the relationship between perceptual phenomena (or categorical judgments) and proposals for their parameterization. Techniques such as principal component analysis and single value decomposition can provide quantitative information about the dimensions on which a data set varies, but

the results provided do not always line up with perceptual data or phenomenological reports. Our methodology provides a way of shortening this gap, evaluating a proposed parameterization of a phenomenon in terms of the decision making processes of human subjects.

Stated in general terms, the approach taken by the present work is as follows. 1) Select some complex perceptual phenomenon. The present study examines emotion as expressed in music and motion, although a number of other candidates present themselves. These candidates may tend toward seemingly universal, abstract ideas, such as the shapes or forms of objects, categorization of movement as implying behavior, and the perception of musical behavior as described by e.g. Smalley (1995), or musical gestalts as described by e.g. Tenney (1986). Other good candidates include judgments of appearance, such as whether someone looks like they're lying, or the apparent intentions of an automobile driver. 2) Determine, by looking to the literature, doing preliminary pilot studies, and consulting domain experts, a candidate parameterization of the phenomenon. For example, whether or not someone appears to be lying may be a function of their facial expression and the speed and pitch contour of their voice. (Note that this approach can only test appearance, not reality. A person may appear to lie but still be telling the truth.) 3) Build a stimulus generation system which allows naïve users to easily create exemplars of the phenomenon under investigation. 4) Analyze the settings chosen by those subjects to determine the relative contribution of each parameter. A variation on this approach, in the case of a set of parameters which are difficult for naive users to understand, or which describe a possibility space of unmanageable size, is to sample the full breadth of the possibility space at some sufficiently small interval, creating a stimulus for each sample. The subjects are then instructed to sort and categorically label each stimulus as they see fit.

Note that this approach cannot reveal whether a phenomenon is universally perceptible, absolutely grounded, authentic, or completely and perfectly decomposable into the proposed parameterization. We can only show how subjects decide to use a given parameterization/generative system pair for encoding comprehensible representations of the phenomenon. Even if the results of an experiment of this type are strong, the generated stimuli are likely to be much simpler than the naturally occurring phenomena they are meant to represent. Any parameterization approaching completeness (if such a thing is possible) would be large enough that a subject-driven stimulus generation system would be too complex to use, and a suitably dense sampling of the possibility space too vast to test in a reasonable amount of time. Further, the parameterization itself may be decomposed such that optimistic interpretations of apparently strong results may be undermined. For example, if there is some implicit factor X which is correlated, but perhaps nonlinearly or unreliably, with a factor Y explicitly accounted for in the model, triumphant statements to the effect of "factor Y is the cause of effect Z" may be completely off the mark. Or we may incorrectly model some parameter as continuous when its perceptual analog is categorical, or vice versa. The way we model various parameters may bear only a superficial relationship to how those parameters are experienced by our subjects. We should not underestimate the complexity of the "folk" notion that, for example, dissonance is important to conveying some emotion in music, understanding that the model of dissonance we build is not (and perhaps can never be) completely isomorphic with any perceptual experience. If our experiment works, we can rightly claim our model has some validity, but this claim is pragmatic, not absolute. We can only say the way we have modeled a given parameter is good enough to give us repeatable, statistically significant, useful results.

4 Bibliography

Agawu, K. (1997). John Blacking and the study of African music. Africa, 67 (3), 491-99.

Alexander, C. (1977). A Pattern Language: Towns, Buildings, Construction. New York, Oxford University Press.

Amaya, K., Bruderlin, A., and Calvert, T. (1996) Emotion From Motion. Proceedings of the Graphics Interface 1996 Conference, May 22-24, 1996, Toronto, Ontario, Canada.

Anolli, L. and Ciceri, R. (1997). The voice of deception: vocal strategies of naive and able liars. Journal of Nonverbal Behavior 21 (4) 259-284

Atkinson, A. P, Dittrich, W. D., Gemmell, A. J., and Young, A. W. (2004). Emotion perception from dynamic and static body expressions in point-light and full-light displays. Perception, 33, 717-746.

Bachorowski, J. (1999). Vocal Expression and Perception of Emotion. Current Directions in Psychological Science 8 (2) 53-57

Badler, N. I., Chi, D. M., and Chopra-Khullar, S. (1999). Virtual Human Animation Based on Movement Observation and Cognitive Behavior Models. Proceedings of Computer Animation, 1999, 128-137. IEEE Publishing.

