INVESTIGATING THE COGNITIVE MECHANISMS UNDERLYING INTERPRETATION OF EMOTIONAL CUES IN MUSIC

Jamie L. Abaied

Thesis Submitted in Partial Fulfillment of the Requirements

for the Bachelor of Arts Degree at Hamilton College

May 7, 2004

Advisor: Penny Yee

Abstract

This study explores the cognitive framework underlying our use of the structural features of music as emotional cues. Thirty-nine college students listened to 36 musical excerpts (18 per each of two melodies), rating the extremity of the emotional tone of each one. In order to determine whether cue-specific cognitive mechanisms exist for individual cues as well as what relationships exist between them, three variables were manipulated at the following levels for each musical excerpt: (1) slow, moderate, or fast tempo, (2) low C, middle C, or high C pitch octave, and (3) a computer generated approximation of either a clarinet or a plucked violin. Results revealed significant main effects for Tempo, Pitch, and Instrument (Timbre), as well as significant interactions between cues. Tempo was the strongest cue to emotion, and a unique association between Tempo and Timbre appears to exist. Overall, results suggest a hierarchal arrangement of cue-specific cognitive mechanisms aimed at preserving cognitive resources.

Investigating the cognitive mechanisms underlying interpretation of emotion in music

Music is an integral, prevalent part of our culture that we encounter in a variety of situations on a regular basis. Music is an undeniably emotional stimulus; it moves us to tears, inspires us to dance, and commonly accompanies emotional events such as celebrations and religious services. But what is it about music that inspires our experience of emotion?

The goal of this study is to examine how we cognitively process physical features of music and use them to form interpretations of the emotional valence of a musical piece. In order to understand how emotion manifests in music, the present introduction outlines (a) developmental patterns of interpretation of emotion in music (b) the psychological approach to studying music and emotion (c) emotion in music's impact on the listener (d) research findings on the specific features of music that serve as emotional cues. This background information on how emotion operates in music leads to the question of *how* our cognitive system interprets the emotion in musical stimuli.

Developmental Patterns

The ability to identify emotion arises early in development. Infants' first experience with emotion recognition most likely lies in facial expressions of caregivers (Montague & Walker-Andrews, 2001). Infants can usually discriminate between visual displays of different emotions by the second half of the first year (Montague & Walker-Andrews, 2001), and infants can recognize emotion in both visual and auditory stimuli as early as seven months (Montague & Walker-Andrews, 2001). Most of the research on infant recognition of emotion in auditory stimuli examines vocal expression rather than music, such as Fernald (1993) and Spangler, Emlinger, Meinhardt, and Hamm (2001). Tsang and Trainor (2002) found that eight-month-old infants could recognize change in "spectral slope," an aspect of timbre that is important in both speech and music. Timbre is a complex auditory concept (full explication of which is unnecessary here) that allows us to tell the difference between two sounds (e.g., two speakers, singers, or musical instruments) that have the same pitch, volume, and duration (in other words, two equal sounds coming from different entities). Although Tsang and Trainor (2002) did not use musical stimuli, the levels of spectral slope in which infants in their study recognized changes apply to both human voices and musical instruments; thus, their findings imply that infants may be capable of perceiving subtle differences in musical as well as vocal sounds. Examining infant's perception of emotion in music from a different route, Rock, Trainor, and Addison (1999) studied mothers' lullabies and play songs. They found that six-month-old infants reacted quite differently depending on which type of song they heard; lullabies led the babies' attention inward, whereas play songs instigated outward attention and interaction with the caregiver who was singing. These findings imply that songs with different moods/tones communicate different messages to infants as young as six-months-old.

Hargreaves (1986) suggests that children possess many of the skills required for an adult-level mastery of music perception and performance, and it is possible that only limited motor skills and lack of experience hold children back. But do the same limitations apply to children's identification of emotion in music? There has been much research showing that preschoolers are capable of interpreting basic emotions (i.e., happiness, sadness, fear, and anger) in music. In a study comparing the interpretation of emotion in music among preschool-aged children and adults, Nawrot (2003) found that, although preschoolers' interpretations were more variable than those of adults, both groups showed similar response patterns. For example, even if a child's response were incorrect (e.g., choosing "happy" rather than "fearful,") he or she would often supply a verbal label with an emotional valence that matched the correct responses made by adults (e.g., "scary," "monsters," "haunted house"). This suggests that preschool-aged children are as capable of making judgments about emotions in music as adults, but they may lack comprehensive knowledge of the categories adults use to define emotions. Dolgin and Adelson (1990) completed a cross-sectional study with four, seven, and nine-year olds to test recognition of emotion in both vocal and instrumental music. They found that preschoolers (four-year-olds) could recognize the affective valence of melodies with better accuracy in vocal rather than instrumental music and happy rather than "frightened-sounding" music. In fact, all three age groups displayed more accuracy with happy or sad music than frightened or angry music, which seems to be a trend throughout research on music and emotion (Juslin & Laukka, 2003). Dolgin and Adelson (1990) also found a significant increase in accuracy with age, which suggests a steady progression in development of ability to recognize the affective valence of music.

Dalla Bella, Peretz, Rousseau, and Gosselin (2001) examined the development of sensitivity to two specific features of music – tempo (i.e., speed) and mode (i.e., major vs. minor) – and their impact on recognition of emotion in music. They too found response accuracy to increase with age; five-year-olds were significantly sensitive to changes in tempo only, whereas six to eight year-olds were sensitive to changes in both tempo and mode, as were adults, only with higher accuracy. However, Kastner and Crowder (1990) found children as young as three years old to be significantly sensitive to changes in major vs. minor mode. Together, this research suggests children begin to use structural features of music as indexes to emotion in the preschool years.

Another way to examine children's understanding of emotional meaning in music is to study their performance as well as their listening skills. Adachi and Trehub (2000) asked 8-10 year-olds (Experiment 1) and 6-7 year-olds (Experiment 2) to sing a familiar tune (e.g. "Twinkle Twinkle") happily or sadly, then asked children in the same age groups as well as adults to identify which performances were happier or sadder. They found a developmental progression in children's ability to both *provide* and *interpret* emotional cues. In their singing, older performers provided better cues to the intended emotion than younger performers, and 8-10 year-olds were better able to interpret performance cues than 6-7 year-olds. Eight to ten year olds showed accuracy levels similar to the adults.

In sum, it appears that from adolescence onward, we consistently interpret certain emotions in music according to set, predictable patterns. The trend of improvement in recognizing emotion in music as age increases suggests that interpretation of musical emotion is a cognitive skill to be developed with time like any other, such as abstract thinking. This in turn warrants an investigation of the nature of interpretation of emotion in music as a cognitive skill, which is the main goal of the present study.

Emotion in Music: Impact on the Listener

The power of music to convey emotional messages is apparent when that power is put into practice. Relying on the assumption that music is an emotional stimulus, specific emotional messages in music are often employed in order to impact the listener in a desired way. Music can be used to alter the listener's own emotional experience, as in mood induction procedures (Pignatiello, Camp, Elder, & Rasar, 1989; Van der Does, 2002; Vaestfjaell, 2002), or music may be used to inspire listeners to associate simultaneous non-auditory stimuli with certain emotions, as in film (Boltz, 2001; Cohen 2001).

Mood Induction. Although the two terms are sometimes used interchangeably, psychologists typically distinguish between *mood* and *emotion*. *Emotions* are short-term, intense, and elicited by specific objects or events in the environment, while *moods* are general, long lasting, and usually not attributable to any concrete object or event (Vaestfjaell, 2002). When listeners interpret *emotion* in music, they may experience *mood* changes as a result, adopting a mood that matches the emotional valence of the music as they listen; happy music induces positive moods, and sad music induces negative moods. In this way, music's emotive power has been applied to research on emotional states with musical mood induction procedures (MIP), which involve manipulation of participants' mood in controlled settings in order to study mood effects on cognitive processes such as memory encoding and recall, interpretation of affect-laden stimuli, and task success vs. failure, among others (Vaestfjaell, 2002). MIPs are beneficial in research on depression in that they eliminate the need for subjects from a clinical population. In addition, they allow clinicians to identify individuals at risk for developing depressive disorders (Van der Does, 2002).

According to Vaestfjaell (2002), researchers criticize the widely used Velten MIP, in which participants read statements to themselves (such as "I have too many bad things in my life" or "This is great, I really do feel good"), as susceptible to demand characteristics. Music MIP was developed as a more sensory response to the Velten technique that would, in turn, be relatively free of demand characteristics and thus be a more valid procedure. However, Vaestfjaell (2002) points out that all MIPs are susceptible to demand characteristics because the nature of these procedures – asking people to enter a certain mood and then complete some sort of task while in that mood – implies the hypotheses of the research to the participants. This implied knowledge of the researcher's goals can then influence participants' performance and prevent researchers from confidently attributing observed effects to the induced mood state. Both Vaestfjaell (2002) and Mayer, Allen, and Beauregard (1995) believe that Music MIP is no exception to this limitation, while the earlier research of Pignatiello *et al.* (1989) suggests that Music MIP is less vulnerable to demand characteristics as originally intended when the procedure was developed. Westermann, Spies, Stahl, and Hesse (1996) take a middle position, acknowledging the complexity of the demand characteristics debate and proscribing further research on the issue.

According to Westermann *et al.*'s (1996) meta-analysis of MIPs, Music MIP usually involves subjects listening to a "mood-suggestive" musical piece after receiving instructions to attempt to adopt the mood of the music. The procedure may also be employed without suggesting any mood change to the listener, but this method is less common. The experimenter usually chooses the music used in the induction procedure, but in some studies, participants choose a piece of music on their own that they believe will best allow them to adopt a particular mood (Westermann *et al.*, 1996). Westermann *et al.* (1996) found Music MIP to yield mood induction effects comparable to other MIP techniques such as "Imagination," "Velten self-statements," "Feedback," and "Interaction" for both positive and negative moods, which supports Music MIP as a useful technique.

