Two Studies on Assessing Emotional Responses to Music and Mode: The Effect of Lowered Pitch on Sadness Judgments, and the Affective Priming Paradigm as an Implicit Measure

Dissertation

Presented in Partial Fulfillment of the Requirements

for the Degree Doctor of Philosophy in the

Graduate School of The Ohio State University

Gary King-Yee Yim, M.A.

Graduate Program in Music

The Ohio State University

2014

Dissertation Committee:

David Huron, Advisor

David Clampitt

Lawrence Feth

Abstract

Music may convey or elicit emotional responses in listeners. After reviewing some of the theories and mechanisms concerning music and emotion, two experiments are presented. In the first experiment, participants rated the sadness of melodies that were constructed using artificial scales. It was found that melodies based on scales with lowered pitches were perceived as sadder than melodies based on scales with raised pitches. A modest explanation for the sadness of the minor mode is then proposed: in comparison to the normative major mode, the minor mode may be perceived as lower than normal. In the second experiment, an affective priming paradigm for music was designed. Specifically, participants categorized words as positive or negative while listening to chord progressions, scoring points for fast and accurate responses. It was found that, on average, positive words were categorized more quickly during the major chord progressions, while negative words were categorized more quickly during the minor chord progressions. However, these differences in reaction time were not statistically significant, and could not be used to predict the progressions that were heard.

KEYWORDS: music, emotion, subjective ratings, implicit measures, minor mode, sadness, affective priming

Acknowledgements

I would like to thank my advisor, David Huron, for his insightful comments and supportive words throughout my studies at Ohio State. I am grateful for my colleagues and friends at the Cognitive and Systematic Musicology Lab, both past and present, for all the sharing together of ideas and food. I would like to thank Parag Chordia, for his role in synthesizing the specified stimuli for the sad scales experiment. I am also appreciative of the invaluable suggestions made by David Clampitt, Larry Feth, and Sarah Boysen in the revision of this paper.

The Affective Norms for English Words (ANEW) materials are used by permission from the NIMH Center for the Study of Emotion and Attention.

Vita

2007 .... B. Music, Queen's University

2008 .... B. Computing, Queen's University

September 2009 to July 2013 .... Graduate Teaching Associate, School of Music, The Ohio State University

2011 .... M.A. Music Theory, The Ohio State University

Fields of Study

Major Field: Music

Table of Contents

List Of Figures

Chapter 1: Emotional Responses to Music

How does music convey or elicit emotions in listeners? Such an inquiry may interest composers and performers who want to understand how listeners respond to their work. It may also be of interest to those scholars studying art, or emotion, or hearing, or other related fields. And, of course, it may simply be of general interest to any thoughtful aficionados of music at large.

In the two subsequent chapters, two studies are presented that investigate musical emotions in listeners. Chapter Two presents an experiment that investigated how listeners' perceptions of sadness were affected by the manipulation of the musical scales used in the melodies. Chapter Three presents an experiment that explored how hearing major- or minor-mode music affected listeners' performance in a concurrent word task. In both studies, novel experimental procedures were devised to address some specific methodological issues that may arise when measuring emotional responses to music.

This present chapter introduces some of the theoretical foundations required by the experiments that follow. First, a brief definition for "emotional response" is given, followed by an overview of how emotions can be measured and modeled. Then, two complementary accounts on how emotions may be conveyed or elicited by music are presented.

Emotions and Their Measurement

In one description (Scherer, 2005), emotion is conceptualized as a relatively short-duration phenomenon arising in response to an internal or external stimulus, which involves some combination of five different components: 1) subjective feelings, 2) physiological changes, 3) motor expression, 4) behavior preparation, and 5) cognitive processes. For example, the experience of sadness may include not only the sense of feeling "down," but may also involve lowered physiological arousal, sad facial expression, a tendency to act sluggishly, and a preoccupation by ruminative thoughts. Although studies on emotion would ideally assess all five of these components, it may not always be feasible to design experiments in this way. Instead, in order to adequately characterize an emotional response, researchers should seek convergent evidence from a variety of studies that together may include each component.

Three types of methods for measuring the emotional components in this conception are discussed below: self-report (usually used for subjective feelings and cognitive processes), physiological measures (for physiological changes), and behavioral measures (for changes in motor expression and behavior preparation, or in some cases as an implicit measure).

Self-Report Methods

The subjective experience of emotions, that is, the phenomenological change in feelings, is arguably the most essential component that identifies an emotional response. Along with concurrent changes in cognitive processes, subjective feelings are usually studied through self-report methods. There is a range of self-report methods, including free response, structured interviews, diary studies, as well as the use of checklists or rating scales. These methods, however, are always limited by the participants' abilities of introspection and articulation: there may be feelings or cognitive processes of which the participant is not aware, or he or she may not be able to adequately express them. Nonetheless, self-report methods can be reliable when appropriately designed (for a review, see Zentner & Eerola, 2010). The study presented in Chapter Two used self-reporting, with participants indicating their perceived sadness of a heard melodic excerpt using a 7-point rating scale.

Self-report methods are also particularly susceptible to the problems of demand characteristics. That is, participants' responses may be unconsciously biased due to their perception of the experimental hypothesis (whether or not that perception is correct). Since self-report methods directly call for participants to reflect on their feelings and thoughts, they may more readily modify their responses according to some perceived version of the hypothesis. In order to help mitigate the effects of demand characteristics, after the experiment is over, researchers can interview participants about the experimental hypothesis and any particular response strategies they may have used (for an example, see Rubin, Paolini, & Crisp, 2010).

Alternatives to Self-Report Methods

Physiological measures, in general, are relatively resistant to demand characteristics. These measures assess the physiological component of emotion, which includes changes in heart rate, skin conductance, hormone levels, as well as muscle or neural activity (for a review, see Hodges, 2010). Such measurements can require specialized equipment or procedures, such as electrocardiograms, sensor electrodes, or biochemical analyses. Although any single measure may be insufficient for characterizing an emotion, coordinated or systematic changes may provide important convergent evidence for identifying response patterns for different emotions.

Lastly, behavioral measures can also be used, which involve assessing both motor responses and preparation for motor responses. Although some motor responses may be directly observable (for example, facial expressions, or some gross motor behaviors such as foot-tapping or head-bobbing), they may still be difficult to interpret. Moreover, for music, cultural norms may cause behavioral expression to be inhibited or suppressed altogether. A special case of behavioral measures is implicit measures: if a change in an emotional state involves motor or cognitive changes, experimental tasks can be designed specifically to detect these changes, and thus infer the underlying emotional changes (Västfjäll, 2010). The study presented in Chapter Three measured changes in reaction time as an implicit measure to assess emotional responses to major- and minor-mode music.

Perceived and Felt Emotions

An important distinction concerns two different ways by which music may be considered "emotional" for listeners. Emotion may be *perceived* in the music, such that an individual may detect that an emotion is embodied within the stimulus, without they themselves independently experiencing that emotion. Or, emotion may be *induced* by music, such that the stimulus causes some emotion to actually be experienced by the individual. Although the precise relationship between perception and induction of emotions remains an active area of research, it seems clear that listening to music is capable of involving both. In instructions to participants for self-report, it is important to ensure that they understand the distinction. The study presented in Chapter Two had participants report on perceived sadness of melodies. However, the study presented in Chapter Three did not explicitly make such a distinction, since it did not involve self-report.

Emotion Models

Two main categories of emotional models have proved fruitful in emotion research, namely discrete models and dimensional models (for a review, see Vuoskoski & Eerola, 2011; also, Eerola & Vuoskoski, 2011).

Discrete Models

In discrete (or categorical) models, a certain number of emotions are construed of as canonical. One advantage of this approach is that a single familiar term (e.g.: “happy”) can be used to capture the meaning of an entire response pattern. But there are some disadvantages to such labeling: the meaning of familiar terms may not be the same for everyone, single terms may not capture nuanced variations, and different terms may, in fact, refer to the same phenomenological experience.

One commonly used discrete model assumes the existence of a small number of “basic emotion” categories, based on the study of facial expressions (Ekman, 1992). While there is not a consensus on what the specific basic emotions are, happiness, sadness, fear and anger are usually included. Other basic emotions sometimes include tenderness, disgust, surprise, or contempt. These basic emotions are supposedly innate and universal. However, note that a discrete model need not have a small number of emotions, and assumptions of innateness or universality are not necessary. The study presented in Chapter Two required listeners to rate specifically how “sad” a melody was perceived to be.

Continuous Models

Continuous models posit either two or three dimensions into which emotions can be categorized. In the circumplex model of affect (Russell, 1980; see also Hevner, 1936), emotions can be categorized along the two dimensions of valence (how positive or negative an emotional state is) and arousal (how energetic or subdued an emotional state is). In some three-dimensional models, “potency” or “dominance” is sometimes included as well. Figure 1 below reproduces Russell’s circumplex model, with emotion labels plotted against valence (x-axis) and arousal (y-axis).

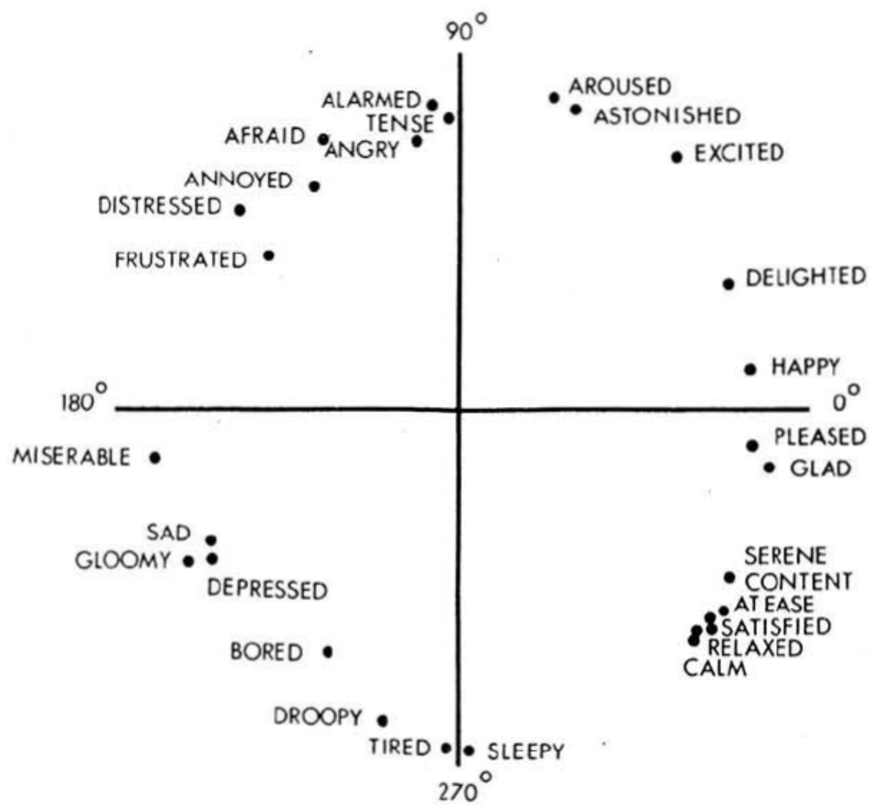

Figure 1. Russell’s (1980) Circumplex Model of Affect. Participants categorized twenty-eight emotion words. Valence is plotted on the x-axis, with negative valence on the left and positive valence on the right. Arousal is plotted on the y-axis, with low arousal at the bottom and high arousal at the top.

While continuous models of emotion are comprehensive, in the sense that they can be used (in principle) to plot any emotional state, they cannot always distinguish between certain emotions. In the circumplex model above, for example, both “angry” and “afraid” are plotted as high arousal, negative valence emotions, although one would suppose that these emotions are quite different from each other at the phenomenological level.

A Note on Terminology

Note the use of the term “affect” in the model above. Although often used interchangeably with the words “emotion” and “feeling,” the meaning of these terms can be differentiated. The usage of these terms will be consistent with the recommendations in the Handbook of Music and Emotion (Juslin & Sloboda, 2010, p. 10). “Emotion,” as mentioned above, is the brief response to a stimulus with multiple coordinated components. “Feeling” is one of these components, describing the subjective and phenomenological experience of an emotion. “Affect” is an umbrella term for all states that may have to do with subjective evaluation, including not only emotion, but also preferences and moods.

Music and Emotion

Various accounts have been proposed to explain the process by which musical sounds can influence emotional responses. Some aesthetic philosophers have considered this problem via two opposed positions: “emotivism,” positing that music directly arouses emotions in listeners; or “cognitivism,” positing that music merely portrays emotions that are recognized by listeners (e.g., Matravers, 2001; Kivy, 1999). An alternative way to orient accounts of musical emotion, given below, is by whether they are based on expectations or on associations (which need not be mutually exclusive).

Expectation

Meyer’s (1956) highly influential account of emotion and meaning in music posits that musical events within a work create expectations about subsequent musical events. These expectations may or may not be realized, but it is in the context of these expectations that emotions are experienced. That is, listeners may experience emotions relating to the uncertainty of future musical events, and also relating to the retrospective evaluation of that uncertainty when it is resolved. Such musical expectations are learned through enculturation, and therefore vary for individuals and for cultures.

In a similar vein, a work relating emotions and expectation is the ITPRA model by Huron (2006), which consists of five different emotional responses relating to expectations: Imagination, Tension, Prediction, Reaction, and Appraisal. In this general expectation model, emotions are connected to expectations as biological adaptations. If accurate prediction about future events is a useful evolutionary trait, emotions may motivate the organism to behave in a certain way. For example, tension is an anticipatory state where an organism prepares its physiological resources to respond appropriately to a future event. Narmour’s Implication-Realization model (1997) is another account that connects experienced emotions with learned musical expectations. In these views, music making may be described as the craft of exploiting the learned or innate connections between expectation and certain emotions.

Association

In addition to these expectation-based models, where musical emotions relate to the temporal relations of music, other accounts recognize that different aspects of music may be readily linked to certain meanings. For example, some pieces of music may remind different individuals of specific memories that have particular significance to them. Similarly, some genres of music, such as dirges and waltzes, may have specific meanings and associated emotions for enculturated listeners. Furthermore, some general attributes of music, such as its loudness or tempo, may also have associative links with emotion. The associations of the latter kind have become a particularly active area of research, where those musical attributes are commonly known as “acoustic cues.”

For example, in a thorough review by Juslin and Laukka (2003), the authors describe musical emotions within a communication model. In this model, performers and composers encode emotions in music using acoustic cues, while listeners use those same acoustic cues to decode the communicated emotion. Acoustic cues are redundant and additive, as any single cue is neither necessary nor sufficient for conveying an emotion. However, the coincidence of multiple coherent cues can convey an emotion more strongly than just a single cue. For example, music that is fast and loud may convey happiness, but it may also convey anger. The inclusion of additional acoustic cues, such as modality or dissonance, may further clarify the intended emotion. The relevant acoustic cues for emotion in music are often compared to the acoustic cues for emotion in speech, an idea further pursued in Chapter Two.