Balch, W., Myers, D. M., and Papotto, C. (1999). Dimensions of mood in mood-dependent memory. Journal of Experimental Psychology: Learning, Memory, and Cognition 25 70-83.

Balkwill, L., and Thompson, W. F. (1999). A Cross-Cultural Investigation of the Perception of Emotion in Music: Psychophysical and Cultural Cues. Music Perception, 17 (1), 43-64.

Bartlett, D. L. (1996). Physiological responses to music and sound stimuli, in Handbook of Music Psychology (2nd edn), D. A. Hodges, ed. San Antonio, IMR.

Berlin, B. (1994). Evidence for pervasive synesthetic sound symbolism in ethnozoological nomenclature. In L. Hinton, J. Nichols, and J. Ohala (Eds.), Sound symbolism (pp. 76–93). New York: Cambridge University Press.

Bernhardt, D and Robinson, P. (2007). Detecting Affect from Non-Stylised Body Motions. In A. Paiva, R. Prada, R. W. Picard (Ed.), Affective Computing and Intelligent Interaction, Second International Conference, ACII 2007, Lisbon, Portugal, September 12-14, 2007, Proceedings: LNCS, 4738. Berlin: Springer-Verlag.

Blacking, J. (1965). The role of music in the culture of the Venda of the northern Transvaal. Studies in Ethnomusicology, 2, 20-53.

Blacking, J. (1973). How Musical is Man? Seattle and London: University of Washington Press.

Blacking, J. (1995). Music, culture, and experience: selected papers of John Blacking. Chicago: University of Chicago Press.

Brunswik, E. (1956). Perception and the representative design of psychological experiments. Berkeley, CA: University of California Press.

Carnap, R. (1950). Logical foundations of probability. Chicago, University of Chicago Press.

Castellano, G., Villalba, S., & Camurri, A. (2007). Recognising human emotions from body movement and gesture dynamics. In A. Paiva, R. Prada, R. W. Picard (Ed.), Affective Computing and Intelligent Interaction, Second International Conference, ACII 2007, Lisbon, Portugal, September 12-14, 2007, Proceedings: LNCS, 4738. Berlin: Springer-Verlag.

Childs, E. (2003). Musical sonification design. Masters thesis, Dartmouth College.

Cohen, A.J. (1993). Associationism and musical soundtrack phenomena. Contemporary Music Review, 9, 163-178.

Csikszentmihalyi, M. and Lefevre, J. (1989). Optimal experience in work and leisure. Journal of Personality and Social Psychology, 56, 815-822.

Davis, R. (1961). The fitness of names to drawings: A crosscultural study in Tanganyika. British Journal of Psychology, 52, 259–268.

Dissanayake, E. (2000). Antecedents of the Temporal Arts in Early Mother-Infant Interaction. In The Origins of Music, Wallin, N. L., Merker, B., and Brown, S. eds. Cambridge, MA, MIT Press.

Eitan, Z., and Granot, R. Y. (2003). Inter-parametric analogy and the perception of similarity in music. Proceedings of the 5th Triennial ESCOM Conference, 116-119.

Eitan, Z., and Granot, R. Y. (2006). How Music Moves: Musical Parameters and Listeners' Images of Motion. Music Perception, 23 (3), 221-247.

Ekman, P., Sorenson, R., Friesen, W. V. (1969). Pan-Cultural Elements in Facial Displays of Emotion. Science, New Series, 164 (3875), 86-88.

Ekman, P., Friesen, W. V. (1971). Constants across cultures in the face and emotion. Journal of Personality and Social Psychology, 17 (2), 124-129.

Ekman, P., Friesen, W. V., & Scherer, K. R. (1976). Body Movement And Violence Pitch In Deceptive Interaction. B Semiotica 16 (1) 23-27.

Ekman, P. (1999). Basic Emotions. In Handbook of Cognition and Emotion, T. Dalgleish and M. Power (eds.). Sussex, U.K.: John Wiley & Sons, Ltd.

Euler, L. (1739). Tentamen novae theoriae musicae ex certissimis harmoniae principiis dilucide expositae. St. Petersburg.