Nonetheless, Vaestfjaell's (2002) meta-analysis suggests that Music MIP is most useful as an enhancement to other techniques. Similarly, Mayer, Allen, and Beauregard (1995) propose that music is most useful in providing a "mood-supportive background state" to point participants in the desired emotional direction while another MIP takes place. They achieved successful mood induction with a combination of Music MIP and guided imagery vignettes. It is possible, however, to successfully induce mood with music alone (Ferraro *et al.*, 2003) or with a combination of music and memory focus (Van der Does, 2002). Although controversial, research using Music MIP demonstrates that emotion in music is potentially strong enough to influence the mood state and subsequent behavior of the listener.

Music in Film. The affective valence of music can have powerful effects on listeners, and the film and theater industries often use music to influence the emotional reactions of audiences. Cohen (2001) points out that emotion characterizes our experience of both film and music. In fact, Cohen lists mood induction (as discussed above) as a vital function of music in film. Filmmakers use music to enhance mood and establish emotional contexts to supplement the dialogue of the film, but, as Cohen (2001) points out, much of this music is now composed with the supposition that the audience will not consciously attend to it. If we paid direct attention to the presence of music in film, particularly in extremely emotional scenes, the music may seem a departure from the reality of the plot. For example, we would perhaps suppose there was an orchestra playing just off screen, even though this makes little sense given the plots of most films. Cohen (2001) suggests "if anything, departures from reality via music make[s] an episode 'more real,' more vivid, more emotionally relevant" (p. 254). Thus music enhances the emotional experience of the moment as we watch the film.

According to Cohen (2001), the affective valence of music can have an extremely salient effect on our interpretation of a film's meaning, directing our attention toward certain objects onscreen and leading us to make certain inferences about those objects. Boltz (2001) expands this concept by suggesting that filmmakers intentionally use music's emotive power to exert certain effects upon the audience, such as establishing context and setting, highlighting a scene's significance and emotional meaning, and, perhaps most importantly, activating schemas that in turn influence our attention to, memory of, and inferences about the film. Boltz (2001) tested this theory by presenting emotionally ambiguous film clips with positive, negative, or no music and asking participants to extrapolate events in the film's plot, as well as fill out a memory test about events in the film clip. Boltz (2001) observed that extrapolations made in the music conditions were limited to interpretations that matched the affective valence of the music. For example, participants were far more likely to believe that one character would eventually harm another if they heard negative music while watching the film. On the memory test, participants recognized more negative words if they heard negative music, and vice versa for positive words and music. Boltz (2001) concluded that manipulating musical affect is in itself enough to systematically bias viewers' cognitive processing of films, and this tends to happen in a mood congruent fashion.

Given that music can impact the audience's thought processes, Cohen (2001) proposes that the composition of music for films is a type of problem solving – composers (in connection with directors, actors, etc.) must know how to manipulate music in order to communicate specific affective messages to match specific scenes in the film in question. What, then, are the 'raw materials' of emotional expression in music; what characteristics can be manipulated in order to achieve the desired emotional effect? The answer, it seems, lies in psychological research on the issue.

Psychological Study of Music and Emotion

The study of music and emotion, which is in a sense a combination of the fields of music and psychology, is rarely addressed in mainstream psychological research, but some psychologists, such as Juslin and Sloboda, have devoted their careers to the study of music and emotion. Juslin and Sloboda (2001) explain the psychological (as opposed to musicological) approach to the study of emotion in music. Psychologists attempt to explain the "how and why" of emotional experience/reaction to music – how and why do we perceive music as expressive of emotion, and how and why does music influence our own emotional experience? Furthermore, these investigations seek to understand the mechanisms operating between the auditory perception of music (i.e., when the sound waves hit the ears) and the emotion a person detects or experiences as a result of hearing the music; this approach is the focus of the present study. In attempts to answer these questions, psychologists turn to research experiments. Experimental designs are fitting because psychologists consider emotional responses to music "material phenomena," the components of which correlate with "observable phenomena in law-governed ways" (Juslin & Sloboda, 2001, p.72). In other words, Juslin and Sloboda (2001) believe our emotional responses to music are observable events that can be traced, in turn, to observable causes according to predictable relationships, thus calling for experimental paradigms that investigate those relationships through variable manipulation. Finally, because research in this context focuses on psychological rather than biological explanations, Juslin and Sloboda (2001) believe research on musical emotion should attempt to explain how these emotions "serve to fulfill psychological goals and outcomes" (p.73) as does most other research on emotion. They urge, then, that we study emotions in music as we would study emotion in other contexts.

Juslin and Sloboda (2001) also elucidate the type of emotional reactions in which psychologists are interested. They propose two types of observable emotion in music: aesthetic and expressed. Aesthetic emotion in this context concerns reactions to music as an art; for example, we may feel positive emotion in appreciation of a musical piece's beauty, or we may feel disgust if we find the piece to be poorly written or performed. Aesthetic emotion, although potentially very powerful, is rarely examined in psychological research. Instead, psychologists tend to focus on the emotions expressed (emotions listeners interpret) or induced (emotions listeners experience themselves) by music as an *auditory stimulus* and not a work of art. This second type of emotion as found in music is studied in experimental research. Juslin and Sloboda, (2001) however, voice the concern that aesthetic and expressed emotions are not necessarily wholly independent of one another, and they suggest researchers take this possibility into account.

Juslin and Sloboda (2001) identify several drawbacks and advantages to studying emotion in music as opposed to other emotional stimuli. One problem, they suggest, is that it is difficult to connect emotional reactions to music with adaptation and biological survival. This prevents explanation of emotion in music within the existing theories of emotion, as most models conceptualize emotional reactions to the environment as adaptive and related to survival. Individual differences in reactions to music, both across time and across individuals, also complicate research on emotional reactions to music, making it difficult to determine whether or not variance may be attributed to experimental manipulations. Third, Juslin and Sloboda (2001) suggest that asking listeners to report about their emotional reactions to music may impact their listening experience to the point of preventing psychologists from concluding that the emotional reaction was indeed prompted by the music itself. This problem is similar to the demand characteristics limitation of mood induction procedures discussed earlier.

Despite these limitations, Juslin and Sloboda (2001) posit that music has the potential to be an ecologically valid stimulus to use in emotion research because people are accustomed to making judgments about and responding emotionally to music. Compared with other tasks in emotion research, such as tests of memory or attention, listening to music is a more familiar, "real-life" task. Research on music also avoids ethical concerns associated with other stimuli

(Juslin & Sloboda, 2001), such as visual arrays depicting disturbing events (e.g., violent scenes) or stimuli inducing depression (e.g., sad vignettes). In addition, musical stimuli hold the advantage of a clear, well-known structure, and knowledge of the structure of an emotional stimulus may better allow researchers to draw conclusions about the emotional reactions it inspires (Juslin & Sloboda, 2001).

Juslin and Sloboda (2001), both of whom have contributed enormously to psychological research on emotion in music, provide a comprehensive overview of the nature of the topic at hand and the ways psychologists must approach it in order to contribute meaningful and effective research on music and emotion. At this point, we may begin examining psychological studies addressing the well-known structure of music, which Juslin and Sloboda (2001) deemed a vital advantage to research on music and emotion.

Emotional Cues in Music

Research on mood induction, music in media, and musical development reveals that the affective power of music is strong enough to influence both our mood and our behavior. From here, it is useful to investigate the local characteristics of music that allow this influence to take place. Krumhansl (2002) argues that music may have inherent emotional meaning because a large amount of research has shown similar patterns of emotional interpretation of music across listeners. She also argues that there appears to be some connection between the cognition of musical structure and emotion. In attempts to account for this phenomenon, Balkwill and Thompson (1999) offer a model of perception of emotion in music in which listeners depend upon a combination of cultural and "universal" or "psychophysical" cues. Cultural cues include distinct features or conventional styles of music that are relatively unique to a certain area of the world; for example, most American and Western European music uses a different tonal system than African and East Asian music. Balkwill and Thompson (1999) define "psychophysical" cues as "any property of sound that can be perceived independent of musical experience, knowledge, or enculturation" (p. 44). These cues are "universal," then, because people from any culture can recognize them and use them to interpret the emotional valence of a musical stimulus. In their model, Balkwill and Thompson (1999) propose that both listeners and composers use a combination of these two types of cues when attempting to communicate or perceive emotional messages in music, and the more cues present, the stronger the emotion communicated or perceived. Furthermore, the model suggests that listeners attend to cultural cues *first* if they are familiar; if the cultural cues are unfamiliar, listeners give more attention to psychophysical cues (e.g., tempo) in order to understand the music's emotional message.

Balkwill and Thompson (1999) note that despite a preferred reliance on cultural cues, listeners are "highly adaptable" to fundamentally unfamiliar music, which means the structural features of music that serve as 'universal' cues to emotion are extremely important to our perception of emotion in music. Juslin and Sloboda (2001) posit that there is "an accumulating body of knowledge that shows that there is a lawful relationship between the intensity of emotional qualities experienced in music and the specific structural characteristics of the music at a particular point in time" (p.91). Indeed, much of the research addressing music and emotion has been devoted to investigating the physical aspects of music and how they contribute to the perception of particular emotions.