Orientation of Present Studies

While the expectation- and association-based accounts contribute explanations of *how* musical sounds may influence emotional responses, the studies presented in the following chapters are instead primarily concerned with *which* musical sounds influence particular emotional responses. Thus, while the experiments were motivated by such theories, they do not directly test or support them. Rather, the existence of a reliable correspondence between certain musical structures and certain emotional responses – a common claim for both accounts – is taken as the point of departure. Specifically, the correspondence between minor-mode-like scales and the perception of sadness is investigated in Chapter Two. Then, the correspondences between major chord progressions and the perception of positive-valence words, and minor chord progressions and negative-valence words, are investigated in Chapter Three.

It may seem like a gross oversimplification to discuss the experience of music listening as mere correspondences between musical structure and their related emotional responses. But it is important to point out that this reductionist approach is not a belief about the nature of music, but simply a strategy for inquiry within complex phenomena. Thus, the elements of music are isolated and manipulated in the following experiments, with the understanding that such isolation does not adequately capture the nuances of actual musical works, and that the manipulation does not resemble any actual musical practice. In the same vein, the reductionist conception of emotions as presented is not a belief that emotions are necessarily reducible thusly (that is, they may not be adequately described using valence and arousal, for example).

Summary

In this chapter, an emotional response to a stimulus was described as a multicomponent phenomenon, involving subjective feelings, physiological changes, motor expression, behavior preparation, and cognitive processes. Various methods for assessing these changes were discussed, including self-report measures and implicit measures. Discrete and continuous models of emotion were discussed, including the circumplex model involving valence and arousal. A distinction was made between induced emotions felt by the listener, and emotions that are only perceived or represented in the sound. Then, various accounts were presented for how emotional responses may arise from music listening. Specifically, emotions were linked with expectations about the musical events, and with associations from the musical features (or “acoustic cues”).

In the ensuing chapters, two studies are presented, both of which relate to emotional responses to major or minor modality. In Chapter Two, a study is presented in which the relationship between the scale structure of various melodies and the self-reported ratings of perceived sadness was investigated. It was found that listeners heard melodies based on scales containing relatively lowered pitches as sounding sadder. In Chapter Three, a study is presented which explored the use of an implicit measure of emotion. Changes in reaction time during an evaluative decision task were used to assess emotional responses to major or minor chord progressions. It was found that task performance on positive-valence words may have been better during the major chord sequences, and likewise for negative-valence words and minor chord sequences.

Chapter 2: Sadness and Lowered-Pitch

In this chapter, a study pertaining to the perceived sadness of the minor mode is presented. First, the similar acoustic cues for sadness in both music and speech are discussed. Special consideration is given to the emotional connotations of the minor mode: unlike other acoustic cues in music, modality seems to have no good parallels in speech. An argument is made that the minor-mode may be particularly suitable or apt for learned associations with sadness. Specifically, the characteristic lowering of pitches in the minor scale is considered, and the concept of “relatively lower pitch” is proposed as a cue for sadness. An experiment is then presented, that tested the hypothesis that relatively lowered pitches are perceived as more sad. The design of the experiment was novel in that the test stimuli were identical for paired participants, with only the immediately prior exposure stimuli controlled and manipulated. The results were consistent with the experimental hypothesis: it was found that melodies using a scale that contained relatively lowered pitches were perceived as more sad than those that contained raised pitches.

Background

In categorical models of emotion, sadness is usually considered as one of the basic emotions (e.g.: Tomkins, 1962; Ekman, 1992; Lövheim, 2012). In dimensional models, sadness is categorized by low arousal and negative valence, and contrasts with happiness, which is characterized by high arousal and positive valence (e.g.: Hevner, 1936; Russell, 1980).

Sad Music and Sad Speech

It seems that music can readily express sadness, and furthermore, music can actually elicit sadness for some listeners. How does music do this? Juslin and Laukka (2003) reviewed 41 studies on emotional music performance and found that sadness in music was reliably communicated by nine factors. Their findings were consistent with the hypothesis that composers and performers encode acoustic cues into the music. Consequently, listeners are able to hear and decode these same acoustic cues from the music, in order to reliably perceive the emotion expressed. They listed nine acoustic cues for sadness: slow tempo, low sound level, low sound level variability, little high-frequency energy, low pitch level, low pitch variability, falling pitch contour, slow tone attacks, and micro-structural irregularities (in frequency, intensity and duration). Additionally, in reviewing 104 studies of vocal expression, they found that these same acoustic cues for sad music were also used for sad speech.

Various relationships between music and speech have been conjectured (e.g.: Wachsmann, 1971; Roederer, 1984). Although similar acoustic cues seem to have similar emotional meaning in both domains, attempts to establish the direction of influence, if any, have been inconclusive. Besides human speech, it has been noted that some animal vocalizations also seem to share these acoustic cues for emotional meaning (Morton, 1977; elaborated by Huron, 2012). Although these kinds of relationships are not central to the premise of this work, they are suggestive of some ways in which the associations between sound and meaning may have developed for music.

The acoustic cues for sadness listed from the Juslin and Laukka (2003) review pertain primarily to how slow, soft and low are the sounds. As noted, the same acoustic cues seem to be shared by both sad music and sad speech (and, in fact, with depressed speech; see Hargreaves, Starkweather, & Blacker, 1965). Additionally, for music, another important cue for sadness is the use of the minor mode. While major-mode music tends to convey happiness or some other cheerful, positive affect, the minor mode is strongly associated with sadness and negative affect, at least for Western-enculturated listeners. In a series of experiments, Huron and his collaborators investigated sadness in music by examining how the minor mode is correlated with other acoustic cues. Indeed, minor-mode music tends to be slower (Post & Huron, 2009), softer (Turner & Huron, 2008), and lower in both pitch and pitch variability (Huron, 2008). Minor-mode music also tends to be characterized acoustically by less high-frequency energy, which is associated with a darker timbre (Shutz, Huron, Keeton, & Loewer, 2008; Huron, Anderson, & Shanahan, in press).

Of course, this is not to say that the minor mode is always sad, or that such acoustic features always accompany it. However, when minor-mode music is slower, softer, and lower, with smaller intervals and darker timbre, sadness may be strongly conveyed or evoked. This is consistent with the expectation that a skilled musician, intending to convey sadness, would employ a variety of cohesive elements to do so. However, if the minor mode is used in conjunction with other contrasting acoustic cues, such as fast tempo or loud dynamics, for example, different emotions such as passion or anger may be conveyed instead (Hevner, 1935).

The Minor Mode and Sadness

Nonetheless, the association between the minor mode and sadness is reliable for Western-enculturated listeners. Among the cues for sadness in music, the use of the minor mode is notable, in that there is no obvious analogue in speech or in animal sounds (with a possible exception in Curtis & Bharucha, 2010). Consequently, one may take special interest in an explanation as to why the minor mode may sound sad, or why it may have sad associations.

With the exception of the minor mode, all of the acoustic cues for sadness may arguably be related to lowered arousal, one of the physiological components of sadness (Huron, 2012). In addition to the subjective feeling component (feeling “down” or “blue”) and the cognitive component (e.g., ruminative thoughts), sadness is also characterized by low physiological arousal, resulting in bodily changes such as decreased heart rate and low muscle tone. Low muscle tone may result in inhibited motor tendencies and a “drooped” facial expression. Additionally, this relaxation of muscles in the vocal apparatus may account for the acoustic features of sadness. For example, voices are quieter when the pulmonary muscles are more relaxed, and vocal pitch is lower when the vocal folds are more relaxed.

However, the minor mode’s association with sadness does not seem to be directly related to physiological arousal, and cannot be explained by appealing to analogues in speech. Instead, the sadness of the minor mode is often explained as an arbitrarily learned cultural phenomenon. After all, minor-mode music is not sad in all cultures (e.g., Rice, 2004). The associations between major mode and positive valence, and minor mode and negative valence, are observed for children as young as three years (Kastner & Crowder, 1990), though apparently not for infants younger than that (Crowder, Reznick, & Rosenkrantz, 1991). But while the sad associations of the minor scale seem to be learned, perhaps the association is not entirely arbitrary: there may be features of the minor mode that make it particularly suitable for learned associations with sadness.

Minor Mode Features

Are there particular features of the minor mode that are amenable to conveying sadness? The existence of such features would be consistent with the interpretation that its learned association with sadness is not arbitrary, but that the minor mode itself is particularly suitable for forming such associations. Recall the acoustic cues of sadness listed previously. Juslin and Laukka (2003) include the following features: slow tempo, low sound level, low sound level variability, little high-frequency energy, low pitch level, low pitch variability, falling pitch contour, slow tone attacks, and micro-structural irregularities (in frequency, intensity and duration). The various studies on minor mode and sadness by Huron and his colleagues (inspired by Kraepelin, 1899/1921) include the following features: slower tempo, softer dynamics, lower pitch, smaller pitch movement, less articulation, and darker timbre.

Recognizing that the minor mode (or minor scale) is an organization of pitches, the pitch-related elements can be extracted from these two lists above (which do overlap considerably). This yields three particular acoustic cues that are to be presently considered: low pitch, low pitch variability, and descending pitch contour. These pitch-related acoustic cues for sadness in both music and speech are discussed below in terms of their relation to the minor mode. Low pitch is discussed last, as it is the one cue that is directly investigated in the current study.

Low Pitch Variability

Sad speech may be characterized as monotonous, with few excursions to high or low fundamental frequencies. For sad music, Huron (2008) found lower pitch variability for minor instrumental themes than for major themes, when pitch variability was operationalized as mean interval size for consecutive pitches. In fact, Huron and Davis (2012) found that the minor mode was optimal for minimizing pitch variability, relative to the major mode. Major mode melodies were sampled (instrumental themes, national anthems and well-known songs). Then, all possible combinations of semitone pitch-alterations were made exhaustively to the diatonic major scale on which the melodies were based. They found that lowering the third and sixth scale degrees resulted in the smallest mean interval size overall for all the melodies. In other words, if semitone pitch changes in any direction were allowed for any notes in the major scale, pitch alterations consistent with the minor mode resulted in the least melodic pitch variability. That is, the intervallic structure of the minor mode may effectively minimize mean interval size for melodies that use Western tonal conventions. This result is consistent with the claim that the minor mode may have features that make it particularly suitable or apt for use as an acoustic cue for sadness, namely by decreasing pitch variability.

Descending Pitch Contour

The effect of both mode and contour on perceived sadness has previously been studied, generally corroborating the finding that both the minor mode and descending contours are associated with negative affect such as sadness (e.g.: Gerardi & Gerken, 1995).

However, there does not yet seem to be research oriented toward finding an explanation for the sadness of the minor mode that involves contour. If descending pitch contour is an acoustic cue of sadness, and if the minor mode is apt for use with descending pitch contours, an argument may be constructed as to why the minor mode may be apt for use (or “useful”) as an acoustic cue for sadness. Whether or not the minor mode is indeed particularly suitable for descending pitch contours remains an open question.

Although not pursued in the present studies, this seems to be a fruitful line of inquiry. For example, do minor-mode melodies exhibit more descending contours than major-mode melodies? It is a promising to note the raising or lowering of the sixth and seventh scale-degrees of the minor scale (specifically, the textbook “melodic minor”) is dependent on whether the scale is ascending or descending. This implies, at the very least, that the pitch organization of the minor mode does interact meaningfully with contour.

Lower Pitch

The last pitch-related acoustic cue for sadness to be considered here is lower pitch. At the onset, there does not seem to be any reason to suppose that the minor mode *per se* would be particularly suitable or apt for being played at lower pitches. However, note that Huron (2008) found that in a sample of over 9,000 instrumental themes, the minor mode themes did have a slightly (but significantly) lower average pitch height, by about one semitone, compared to the major mode ones.

An alternative approach, taken in the present study, is to consider lower pitch in the context of corresponding scale degrees. In this conception, a minor scale may be considered as a “lowered version” of its parallel major scale. For example, the canonical harmonic minor scale may be conceived of as a major scale, but with the third and sixth scale degrees lowered. Although lowered overall pitch has been shown to be an appropriate cue for sadness, it remains to be seen whether only lowering certain pitches is also such a cue. If it were, it would provide another explanation for the perceived sadness of the minor mode.

To make this distinction explicit, the terms “absolute” and “relative” will be used. Something is “absolutely” lower in pitch when its overall pitch height is lower than that of another. For example, a melody that is transposed downward is absolutely lower in pitch than the original melody. As another example, the repertoire of tuba music is absolutely lower in pitch than the repertoire of piccolo music. In contrast, being “relatively” lower in pitch is a more limited term. It is only applicable when the objects being compared have a comparable pitch reference, such as a tonic (which may be different). Additionally, they must have a corresponding ordering of pitches. Then, if, after normalizing their respective pitch references, one of the objects has some pitches that are lower than the corresponding pitches of the other, without any pitches that are higher, it may be said to be “relatively” lower in pitch. Thus, a minor-mode melody would always be relatively lower in pitch than a major-mode melody, even if it were absolutely higher in pitch. For example, the melody that results from upwardly transposing a major-mode melody (say, in D major) into a minor-mode melody (for example, up a diatonic second to some kind of E minor) would be considered relatively lower in pitch than the original major melody.

The Lower-Than-Normal Hypothesis

The hypothesis to be tested in the current study is that relatively lower pitch is a cue for sadness. This would be consistent with the claim that the minor mode is suitable for learned associations with sadness. Note that the perception of "relatively" lower pitch is only possible by a suitable comparison to something else. To what might the minor mode be compared? Huron (2006) reported that 65 to 75 percent of Western music is in the major mode, and that listeners tend to assume that unfamiliar melodies are in the major mode by default (until they have reason to believe otherwise). Thus, it is also conjectured here that Western listeners may perceive the minor mode or scale as being relatively lower in pitch, by comparison to the more prevalent and prototypical major mode or scale. It is hypothesized, then, that the sad connotations of the minor mode may be due to its perception as being "lower than normal."

An analogous case of "normalization" can be given for the case of lowered sad speech. Although lower pitch is a cue for sadness in speech, and although men tend to have lower-pitched voices than women, it does not, however, seem to be the case that men tend to sound sadder than women. Presumably, when assessing the emotion of a person speaking with a low voice, there is an unconscious comparison made with that person's normal voice. Thus, an expression of sadness may only be perceived if their voice sounds "lower than normal" for them.