Fontaine, J. R. J; Scherer, K. R. et al. (2007). The World of Emotions Is Not Two-Dimensional. Psychological Science, 18 (12), 1050-1057.

Fernald, A. (1991). Prosody in speech to children: prelinguistic and linguistic functions. Ann. Child Developmentm 8 43–80.

Fernald, A. and Kuhl, P.K. (1987). Acoustic determinants of infant preference for motherese. Infant Behavior & Development, 10 279–293.

Forte, A. (1973). The Structure of Atonal Music. New Haven: Yale University Press.

Fritz, T., Jentschke, S., Gosselin, N., Sammler, D., Peretz, I., Turner, R., Friederici, A. D., and Koelsch, S. (2009). Universal Recognition of Three Basic Emotions in Music. Current Biology, 19, 1-4.

Griffiths, P. E. (2004). Is Emotion a Natural Kind? in Philosophers on Emotion, Solomon, R., ed. New York, Oxford University Press.

Gunns, R. E., Johnston, L., and Hudson, S. M. (2002). Victim selection and kinematics: A point-light investigation of vulnerability to attack. Journal of Nonverbal Behavior, 26, 129-158.

Haslam, N. (1995). The Discreteness of Emotion Concepts: Categorical Structure in the Affective Circumplex. Personality and Social Psychology Bulletin, 21 (10) 1012-1019.

Helmholtz, H. (1912). On the sensations of tone. Ellis, A. J., trans. London: Longman's, Green, and Co.

Hevner, K. (1935). The Affective Character of the Major and Minor Modes in Music. The American Journal of Psychology, 47 (1), 103-118.

Huron, D. (1994). Interval-Class Content in Equally Tempered Pitch-Class Sets: Common Scales Exhibit Optimum Tonal Consonance. Music Perception, 11 (3), 289-305.

Hursch, C. J., Hammond, K. R., and Hursch, J. L. (1964). Some methodological considerations in multiple-cue probability studies. Psychological Review, 71, 42-60.

Hutchinson, W, and Knopoff, L. (1979). The significance of the acoustic component of consonance in Western triads. Journal of Musicological Research, 3, 5-22.

Iyer, V. (2002). Embodied Mind, Situated Cognition, and Expressive Microtiming in African-American Music. Music Perception (2002) vol. 19 (3) pp. 387-414.

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. Perception & Psychophysics, 14, 201-211.

Juslin, P.N. (2000). Cue Utilization in Communication of Emotion in Music Performance: Relating Performance to Perception. Journal of Experimental Psychology, 26 (6), 1797-1813.

Juslin, P.N. (2001). Communicating emotion in music performance: a review and a theoretical framework. In Music and Emotion, Juslin, P. and Sloboda, J. A., eds. New York, Oxford University Press.

Juslin, P. N. and Laukka, P. (2003) Emotional Expression in Speech and Music: Evidence of Cross-Modal Similarities. Annals of the New York Academy of Sciences, 1000 (1), 279-282.

Juslin, P. N. and Laukka, P. (2004). Expression, Perception, and Induction of Musical emotions: A Review and a Questionnaire Study of Everyday Listening. Journal of New Music Research, 33 (3), 217-238.

Juslin, P. N. (2005). From mimesis to catharsis: expression, perception, and induction of emotion in music. In D. Miell, R. MacDonald, and D. J Hargreaves (eds.), Musical communication 85-115. New York, Oxford University Press.

Kameoka, A., and Kuriyagawa, M. (1969) Consonance theory, Part I: Consonance of dyads. Journal of the Acoustical Society of America, 45, 1451-1459.

Kenealy, P. (1988). Validation of a music mood induction procedure: some preliminary findings. Cognition and Emotion 2 41-48.

Kivy, P. (1989). Sound sentiment: an essay on the musical emotions, including the complete text of The Corded shell. Philadelphia, Temple University Press.

Köhler, W. (1929). Gestalt Psychology. New York: Liveright.

Kolinski, M. (1967). Recent Trends in Ethnomusicology. Ethnomusicology, 11 (1), 1-24.

Køppe, S., Harder, S., and Væver, M. (2008). Vitality affects. International Forum of Psychoanalysis, 17, 169-179.

Koriat, A., and Levy, I. (1979). Figural symbolism in Chinese ideographs. Journal of Psycholinguistic Research, 8, 353–365.