In a study comparing the communication of emotion in music and vocal expression, Juslin and Laukka (2003) provide the most recent meta-analysis of research studies exploring the relationships between structural features of music and listeners' perceptions of emotion. They include a comprehensive review of 41 psychological studies addressing relationships between specific structural features of music and specific emotions, mapping out every musical cue studied and providing summaries of the observed emotional effects associated with each cue. Gabrielsson and Lindstrom (2001) assembled a similar although less systematic meta-analysis. I will describe the combined findings of these two meta-analyses here. Specifically, I will define several physical aspects of music, explain how they are measured, and outline their relationships to emotional interpretations as supported by research included in the two meta-analyses. All associations between the physical aspects of music and emotional messages (e.g., fast tempo is associated with happiness) have been observed by a number of studies included in the meta-analyses, all of which used experimental designs in which the physical aspects of music were manipulated and listeners' emotional responses were recorded.

Pitch. Juslin and Laukka (2003) record three dimensions of pitch. They define the first, *F0,* as "the lowest periodic cycle component of the acoustic waveform" (p.790). In other words, *F0* is the baseline frequency for a musical note played *in tune*, and the pitch may deviate above *F0* to become sharp or deviate below *F0* to become flat. Thus we may examine the *macro pitch level* of the music, meaning the pitches prescribed by the written music and measured in semitones, and "the micro intonation of the performance" (p. 790), measured in deviations from the macro pitch notated in the musical score. High macro pitch levels are associated with happiness, anger, and fear (Juslin & Laukka, 2003), as well as excitement, and serenity (Gabrielsson & Lindstrom, 2001). Low macro pitch levels are associated with sadness, tenderness (Juslin & Laukka, 2003), solemnity, and vigor (Gabrielsson & Lindstrom, 2001). Sharp micro-intonation is associated with anger and happiness, while flat micro-intonation is associated with sadness; however, fewer studies addressed micro-intonation effects than macro-intonation effects. The second dimension, *F0 contour,* is "the sequence of *F0* values" (Juslin & Laukka, 2003, p.790) as written in the music – intonation thus refers to the way the performer maintains (or deviates from) the prescribed pitches of notes in the score. Gabrielsson and Lindstrom (2001) conceptualize this as "pitch variation" (p. 240). If the *F0 contour* moves upward, it is associated with anger and fear, and if it moves downward, it is associated with sadness and tenderness. Lastly, v*ibrato* is an intermittent change in pitch or loudness of a tone, which can be measured manually with an *F0* trace of the acoustic waveform. Although Juslin and Laukka (2003) included it in their meta-analysis, vibrato yielded inconsistent associations with emotions.

Intensity. Intensity, or "loudness" represents the energy of an acoustic signal and is measured according to the amplitude of the acoustic waveform. Louder musical pieces elicit anger (Juslin & Laukka, 2003), intensity, or joy (Gabrielsson & Lindstrom, 2001), whereas softer music elicits sadness or tenderness. *Attack* refers to the speed with which a performer increases the intensity of each individual note, and it is measured as the rate at which the amplitude of the note rises (Juslin and Laukka, 2003). Fast tone attacks are associated with anger and happiness, and slow tone attacks are associated with fear, sadness, and tenderness (Juslin & Laukka 2003). Additionally, Gabriellson and Lindstrom (2001) report that large variations in intensity or loudness can communicate fear, while small variations in loudness may suggest happiness.

Temporal aspects. Juslin and Laukka (2003) address three temporal aspects of music: *Tempo*, *articulation*, and *timing*. *Tempo,* meaning the speed or velocity of the music, can vary within a musical piece, but the mean tempo is calculated by "dividing the total duration of the performance . . . by the number of beats," which yields a number of "beats per minute" (Juslin

& Laukka, 2003, p. 791). According to Gabriellson and Lindstrom (2001), music psychologists tend to consider tempo the most important structural feature of music that affects listeners' emotional interpretations. Fast tempos convey fear, anger, happiness (Juslin & Laukka, 2003), excitement, potency, and surprise (Gabrielsson & Lindstrom, 2001), while slow tempos convey sadness, tenderness (Juslin & Laukka, 2003), serenity, solemnity, and disgust (Gabriellson & Lindstrom, 2001). In Juslin and Laukka (2003), *articulation,* or "proportion of sound to silence in successive notes" (p. 791), is measured by a ratio of two durations: (1) from one tone's onset to the next tone's onset (d_{ii}) (2) from one tone's onset to it's offset (d_{io}). The d_{ii}:d_{io} ratios for each note in a piece are then averaged together and reported as a percentage. 100% denotes to *legato* articulation, which is associated with sadness and tenderness, and 70% or lower denotes *staccato* articulation, which is associated with anger, fear, happiness (Juslin and Laukka, 2003), energy, and gaiety (Gabriellson & Linstrom, 2001). *Timing,* or variation from tempo and/or rhythm, is described as deviation from the piece's musical notation (as is intonation.) *Timing* is measured in two ways. First, an overall measure is "obtained by calculating the number of notes whose deviation is less than a given percentage of the note value" (Juslin & Laukka, 2003, p.791). In this case, the researcher selects the "given percentage" as a standard for comparison. Or, one may contrast durations between short vs. long notes; a "sharp" durational contrast represents increased variation of durations between notes as compared to the music notation, and a "soft" durational contrast represents a smaller ratio of variation. Musical pieces with the most timing variability convey fear or anger, and happiness is associated with pieces with the least timing variability (Juslin & Laukka, 2003).

Timbre. Timbre, which allows listeners to distinguish between two people or instruments singing or playing the same pitch at the same volume simultaneously (Tsang & Trainor, 2002), is a bit harder to define than the other musical cues accounted for in Juslin and Laukka (2003); however, they do mention two dimensions of timbre. *High-frequency energy* is "the relative proportion of total acoustic energy above versus below a certain cut-off frequency" (Juslin & Laukka, 2003, p. 791), and this cut-off is different for specific instruments. High levels of high-frequency energy convey anger, medium levels convey happiness, and low levels convey fear, sadness, or tenderness. Despite these measurements, Juslin and Laukka (2003) provide no information about which emotions are associated with specific timbres in music. Gabriellson and Lindstrom (2001) conceptualize timbre slightly differently. They report that musical tones with "amplified higher harmonics" can communicate anger, tones with "few, low harmonics" can suggest boredom, happiness, or sadness, and tones with "suppressed high harmonics" can suggest sadness and tenderness (Gabrielsson & Lindstrom, 2001, p. 241).

With extensive research suggesting stable associations between the physical features of music and the listener's emotional experience, the question remains: how do we go about processing those features in order to make judgments about the emotional valence of the music we hear?

Cognitive Processing of Emotional Cues

With a working knowledge of which musical cues are most likely to elicit certain emotions, we may go on to investigate the cognitive mechanisms behind our interpretation of these cues. A fundamental question therein concerns the nature of music processing: is there a cognitive framework for our comprehension of emotion in music, and if so, what are its components?

There is a large body of research addressing similarities between language and music; in fact, Juslin and Laukka (2003) include a meta-analysis of studies on structural cues to emotion in vocal expression as well as studies on structural cues in music. They, as well as Juslin and Sloboda (2001), speculate that language and music may have common origins, or that music may have "evolved" or developed out of language tone and prosody. Language research has yielded evidence supporting the presence of cue-specific cognitive mechanisms devoted to processing distinct components of language, such as syntax and phonemes (Forster, 1979). Could the same be true of music processing? Besson and Friederici (1998) argue that music and speech are composed of "the same raw material, that is, wide pitch bands combined in spectral space (e.g., timbre or phonemes)" (p. 4). If this is true, there is reason to believe that specialized mechanisms exist for music much in the same way they exist for language.

There are several possible models that may count account for our processing of emotional cues in music. First, there may be separate mechanisms that address each emotional cue independently; for example, there may be a processor for tempo, one for pitch, one for timbre, etc. Such cue-specific mechanisms are probable in light of the research compiled on the structural features of music and emotion; since each feature/cue has a unique relationship to our emotional interpretations, it is likely that we process these cues independently at some level. In such a model, several processors, each addressing only one musical cue, could all contribute separately, additively, and equally (if the cue is present) to our overall judgment of the emotional valence of the music. The more cues present for a particular emotion, the easier it would be to interpret that emotion, as Balkwill and Thompson (1999) suggest.

An alternate explanation is that there may be more complex relationships operating among musical cues to emotion. Certain cues may be more important or more useful to our emotional judgments than others, implying a hierarchy of processors, each still devoted to only one cue, which we work through serially when we listen to music. In other words, there may be specific cues that we process first because they serve as better indicators of emotion than other cues. We may then continue down the line, so to speak, referring to less useful cues until we glean enough information in order to draw a conclusion about the music's emotional valence. A hierarchy of cue usefulness is implied in Juslin and Laukka's (2003) meta-analysis, which reveals that some cues, such as tempo and sound level, show more robust effects on interpretation of emotions than others, such as F0 contour and attack.

Another possibility is that within this hierarchy, unique associations exist between certain cues but not others; for example, tempo and pitch may be associated, while tempo and rhythm may not. This would be implied if certain combinations of cues prove to be more powerful indicators of emotion than other combinations. If certain cues are associated (e.g., tempo and pitch), we may process them together with a single mechanism, and the simultaneous presence of these associated cues may have a different effect on our overall judgment of the music's emotional valence than if either cue were presented alone. This differential effect may be more than a simple increase in strength of the emotional signal gleaned from the cue; simultaneous presentation of associated cues may yield a different relationship between the cues and the resulting emotion than individual presentation.

It is possible that no such cue-specific mechanisms exist in the cognitive system; our interpretation of emotion in music may not be as systematic as these previous models suggest, and emotional cues in music may not be psychologically unique. However, the wealth of research on differential effects associated with specific cues suggests that some sort of cue-specific processing is operating when we interpret emotions in music.

The Present Study

This study will attempt to elucidate the cognitive framework underlying the way we process emotion in musical stimuli. Previous research has shown existing associations between certain physical features of music and certain emotions, but what cognitive mechanisms are operating when we process those cues and access those associations? Two other studies have used similar paradigms (Dalla Bella *et al.*, 2000; Peretz, Gagnon & Bouchard, 1998), finding significant main effects for tempo and mode (i.e., major vs. minor), but neither drew conclusions about the cognitive mechanisms underlying those effects.