In summary, the association of the minor mode or scale to sadness is probably learned through culture. The association may not be arbitrary, however, if the minor mode has features which themselves act as cues for sadness. Three such potential features were considered. First, evidence was presented that is consistent with the conjecture that the minor mode is more effective than the major mode in minimizing interval size (i.e., pitch variability) in conventional melodies. Second, an untested conjecture was made that the minor mode may be more suitable than the major mode for use with descending contours. Third, the conjecture was made that the minor mode is heard as "relatively lower" than the major mode, and is therefore perceived as sadder – this is the claim that is tested in the current study. Note that all of these features are defined by contrast with the major scale, which is assumed to be normative. For cultures where the major scale is not normative, the minor scale may not be perceived as having relatively smaller intervals or relatively lower pitches, and would therefore not necessarily be associated with sadness.

Method

Overview

In brief, the experiment consisted of first exposing listeners to unfamiliar melodies based on unfamiliar scales in the exposure phase. In the subsequent test phase, listeners rated the sadness of melodies that used scales that were either relatively higher or relatively lower in pitch than the scales used in the exposure phase. To anticipate the results, the test melodies that were relatively lower than those in the exposure phase tended to be rated as sadder by listeners. The following sections describe 1) the novel use of a between-subjects design, 2) the synthesis and modifications of the musical stimuli, and 3) the experimental procedure itself.

1) Between-subjects design

To test the effects of relatively lower pitch, it was necessary to modify pitches (i.e., raise or lower them) in the context of a scale or melody. However, whenever a pitch or note in a scale or melody is modified, whether by raising or lowering it, its intervallic relationships with all of the other notes in that scale or melody are also changed. This may cause, for example, unintended changes in overall consonance and dissonance, or in particular harmonic implications. One way to control for this is to use an identical stimulus for both conditions, and manipulate only the immediately preceding listening experience. Thus, a between-subjects design was used for the experiment, where both participants heard the same melodies as each other in the test phase, after they had heard different melodies in the exposure phase. The exposure phase melodies were created differently for each participant, such that the identical test melodies were heard as either relatively lower for one participant in each pair, or relatively higher for the other.

Thirty exposure stimuli (ES), which were different for each participant, were heard first in the exposure phase. Then, fifteen test stimuli (TS), which were identical for both participants in each pair, were heard in the test phase. In addition to the 15 TS, the test phase was interspersed randomly with 15 additional ES, in order to ensure that the ES were heard as normative. In this way, although the TS were in fact identical for both participants in each pair, one participant heard them as relatively lower (“low TS”), and the other heard them as relatively higher (“high TS”), due to differences in the exposure phase. That is, the low TS melodies were “lower than normal” due to the higher ES, while the high TS melodies were “higher than normal” due to the lower ES. Melodies were uniquely generated for each pair of participants, and each participant rated the sadness of all 60 melodies they heard. Thus, each ES melody was heard by a single participant, while each TS melody was heard by a pair of participants. The presentation of the 45 exposure stimuli and 15 test stimuli within the exposure and test phases for a pair of participants is summarized in Figure 2 below.

	Exposure phase	**Test phase**
Participant A	30 High ES	15 High ES + 15 “Low” TS
Participant B	30 Low ES	15 Low ES + 15 “High” TS

Figure 2. Schematic of between-subjects design and exposure and test phases. Each participant heard 60 melodies, and the test stimuli (TS) heard only in the test phase were identical for each pair. The exposure stimuli (ES) heard in both the exposure and test phases were different, either higher or lower, for each participant within a pair. As a result, the identical TS were either relatively higher than the ES (for one participant) or relatively lower than the ES (for the other).

While the 15 TS melodies were identical for the participants within each pair, the 45 ES melodies were the same melodies based on slightly different scales. For one participant in each pair, these ES melodies were based on a scale that was relatively higher in pitch than the scale used for the TS. For the other participant in the pair, the ES melodies were based on a scale that was relatively lower in pitch than the scale used in the TS. Details of scale construction are given in the Stimuli section below.

It was hypothesized that relatively lower pitch (or "lower than normal" pitch) is a cue for sadness. Thus, it is predicted that for each TS melody, rated by a pair of participants, the participant who heard it as relatively lower TS (due to the higher ES in the exposure phase) would rate it as sadder. To anticipate the results, this finding was indeed observed.

2) Stimuli

In this section, the construction of the 60 melodies for each participant is described. For each pair of participants, three sets of scales were generated: one for the test stimuli (TS scale), which was shared by both participants; one for the low exposure stimuli (low ES scale), which was heard by only one of the participants; and one for the high exposure stimuli (high ES scale), which was heard only by the other participant. The three scales were used in conjunction with selected melodies and a synthesized timbre to create a unique set of stimuli for each pair of participants. It will be seen that there are a variety of constraints and desiderata that led to the peculiar method of stimuli generation as described here.

A naïve approach to testing the hypothesis would be to simply use existing Western melodies as stimuli, then modifying (i.e., raising or lowering) a random selection of pitches. However, since the stimuli would compare music in Western scales with music in modified scales, the most salient feature would probably relate to familiarity, rather than relative pitch height. Thus, in order to avoid confounds due to Western musical conventions, stimuli were explicitly designed so as to be unfamiliar to Western-enculturated listeners. At the same time, the stimuli should not be so unfamiliar (or unconventional) that they seem unmusical. This may be important if the participants are expected to engage emotionally with the music. Therefore, while the use of scales and timbre were designed to be unfamiliar, the rest of the listening experience was designed to be as musically conventional as possible.

In sum, the requirements for the stimuli to be created were that: (i) the created scales should sound unfamiliar, (ii) the scales should allow modifications to produce "higher" and "lower" versions, (iii) the timbre should sound unfamiliar, (iv) the timbre should sound musical, (v) the melodies should sound musical, and (vi) the scales should be average pitch height equivalent.

Stimuli were designed with these desiderata in mind. Artificial scales were created, along with an unusual timbre, in order to subvert Western listening habits. Monophonic

Germanic folksongs were selected, to have musically conventional melodic structure. The final stimuli were created by combining the folksongs with the artificial scales and timbre, as detailed below.

(i) "The created scales should sound unfamiliar"

Since the participants were all enculturated to Western music, any resemblances to the major/minor scale system might be expected to confound the results. Accordingly, the test stimuli (TS) scales were generated in a way that subverted Western listening habits by avoiding familiar intervals, avoiding equal temperament, and avoiding octave equivalence. The number of tones in the scale was randomly set in a range from five to nine notes. Then, the distance between successive scale steps was set to a random value between 80 and 350 cents (hundredths of semitones). In order to avoid an excessively small or large pitch compass, a further condition was that the span of the complete scale must lie between 800 and 1400 cents (i.e. between 8 and 14 semitones). This method was used to produce a single TS scale for each pair of participants.

(ii) "The scales should allow modifications to produce higher and lower versions"

After a TS scale was generated, two exposure stimuli (ES) scales were derived from it. Recall that in the exposure phase, melodies were based on a scale that was either "relatively higher" or "relatively lower" than the TS scales. To create an ES scale, a random number of tones was modified from the TS scale: one tone was modified at a minimum, and one-third of the TS scale tones were modified at a maximum. (It was assumed that the majority of scale tones must be preserved in order to encourage listeners to hear the test scale as a modified version of an exposure scale, rather than as a wholly novel scale.) For the high ES scale, randomly selected tones from the TS scale were raised in pitch. For the low ES scale, those same scale tones were lowered in pitch. The "tonic" was never modified in this way, and the size of the pitch modification (in cents) was always the same for both the high and low ES scales.

In determining the amount of pitch modification, it was important to maintain clear categorical boundaries of pitches. For example, if a modified pitch should be heard as a raised second scale tone, several kinds of confusion should be avoided. That is, it should not be heard as (a) an unmodified second scale tone, (b) a lowered third scale tone, or (c) a new scale tone altogether (increasing the total notes in the scale).

In order to minimize (a), a minimum pitch modification of 50 cents (i.e., a quarter tone) was established. This minimum distance is in the region of the perceptual limen for pitch interval judgments for naïve Western listeners (Burns & Ward, 1978). In order to minimize possible confusions pertaining to (b) and (c), the amount of pitch modification was restricted so that the modified scale tone was always closer to the original scale tone than any neighboring scale tone. Pitch proximity is known to have a powerful perceptual grouping effect (Bregman, 1990), so there was reason to suppose that proximity to the original scale tone would ensure the appropriate perception.

(iii) & (iv) "The timbre should sound unfamiliar" and "sound musical"

Post-doctoral fellow Parag Chordia synthesized stimuli, in order to create a sound that seemed both "foreign" and "natural." To this end, the sounds were modeled on the Kalimba (African thumb-piano), using the specifications for pitch and timing provided by the author.

(v) "The melodies should sound musical"

Although the melodies used a foreign-sounding timbre and employed artificially generated scales, the construction of the melodies was intended to sound musically conventional. Thus, all of the stimuli were based on real melodies, preserving their pitch contour, rhythms, and other tonal elements. Specifically, Germanic folksongs in the Essen Folksong Collection were used as melodic templates for the stimuli. This electronic collection contains over six thousand traditional Germanic folksongs assembled from various sources. Each folksong in the Essen collection is encoded in terms of diatonic scale degrees with possible chromatic alterations.

Melodies were randomly selected for each pair of participants, without replacement. Folksongs containing chromatic tones were excluded from selection, as were songs whose lowest note was not the tonic. Of the 6,255 folksongs in the database, there were 1,685 that fit these criteria. For each pair of participants, 15 randomly selected melodies were mapped to the test stimuli (TS) scale, and 45 randomly selected melodies were mapped to both the high and low exposure stimuli (ES) scales. In mapping the melodies, the tonic (which was always the lowest note) in the original folksong was always mapped to the lowest scale tone in the novel scale, which was not modified. This was done to facilitate the perception of a tonal center, which may help listeners in forming a sense of higher or lower than normal.

The folksong melodies were mapped to the novel scales (whether test, high exposure or low exposure) as follows. A folksong with the same number of unique scale tones as the novel scale was randomly selected, using the criteria above. For example, if a test scale contained six tones, then, the corresponding randomly selected folksong also had to contain six unique scale tones. The lowest scale tone in the folksong was then mapped to the lowest tone in the novel scale, the second lowest to the second lowest, and so on. This permitted a consistent one-to-one mapping between the scale tones in the folksong and the novel scale.

In this way, "homologous" melodies were created in which the "tonic" pitches were the lowest pitch in both the source folksong melodies and in the mapped artificial-scale versions. This approach preserved whatever correlations may exist between tonic pitches and rhythmic or phrase-related patterns in the original folksong. For example, if tonic pitches tend to end phrases, then the designated "tonic" for the novel scales would also tend to end phrases.

(vi) "The scales should be average pitch height equivalent"

One final consideration relates to the absolute pitch height of a passage. Huron, Kinney and Precoda (2006) found that exact transposition of a melody influenced perceived affect. Namely, they found that transposing a melody down an octave caused it to be judged as more negatively valenced. In the current experimental design, note that the modification of pitches in a scale will very slightly increase or decrease the overall average pitch height. That is, manipulating relative pitch will also result in a slight change of absolute pitch, which may potentially influence affective judgments. In order to adjust for this possible confound, the ES scales were transposed upward or downward in order to have the same average pitch height as the TS scale from which they were derived. Pitch compensation was applied to the scale as a whole, rather than to the individual melodies. In practice, these compensating transpositions were quite small (typically less than 30 cents). The compensating transpositions did not seem easily detectable to the experimenters, even though they were conscious of the manipulations.

The eight sets of generated scales used in the experiment are given in Appendix A, while Figure 3 below summarizes the construction of the TS scales and the derived ES scales.

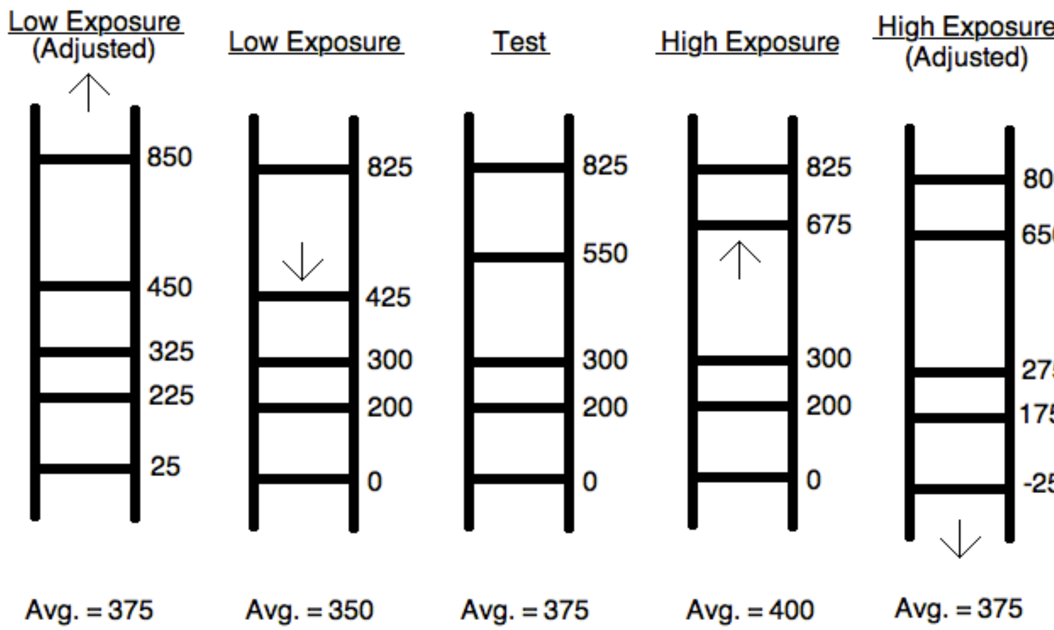

Figure 3. Schematic for scale construction. Each scale is represented by a ladder, with each scale tone represented by a rung. The central ladder, the TS (test stimuli) scale, is created first, with random values (in cents) for each rung, or note. To create the ES (exposure stimuli) scales, one of the rungs from the TS is randomly selected and then raised or lowered by a random amount, to create the high ES scale or the low ES scale. Then, both ES scales are slightly adjusted to maintain the same overall average pitch height. Only the TS scale and the adjusted ES scales are used for the melodies in the experiment.

3) Procedure

Sixteen participants were recruited from the Ohio State University School of Music subject pool. This was one of several experiments that could be selected by participants in order for them to receive partial course credit in a sophomore aural skills class. In the recruitment materials, participants were informed that the task would involve judging musical sadness in non-Western melodies. Participants were undergraduate music majors, 7 males and 9 females. All participants reported that they did not possess absolute pitch.