Kozlowski, L. T., & Cutting, J. E. (1977). Recognizing the sex of a walker from a dynamic point-light display. Perception and Psychophysics, 21, 575-580.

Krumhansl, C. L. (1990). Cognitive foundations of musical pitch. New York, Oxford University Press.

Krumhansl, C. L. (1997). An exploratory study of musical emotions and psychophysiology. Canadian Journal of Experimental Psychology 51 336-352.

Laban, R. (1960). The mastery of movement. 3rd ed. London, Macdonald & Evans.

Leibniz, G. W. (1714). The principles of nature and grace, based on reason. In Philosophical Papers and Letters (2nd ed.), Leibniz, G. W., and Loemker, L. E. (trans., ed.). Dordrecht, D. Reidel Publishing Co.

Lewis, M.M. (1951). Infant Speech. London, Routledge & Kegan Paul.

Liu, D., Lu, L., Zhang, H. (2003). Automatic Mood Detection from Acoustic Music Data. Proceedings of 4th International Conference on Music Information Retrieval, ISMIR 2003, 81-87.

Lundin, R. W. (1947). Toward a Cultural Theory of Consonance. Journal of Psychology, 23, 45-49.

Makeig, P. (2001). Sensitivity to kinematic specification of emotion and emotion-related states. Unpublished master's thesis, Canterbury University, New Zealand.

Malmberg, C. F. (1918). The perception of consonance and dissonance. Psychological Monographs, 25 (2), 93-133.

Mampe, B., Friederici, A. D., Christope, A., and Wermke, K. (2009). Newborns' Cry Melody is Shaped by Their Native Language. Current Biology, 19 (23), 1994-1997.

Mancia, G., Ferrari, A., Gregorini, L., Parati, G., Pomidossi, G., Bertinieri, G., Grassi, G., di Rienzo, M., Pedotti, A., and Zanchetti, A. (1983). Blood pressure and heart rate variabilities in normotensive and hypertensive human beings. Circulation Research, 53 (1), 96-104.

Marks, L. E. (1974). On associations of light and sound: The mediation of brightness, pitch, and loudness. American Journal of Psychology, 87, 173–188.

Martin, M. A. and Metha, A. (1997). Recall of early childhood memories through musical mood induction. Arts in Psychotherapy, 25, 447-454.

McDermott, J., and Hauser, M. (2005). The Origins of Music: Innateness, Uniqueness, and Evolution. Music Perception, 23 (1), 29-59.

Merriam, A. P. (1964). The anthropology of music. Evanston, Illinois: Northwestern University Press.

Meyer B. M. (1956). Emotion and Meaning in music. Chicago, University of Chicago Press.

MIDI Manufacturers Association (1996). The Complete MIDI 1.0 Detailed Specification. Los Angeles, CA: The MIDI Manufacturers Association.

Nettl, B. (1956). Music in primitive culture. Cambridge, MA: Harvard University Press.

Nettl, B. (1983). The study of ethnomusicology: Twenty-nine issues and concepts. Urbana: University of Illinois Press.

Norman, K., Polyn, S., Detre G., and Haxby, J. (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. Trends in Cognitive Sciences 10 (9) 424-430

Papoušek, M. (1992). Early ontogeny of vocal communication in parent-infant interactions. In Nonverbal Vocal Communication: Comparative and Developmental Approaches. H. Papoušek, V. Jürgens & M. Papoušek, Eds. 230–261. Cambridge University Press. Cambridge.

Peretz, I, et al. (1998). Music and emotion: perceptual determinants, immediacy, and isolation after brain damage. Cognition 68 111-141

Phillips-Silver, J., and Trainor, L. (2007) Hearing what the body feels: Auditory encoding of rhythmic movement. Cognition, 105 (3), 533-546.

Plato. (1992) The Republic. G.M.A. Grube, trans. Revised by C.D.C. Reeve. Indianapolis, Hackett.

Pollick, F. E., Paterson, H. M., Bruderlin, A., and Sanford, Anthony J. (2001). Perceiving affect from arm movement. Cognition, 82, B51-B61.

Puckette, M. (1991). Combining Event and Signal Processing in the MAX Graphical Programming Environment. Computer Music Journal, 15 (3), 68-77.