I will look at different paired combinations as well as individual effects of three cues – tempo, pitch, and timbre – on listeners' interpretations of emotion in short musical pieces. Individual main effects of each cue as compared to the combinations of two or more cues will provide insight into the cognitive processes behind judgments about emotion based on these cues. For example, an additive effect, or no change in the nature of a cue's effect when considered in combination with another cue, would imply independent processing of each cue. An interaction effect, however, as in combinations of two or more cues revealing different relationships between cues and emotion, would imply that possible shared mechanisms and/or associations between cues are at work. If there are clear quantitative differences between the strength or extremity of each cue's effect on emotional interpretations, this will imply a cue hierarchy, with the cues having the strongest effect on emotion ratings at the highest level.

It is predicted that each musical cue (independent conditions) will yield an interpretation of emotion (happy vs. sad) in accordance with the previous research discussed above: faster tempo and higher pitch will elicit interpretations of happiness, and slower tempo and lower pitch will elicit interpretations of sadness. In addition, it is hypothesized that the data will reflect a hierarchical framework of musical cue processors; specifically, based on previous data outlined in Juslin and Laukka (2003), tempo is expected to be the strongest indicator of emotion (yielding the most extreme mean emotional interpretations), followed by pitch. No specific hypotheses were formed concerning associations between cues or timbre. No previous research has been conducted on the emotional valence of the specific instruments used in this study, both of which were electronically designed; this study is largely exploratory in this respect.

Method

Participants

Thirty-nine undergraduate college students, nine males and 30 females, were recruited through flier advertisements and visits to introductory psychology classes. Ages ranged from 18-22, with a mean age of 19.03. No participants reported a major hearing disability.

Design

This study was completed with a 2 x 2 x 3 x 3 within subjects design. Each condition was manipulated to include one of two melodies, one of two instruments or timbres (Plucky or Woody), one of three levels of tempo (slow, moderate, or fast), and one of three levels of pitch (low, medium, or high). This yielded a total of 36 conditions, which were presented to participants in three fixed random orders in order to counterbalance any effects the order of presentation may have had on participant responses. The only dependent variable in this experiment was participants' emotional interpretations, as measured by a forced choice self-report.

Stimuli and Apparatus

With input from musical experts (all of whom were members of the Hamilton College faculty), brief melodies were selected from collections of American folk music. Extremely familiar melodies, especially those with widely known lyrics, were avoided in order to prevent previous affective associations from impacting participant responses. Two melodies with a fairly neutral emotional valence were selected: "In the Good Old Colony Days" (composer unknown) and "Bicycle Built for Two" by Harry Dacre (Forcucci, 1984). Thus, each melody could be potentially manipulated to express both negative and positive emotions. Musical notations for each musical piece appear in Figures 1 and 2.

Based on these two pieces, musical stimuli were developed on both Macintosh and Windows computers with Csound, a program used in virtual music composition and performance. Csound allowed precise control of all structural features of the music note by note, including timbre, tempo, note length/duration, and attack, as well as the amplitude, frequency, and oscillation of the auditory waveform. Csound also allowed consistent manipulations within each musical cue (tempo, pitch, and timbre).

As shown in Table 1, tempo and pitch were varied on three levels at constant intervals. Csound's default tempo of 60 beats per minute was used as the middle tempo because it kept the melody at a moderate pace; halving and then doubling this tempo in turn fit best with the melody, as any slower may have been quite tedious and any faster may not have allowed the participants enough time to respond. As discussed earlier, three different dimensions of pitch exist in music, but in this study only macro-pitch was manipulated (in other words, the musical notation was changed between conditions rather than the in-tuneness or vibrato of the notes.) The three pitch octaves chosen worked well within the range of the two electronic instruments used in the study; any higher or lower may have been out of the instruments' range and difficult to hear. Timbre was varied with two computer-generated instruments developed in Csound: "Plucky," which approximates the sound of a plucked violin, and "Woody," which approximates the sound of a clarinet. The manipulations for timbre were extremely complex; the definitions for these two instruments may be found in the full text files used to create each condition with Csound (see the Appendix). Each condition was created with both melodies and both instruments, yielding a total of 36 conditions. All excerpts were converted into mp3 files and recorded as individual tracks on a compact disc. Each participant heard all 36 conditions through stereo headphones on a CD player.

Procedure

Participants were seated at a desk in a private room and asked to read the following instructions:

> This is a study about music and emotion. You are about to hear 36 short musical excerpts in succession, each of which appears as a separate track on the CD player display. This packet is where you will record your responses to each excerpt. Please record your *immediate* impression of the emotional tone of each *individual* track. This involves deciding the overall emotional tone of the music (happy or sad) and the strength or intensity of that emotion. At the conclusion of each track, first circle either 'HAPPY' or 'SAD' (you must choose one). Then, on the one to ten point scale provided, record the strength of the emotion you selected. For example, if the melody sounded *extremely* happy, you would circle 'HAPPY' and circle a ten on the scale. On the other hand, if the melody sounded only *slightly* sad, you would circle 'SAD' and circle a one on the scale. Please respond as quickly as possible with your *immediate* interpretation of each melody's emotional tone.

The example scale to which the instructions refer appears in Figure 3.

Next, the experimenter emphasized verbally the importance of recording a response as soon as the participant felt that he or she had heard enough of the musical excerpt in order to make a decision about its emotional tone. Participants were then instructed to adjust a set of stereo headphones to fit comfortably. After addressing the participant's questions (if any), the experimenter pressed "play" on the stereo CD player to begin the first melody and left the room. Participants then listened to the tracks in one of three fixed random orders. If participants had already recorded a response to a track after 15 seconds, they were allowed to skip ahead to the next track as often as they chose; they were also allowed to return and re-listen to a track if it went by too quickly for them to respond. Once participants finished, the experimenter outlined verbally the purpose and hypotheses of the experiment.

Results

For each participant, the mean ratings were computed by averaging across melody. Means were then converted to a one to 20 pt scale, one being saddest (and equal to "one" on the original ten point scale when participants circled "sad") and 20 being happiest (and equal to "ten" on the original ten point scale when participants circled "happy"). Thus means up to nine denote sadness, means from 10-11 suggest a neutral tone, and means from 12 to 20 denote happiness. To determine whether the manipulated cues yielded independent or interactive effects, mean emotion ratings were then analyzed with a 3 x 3 x 2 (tempo x pitch x timbre) repeated measures analysis of variance. As seen in Table 2, mean emotion ratings of musical excerpts significantly increased (became happier) as tempo increased, $F\ (1, 38) = 680.31, p < .001$, and as pitch increased, $F\ (1, 38) = 23.24, p < .001$, all of which is consistent with hypotheses. In addition, participants rated musical excerpts recorded with the instrument "Woody" as slightly sadder than melodies recorded with "Plucky," and this difference in effect of timbre was significant, $F\ (1, 38) = 49.30, p < .001$. Among the means for each emotional cue that appear in Table 2, the larger the difference between manipulated levels (e.g., slow, medium, and fast), the more extreme the response and the stronger the effect of that cue on listener responses. Given that tempo yielded the largest differences in mean emotion ratings between slow, medium and fast, tempo appears to have had the strongest effect on listeners' emotional ratings of the music as predicted. However, contrary to predictions, pitch's effect appears to be weaker than that of timbre, as the mean response differences between pitch levels, although significant, were slightly smaller than the difference between Plucky and Woody.

A significant two-way interaction was found between timbre and tempo, $F(1, 38) = 9.236$, $p < .001$. As seen in Figure 4, listener ratings for Plucky remained happier than ratings for Woody at all three tempos, but Plucky's effect was greatest when the tempo was 60 beats per minute, the medium/moderate tempo level in this experiment. A significant two-way interaction was also observed between tempo and pitch, $F(1, 38) = 3.412$, $p < .05$. The means as they appear in Figure 5 suggest that at the medium/moderate tempo of 60 beats per minute, the effect of the lowest pitch octave, which elsewhere communicated a sadder emotional tone than the middle pitch octave, nearly failed to do so. In addition, at a tempo of 30 beats per minute, the mean for the highest pitch octave was much higher than the means for the two lower octaves, suggesting that the high pitch octave's predicted association with happiness was greatest at the slowest tempo.

A significant three-way interaction was found between timbre, tempo, and pitch, $F(1,38) = 5.084$, $p < .001$. All mean patterns, as seen in Figure 6, follow predicted directions with two exceptions. First, at a tempo of 60 beats per minute, the mean rating for Plucky at the lowest pitch octave was happier than that of the middle pitch octave, which is contrary to the hypothesis stating that a *higher* pitch will lead to a happier emotional rating. Second, at a tempo of 30 beats per minute, the mean rating for Woody at the lowest pitch level was happier than the mean rating for the middle pitch level, which is contrary to predictions for the same reasons as the previous exception discussed above. No other observed interaction effects were significant.

Discussion

Given the large body of research establishing reliable associations between certain physical aspects of music, or *cues,* and emotional interpretations of listeners, the present study sought to investigate the nature of the cognitive processes by which listeners use these cues to make judgments about the emotional valence of a musical piece. Specifically, tempo, pitch and timbre were manipulated, and different levels of each of these three cues were presented to listeners in different combinations in order to determine how the cues influenced emotional interpretations of music, both individually and in combination. The observed effects of tempo and pitch on emotion ratings of music were consistent with hypotheses; slower tempo and lower pitch led to sadder emotional interpretations of the music, while faster tempo and higher pitch led to happier emotional ratings. These findings were also consistent with previously conducted meta-analyses of research on specific emotional associations with the structural features of music (Gabrielsson & Lindstrom, 2001; Juslin & Laukka, 2003). In addition, results suggested that the computer-designed instrument Plucky, which imitates the sound of a plucked violin, has a slightly stronger association with happiness than Woody, a computer-designed instrument emulating the sound of a clarinet. As no previous research of this nature has been conducted with these two electronic instruments, no hypotheses were formed regarding their associations with emotion, but this result is intriguing and should perhaps be taken into consideration in performances involving ‘real’ violins and clarinets.