Participants viewed the following written instructions while the text was simultaneously read aloud by the experimenter:

INSTRUCTIONS

In this experiment, you will hear 60 short melodies. Your task is to rate how sad each melody seems to you. If you think the melody is very sad, you should rate it towards the right side of the scale. If it is not sad at all, you should rate it towards the left. After the melody is played twice, there will be an indication on the screen prompting you to respond.

Do you have any questions?

I will stay with you while you rate the first melody in case you need any help.

Of the 60 total melodies heard by each participant, 45 were exposure melodies; 30 of these exposure melodies were heard in the initial exposure phase – lasting roughly 15 minutes. In the subsequent test phase, the remaining 15 exposure melodies were randomly interspersed with the 15 test melodies – also lasting roughly 15 minutes. Paired participants heard the 60 melodies individually, which were about 30 seconds in duration. The experiment lasted about 30 minutes in total. Participants were tested in an Industrial Acoustics Corporation sound attenuation room. Stimuli were presented using a computer connected to loudspeakers. Participants responded after two presentations of each melody, using a mouse to click on the rating scale on a computer monitor. The linear rating scale had 7 discrete points, and only the end-points were labeled, with "Very sad" on the right and "Not sad at all" on the left.

Results

Since different participants may not have used the rating scale in a comparable way, normalized ratings were obtained. Ratings for the test stimuli (TS) were Z-normalized for each participant, using their means and standard deviations for ratings of the exposure stimuli (ES). 15 TS melodies were rated by each of the 16 participants, for a total of 240 TS sadness ratings.

Since the normalization was done using the ES melodies, the normalized mean sadness rating for all participants for the 480 (16 x 30) ES melodies was (obviously) zero. The mean sadness rating for all participants for the 240 TS melodies was -0.218. It appears, then, that regardless of the exposure condition, participants rated the test melodies as less sad than the exposure melodies. A t-test for independent samples with equal variances showed that this difference was significant ($t = -2.6186$, $df = 718$, $p = 0.009$). That is, melodies heard in the test phase were rated as less sad than melodies in the exposure phase, regardless of the condition.

In order to test the experimental hypothesis, low TS and high TS ratings need to be considered separately. Of the 240 sadness ratings for the TS melodies, 120 ratings were for low TS and 120 were for high TS. According to the experimental hypothesis, the low

TS melodies should be rated as more sad, while the high TS melodies should be rated as less sad. Indeed, the mean normalized sadness rating for the low TS melodies was -0.026, and the mean normalized sadness rating for the high TS melodies was -0.411. Thus, it appears that, consistent with the experimental hypothesis, participants who heard a higher exposure scale rated the test melodies as more sad than those who heard a lower exposure scale. A t-test for independent samples with equal variances showed that this difference was significant ($t = 2.5485$, $df = 238$, $p = 0.011$), with a 95% confidence interval of 0.087 to 0.682.

However, recall that the 240 TS ratings were not for unique melodies. Since the stimuli were uniquely created for each pair, there were only 120 unique TS melodies (15 x 8). That is, the TS melodies were identical for each participant within a pair. However, for the high ES participant, the identical TS melodies were relatively lower than the exposure (hence, low TS). For the low ES participant, the identical TS melodies were relatively higher than the exposure (high TS). Since the two sets of ratings correspond in this way, a paired t-test was conducted. The results were significant ($t = 2.7733$, $df = 119$, p-value = 0.006), with a 95% confidence interval of 0.110 to 0.659, with a mean difference of 0.385, consistent with the prediction that the low TS melodies would be rated as sadder.

In order to determine whether the assumptions of the t-tests above were appropriate, a Shapiro-Wilk normality test was performed. However, the null hypothesis that the TS ratings were normally distributed was rejected ($p < 0.0001$ for high ES participants; p <

0.01 for low ES participants). Furthermore, since each participant provided 15 ratings, the data might not be considered as independent. Consequently, an ANOVA was conducted.

First, a one-way, between-subjects ANOVA ($df = 1, 238$) was conducted for normalized sadness ratings and exposure condition (i.e., high or low). The results showed a significant difference between the two exposure differences ($F = 6.495$, $p < 0.01$). Next, a between-subjects ANOVA was conducted for normalized sadness ratings, using both exposure condition and subject as factors. The results showed a significant effect of the exposure condition ($F = 7.02$, $p < 0.008$), but also showed a significant effect by subject ($F = 2.37$, $p < 0.004$).

Discussion

An experiment was conducted pertaining to the perception of sadness in melodies. Specifically, the effect of relative pitch height was examined. Relative pitch height was differentiated from absolute pitch height, using a definition that effectively compared the scales or pitch collections for two melodies. If the scales were transposed such that they have the same tonic, the one with the lower corresponding scale tones was considered relatively lower. In this way, any minor scale could be considered relatively lower than any major scale. In Western cultures, where the major scale is normative, the minor scale may then be said to be relatively lower than normal. It was proposed that this relationship accounts for the sadness of the minor mode: it is lower in pitch than what is expected, and low pitch is an acoustic cue for sadness that correlates with low physiological arousal.

A study was designed to investigate this hypothesis. In order to avoid the possible confounds of enculturation, artificial scales were used instead of Western major or minor scales. Participants heard two kinds of melodies: exposure scale (ES) melodies and test scale (TS) melodies, and rated each melody for perceived sadness. ES melodies were heard in the first part of the experiment, in order to establish an expectation or norm. TS melodies were heard subsequently, and were either relatively higher or relatively lower than the ES melodies. It was predicted that the low TS melodies would be rated as sadder. A between-subjects design was used, in order to minimize confounds related to pitch manipulations. That is, the TS melodies were actually identical, but the ES melodies were manipulated such that the TS melodies were heard as either relatively higher or relatively lower for either participant in a pair.

After the TS ratings were Z-normalized for participants using their ES scores, results were obtained that were consistent with the main hypothesis. The relatively lower TS melodies were rated as sadder than the relatively higher TS melodies, with a significant average Z-score difference of 0.39. Each of the 240 TS melodies was rated twice, once in the low ES condition and once in the high ES condition. However, the mean Z-score difference is significant whether the data is paired or not.

Interestingly, all TS melodies were rated as less sad than the ES melodies on average, regardless of whether they were relatively higher or lower. Unexpectedly, the effect size was greater for the high TS melodies: the mean normalized sadness rating was -0.411 for the high TS melodies and -0.026 for the low TS melodies. Thus, the results were consistent with the conclusion that relatively higher pitched melodies are less sad, which is arguably a *post hoc* corollary to the original hypothesis. The results were also consistent with the original hypothesis, that relatively lower pitched melodies are more sad, but only when comparing the low TS melodies to the high TS melodies (that is, the *a priori* between-subjects test). However, when comparing the low TS melodies to the ES melodies for the same participant (a *post hoc* within-subjects test), the relatively lower-pitched melodies (the low TS) were less sad, contrary to the original "lower than normal" hypothesis.

It may be argued that the *a priori* between-subjects test (which was consistent with the hypothesis) should be given more consideration than the *post hoc* within-subjects test (which was not consistent with the hypothesis). It is important to note that for the between-subjects test, identical TS melodies were compared, such that any effects other than relative pitch height can be ruled out. Participant who heard the TS as relatively lower rated them as significantly sadder than those who heard the same TS as relatively higher (mean Z-score difference of 0.395). However, for the within-subjects test, the low TS melodies were compared collectively with the relatively higher ES melodies, which were entirely different melodies. Thus, factors other than relative pitch height may have been responsible for the different sadness ratings. That is, the effects of factors unique to each individual melody cannot be ruled out. Finally, note that this difference in sadness rating for low TS melodies and high ES melodies, although significant, was quite small (mean Z-score difference of 0.026).

Generally, it can be concluded that the results of the experiment are consistent with the hypothesis that melodies that have relatively lower pitch are heard as more sad. Since the results were obtained by using an exposure phase that established a pitch norm, it may be added that the results were consistent with the hypothesis that “lower than normal” melodies are heard as sadder. Lastly, the results were also consistent with the corollary that melodies with relatively higher pitch are heard as less sad.

Considering “Higher is Happier”

The above conclusions notwithstanding, the obtained differences in effect size for raised pitch versus low pitch in the *post hoc* within-subjects test were unexpected. There was no *a priori* reason to expect that the effect size for high TS would be an order of magnitude greater than the effect size for low TS. In retrospect, at least two accounts may be given for this difference in effect size magnitude. The first concerns asymmetries in the rating scale, and the second, asymmetries in pitch perception.

First, consider that participants were prompted to judge “how sad this melody seems to you,” on a response scale that was marked from “Not sad at all” to “Very sad.” The results above may thus be reframed: high TS test melodies were rated “not sad at all” much more often than low TS test melodies. Critically, note that the “Very sad” end of the scale identifies the emotion perceived, but the “Not sad at all” end, being a negation, does not. Thus, there is an asymmetry in terms of specificity: while the “Very sad” label points to a specific emotion, the “Not sad at all” label is nonspecific.

As noted previously, the minor mode is not only associated with sadness. A variety of other emotions, such as such as “soberness,” “mysteriousness,” or “seriousness” are also associated, although to a lesser degree (Hevner, 1935). It seems reasonable to assume that relatively lower pitch, in a similar way, may convey a variety of emotions in addition to sadness. If this is the case, by specifying “Very sad” on the scale, the effect of relatively lower pitch may have been muted. That is, participants may not have rated the melodies as “sad,” if instead they were perceived as “mysterious” or “serious,” even though this perception might be an effect of relatively lower pitch.

By contrast, this issue does not arise for the other “Not sad at all” label. While it could have been interpreted specifically as “less sad,” that the specific emotion of sadness was diminished, it could also have been interpreted as “more something else.” Thus, “Not sad at all” may have served as a catchall category for any perceived emotions other than sadness. Thus, it was less likely to fail to capture any effects of relatively higher pitch, which may include emotions such as “happy” or “exhilarated,” but also affects such as “mysteriousness” or “seriousness.” Under this first account, then, the asymmetry of the results could be due to the asymmetry of the rating scale: the more inclusive label was simply more applicable, even if the actual effects of pitch change were comparable in either direction. “Not sad at all” simply describes more emotions than “Very sad.”

A second account considers the asymmetry in the way the altered melodies were perceived. Indeed, the manipulations of the scales were symmetrical when measured in cents: whether they were raised or lowered, the altered notes were the same notes, and they were altered by the same amount in either direction. However, it does not follow that the alterations were perceived in the same way – perceptual salience need not be identical. If raised pitches were more salient than lowered pitches, this may account for the greater observed effect size of the high TS melodies. There may be reason to consider this asymmetry in pitch perception: in speech, pitches tend to decline as a natural consequence of exhaling, whereas a rising pitch is a relevant auditory cue. Thus, the greater effect size for high TS melodies over low TS melodies may be due to differences in the perception of the direction of pitch alteration, where the reduced effect of lowered pitches is simply due to their lesser salience.

Accounting for the Sadness of the Minor Mode

In this chapter, an explanation was offered for the association of the minor mode with sadness. Within the communication model, it was one of several acoustic cues that could convey sadness. Like the rest of the acoustic cues, its role is probabilistic: the minor mode is neither necessary nor sufficient for conveying sadness, but, especially when used in conjunction with the other cues, can reliably convey sadness for enculturated listeners. Unlike other acoustic cues, the minor mode had not been previously linked to low arousal as an explanation for its association with sadness. Three such factors of the minor mode were conjectured.

The first pertained to average size: the construction of the minor scale, in relation with typical melodic paradigms of tonal music, tends to reduce the average interval size of those melodies. Since a reduced average interval size, or low pitch variability, is a characteristic of sad speech, it is possible that this factor contributes to the suitability of the minor mode to convey sadness. The second factor, though not explored in this paper, pertained to descending intervals: if descending contour is a characteristic of sad speech, and if the minor scale is apt for scalar descent, this may be another contributing factor. The third factor, the topic of the current study, pertains to relatively lower pitch, or lowered scale degrees. All of these potential factors are in relation to the normative major mode.

The present experimental results were consistent with the hypothesis that the relatively lowered pitch of the minor scale may contribute to its suitability for learned associations with sadness. However, alternative interpretations can be proposed. At one extreme, there may be no relationship at all: even if lowered pitches did elicit an emotional response of sadness, the sadness of the minor mode may be due to something entirely different. At another extreme, there may be a causal relationship: perhaps lowered pitch can be directly perceived in the minor mode, which causes the perception of sadness. Alternatively, it may be the case that lowered pitch is only involved with the historical development of such an association, as a "bootstrap" that is no longer perceived by modern listeners. Since the role of learning and culture is certainly central, a relatively moderate conclusion is suggested here. The lower than normal pitch of the minor mode may be one factor that facilitated its culturally developed associations with sadness.

Other researchers have approached this topic in a different way. Temperley and Tan (2013) conducted an experiment where listeners judged the happiness of melodies in different church modes. Although happiness is not necessarily the opposite of sadness, they are at least negatively correlated. They found that the Ionian mode was judged the happiest sounding. As the number of flats was increased, starting with the Mixolydian mode and ending with the Phrygian mode, happiness ratings decreased. At face value, this result is consistent with the results and conclusions above. However, Temperley and Tan do not favor the "relative pitch height" account. (They refer directly to an earlier conference presentation on this topic by Huron, Yim, & Chordia, 2010).

First, they noted that the Lydian mode, which is the only mode with a sharp, and which is relatively higher than the Ionian, should be rated as even happier than the Ionian mode. However, it was judged as less happy than the Ionian mode, and approximately as happy as the Mixolydian mode. To account for this, they posit the familiarity effect: the most familiar mode, Ionian, is judged as most happy. Thus, since the Lydian mode (as well as the other modes) is less common and therefore less familiar, it was judged as less happy. However, they noted that the familiarity account was dissatisfying, because the Lydian mode was rated more happy than would be expected based on its relatively rarer frequency.

Temperley and Tan's preferred account for the findings is the “line-of-fifths” representation of the diatonic scale (Temperley, 2001), where the diatonic collection can be represented by seven adjacent pitches in the circle of fifths. Each diatonic mode specifies which of the seven pitches should be tonic, in the order Lydian, Ionian, Mixolydian, and so on to Phrygian – this is, of course, the same ordering as “adding flats.” Temperley and Tan (2013) note that this account makes the exact same predictions as the relative pitch height account, and similarly fails to account for the decreased happiness of the Lydian mode.