Quine, W. V. (1969). Natural Kinds. Ontological Relativity and Other Essays. New York, Columbia University Press.

Ramachandran, V. S., & Hubbard, E. M. (2001). Synethesia: A window into perception, thought, and language. Journal of Consciousness Studies, 12, 3–34.

R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org

Reas, C., and Fry, B. (2006). Processing: programming for the media arts. AI & Society, 20 (4), 526-538.

Rost, R. J. (2004). OpenGL Shading Language. Boston, MA: Pearson Education.

Rubin, E. (1915). Visuell wahrgenommene Figuren. In D. C. Beardslee & M. Wertheimer (Eds.), Readings in perception (pp. 194–203). Princeton, NJ: D. Van Nostrand Company, Inc.

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

Saenz, M. and Koch, C. (2008). The sound of change: visually-induced auditory synesthesia. Current Biology, 18 (15), R650-R651.

Saffran, J.R., Johnson, E.K., Aslin, R.N., & Newport, E.L. (1999). Statistical learning of tone sequences by adults and infants. Cognition, 70, 27-52.

Saffran, J. R., Hauser, M., Seibel, R., Kapfhamer, J., Tsao, F., and Cushman, F. (2008). Grammatical pattern learning by human infants and cotton-top tamarin monkeys. Cognition, 107 (2), 479-500.

Sauter, D. A., Eisner, F., Ekman, P., and Scott, K. (2010). Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. Proceedings of the National Academy of Sciences, 107 (6) 2408-2412.

Scarantino, A. (2005). Explicating Emotions. Dissertation, University of Pittsburgh.

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

Scherer, K.R., Banse, R., and Wallbott, H.G. (2001). Emotion inferences from vocal expression correlate across languages and cultures. Journal of Cross-Cultural Psychology, 32 (1), 76-92.

Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. Speech Communication 40 227-256

Shaver, P. et al. (1987). Emotion Knowledge: Further Exploration of a Prototype Approach. Journal of Personality and Social Psychology, 52 (6), 1061-1086.

Sloboda, J. A. and O'Neill, S. A. (2001). Emotions in everyday listening to music. In Music and Emotion, Juslin, P. and Sloboda, J. A., eds. New York, Oxford University Press.

Smalley, D. (1995). The Listening Imagination: Listening in the Electroacoustic Era. In Paynter, J. et al. (eds.) Companion to Contemporary Musical Thought: Volume 1, 514-554. London: Routledge.

Spector, F. and Maurer, D. (2009). Synesthesia: A new approach to understanding the development of perception. Developmental Psychology, 45 (1), 175-189.

Stern, D. N. (1985). The Interpersonal World of the Infant. New York, Basic Books.

Stumpf, C. (1890). Tonpsychologie. Leipzig: Hirzel.

Tenney, J. (1986). META+HODOS: A Phenomenology of 20th Century Musical Materials and an Approach to the Study of Form, and META Meta+Hodos. Hanover, NH: Frog Peak Music.

Trehub, S. E. (2000). Human Processing Predispositions and Musical Universals. In The Origins of Music, Wallin, N. L., Merker, B., and Brown, S. eds. Cambridge, MA: MIT Press.

Trehub, S. E. (2001). Musical predispositions in infancy. Annals of the New York Academy of Sciences, 930, 1-16.

Ververidis, D., Kotropoulos, C., Pitas, I. (2004). Automatic emotional speech classification. Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP 2004), Vol. 1. Montreal. 593–596.

Wittgenstein, L. (1953). Philosophical Investigations (2nd ed). Oxford, Blackwell.

Wundt, W. (1897). Outlines of psychology (trans. C. H. Judd). Leipzig, Englemann.

Yik, M.S.M., Russell, J.A., and Feldman-Barrett, L. (1999). Structure of self-reported current affect: Integration and beyond. Journal of Personality and Social Psychology, 77, 600–619.

Zentner, M. R., Meylan, S., and Scherer, K. R. (2000). Exploring 'musical emotions' across five genres of music. Paper presented at the Sixth Annual Conference for the Society for Music Perception and Cognition (ICMPC), 5-10 August 2000, Keele, UK.

Zicarelli, D. (1998). An extensible real-time signal processing environment for Max. Proceedings of the 1998 International Computer Music Conference. Ann Arbor, USA.