That each manipulated cue – tempo, pitch, and timbre – had an independent significant main effect on listeners' emotional ratings of the musical stimuli suggests that individual, cue-specific mechanisms exist in the cognitive system devoted to interpreting emotion in music. In listening to music, our cognitive system seems capable of singling out specific cues and using them in different ways when interpreting emotion. To determine the relationships between these cue-specific mechanisms and the shape of the full cognitive model, we must look to the relative strength or salience of each cue, as well the statistical interactions between cues.

As predicted, tempo appears to be the most salient cue to emotion in music, in light of both present findings and previous research (Gabrielsson & Lindstrom, 2001; Scherer & Oshinsky, 1977). Mean emotion ratings in this experiment fell between two extremities, namely happy vs. sad. The more 'space' existing between mean emotion ratings for different manipulated levels, the more *extreme* the emotional interpretations. Another way to conceptualize a cue's strength is by proximity of mean emotion ratings to the extremes of the 20-point scale. Tempo's effect on emotion ratings appears strongest in that the differences between mean ratings for each manipulated level – slow, moderate, and fast – were much larger than differences between mean ratings for each pitch octave or each instrument/timbre. Hence the mean emotional ratings based on tempo alone are much more extreme than ratings based on pitch or timbre alone. Tempo's presence seems to loom over the entire experiment, as tempo was the only variable to interact with all other variables present. Gabrielsson and Lindstrom (2001) acknowledge that music psychologists tend to consider tempo the most important structural cue to emotion in music. It is also, they point out, one of the most important factors for composers, who nearly always denote the intended tempo of a piece in the written score. Developmental research also supports tempo as one of the most important structural cues to emotion in music; tempo is one of the first features of music that children are able to use as an index of emotion (Adachi and Trehub, 1998; Dalla Bella *et al.*, 2001). Thus, clearly tempo is at the forefront of listeners' emotional interpretations of music.

Nevertheless, Gabrielsson and Lindstrom (2001) caution that while tempo is arguably the strongest cue to emotion, other strong factors may override its effect, such as mode (i.e., major vs. minor). In the Timbre x Tempo interaction, the mean emotional ratings for the slow and fast tempos were quite similar between the two timbres; means for Plucky in these conditions were approximately one point higher (happier) than those for Woody, reflecting Plucky's slightly happier valence as compared to Woody. But the difference in emotional ratings between timbres when the tempo was moderate was much larger – here Plucky's rating was almost three points higher (happier) than that of Woody.

The moderate tempo of 60 beats per minute was expected to communicate a fairly neutral emotional tone, as it lies at the exact middle between the slow and fast tempos, both of which were associated with extreme emotional ratings of sadness and happiness respectively. All previous research on tempo seems to elucidate emotional associations with *only* slow or fast tempos (Gabrielsson & Lindstrom, 2001; Juslin & Laukka, 2003). The mean rating for the moderate tempo in this experiment hovered around the neutral rating at 10.94. All of this suggests that moderate tempos tend to communicate neutral emotional messages. In the Tempo x Timbre interaction, timbre seemed to have the most effect on emotion ratings when the tempo communicated a neutral emotional tone. This suggests that participants used timbre as a 'back-up' cue when tempo was insufficient in communicating the music's emotional valence. It seems that in order to preserve cognitive resources, we first turn to tempo, the most salient cue available. If tempo fails to provide adequate emotional information, we move on to access timbre's emotional message. This pattern supports a hierarchal model of emotional cues in music, tempo being highest because we access it first and most often, and timbre being just beneath it.

But where does pitch fit into the hierarchy? Despite its significant main effect, pitch appears to be a "weakest link" of sorts, having the smallest impact on participants' emotional interpretations of the music and failing to interact with timbre. In its interaction with tempo, we could expect pitch to fill a role similar to timbre, namely serving as a 'back-up' cue when tempo is neutral. However, in the Tempo x Pitch interaction, when tempo was neutral, only the highest pitch condition affected emotional ratings in the expected direction; the mean rating for the lowest pitch octave was barely different from that of the middle pitch octave. This may be because in this experiment, pitch lay on the same type of continuum as tempo, with the middle octave communicating a neutral emotion relative to the high and low octaves. Thus the nature of pitch differences and the emotional associations therein perhaps weakened pitch's salience as an emotional cue in supplement to tempo. Overall, results suggest that pitch is below both tempo and timbre in the cue hierarchy. Pitch's effect on emotional ratings of music was nowhere near as strong as that of tempo, and pitch wasn't as reliable a cue to emotion as timbre. Participants seemed to use pitch as a 'last resort' when forming emotional judgments of the music, preferring to look to other, more salient cues to emotion and avoid wasting cognitive resources on the least fruitful source of emotional information. Collectively, the results suggest that pitch holds a low placement in the cognitive hierarchy relative to timbre and tempo.

The three-way interaction between tempo, timbre, and pitch is a testament to Gabrielsson and Lindstrom's (2001) insistence that all structural features of music interact with one another in influencing our perception of emotion. As seen in Figure 6, the pattern observed in the Timbre x Tempo interaction reappeared here. When tempo was slow or fast, there was a slight difference in emotional ratings between Plucky and Woody, with Plucky yielding slightly happier ratings. When tempo was moderate/neutral, the difference in emotion ratings for Plucky vs. Woody was much larger. These patterns provide further support for the notion of timbre as a reliable 'back-up' cue, which we access for emotional information when tempo is insufficient. This also further supports timbre's place just below tempo in the cue hierarchy. Pitch's pattern of limited usefulness remerged in the three-way interaction. Pitch was the only cue that seemed to disrupt the response patterns observed elsewhere; in two cases, the lowest pitch actually led to a happier emotional rating than the medium pitch after tempo and timbre were taken into account. No such anomalies occurred in the Timbre x Tempo interaction. Again, this suggests that pitch was the least reliable source of emotional information in this experiment; pitch's weakness in interactions with other cues is in harmony with its weakness as a main effect relative to tempo and timbre, all of which again supports pitch's place at the bottom of the cognitive hierarchy.

A summation of the proposed cognitive model appears in Figure 7. The hierarchal pattern that emerged in the way participants used the three manipulated cues is similar to the pattern Balkwill and Thompson (1999) observed in their model of perception of emotion in music. Both models reflect the cognitive system's attempts to preserve resources; we look to certain preferred cues first, and if these cues are absent or insufficient, we move on to a different type of cue in order to interpret the music's emotional message. In Balkwill and Thompson's

(1999) model, listeners look to familiar cultural cues first, followed by structural features of music if cultural cues are absent. Similarly, in the present model, we look to the most reliable cue first (tempo), followed by less reliable cues (timbre and pitch) if tempo sends an unclear message. Thus we observed an analogous pattern between structural cues alone, as opposed to cultural vs. structural cues. Since the present study only measured what Balkwill and Thompson (1999) call "universal" cues, and since all the musical excerpts used in this study contained identical familiar cultural cues, the two models perhaps compliment one another.

There may be a way to interpret the present study's findings in terms of cue associations rather than a hierarchal model. Pitch's failure to supplement tempo in the same manner as did timbre may leave us with evidence for a unique association between tempo and timbre. Our cognitive system seems to use tempo and timbre together in a relationship that did not repeat itself elsewhere in the findings; timbre's apparent status as "back-up" for timbre in this experiment seems exclusive. However, this unique association is limited to one specific condition, namely the moderate/neutral tempo played with the instrument Plucky (see Figure 4). Furthermore, this apparent relationship may simply be a function of the limits of the paradigm; only three cues were manipulated, which leaves little room for such associations to emerge. With these limitations in mind, the hierarchal model of cue-specific mechanisms remains the most plausible explanation for the observed patterns.

The question remains – why is tempo so important? Why do we rely on tempo so heavily when interpreting music's emotional message? In the context of the present study, we may simply be more accustomed to attending to the *timing* of stimuli in our environment rather than pitch or timbre. The latter two cues are specific to auditory stimuli, but tempo varies in a wider variety of sensations. In addition to sound, we experience visual and tactile stimuli in motion at different speeds or tempos; thus our environment requires us to attend to timing much more often and on more levels than pitch or timbre. As a result, our cognitive systems are perhaps especially attuned to timing, making tempo an especially salient and useful cue to emotion in music.

Limitations and Future Directions

There are many variables that may affect the way we interpret emotion in music that were not addressed in the present study. First, there may be differences in the ways musicians vs. non-musicians process the structural features of music as cues to emotion. Unsurprisingly, musicians have been found to respond differently to music than non-musicians, with musicians tending to show better awareness and understanding of different aspects of musical stimuli (Halpern, Bartlett, & Dowling, 1998; Wolpert, 2000). These differences seem to exist at all levels of processing. Koelsch, Schroeger and Tervaniemi (1999) used EEG measurements and found musicians were better able than non-musicians to process music at the pre-attentive level; they attributed this superior processing to long-term training in music performance. In light of a musician's probable enhanced awareness of the structural features of music, it may be beneficial in the future to record participants' musical training/ability (or lack thereof) and compare systematically the results of listeners with different levels of music instruction and/or performance experience.