They then provide an argument for the advantage of the line-of-fifths model over the relative pitch height model. Even though both accounts make the same predictions for happy or sad connotations of the diatonic modes, they make different predictions for the pentatonic modes. In Figure 4 below, reproduced from Temperley and Tan (2013), the five pentatonic modes are listed. In the line-of-fifths account, moving from left (Ionian) to right (Phrygian) is equivalent to “adding flats,” and should have decreasing perceived happiness (or increasing perceived sadness). This is consistent with musical intuition.

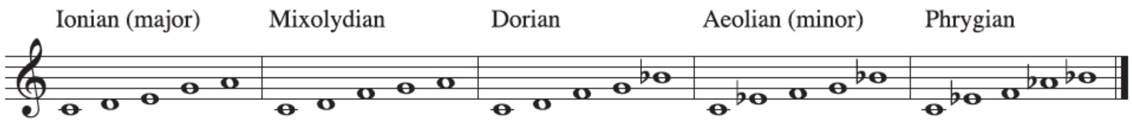

Figure 4. Five pentatonic modes. Modes left-to-right represent decreasing perceived happiness according to the line-of-fifths model (Temperley & Tan, 2013).

However, they argued that the relative pitch height account would predict the exact opposite. For example, the second pentatonic mode (Mixolydian) may be considered a relatively higher version of the first pentatonic mode (Ionian), by raising the third scale scale degree, from E-natural to F-natural. Thus, they argued that moving from left to right would be an increase of relative pitch height, and consequently, the relative pitch height model would predict an increase in perceived happiness (or decrease in perceived sadness) from left to right. Since this prediction clearly goes against musical intuition, Temperley and Tan argue that their account is preferred.

Two rebuttals are offered for this somewhat convincing argument. First, the account of relative pitch height, as described in this chapter, does not necessarily make the opposite prediction as was claimed. Recall that it was argued that the sadness of the minor mode was dependent on the culturally normative major mode. If the major pentatonic scale is not normative, it would not be expected that any other pentatonic scale would be compared to it, and thus no prediction would be made at all concerning their relative happiness or sadness. In fact, since the diatonic major scale is normative for Western listeners, one might expect that the pentatonic modes would be heard relative to the diatonic major scale instead. Then, moving from left to right would result in decreasing relative pitch height, resulting in predictions that would be the same as those of the line-of-fifths account.

Second, Temperley and Tan acknowledge that while there is evidence for the line-of-fifths account in general, there is no empirical or theoretical work linking it to emotions. In other words, while this account may be predictive, it does not necessarily explain much about emotional responses. In contrast, the present work associates the minor mode with other known acoustic cues for sadness, namely lower pitch. Lower pitch, along with a variety of other cues including quieter sound and slower tempo, can be linked to low physiological arousal, which is a component of sadness.

Summary

In this chapter, the acoustic cues of sadness were considered, including slower tempo, softer dynamics, and lower pitch. Additionally, it was noted that while the minor mode was a cue for sadness in music, it could not easily be related to lowered arousal or sad speech. It was proposed that the minor mode may have certain features that make it particularly suitable or apt for learned associations with sadness: minimizing pitch variability, a (conjectured) usefulness for descending contour, and lower-than-normal pitches – when compared to a culturally normative major mode.

In the present study, an experiment was conducted to investigate whether relatively lower pitches could elicit perceptions of sadness. Artificial scales were generated, and a between-subjects paradigm was used. An exposure phase was used to establish a pitch norm. It was found that test melodies that were relatively lower than exposure melodies were perceived as sadder. The results are consistent with the hypothesis that the minor mode may elicit sad associations because it sounds “lower than normal.”

Chapter 3: Affective Priming Study

The experiment described in the previous chapter relied on the subjective responses of participants. Such introspection is rather indirect: it requires participants to observe their own emotions, and then convey these observations to the experimenter. Can emotional changes somehow be observed more directly? In this final chapter, an alternative to subjective responses is explored. The affective priming paradigm will be introduced, and then considered for use as an implicit measure for assessing emotional responses to music. An exploratory study is presented that uses major and minor chord progressions as priming stimuli for an evaluative decision task: task response times are used as an implicit measure for emotional response to the chord progressions. It was found that major and minor chord progressions did not significantly facilitate performance in the evaluative decision task, although the results were skewed in the predicted direction. Furthermore, the present experimental design did not seem to be a reliable means of evaluating emotional responses.

Background

As discussed in Chapter One, implicit measures make use of observable cognitive or behavioral processes to assess mental states that cannot be directly observed. For example, Riener, Stefanucci, Proffitt, and Clore (2011) found that an observer's estimatation of the incline of a hill could be biased by affective state. Observers induced with sad mood (incidentally, using music) estimated that hills were steeper than those in the control group. In this example, the estimates of hill incline were an implicit measure for the induced sadness of the observers. A thorough review of implicit measures in the context of music is given by Västfjäll (2010).

Affective Priming

The paradigm for affective priming (also known as "automatic attitude activation") is not usually described as an implicit measure. However, after introducing how the effect operates, it will be argued that the affective priming paradigm is potentially useful as an implicit measure, to be used as an alternative to subjective self-reports for assessing emotional responses to music.

The affective priming effect is related to the better-known semantic priming effect, first described by Meyer and Schvaneveldt (1971). A pair of letter strings were simultaneously presented visually to participants, who had to press a certain key if both strings were English words, and another key if they were not. It was found that participants responded to semantically related words both faster and more accurately. That is, task performance was better on word pairs such as "bread / butter" than on pairs such as "bread / doctor." (For a review of semantic priming, see Neely, 1991).

Affective priming studies are similar to the semantic priming studies, except that the facilitating words are affectively related, rather than semantically related. This affective relation between words lies along the valence dimension, usually with a "good versus bad" or "like versus dislike" dichotomy. For example, in affective priming, task performance on a word pair such as "cockroach / ugly" (negative and negative) would be better than on a word pair such as "cockroach / comfortable" (negative and positive). Note that this effect is observed even though the primes are not informative about the subsequent targets: positive or negative targets are equally likely, regardless of the target that precedes them.

Another difference is that, instead of simultaneous word presentation, pairs of words are presented in sequence. The first word, the "priming stimulus," is presented for a short duration (less than 500 ms). The second word, the "target stimulus," would then appear after a short interval (between 300 to 500 ms). The task involves an evaluation of the target only, although several different tasks may be used. In the evaluative decision task, participants must categorize target words as having positive or negative valence (i.e., "good" versus "bad", or "like" versus "dislike.") In the lexical decision task, targets could be letter strings or words, and participants must categorize the target strings as words or non-words. In the pronunciation task, participants simply had to say the target word aloud. An affective priming effect is observed when task performance is faster and more accurate whenever the prime and target words are affectively congruent. Klauer (1997) and Fazio (2001) provide extensive reviews on affective priming paradigms.

The results of numerous affective priming studies have contributed indirect evidence to support the view that affective judgments are fast, automatic and unconscious (e.g. Hermans, De Houwer, & Eelen, 1994). This inference is based on the observation that the affective priming effect is only observed when the processing time is very short, making it impossible to employ conscious task strategies. For example, researchers found that presentations of the prime that were as short as 4 ms could produce the affective priming effect. However, when the prime presentation was lengthened to 1000 ms, the affective priming effects were not observed (e.g.:, Zajonc, 1980; Rotteveel, Groot, Geutskens, & Phaf, 2001). Additionally, while the affective priming effects did not persist with long primes, semantic priming effects did persist, consistent with the interpretation that affective judgments were fast, automatic and unconscious (e.g. Murphy & Zajonc, 1993).

Similarly, previous research has reported that the duration of the interval between prime and target was required to be short, if an affective priming effect were to be observed (Hermans, De Houwer, & Eelen, 2001). They found that the priming effect was robust for interstimulus intervals of less than 150 ms, but not for intervals of 1000 ms. Again, this is consistent with an argument for fast, automatic, and unconscious affective processing. In the same vein, the response interval (the time between the target presentation and the evaluative response) is also short whenever the affective priming effect is observed.

Mechanism for Affective Priming

There are two main accounts by which the affective priming effect is explained. In the first account, the effect is caused by "spreading activation" of associated concept nodes (e.g. Collins & Loftus, 1975; Bower, 1981; Murphy & Zajonc, 1983). In this view, the mental activation of the concept represented in the prime will facilitate the subsequent activation of congruent target concepts. In other words, this view assumes that valence-congruent concepts have overlapping mental representations (at least to a greater extent than valence-mismatched concepts).

In the second account, the affective priming effect is attributed to differences in perceptual fluency (e.g.: Reber, Winkielman, & Shwarz, 1998; De Houwer, Hermans, Rothermund, & Wentura, 2002; Logeswaran & Bhattacharya, 2009). In this view, both the prime and target stimuli induce a response tendency for "positive" or "negative" valence. Therefore, non-congruent stimuli cause two conflicting responses, which require an increased cognitive load to resolve, resulting in a slower overall response. This explanation is analogous to the one for the Stroop task, where the need to resolve conflicting ink color and color word may cause decreased reading speed (Stroop, 1935).

Other Modalities for Affective Priming

Affective priming effects have also been observed when stimuli other than words were used, as either the prime or target stimuli. For example, Spruyt, Hermans, De Houwer, and Eelen (2002) found that pictures of objects were actually more effective prime stimuli than words. They argued that this was because pictures were less likely to be processed semantically (and therefore more likely to be processed affectively). As another example, a cross-modal affective priming effect was observed by Hermans, Bayens, and Eelen (1998) using pleasant and unpleasant odors as prime stimuli (although the effect was only observed for females). Notably, this is an example of an affective priming effect obtained using a longer duration prime: the odors were circulated in the air for 10 seconds before presentation of the target words.

In the auditory domain, Goerlich, Witteman, Schiller, Heuven, Aleman, and Martens, (2012) observed the affective priming effect using 600 ms excerpts of either music or speech, which were used as either priming or target stimuli. For music specifically, two studies observed the affective priming effect when musical chords were used as prime stimuli. Sollberger, Reber, and Eckstein (2003) used consonant or dissonant chords as primes in an evaluative decision task, and found that consonant and dissonant chords improved task performance for positive and negative words respectively. Their results are reproduced in Figure 5 below. Similarly, Steinbeis and Koelsch (2010) observed the affective priming effect for consonant and dissonant chords, and also for major and minor chords (congruent with positive and negative words, respectively), and also for pleasant and unpleasant timbres.

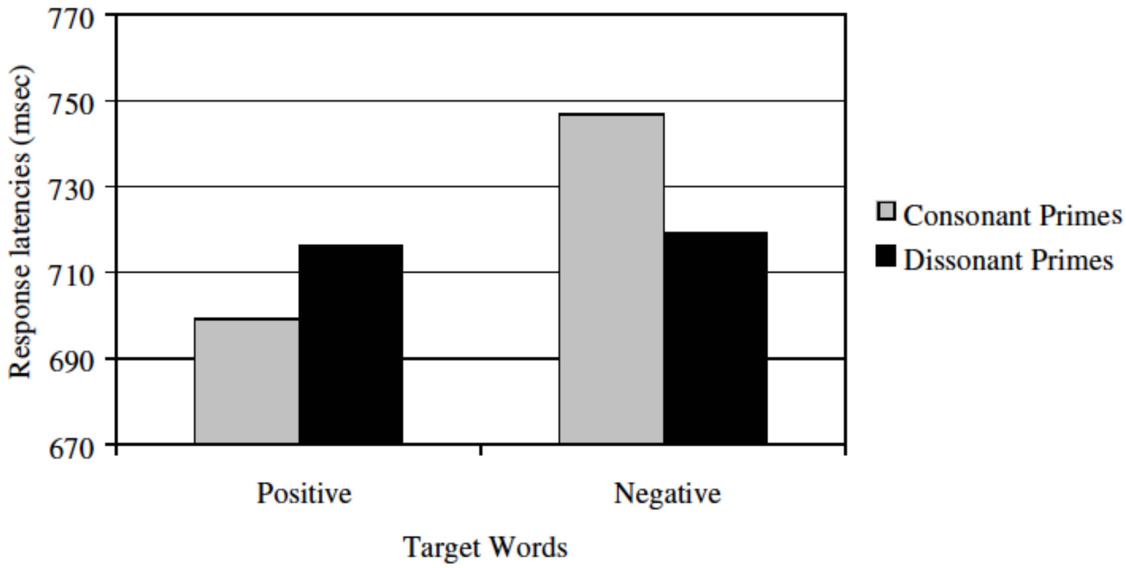

Figure 5. Results of Sollberger et al (2003). Mean response times are shown for consonant versus dissonant primes, and positive versus negative targets. Response times are faster in congruent conditions (consonant/positive and dissonant/negative) than in non-congruent conditions (consonant/negative and dissonant/positive).

Using Affective Priming as an Implicit Measure

The demonstrated robustness of the affective priming effect in numerous and varying studies is suggestive of a further application. It has been shown that task performance depended on whether or not the prime and target stimuli were congruent. Thus, good performance on the task may be used to infer a congruent prime, while poor performance on the task may be used to infer a non-congruent prime. In this way, the affective priming paradigm may be construed as an implicit measure. That is, response times and accuracy of the target stimuli in an evaluative task can be observed and used to infer the unobserved affective response to the prime stimuli. Although this is not a common application, the use of affective priming as an implicit measure in this way has proved successful in a number of studies (e.g.: Hermans, Vansteenwegan, Crombez, Baeyens, & Eelen, 2002; Hermans, Spruyt, & Eelen, 2003).

In considering affective priming for music, a critical methodological issue may be foreseen. Recall that effective prime stimuli for affective priming have generally been found to be shorter than 1000 ms. Since longer primes were effective for semantic priming, it has been argued that affective processing is fast, automatic and unconscious. However, interesting musical sounds tend to be much longer than this, on the order of seconds or minutes. Hence, musical excerpts may not be successful for affective priming: since they are (very) long primes, conscious, non-affective processes may control response patterns rather than affective processes. Note that for the previously described affective priming studies that used musical sounds, only very short excerpts (600 ms) or single chords were used as primes.

The requirement for short primes suggests that automatic affective processes are involved, and it may be the case that these processes are disrupted when enough time is allowed for conscious, non-affective processing. Thus, one may conjecture that the affective priming effect may be obtained for music, if only there were a means to prevent conscious contemplation during the presentation of the musical priming stimuli. To this end, an experimental procedure was designed, with the goal of producing cognitive overload during presentation of the musical priming stimuli. Rather than evaluating a single word after the prime presentation, multiple words were to be quickly evaluated in sequence during the prime presentation.

This approach has other advantages, pertaining to the need for a short interstimulus interval and a short response interval in affective priming. The need for a short interstimulus interval is circumvented, since the target stimuli (the words) are presented simultaneously with the prime stimuli (the music). Additionally, the continuous sequence of target words, presented in rapid succession over the duration of the prime, may encourage participants to respond as quickly as possible, minimizing response intervals.

Implicit Measures versus Subjective Measures

For the purpose of assessing emotional responses to music, the use of implicit measures (such as the proposed affective priming paradigm) may serve to complement the use of subjective measures. First, it was noted previously that subjective measures are limited by participants' abilities of introspection and articulation. Implicit measures avoid this limitation by not requiring either. On the other hand, it presently seems unclear whether implicit measures are assessing perceived or felt affect. Convergent evidence from other sources, such as subjective (or physiological) measures, may resolve this issue.

Second, it was noted previously that subjective measures are particularly susceptible to demand characteristics. Implicit measures, by contrast, tend to be resistant to demand characteristics. If participants are asked to respond as quickly as possible (as in the affective priming paradigm), unusual response strategies may be difficult to execute. This can be important especially if the experimental hypothesis is easily discernible.

Third, results from subjective ratings may not be easily verifiable. Since there may not be a "ground truth" for subjective emotion ratings, it can be difficult or impossible to determine whether an unexpected result is a genuine anomalous outlier, or simply an error or other artifact of inattention. These limitations may be particularly relevant for less motivated volunteers who may not devote as much effort or attention to an experimental task. Using implicit responses, however, objective *a priori* exclusion criteria can be employed. For example, with affective priming, unusually slow response times can be used to reject individual responses from otherwise good response sets, while unusually low accuracy scores can be used to reject the response sets of inattentive participants.

Exploratory Study

The goal of the exploratory study was to design an implicit measure that could be used to complement subjective measures for emotional responses to music. The procedure was based on an affective priming paradigm. In order to accommodate the typically lengthier nature of musical excerpts, it was necessary to encourage fast processing and fast responses. Thus, during the presentation of the prime stimuli (which were musical sounds), the target stimuli (words) were presented in a continuous sequence. In order to further encourage fast responding, the entire procedure was framed as a game, where faster and more accurate responses scored more points. Thus, it was hoped that the required fast categorization of the words (into positive, negative, or neutral categories, i.e., the evaluative decision task) would draw the participants away from conscious contemplation of the musical prime stimuli, which might be construed of as background music for the game.

Since the primary aim of the study was to explore the applicability of affective priming as an implicit measure, a well-documented and reliable association was tested. As previously noted, an affective priming effect was found for chords as primes, for the congruent conditions of major/positive and minor/negative (Steinbeis & Koelsch, 2010). However, the Steinbeis and Koelsch study used short single chords, which are not representative of the length of actual musical excerpts. For this reason, chord progressions were used as the musical primes instead.

Participants

Thirty-three participants (14 male, 19 female) received partial course credit in a second-year aural skills class for their participation in the study.

Materials

Experimental materials consisted of 1) musical sounds used for the affective priming (prime stimuli), and 2) word list used for the word categorization task (target stimuli).

1) Musical Sounds (Priming Stimuli)

There were five kinds of musical sounds used as the priming condition: 1) No music (N), 2) ordered chord sequences in a major key (O+), 3) ordered chord sequences in a minor key (O-), 4) random chord sequences using only major chords (R+), and 5) random chord sequences using only minor chords (R-).

The "no music" (N) condition was used as a control condition. The ordered chord conditions used typical chord progressions in either C major or C minor, while the random chord conditions used scrambled sequences of chords not intended to convey any strong tonal center. Each condition was comprised of four different chord sequences of the appropriate type, and each chord sequence was seven chords long, with the last chord held for two beats. Since each sequence was performed at 60 bpm, each sequence lasted 8 seconds. Each condition, therefore, lasted 32 seconds, including the "no music" condition. Chord sequences were played in MIDI format using an acoustic piano sound. Participants heard the sounds via headphones, and they set the volume for themselves at a comfortable listening level. The resulting loudness levels were not measured or recorded.

The ordered chord sequences, either major (O+) or minor (O-), were composed with the goal of evoking major or minor tonal centers. In order to counter-balance the stimuli, a reciprocal procedure was used. First, 5 progressions in C major were composed, and then transposed to C minor. Similarly, 5 C-minor progressions were composed then transposed to C major. Overall, 10 C-major progressions (O+) and 10 C-minor progressions (O-) were created. All of the chord progressions are listed in Appendix B.

From these 20 ordered chord progressions, 20 randomized chord sequences were then created. For these scrambled sequences, which were not intended to strongly convey a tonal center, chords were selected randomly from the 20 ordered progressions. The ordered major and the ordered minor progressions, of course, included both major and minor chords. Together, the 20 ordered (major and minor) chord progressions included 84 chords with a major triad on the chord root (i.e., major triads, dominant sevenths, and their inversions), and 56 chords with a minor or diminished triad on the chord root (i.e., minor triads, minor sevenths, and half-diminished chords), for a total of 140 chords. Each of the ten random major chord sequences was created by randomly selecting seven chords, without replacement, from the 84 “major” chords. Since 70 minor chords (10 sequences x 7 chords) were required, the 56 “minor” chords were duplicated to create 112 chords. Each of the ten minor chord sequences was then created by randomly selecting seven chords, without replacement, from these 112 “minor” chords.

Figure 6 below shows the four kinds of sequences created:

Figure 6. Sequences used in priming conditions. The ordered sequences were typical chord progressions in C major or C minor. The random sequences used random major or random minor chords, which were sampled from the ordered sequences. The fifth condition, silence, is not represented here.

Four sequences of the same kind were concatenated to create one priming condition, and each priming condition was presented five times. The sequences were selected at random, such that each sequence was never played consecutively, and each sequence was selected twice. Again, there were five priming conditions:

1) No music (N)

2) Ordered major (O+), each with four randomly selected C-major progressions

3) Ordered minor (O-), each with four randomly selected C-minor progressions

4) Random major (R+), each with four random sequences of “major” chords

5) Random minor (R-), each with four random sequences of “minor” chords

Each priming condition was presented five times, for a total of 25 trials, each lasting 32 seconds.

2) Word List (Valenced Target Stimuli)

Target words were selected from the Affective Norms for English Words list (Bradley & Lang, 1999). The ANEW list includes valence scores for English words, collected by subjective rating surveys. At the time the database was accessed, there were 2,474 words in the list, with ratings for males and females available separately. Although arousal and dominance ratings for the words were also available, they were not used. The objective was to use the ANEW list to assemble target words that could be strongly categorized as either positive, negative, or neutral.

Due to possible gender differences in perception of word valence, two target word lists were assembled. The word selection procedure was done separately, for male ratings and for female ratings: first, the word list was sorted by valence ratings. Then, the 150 words with the highest valence rating were selected as potential positive words, while the 150 words with the lowest valence rating were selected as potential negative words. Finally, the 150 words in the middle of each sorted list were selected as potential neutral words. Thus, two sets of 450 potential words (having 229 words in common between them) were selected.

A screening process was used to ensure that only words that were readily categorized as positive, negative or neutral would be included as target stimuli. Additionally, as a precaution, words that may be potentially offensive to some students were eliminated. To this end, two people of the corresponding gender evaluated each of the potential word lists. They were asked to categorize each of the 450 words as positive, negative or neutral, using their “initial feelings of the word” and “[not thinking] too much about any single word.” Additionally, in order to conform to our Human Subjects (IRB) Protocol prohibiting degrading or offensive stimuli, they were instructed to flag any words that may seem offensive either to them, or to undergraduate students.

To minimize word ambiguity, any words that were categorized incorrectly by any evaluator were excluded (e.g.: “god,” “rebuff,” and “rollercoaster”); in this way, 109 words in total were excluded. Additionally, any words that were flagged as potentially offensive were also excluded; in this way, 18 words were excluded. The excluded words are listed in Appendix C.

For both males and females, there were at least 100 words remaining for each valence. Thus, 100 words of each valence were selected for both, totaling 300 words each. The 100 positive words with the highest valence, the 100 negative words with the lowest valence, and the 100 neutral words with the nearest-to-middle valence were included. For both males and females, positive words included “SUCCESS” and “VICTORY,” and negative words included “DEATH” and “POVERTY.” For males, neutral words included “ACRE” and “ZIPPER.” For females, neutral words included “ADHERE” and “YEAR.” All of the selected words are listed by valence and gender in Appendix C.

Procedure

Participants were instructed to categorize the selected words into valence categories as quickly and accurately as possible. Different musical sounds were presented while the participant engaged in the categorization task. The task was presented on a computer, which coordinated the musical stimuli presentation with the word categorization task.

Musical Chord Sequences (Priming Conditions)

The experiment involved five different priming conditions: no music (N), ordered major key sequence (O+), ordered minor key sequence (O–), random major chord sequence (R+) or random minor chord sequence (R–). Each priming condition was presented five times, for a total of 25 presentation blocks, each 32 seconds long. Twelve words were presented during each block, which participants categorized as positive, negative or neutral using the computer keyboard (described below). If the words were categorized before the full duration of the presentation block, the block ended. If the words were not yet categorized after 32 seconds, the testing continued in silence. The sequence of the segments was randomized for each participant, with the first segment always N (no music), and consecutive segments were never the same.

The segments for condition N had no music, and therefore consisted of simply 32 seconds of silence. In the other priming conditions (either O+, O–, R+ or R–), four different sequences of the appropriate kind were played consecutively. Each sequence was 8 seconds long (i.e.: seven-chord sequences at 60 bpm, with the last chord held for two beats). Therefore, these segments were also each 32 seconds long. The same sequence was not played more than once in any given segment. In between segments, participants were given feedback on their performance for that segment only (described below). Whenever they were ready, they could press a key to go on to the next segment.

Word Categorization Task (Targets)

The word categorization task was carried out simultaneously with either the presentation of musical sounds or silence, for the 25 testing segments. Each word was categorized only once by each participant. There were two word lists, one for males and one for females. Either list contained 300 prescreened words from the ANEW word list: 100 positive, 100 negative, and 100 neutral. In each of the 25 segments, the participant categorized 12 words as positive, negative or neutral. There were 3, 4, or 5 words of each valence type per segment, but for each of the 5 conditions (presented 5 times each), there were a total of 20 positive words, 20 negative words and 20 neutral words.

Participants were instructed to classify each word that appeared on the screen as positive, negative or neutral, and to do so as quickly and accurately as possible. Each segment started by awaiting a key press from the participant to ensure that they were ready, with the text "Press LEFT, DOWN, or RIGHT to begin." When the key was pressed, the appropriate musical sequence was played (for the O+, O–, R+, or R– conditions) or there was silence (N condition). While the sequence was still playing (or during silence in the N condition), a sequence of 12 words appeared on the screen to be categorized by the participant.

A trial consisted of categorizing one of the 12 words in each segment. Each word was presented in the center of an otherwise blank screen. Participants decided how each word should be categorized, then responded using the keyboard. If the word was positive, they were to press the right arrow key; if negative, the left arrow key; and if neutral, the down arrow key. They were instructed to make this categorization as quickly and accurately as possible. Response time was measured as the difference between the onset of the word and the timing of the key press. Responses were counted as correct if the first key pressed corresponded to the designated valence. As soon as a response was made, the current trial ended, and the word was cleared from the screen. The trial also ended if no response was given after 2000 ms (although only response times within 300 ms to 1500 ms were used in the analysis). To minimize entrainment effects, the interval between each trial varied randomly in duration from 1000 to 2000 ms. Thus, in each segment, one of 5 priming conditions was presented, while participants categorized 12 words as positive, negative or neutral. Each priming condition was presented five times; thus, there were 300 words in total presented over 25 segments, with response times and response correctness collected for each word.

Scoring System

In order for reaction times to be useful implicit measures, they should not be too fast or too slow. Reaction times that are too fast may indicate guessing, as may those that are too slow. Reaction times that are too slow may also indicate non-affective task strategies. Thus, it was necessary to motivate participants to perform the word categorization task as quickly and accurately as possible. To this end, the experiment was framed as a game in which the goal was to achieve the highest score. Points were earned when words were categorized correctly, but more points were earned when this was done quickly as well.

If a correct response was given within 2000 ms of word onset, points were awarded. If the correct response was given at 400 ms or earlier, the maximum of 100 points was awarded. After 400 ms, the points awarded decreased by a fractional amount (9/160 points per ms), such that if the correct response was given at 2000 ms, the minimum of 10 points was awarded. In summary, the points awarded for a correct response was given by the following equation, where x was the amount awarded, and t was the amount of time (in ms) that had elapsed since the word had appeared:

$$\text{for } t <= 400: \quad x = 100$$

$$\text{for } t > 400: \quad x = 100 - (9 * (t - 400) / 160)$$

However, if the categorization was incorrect, 150 points were deducted. Note that points could be awarded even if the response was made outside of the 300 to 1500 ms window used for data analysis, as the scoring system was only intended to encourage participants to respond as quickly as possible.

Prior to the experiment, there was a practice round to help participants familiarize themselves with the task as well as with the scoring system. During the practice round, participants got feedback after each response, with the points earned or lost displayed on the screen. During the actual experiment, however, feedback was only given after each segment. Participants viewed a screen with the following information: the average points gained for correct words for that segment (presented as “Average Speed %”), the number of correct responses out of 12 (“Accuracy”), the net points gained or lost that segment (“Round score”), and the cumulative points earned for all the rounds thus far (“Total score”).

Instructions

All instructions were displayed on the computer screen and read aloud to the participants. The initial instruction concerned setting the volume level:

```
We will begin by setting a comfortable volume.
Press SPACEBAR, put on the headphones, then adjust the volume.

After you have finished adjusting the volume, remove your headphones, then
press SPACEBAR.
```

As soon as the spacebar was pressed, a simple chord progression was played using the same settings as the experimental stimuli. This was important because, while participants should be able to set a comfortable listening level, they were not permitted to turn the volume down later in order to make the task easier (for example, if they found the music too distracting).

Next, the instructions for the task were given, again presented on the screen and read aloud by the experimenter:

```
INSTRUCTIONS
This experiment studies how listening to music affects your mental abilities.
You will be playing a word categorization game in which you earn points.
For each word, you must decide whether its meaning is associated with positive
emotions, negative emotions, or neutral emotions (or neither).
(Press SPACEBAR.)

EXAMPLES
Positive = friendly, hopeful, congratulations
Negative = hurt, disgusting, wreckage
Neutral = cylinder, decade, orange
(Press SPACEBAR.)

RESPONSE
Positive = Press RIGHT arrow key
Negative = Press LEFT arrow key
Neutral = Press DOWN arrow key
(Press SPACEBAR.)

SCORING
Each correct response earns from 10 to 100 points, with more points being
earned for faster responses.
If you guess incorrectly, you will lose 150 points. If you don't respond after
2 seconds it will also count as incorrect.
(Press SPACEBAR.)
```

WORD EVALUATION

In order to maximize your score, it is important that you categorize the words as quickly as possible.