The present experiment only captured a small portion of the vast array of structural features that make up an acoustic waveform, and the cognitive processing of other cues such as articulation, mode, and loudness has yet to be explored. Systematic manipulation of the other structural features of music outlined by Gabrielsson and Lindstrom (2001) and Juslin and Laukka (2003) should be repeated with the same paradigm used in the present study in efforts to explore (a) the presence of cue-specific processors (b) their placement in the cognitive hierarchy of cue processors (c) further evidence for or against the hierarchical model proposed here. Investigation of a variety of structural features of music will most likely reveal complex interactions between cues and perhaps change the nature of the proposed model.

The present study was limited not only in terms of the structural features of music manipulated, but also in terms of the emotions measured. The paradigm only targeted a small portion of the emotional spectrum with a forced choice design, which simplified emotional responses to happy vs. sad. Structural features of music have been found to suggest any number of emotions, such as surprise, boredom, agitation (Gabrielsson & Lindstrom, 2001), contempt, anxiety, or pride (Juslin & Laukka, 2003). A problem lies in the subtle nuance of emotion as it increases in context specificity; the current study limited emotional ratings to a positive/negative continuum in order to investigate the overall cognitive system operating in emotional interpretations of music. Extreme emotions were expected to have robust effects and hence provide insight into the cognitive mechanisms underlying those effects. Once further support for the hierarchical model has been established and additional cues have been placed within that system, more specific types of positive vs. negative emotion should be explored with the present paradigm. Since the structural cues of music tend to interact with one another and suggest several specific types of emotion depending on context, the relationships between cues and emotions will likely become more complex as more cues and emotions are tested.

Developmental studies may provide further support for the cognitive hierarchy underlying interpretation of emotion in music. As previously discussed, when children interpret the emotional valence of music (when listening or performing) they use those cues that Juslin and Laukka (2003) found to consistently produce *robust* relationships with certain emotions, such as tempo, mode, and loudness (Dalla Bella *et al.*, 2001, Kastner & Crowder, 1990). These findings suggest that cue-specific cognitive mechanisms existing at the top of the hierarchy and tapping into robust, reliable indexes of emotion may develop prior to processors for less reliable cues located lower on the hierarchy. Cross-sectional research would reveal patterns of development of the cognitive system at work in music and emotion, shedding further light on the proposed hierarchal model. It would be helpful to investigate at what point in development the cue hierarchy solidifies, as well as whether or not the cue hierarchy changes between childhood and adulthood (e.g., which cues we rely on the most/least).

In emotion research, the lines between emotion recognition vs. emotion experience can become blurred, which is potentially problematic. Peretz (2001) points out that "in theory, all emotions, including fear, may be perceived, recognized, known, described, and mimicked, without being felt" (p.125). However, in the context of emotion in music, it is difficult to ensure that respondents report only the facet of emotion the researcher is looking for (i.e., recognition or experience). In attempts to solve this problem, Peretz (2001) suggests a combination of direct and objective measures of emotion. Direct measures include asking participants to self-report either how they feel or what emotion the stimulus suggests, and objective measures include physiological indexes of emotion experience (e.g., heart rate, respiration, skin temperature). However, physiological measures are objective but not specific – we cannot determine with confidence which emotion the participant is experiencing (Peretz, 2001). Also, because they measure an individual's arousal level, physiological measures are most useful in research on emotion experience rather than emotion recognition. For this reason and also due to the short duration and computerized nature of the stimuli in the present study, I decided against physiological measures; I suspected I would be unable to observe any significant effects. Nevertheless, future research with musical stimuli of longer duration may attempt to supplement subjective self-report measures with some objective physiological measure to gain a more complete picture of the emotions of listeners. Future research using the present paradigm may also seek to measure emotion experience rather than recognition, in which case physiological measures would be particularly appropriate.

Although familiarity was not systematically recorded in this experiment, upon finishing the paradigm, some participants mentioned the melody "A Bicycle Built for Two" sounded familiar, and a few participants even knew some of the lyrics. This was unfortunate in that the melodies were originally selected with the hopes that none of the listeners would have pre-existing emotional expectations associated with the music. The simplicity of the task may have allowed participants to overcome their pre-established emotional associations with the song, but future research should require participants to record their level of familiarity with each melody and in general attempt to avoid well-known songs.

The musicality of electronic music in relation to human performance is always questionable. Because the musical stimuli sounded a bit different than most music we are used to hearing, it is possible that the present paradigm did not fully tap into the cognitive processes behind interpretation of emotion in music. Many of the results compiled in Gabrielsson and Lindstrom (2001) and Juslin and Laukka (2003) were found with musical stimuli recorded from human studio performances, and effects of emotional cue utilization have been found in the specific context of live performance (Juslin, 2000). However, Csound allowed more systematic control of the structure of the music than any live performance ever could, thereby increasing the likelihood that the observed effects were due to the manipulated parameters and making Csound's benefits outweigh its drawbacks. Additionally, most people who use Csound are musicians rather than research scientists, which speaks of Csound's musical authority.

The findings of the present study, in conjunction with the research compiled by Gabrielsson and Lindstrom (2001) and Juslin and Laukka (2003), have implications for any medium seeking to communicate a particular emotional message through music, such as mood induction research, music therapy, advertising, film, and music composition, conductance, and performance. When attempting to manipulate music's emotive power, we need to consider not only the features of music that influence judgments of listeners but also the cognitive processes listeners employ upon forming those judgments. In light of this study's findings, these processes seem to follow predictable hierarchal patterns, but further research is required in order to discover the full extent of the cognitive system underlying our interpretation of emotional messages in music.

References

Adachi, M. & Trehub, S. E. (1998). Children's expression of emotion in song. *Psychology of Music, 26*, 133-153.

Adachi, M. & Trehub, S. E. (2000). Decoding the expressive intentions in children's songs. *Music Perception, 18*, 213-224.

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

Besson, M. & Friederici, A.D. (1998). Language and music: A comparative view. *Music Perception,* 16, 1-9.

Boltz, M. G. (2001). Musical soundtracks as a schematic influence on the cognitive processing of filmed events. *Music Perception, 18*, 427-454.

Cohen, A.J. (2001). Music as a source of emotion in film. In P.N. Juslin & J.A. Sloboda (Eds.), *Music and Emotion: Theory and Research* (pp. 249-272). Oxford: Oxford University Press.

Dalla Bella, S., Peretz, I., Rousseau, L., & Gosselin, N. (2001). A developmental study of the affective value of tempo and mode in music. *Cognition*, *80*, B1-B10.

Dolgin, K. G. & Adelson, E. H. (1990). Age changes in the ability to interpret affect in sung and instrumentally-presented melodies. *Psychology of Music, 18*, 87-98.

Fernald, A. (1993). Approval and disapproval: Infant responsiveness to vocal affect in familiar and unfamiliar languages. *Child Development, 64*, 657-674.

Ferraro, F. R., King, B., Ronning, B., Pekarski, K., & Risan, J. (2003). Effects of induced emotional state on lexical processing in younger and older adults. *Journal of Psychology, 137*, 262-272.

Forcucci, Samuel L. (1984). *A Folk Song History of America.* New Jersey: Prentice-Hall, Inc.

Forster, K. I. (1979). Levels of processing and the structure of the language processor. In W. E. Cooper & E. C. T. Walker (Eds.) *Sentence Processing: Psycholinguistic studies presented to Merrill Garrett.* Hillsdale, NJ: Erlbaum.

Gabrielsson, A. & Lindstrom, E. (2001). The influence of musical structure on emotional expression. In P.N. Juslin & J.A. Sloboda (Eds.), *Music and Emotion: Theory and Research* (pp. 223-248). Oxford: Oxford University Press.

Halpern, A.R., Bartlett, J.C., & Dowling, J. (1998). Perception of mode, rhythm and contour in unfamiliar melodies: Effects of age and experience. *Music Perception, 15*, 335-355.

Hargreaves, David J. 1986. *The Developmental Psychology of Music.* Cambridge: Cambridge University Press.

Juslin, P. N. (2000). Cue utilization in communication of emotion in music performance: Relating Performance to perception. *Journal of Experimental Psychology, 26,* 1797-1813.

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

Juslin, P. N. & Sloboda, J. A. 2001. *Music and Emotion: Theory and Research.* Oxford: Oxford University Press.

Kastner, M. P. & Crowder, R. G. (1990). Perception of the major/minor distinction: Emotional connotations in young children. *Music Perception*, *8*, 189-202.

Koelsch, S., Schroeger, E., & Tervaniemi, M. (1999). Superior pre-attentive auditory processing in musicians. *NeuroReport, 10*, 1309-1313.

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

Mayer, J. D., Allen, J. P., & Beauregard, K. (1995). Mood inductions for four specific moods: A procedure employing guided imagery vignettes with music. *Journal of Mental Imagery, 19*, 133-150.

Montague, D. P. F. & Walker-Andrews, A. S. (2001). Peekaboo: A new look at infants' perception of emotion expressions. *Developmental Psychology, 37*, 826-838.

Montague, D. P. F. & Walker-Andrews, A. S. (2001). Mothers, fathers, and infants: The role of person familiarity and parental involvement in infants' perception of emotion expressions. *Child Development, 73*, 1339-1352.

Nawrot, E. (2003). The perception of emotional expression in music: evidence from infants, children and adults. *Psychology of Music, 31*, 75-92.

Peretz, I. (2001). Listen to the brain: A biological perspective on musical emotions. In P.N. Juslin & J.A. Sloboda (Eds.), *Music and Emotion: Theory and Research* (pp. 105-134). Oxford: Oxford University Press.

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

Pignatiello, M., Camp, C.J., Elder, S.T., & Rasar, L.A. (1989). A psychophysiological comparison of the Velten and musical mood induction techniques. *Journal of Music Therapy, 26*, 1989.

Rock, A. M. L., Trainor, L. J., & Addison, T. L. (1999). Distinctive messages in infant-directed lullabies and play songs. *Developmental Psychology, 35*, 527-534.