Your decision should be based on the emotions that are most commonly associated with that word, by people of the same gender as you.

(Press SPACEBAR.)

MORE EXAMPLES

Positive = reward, delicious, freedom

Negative = humiliate, injury, worry

Neutral = bag, measure, owl

(Press SPACEBAR.)

AUDIO

Throughout the experiment you will hear various chord progressions or silence.

The purpose is to see how different background sounds affect your mental abilities.

However, in order to minimize bias, you should focus on the words, not the music.

(Press SPACEBAR.)

RESPONSE

Positive = Press RIGHT arrow key

Negative = Press LEFT arrow key

Neutral = Press DOWN arrow key

(Press SPACEBAR.)

Do you have any questions?

There will be a Practice Round first.

When you are ready, put on the headphones and press SPACEBAR.

After the instructions were given, participants could ask clarifying questions. There was a short practice round, where the experimenter observed the participant to ensure that they understood the task. The practice round was similar to the main experiment, except that only one ordered major progression and one ordered minor progression was presented. Only six words that were not used in the main experiment were used in the practice session, and the data for the practice round was not analyzed. Additionally, participants got feedback after each response during the practice round, whereas they got feedback only after each round (i.e.: each segment, after 12 words) during the actual experiment. After the experiment, participants were asked general questions concerning the hypothesis of the experiment and their task strategy, before being debriefed.

Predictions

The general affective priming paradigm predicts that reaction times for congruent stimuli are faster than for non-congruent or mismatched stimuli. In this study, the prediction was that major key progressions (O+) and random major chords (R+) would facilitate task performance with the positive valence words. Similarly, it was predicted that minor key progressions (O–) and random minor chords (R–) would facilitate task performance with negative valence words. That is, reaction times for positive words would be fastest for O+ and R+ conditions, while reaction times for negative words would be fastest for O– and R– conditions. Similarly, the accuracy for positive words would be highest for O+ and R+ conditions, while accuracy for negative words would be highest for O– and R– conditions.

Results

Participants had categorized words as either positive, negative or neutral, in five different conditions: 1) silent (N), 2) ordered major chord progression (O+), 3) ordered minor chord progression (O-), 4) random major chords (R+) or 5) random minor chords (R-). The mean response time was 823 ms, and mean hit rate (i.e. proportion of correctly categorized words) was 90.3%. Acccording to *a priori* exclusion criteria, response times faster than 300 ms or slower than 1500 ms were excluded, and incorrectly categorized words were also excluded. The mean response time after the exclusions was 788 ms.

In post-experiment interviews, participants were asked if the experiment was difficult. Some participants commented that categorizing neutral words was the most challenging aspect of the experiment. Indeed, it was found that neutral words had the slowest response time (mean reaction time for positive words was 761 ms; negative words, 771 ms; neutral words, 832 ms). Thus, the neutral condition was likely an inappropriate control condition.

It was possible that the presence of the neutral words in the task may have made evaluation of all the words more difficult, making it more likely that participants would use non-affective processing. Therefore, additional *post hoc* experiments were carried out, which were identical to the main experiment, except that words were only categorized as positive versus negative, and not neutral. Results for this two-alternative forced choice task ("2AFC," N = 9; 5 males, 4 females) will be analyzed concurrently with the three-alternative forced choice task ("3AFC," N = 33). Note that the 2AFC task also used 300 words in total, but since the available prescreened word lists only had 100 words for positive valence and 100 words for negative valence, half of the positive and negative words were presented twice instead of once, for a total of 150 positive words and 150 negative words. Response times were faster (637 ms vs. 788 ms) and more accurate (94.5% vs. 90.3%) in the 2AFC task. The mean response times for the 3AFC and the 2AFC tasks are given in Figure 7 below.

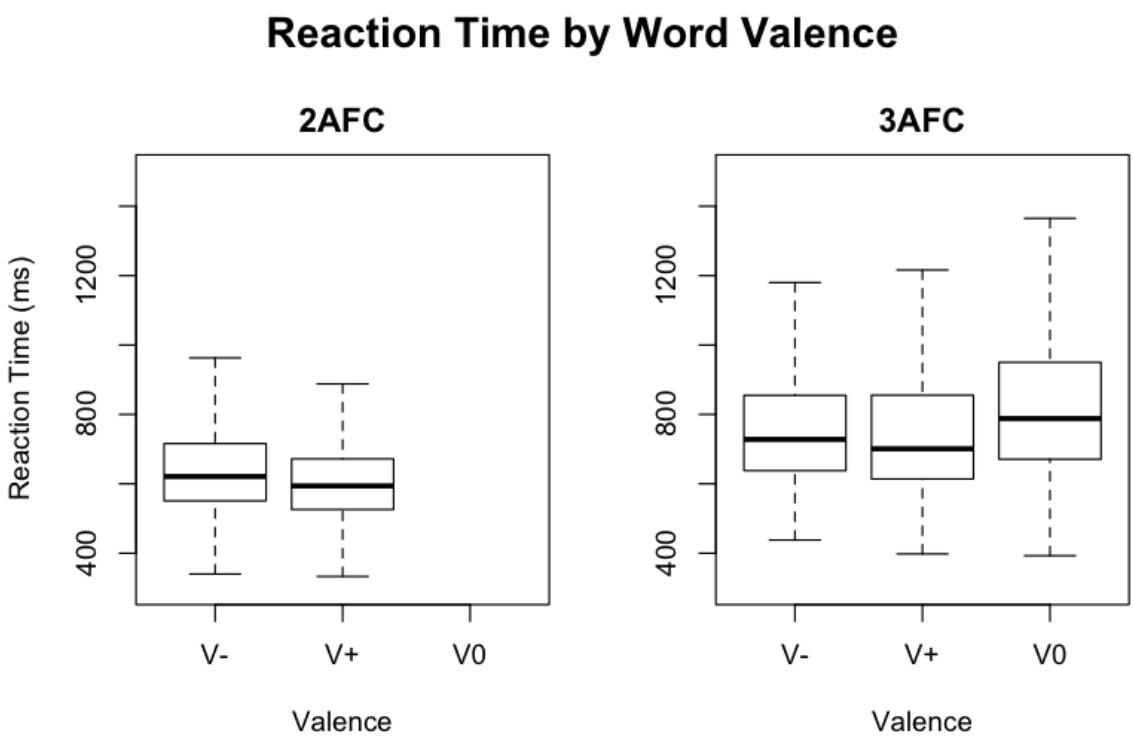

Figure 7. Reaction time by word valence and task type. Aggregate reaction times are displayed for negative (V-), positive (V+) and neutral (V0) valence words, in the two- or three-alternative forced-choice (2AFC, 3AFC) conditions.

Given that the neutral-valence words seemed to make the task more difficult for participants, they were not included in the analysis. Since there was no experimental hypothesis for the neutral-valence words, and since they were included only as controls, this *post hoc* analysis decision seemed justified. However, since the reaction times and hit rates for the 3AFC and 2AFC tasks were different, (consistent with the task-difficulty due to the neutral words), they were analyzed separately and not combined.

To test the priming effect, the silent condition was used as a control. For each participant, the mean reaction time was calculated separately for positive- and negative-valence words in the silent condition. This mean was subtracted from each of the reaction times for positive- and negative-valence words respectively in the other conditions (i.e., major ordered, minor ordered, major random, minor random). Thus, a response time difference greater than zero indicated a word that was slower than the same-valence word in the silent condition, and a response time difference less than zero indicates one that is faster.

The affective priming effect predicts that response times will be faster for words in congruent conditions than in non-congruent conditions. Recall that a congruent condition is a positive valence word in either the major ordered or major random condition, or a negative valence with minor ordered or random, and vice versa for the mismatched condition. For both the 3AFC and 2AFC tasks, the results were skewed in the predicted direction, with matched conditions non-significantly faster than mismatched conditions. In the 3AFC task, the mean reaction time was -48.1 ms for matched conditions, and -39.7 for mismatched conditions; the difference of means was not significant ($t = -1.4674$, $df = 4648.3$, p-value = 0.14), with a 95% confidence interval of -19.4 to 2.8 ms. In the 2AFC task, the mean reaction time was -21.9 ms for matched conditions, and

-14.4 ms for mismatched conditions; the difference of means was not significant ($t = -1.1628$, $df = 2035.9$, p-value = 0.245), with a 95% confidence of interval of -20.1 to 5.13 ms.

Although a significant affective priming effect was not observed, the results were skewed in the predicted direction. However, the main purpose in conducting the study was to explore the use of affective priming as an implicit measure in assessing emotional responses. Thus, a *post hoc* procedure was devised to determine if reaction time data within a segment could be used to predict the priming condition of that segment. Since there were no predicted differences for ordered versus random conditions, the test was simply whether the reaction time data could be used to predict the mode of the priming condition, as either major or minor.

First, the reaction times for each participant were Z-normalized separately for each word valence. Thus, for each participant, the mean normalized reaction time was zero for positive words, for negative words, and for neutral words. Then, for each of the priming segments, the mean normalized reaction time for positive words within that segment was compared with that for negative words within that segment. If positive words were rated faster, it would be predicted that the priming condition was major, while if negative words were rated faster, it would be predicted that the priming condition was minor. Since there was no *a priori* prediction for the silent conditions, they were excluded from the analysis. By this method, the correct mode was correctly predicted 52.4% of the time for the 3AFC task, and 53.8% of the time for the 2AFC task. Neither result was significantly better than the 50% accuracy expected by chance.

Discussion

The affective priming effect was not significant, although it was skewed in the predicted direction: in congruent conditions (major musical prime and positive target words, or minor musical prime and negative target words), a slightly faster response time was observed than in mismatched conditions (a mean difference of 8.4 ms and 7.5 ms for the 3AFC and 2AFC tasks, respectively). Next, reaction time measures were used to predict the mode of the priming condition of each segment (major or minor). However, the accuracy rate for the predictions was not significantly better than chance. Therefore, the affective priming effect did not seem to be an effective implicit measure.

When considering the failure of reaction times to provide accurate predictions of mode, several possibilities should be considered. Perhaps the affective priming effect is simply not an effective predictor. Alternatively, the experiment may need to be redesigned in order to observe a priming effect. Two particular considerations for redesign are worth mentioning.

First, the musical stimuli may not have been sufficiently evocative of emotion. Compared with recordings of actual music, or even artificial renditions of any melodic material, chord sequences are not usually considered to be emotionally stirring. Additionally, some participants reported that their strategy in the task was to try to ignore the sounds altogether and focus on the words. Thus, as “musical stimuli,” chord sequences may not have been moving or captivating enough – which does not seem surprising in retrospect. In future studies, it may prove more successful to instead employ real music already known to elicit strong emotional responses, rather than the simplified stimuli used in this experiment.

Second, despite the reported difficulty of the 3AFC task, accuracy was quite high for both tasks (90.3% for the 3AFC and 94.5% for the 2AFC). This suggests that ceiling effects may have been a factor. This may also be indicative of a task strategy where participants delay responses until they feel certain that they are correct. To address this, the response window could be controlled in a way that forces participants to respond earlier, even if they do not feel confident in the response. Such designs have proved effective in eliciting more robust priming effects (Draine & Greenwald, 1998).

Even though a significant priming effect was not obtained with the present experimental design, it may be possible to do so with another version of this procedure. If such a procedure could be devised, it may prove useful for research in music and emotion. Not only would it provide an implicit measure that avoids the limitations of subjective self-report, it would also provide a diachronic measure of emotion in music. That is, the changes in emotion experienced during the course of a musical excerpt might possibly be inferred from the changing strength of the priming effect throughout the excerpt’s duration.

Summary

The affective paradigm was introduced, and the utility of the paradigm for evaluating affective responses to musical sounds was considered. Since musical stimuli may be too lengthy to act as an effective affective prime, a procedure was designed to discourage active contemplation of the music. For the exploratory study, the music used was major or minor chord sequences. Listeners were required to categorize a rapid succession of words while the chord sequences were played, and the experiment was framed as a game where points could be scored for fast and accurate responses. When the words were categorized as positive, negative or neutral (3AFC task), the overall mean reaction time was about 150 ms slower than when the words were categorized only as positive or negative (2AFC task). A significant priming effect was not obtained in either task, although the results were skewed in the predicted direction: there was a reaction time difference in matched versus mismatched conditions of 8.4 ms for the 3AFC task, and 7.5 ms for the 2AFC task. In the *post hoc* analyses, it was found that these reaction time differences could not be used to predict the modality of the music being played. Thus, the affective priming paradigm using musical stimuli did not seem to provide an effective implicit measure.

Appendix A: Artificial Scales

Scale Group	Scale Variant	Scale Degree								
		#1	#2	#3	#4	#5	#6	#7	#8	#9
1	Test	0	181	320	421	681	903	1083	1225	1325
	Low	15	196	285	436	696	828	1098	1240	1340
	High	-155	166	355	406	666	978	1068	1210	1310
2	Test	0	146	255	522	747	932	1079	1186	-------
	Low	20	112	275	430	767	954	1099	1206	-------
	High	-20	118-	235	614	727	914	1059	1166	-------
3	Test	0	267	492	680	825	933	1199	-------	-------
	Low	10	277	502	618	835	943	1209	-------	-------
	High	-10	257	482	742	815	923	1189	-------	-------
4	Test	0	171	299	591	840	1047	1216	-------	-------
	Low	12	183	311	603	852	1059	1144	-------	-------
	High	-12	159	287	579	828	1035	1288	-------	-------
5	Test	0	278	514	712	868	986	-------	-------	-------
	Low	26	304	441	738	835	1012	-------	-------	-------
	High	-26	252	587	686	901	960	-------	-------	-------
6	Test	0	103	364	588	769	913	-------	-------	-------
	Low	18	121	270	606	787	931	-------	-------	-------
	High	-18	85	458	570	751	895	-------	-------	-------
7	Test	0	132	422	670	881	-------	-------	-------	-------
	Low	21	153	443	691	797	--------	-------	-------	-------
	High	-21	111	401	649	965	-------	-------	-------	-------
8	Test	0	127	418	667	878	-------	-------	-------	-------
	Low	21	148	439	667	794	-------	-------	-------	-------
	High	-21	106	397	646	962	-------	-------	-------	-------

Figure 8. Relative cents values for experimental scales described in Chapter Two. Eight groups of generated scales are represented, with each group having a test scale, a relatively "low" exposure scale and a relatively "high" exposure scale. Each group of scales had five to nine notes per scale. The relative cents values, with 0 as middle C, is given for each of the notes.