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

Spangler, G., Emlinger, S., Meinhardt, J., & Hamm, A. (2001). The specificity of infant emotional expression for emotion perception. *International Journal of Psychophysiology, 41*, 155-168.

Tsang, C. D., & Trainor, L. J. (2002). Spectral slope discrimination in infancy: Sensitivity to socially important timbres. *Infant Behavior & Development, 25*, 183-194.

Van der Does, W. (2002). Different types of experimentally induced sad mood. *Behavior Therapy, 33*, 551-561.

Vaestfjaell, Daniel (2002). Emotion induction through music: a review of the musical mood induction procedure. *Musicae Scientiae, Special Issue: Current Trends in the study of music and emotion,* 173-211.

Westermann, R., Spies, K. Stahl, G., & Hesse, F. W. (1996). Relative effectiveness and validity of mood induction procedures: a meta-analysis. *European Journal of Social Psychology, 26*, 557-580.

Wolpert, R.S. (2000). Attention to key in a nondirected music listening task: Musicians vs. nonmusicians. *Music Perception, 18*, 225-230.

Appendix

A Csound file is made up of an orchestra, which defines the parameters of the computer-generated instrument, and a score, which defines the parameters of the melody the instrument will play. Below are the orchestra definitions for "Plucky" and "Woody," the two instruments used in the present study. The instrument definitions remain the same throughout each condition – manipulations in this study were limited to the score.

PLUCKY

```
<CsoundSynthesizer>

<CsOptions>
</CsOptions>

<CsInstruments>

;header
sr      =       44100
kr      =       4410
ksmps   =       10
nchnls  =       1

;instrument definitions
                instr 1
;initialization block
idur    =       p3
iamp    =       ampdb(p4)
ifrq    =       cpspch(p5)
iatt    =       p6
idcy    =       p7
kfrq    =       cpspch(p5)

kenv    linen   iamp, iatt, idur, idcy
a1              pluck   kenv, kfrq, ifrq, 0, 1
                out             a1
                endin

</CsInstruments>
```

WOODY

```
<CsoundSynthesizer>

<CsOptions>
</CsOptions>

<CsInstruments>

;header
sr      =       44100
kr      =       4410
ksmps   =       10
nchnls  =       1
```

```
;instrument definitions
                instr 1
;initialization block
idur    =       p3
iamp    =       ampdb(p4)
ifrq    =       cpspch(p5)
iatt    =       p6
idcy    =       p7
ifun    =       p8

kenv    linen   iamp, iatt, idur, idcy
a1              oscil    kenv, ifrq, 3
                out             a1
                endin

</CsInstruments>
```

Below are the scores for the musical excerpts as used in the present study. Within these examples, tempo and pitch manipulations are both at the moderate or neutral levels – pitch (column p5) is based on the middle C octave, and tempo (t 0 100) is at Csound's default setting, 60 beats per minute. Periods (.) denote a repetition of the value listed above in the column.

"Bicycle Built for Two" by Harry Dacre

```
<CsScore>

;functions

t 0 100
;p1     p2      p3      p4      p5      p6      p7

i1      0       1.5     70      9.02    .1      .1
i1      1.5     .       .       8.11    .       .
i1      3       .       .       8.07    .       .
i1      4.5     .       .       8.02    .       .
i1      6       .5      .       8.04    .       .
i1      6.5     .       .       8.06    .       .
i1      7       .       .       8.07    .       .
i1      7.5     1       .       8.04    .       .
i1      8.5     .5      .       8.07    .       .
i1      9       3       .       8.02    .       .
i1      12      1.5     .       8.09    .       .
i1      13.5    .       .       9.02    .       .
i1      15      .       .       8.11    .       .
i1      16.5    .       .       8.07    .       .
i1      18      .5      .       8.04    .       .
i1      18.5    .       .       8.06    .       .
i1      19      .       .       8.07    .       .
i1      19.5    1       .       8.09    .       .
i1      20.5    .5      .       8.11    .       .
i1      21      3       .       8.07    .       1

</CsScore>

</CsoundSynthesizer>
```

"In the Good Old Colony Days" (Colonial Song, lyricist/composer unknown)

```
<CsScore>

;functions

t 0 100

;p1     p2      p3      p4      p5      p6      p7

i1      0       .5      70      8.00    .1      .1
i1      .5      .       .       8.0     .       .
i1      1       1       .       8.05    .       .
i1      2       1       .       8.05    .       .
i1      3       .5      .       8.07    .       .
i1      3.5     .       .       8.05    .       .
i1      4       1       .       8.04    .       .
i1      5       3       .       8.05    .       .
i1      8       .5      .       8.05    .       .
i1      8.5     .       .       8.07    .       .
i1      9       1       .       8.09    .       .
i1      10      .       .       8.09    .       .
i1      11      .5      .       8.10    .       .
i1      11.5    .       .       8.09    .       .
i1      12      1       .       8.07    .       .
i1      13      3       .       8.09    .       .
i1      16      .5      .       8.05    .       .
i1      16.5    .       .       8.07    .       .
i1      17      1       .       8.09    .       .
i1      18      .       .       8.09    .       .
i1      19      1.5     .       9.00    .       .
i1      20.5    .5      .       9.00    .       .
i1      21      .       .       8.05    .       .
i1      21.5    .       .       8.05    .       .
i1      22      1       .       8.05    .       .
i1      23      1.5     .       8.09    .       .
i1      24.5    .5      .       8.09    .       .
i1      25      1       .       8.07    .       .
i1      26      .       .       8.05    .       .
i1      27      .5      .       8.04    .       .
i1      27.5    .       .       8.05    .       .
i1      28      1       .       8.07    .       .
i1      29      3       .       8.05    .       .1

e

</CsScore>

</CsoundSynthesizer>
```

Pitch manipulations may be achieved by altering values in the p5 column. The number to the left of the decimal point denotes the pitch octave. 7 denotes low C, 8 denotes middle C, and 9 denotes high C. Notes after the decimal point denote the number of half steps above C; for example, 8.01= C#, 8.02=D, 8.03=D#, and so on. In this experiment, I only altered the octave.

Tempo manipulations may be achieved in one of two ways. First, the tempo definition located at the top of the score (i.e., t 0 100) may be altered. In the present experiment, the slow tempo conditions were recorded as t 0 50, with moderate tempo at t 0 100 and fast tempo at t 0 200. However, occasional acoustic anomalies occurred when using the slowest tempo. To alleviate the problem, I instead altered the score itself, recording time values for each note as twice as long as in the moderate tempo conditions. Differences appear in columns p2 and p3. An example of this type of alteration appears below. (Please compare to this melody's score as it appears above with a moderate tempo.)

"Bicycle Built for Two" at slow tempo with alterations in note length

t 0 100

;p1	p2	p3	p4	p5	p6	p7
i1	0	3	55	9.02	.1	.1
i1	3	.	.	8.11	.	.
i1	6	.	.	8.07	.	.
i1	9	.	.	8.02	.	.
i1	12	1	.	8.04	.	.
i1	13	.	.	8.06	.	.
i1	14	.	.	8.07	.	.
i1	15	2	.	8.04	.	.
i1	17	1	.	8.07	.	.
i1	18	6	.	8.02	.	1
i1	24	3	.	8.09	.	.1
i1	27	.	.	9.02	.	.
i1	30	.	.	8.11	.	.
i1	33	.	.	8.07	.	.
i1	36	1	.	8.04	.	.
i1	37	.	.	8.06	.	.
i1	38	.	.	8.07	.	.
i1	39	2	.	8.09	.	.
i1	41	1	.	8.11	.	.
i1	42	6	.	8.07	.	2

e

Author note

I would like to thank my advisor Penny Yee for her indispensable mentoring at all stages of this project's development, Sam Pellman for generously allowing me access to the Csound program (as well as teaching me how to use it!) and Rob Hopkins and Peter Rabinowitz for their advice on the musical aspects of my experiment. I would also like to thank the members of my thesis group (Nadia Dovi, Nesa Wasarhaley, Adam Lalor, Liz Casey, and Jessalyn Geller) for their feedback, advice, and encouragement.

Table 1

Numeric Manipulations of Tempo and Pitch in Csound

Condition	Beats per Minute	Csound Notation
Tempo		
Slow	30	(0 50)
Moderate	60	(0 100)
Fast	120	(0 200)
Pitch	Octave	Csound Notation
Low	Low C	(7.00)
Middle	Middle C	(8.00)
High	High C	(9.00)

Note. Csound notations are better understood in the context of entire Csound text files, which appear in the Appendix.

Table 2

Mean Emotion Ratings as a Function of Musical Cue and Averaged Across Two Melodies

Condition	Emotion Rating	
Tempo		
Slow	M = 6.36	SD = .29
Medium	M = 10.94	SD = .28
Fast	M = 17.16	SD = .17
Pitch		
Low	M = 10.85	SD = .19
Medium	M = 11.25	SD = .22
High	M = 12.35	SD = .27
Timbre		
Woody	M = 10.79	SD = .22
Plucky	M = 12.18	SD = .20

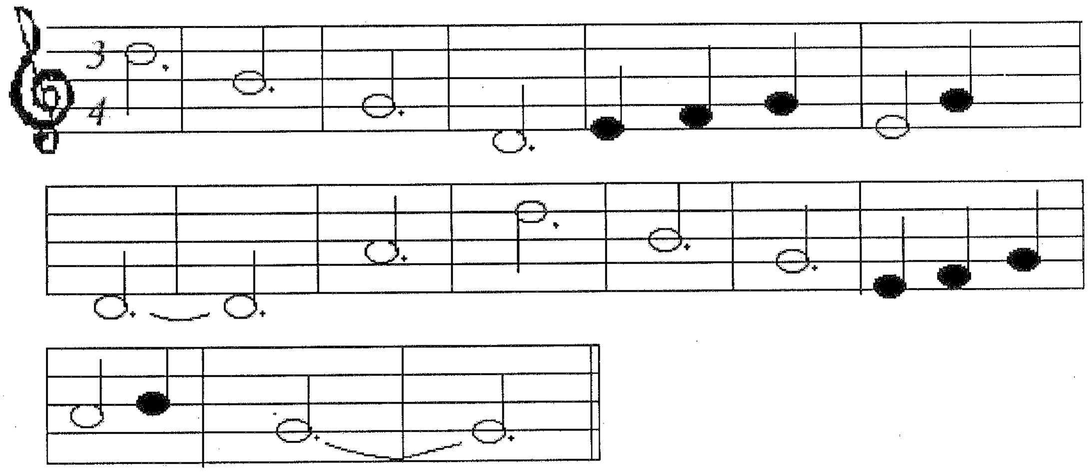


Figure 1. Musical notation of melody from "A Bicycle Built for Two" by Harry Darce, as used in the stimuli created with Csound.

Figure 2. Musical notation of the melody from "In the Good Old Colony Days," author unknown, as used in the stimuli created with Csound.

TRACK 1

This track sounded (circle one):

HAPPY SAD

The strength of this emotion was:

1 2 3 4 5 6 7 8 9 10

slight/weak moderate extreme

Figure 3. Participants used this response scale to record their emotional interpretation of each musical excerpt. The number in the upper left corner indicates each track's place in the order in which each participant heard the excerpts.

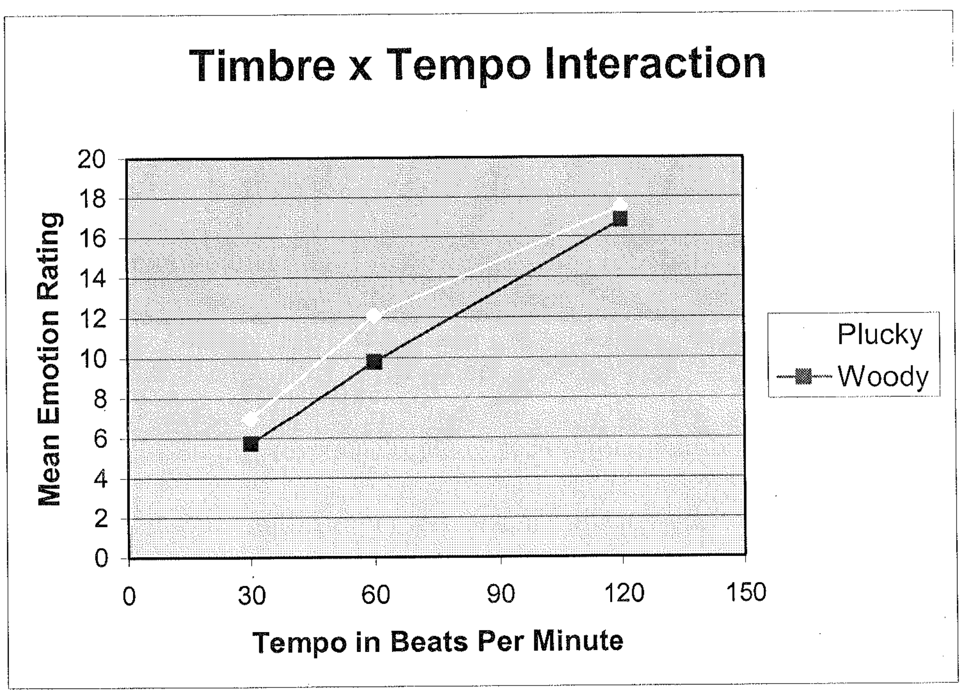


Figure 4. Plot of mean emotional ratings of musical excerpts as a function of timbre and tempo.

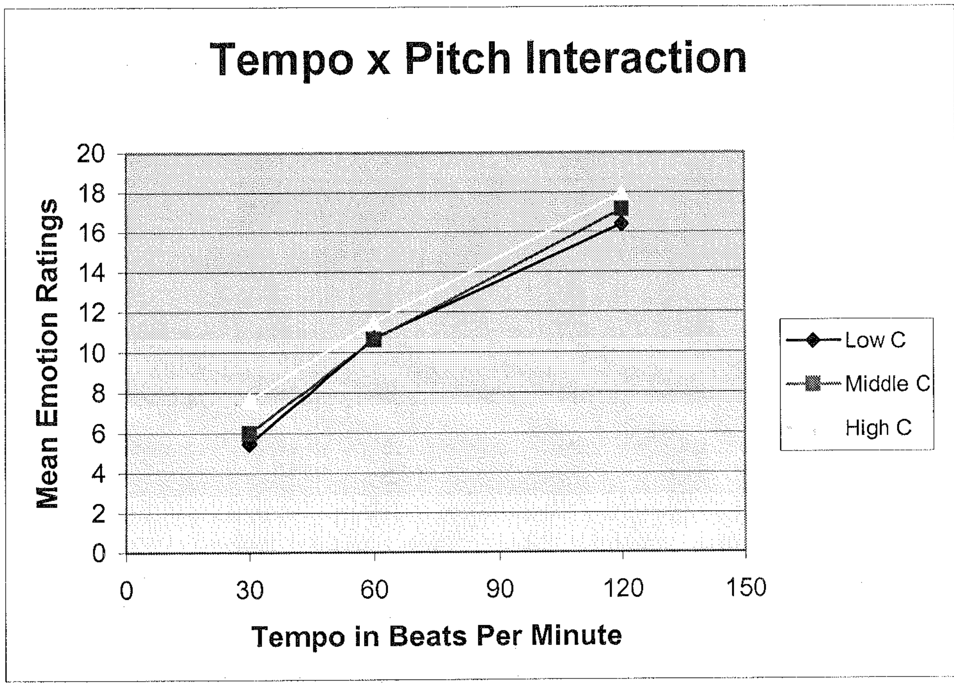


Figure 5. Plot of mean emotional ratings of musical excerpts

as a function of tempo and pitch.

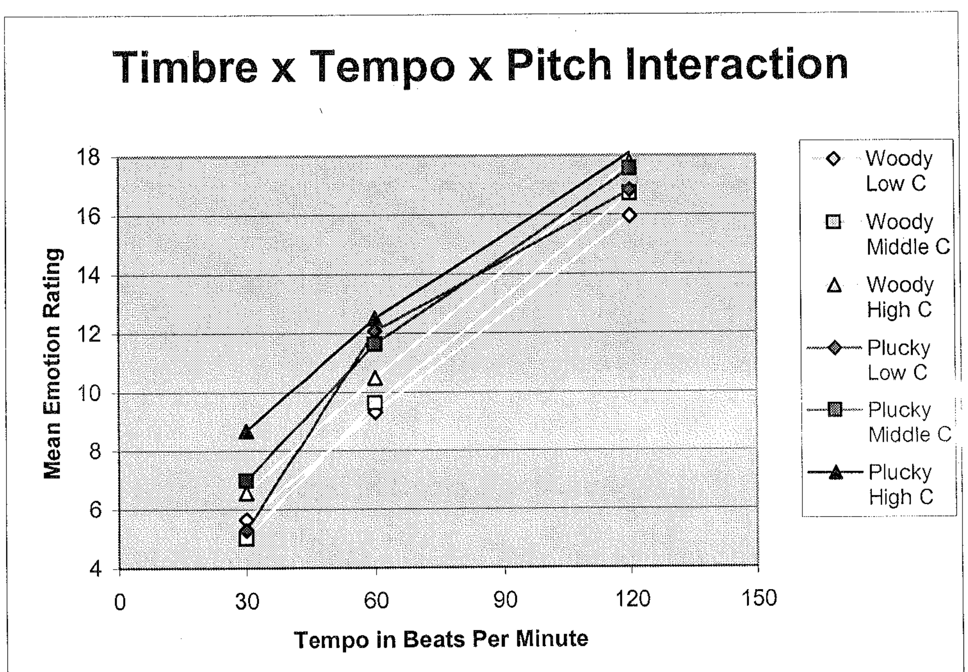
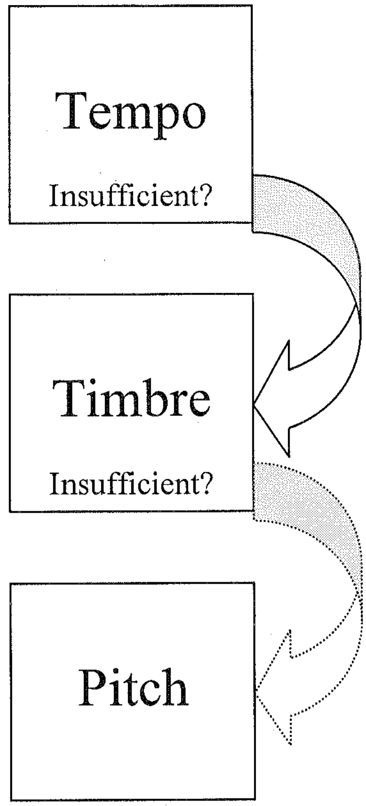


Figure 6. Plot of mean emotion ratings of musical excerpts

in a three-way interaction between timbre, tempo, and pitch.



We access the most salient cue first. This often provides sufficient information for emotional interpretation.

If the highest-level cue provides insufficient emotional information, we access the second-most salient cue available.

If the second cue is also insufficient, we may move on to the third. However, because cues at the bottom of the hierarchy are the weakest indexes of emotion, we may not use them in the same way as the cues above.

Figure 7. Hierarchal model of cue-specific cognitive mechanisms devoted to the interpretation of emotion in music. This model is limited to the three cues manipulated in the present study, but other cues may eventually be incorporated within it.