Appendix B: Ordered Chord Progressions

Figure 9. Ordered chord progressions in major and minor.

Appendix C: Word Lists

Positive Words for Males (n = 100)

ACCEPTANCE, ACHIEVEMENT, ADMIRED, AFFECTION, AMUSEMENT, AROUSE, AWARD, BLOSSOM, BONUS, CAREFREE, CARESS, CHAMPION, CHEER, CHEERFUL, COMEDY, CONFIDENT, COZY, CURE, DELIGHT, DESIRE, ECSTASY, ECSTATIC, ENJOY, ENJOYMENT, ENTHUSIASTIC, EXCEL, EXCELLENCE, FAME, FOOD, FREE, FREEDOM, FRIENDLY, FUN, FUNNY, GLAD, GRADUATE, HAPPINESS, HAPPY, HONESTY, HONOR, HUG, HUMOR, INTELLIGENCE, INTIMATE, JOKE, JOY, JOYFUL, KISS, LAUGH, LAUGHTER, LIBERTY, LOVE, LOVED, LOVING, LUCKY, LUSCIOUS, MERRY, MILLIONAIRE, MIRACLE, MOTHER, PARADISE, PEACEFUL, PERFECTION, PLAY, PLEASURE, POSITIVE, PRETTY, PROFIT, PROMOTION, PROUD, RAINBOW, REFRESHMENT, RELAX, RESCUE, ROMANTIC, SAVIOR, SEX, SEXY, SMART, SMILE, SPOUSE, SUCCESS, SUNRISE, SWEETHEART, TEAM, TERRIFIC, THRILLED, THRIVE, TREASURE, TRIUMPHANT, TROPHY, TRUTH, UNTROUBLED, VACATION, VALENTINE, VICTORY, WEALTH, WEALTHY, WIN, WINNER

Negative Words for Males (n = 100)

ACCIDENT, AFRAID, AGGRAVATION, AWFUL, BANKRUPT, BETRAY, CANCER, CASTRATE, DEATH, DEBT, DEFEATED, DEPRESSED, DESPISE, DISASTER, DISCOMFORT, DISEASE, DISLOYAL, DISTRESSED, DIVORCE, DREADFUL, DROWN, DROWN, ENRAGED, FAIL, FAILURE, FAMINE, FEARFUL, FROWN, FUNERAL, FURIOUS, GLOOM, GRAVE, GRIEF, HEADACHE, HELPLESS, HOMELESS, HOSTAGE, HUMILIATE, HUMILIATION, INFECTION, INSECURE, ISOLATION, JAIL, LEPROSY, LONELINESS, LONELY, LOSS, MAD, MAGGOT, MALARIA, MALNUTRITION, MASSACRE, MISERABLE, MISERY, MOLEST, MORGUE, MOURN, MURDERER, MUTILATE, PARALYSIS, POISON, POLLUTE, POUT, POVERTY, PRISON, PUNISH, PUNISHMENT, RABIES, REGRETFUL, REJECTED, REJECTION, REMORSE, REPULSED, SAD, SADNESS, SEASICK, SICK, SLAUGHTER, SLAVE, SLUM, SOB, STARVATION, SUFFER, SUFFOCATE, SUICIDE, SYPHILIS, TERRIBLE, TERRIFIED, TERRORIST, TOOTHACHE, TORTURE, TOXIC, TRAGEDY, TRAITOR, ULCER, UNFAITHFUL, UNHAPPY, UPSET, VIRUS, VOMIT

Neutral Words for Males (n = 100)

ACRE, ALCOHOL, ALERT, ARM, ARROW, BASKET, BELL, BICYCLIST, BLOUSE, BOLT, BOOTS, BRIDGE, BUCKET, BUILDING, CAVERN, CHESS, COLONY, CONVEY, CORN, COUNTRY, COYOTE, CRAB, CRACKERS, CUSTOM, DEW, EXERT, FACT, FACTORY, FOAM, FORK, FUR, FURNITURE, GARLIC, GARMENT, GATE, GLOSS, GROVE, GYM, HAIR, HAT, HERD, HERRING, HILL, JUG, KERCHIEF, KNOB, LAMB, LANTERN, LIMBER, MAGAZINE, MALT, MANE, MANOR, MEN, MOMENT, MUG, MUSEUM, NEWS, OVEN, PARLOR, PASSAGE, PIGEON, POSITION, POSTER, PRESIDENT, QUART, QUILT, RATTLE, REACT, REPTILE, RESPONSE, RETAIN, RIFLE, ROAD, ROOF, ROOM, SCARF, SEAGULL, SHELF, SHELTERED, SKIP, SNATCH, SOAK, SPHERE, SPINE, SPRAY, SQUASH, STOVE, STRING, STUN, SUBURB, SUGGESTION, SYSTEM, TAXI, THUMB, TOOL, TOWER, WAREHOUSE, WOLF, ZIPPER

Positive Words for Females (n = 100)

ACCEPTANCE, ACHIEVEMENT, ADMIRE, ADORABLE, ADVENTURE, AFFECTION, AFFECTIONATE, AMUSEMENT, ASSURE, AWARD, BEAUTIFUL, BEAUTY, BIRTH, BONUS, CHAMPION, CHEER, CHEERFUL, CHILD, CLEAN, COMEDY, COMFORTABLE, CONFIDENT, COOKIE, CURE, DAISY, DELIGHT, DELIGHT, ECSTATIC, ENJOY, ENTHUSIASTIC, EXCELLENCE, FREE, FRIEND, FRIENDLY, FUN, FUNNY, GIFT, GRADUATE, HANDSOME, HAPPINESS, HAPPY, HERO, HONEST, HONESTY, HUG, HUMOR, INTELLIGENCE, JOKE, JOY, JOYFUL, KINDNESS, KISS, LAUGH, LAUGHTER, LIBERTY, LOVE, LOVED, LOVING, LUCKY, LUXURY, MERRY, MIRACLE, MOTHER, OPTIMISTIC, PARADISE, PARTY, PEACE, PLAY, PLEASED, PLEASURE, POSITIVE, PRAISE, PROGRESS, PROMOTION, PROTECTED, PROUD, RAINBOW, ROMANCE, ROMANTIC, SATISFACTION, SATISFIED, SECURE, SMILE, SPRING, SUCCESS, SUN, SUNSET, SWEET, SWEETHEART, TERRIFIC, THRILLED, TREASURE, TRIUMPH, TRIUMPHANT, VACATION, VALENTINE, VICTORY, WEDDING, WIN, WINNER

Negative Words for Females (n = 100)

ABUSE, ACCUSE, AFRAID, ANGUISHED, BANISH, BANKRUPT, BETRAY, BURGLAR, BURIAL, CANCER, CRISIS, CRUEL, CRUSHED, DANGER, DEAD, DEATH, DEMON, DEPRESSED, DEPRESSION, DIE, DISAPPOINTED, DISASTER, DISLOYAL, EXECUTION, FAILURE, FAMINE, FAT, FEAR, FLABBY, FROWN, FUNERAL, FURIOUS, GLOOM, GRIEF, GUILLOTINE, GUILT, HATE, HATRED, HELL, HOMELESS, HORRIBLE, HUMILIATION, HURT, INFECTION, INJURY, INSULT, KILLER, LONELINESS, LOSS, MALNUTRITION, MISERABLE, MISERY, MORGUE, MOURN, MURDERER, MUTILATION, NIGHTMARE, OVERWEIGHT, PAIN, PARALYSIS, POISON, POLLUTE, POVERTY, PUNISH, RABIES, RASH, REJECTED, REJECTION, ROT, SAD, SHAME, SICK, SLAUGHTER, SLAVE, SOB, SORROW, STARVATION, STENCH, STING, SUFFER, SUFFOCATE, SUICIDE, SYPHILIS, TERRIBLE, TERRIFIED, TERROR, TERRORIST, THIEF, THREAT, TORTURE, TRAFFIC, TRAGEDY, TRAUMA, TUMOR, ULCER, UNFAITHFUL, UNHAPPY, VENOM, VIOLENT, WAR

Neutral Words for Females (n = 100)

ADHERE, ADJUST, ALTER, BANNER, BARK, BEDSPREAD, BODY, BOWL, BUTTON, CANYON, CARD, CHESS, CHIMNEY, CLOCK, CLOSET, CLOTH, CLOWN, COMPEL, CONTINUATION, COUNSELOR, COURTYARD, COW, DEN, DETAIL, DOCTOR, EAGLE, EGG, FABRIC, FEAT, FISHERMAN, FREEWAY, GARTER, GATE, GLACIER, GOLFER, GREYHOUND, HAM, HAT, HAWK, HEADLIGHT, HIKER, HOCKEY, HOTEL, HOUR, INTENT, LAW, LEGS, LION, LOCKER, LOOSEN, MACHINE, MAP, MARKET, MEMBER, METEOR, MILITARY, MILK, MOUTH, MUG, MULE, MUSHROOM, PEA, PLATE, PUDDLE, QUART, QUIET, RAFT, REPENT, ROOM, RUGBY, RULER, RUM, RYE, SCENE, SECRETARY, SHAWL, SIDLE, SKIER, SOCCER, SOLDIER, SPRAY, STRIDE, SUPPLY, TEETER, TEMPO, TIME, TITLE, TOURIST, TOWER, TRAIN, TREAD, TROUT, UNIT, UTENSIL, VEST, WAGON, WANT, WATCH, WOOD, YEAR

Excluded Words (Incorrect) for Both Males and Females (n = 109)

ACCORD, ACE, ADAPT, AGGRESSIVE, AMBITIOUS, ANCESTOR, AROMA, AROUSED, ASSAULT, BABY, BEACH, BED, BEER, BIKINI, BIRTHDAY, BOAT, BOOK, BOOM, BOUNTY, BUSYBODY, CAR, CARROT, CASH, CASTLE, CHEESEBURGER, CIGARETTE, CINEMA, CLAY, COMPLETE, COMPLEX, CONCENTRATE, COOKOUT, COUPLE, CUDDLE, DEFORMED, DEFY, DIAMOND, DIPLOMA, DOG, DUSK, DWARF, EAT, ENABLE, ENGAGED, EROTIC, FAD, FAITH, FEMALE, FINISH, FIREWORKS, FLARE, FOOTBALL, GAME, GIGGLE, GIRL, GOD, GOLD, HARVEST, HOME, ISLAND, JUSTICE, KETCHUP, KNOWLEDGE, MARRY, ME, MEAL, MEDICINE, MODEL, MONEY, MUSIC, NUDE, NUDIST, OASIS, OUTLET, OUTSTANDING, PASSION, PILLOW, POOL, PROSE, REAP, REBUFF, REVERENT, ROLLERCOASTER, SAILBOAT, SALUTE, SECURITY, SENSITIVE, SHOPPING, SISTER, SLEEP, SMIRK, SNUGGLE, SPA, STATUS, STEAK, STRESS, SUNLIGHT, SURPRISE, SUSHI, SYMPATHY, THRIFT, THRILL, TRIP, WANDER, WART, WATERFALL, WEB, WEEP, WINTER

Excluded Words (Offensive) for Both Males and Females (n = 18)

BREAST, CHRISTMAS, CLEAVAGE, COMMUNISM, EJACULATE, INTERCOURSE, NIPPLE, ORGASM, PEE, PENIS, RAPE, RETARD, SEMEN, TESTICLES, THONG, TOPLESS, URINATE, WHORE

References

Bower, G. H. (1981). Mood and memory. *American Psychologist, 36* (2), 129-148.

Bradley, M., & Lang, P. (1999). Affective norms for English words (ANEW): Stimuli, instruction manual and affective ratings. Technical report C-1, Gainesville, FL. The Center for Research in Psychophysiology, University of Florida.

Bregman, A. S. (1990). *Auditory scene analysis: the perceptual organization of sound.* Cambridge: MIT Press.

Burns, E. M., & Ward, W. D. (1978). Categorical perception – phenomenon or epiphenomenon: Evidence from experiments in the perception of melodic musical intervals. *Journal of the Acoustical Society of America, 63* (2), 456-468.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review, 82* (6), 407-428.

Crowder, R. G., Reznick, J. S., & Rosenkrantz, S. L. (1991). Perception of the major/minor distinction: V. Preferences among infants. *Bulletin of the Psychonomic Society, 29* (2), 187-188.

Curtis, M. E., & Bharucha, J. J. (2010). The Minor Third Communicates Sadness in Speech, Mirroring Its Use in Music. *Emotion, 10* (3), 335-348.

De Houwer, J., & Randell, T. (2004). Robust affective priming effects in a conditional pronunciation task: Evidence for the semantic representation of evaluative information. *Cognition & Emotion, 18* (2), 251-264.

De Houwer, J., Hermans, D., Rothermund, K., & Wentura, D. (2002). Affective priming of semantic categorisation. *Cognition & Emotion, 16* (5), 643-666.

Draine, S. C., & Greenwald, A. G. (1998). Replicable Unconscious Semantic Priming. *Journal of Experimental Psychology: General, 127* (3), 286-303.

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

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

Fazio, R. H. (2001). On the automatic activation of associated evaluations: An overview. *Cognition & Emotion, 15* (2), 115-141.

Gerardi, G. M., & Gerken, L. (1995). The Development of Affective Responses to Modality and Melodic Contour. *Music Perception, 12* (3), 279-290.

Goerlich, K. S., Witteman, J., Schiller, N. O., Heuven, V. J., Aleman, A., & Martens, S. (2012). The nature of affective priming in music and speech. *Journal of Cognitive Neuroscience, 24* (8), 1725-1741.

Greenwald, A. G., Draine, S. C., & Abrams, R. L. (1996). Three Cognitive Markers of Unconscious Semantic Activation. *Science, 273*, 1699-1702.

Hargreaves, W. A., Starkweather, J. A., & Blacker, K. H. (1965). Voice quality in depression. *Journal of Abnormal Psychology, 70*, 218-220.

Hermans, D., Baeyens, F., & Eelen, P. (1998). Odours as Affective-processing Context for Word Evaluation: A Case of Cross-modal Affective Priming. *Cognition & Emotion, 12* (4), 601-613.

Hermans, D., De Houwer, J., & Eelen, P. (1994). The Affective Priming Effect: Automatic Activation of Evaluative Information in Memory. *Cognition & Emotion, 8* (6), 515-533.

Hermans, D., De Houwer, J., & Eelen, P. (2001). A time course analysis of the affective priming effect. *Cognition & Emotion, 15* (2), 143-165.

Hermans, D., Spruyt, A., & Eelen, P. (2003). Automatic affective priming of recently acquired stimulus valence: Priming at SOA 300 but not at SOA 1000. *Cognition & Emotion, 17* (1), 83-99.

Hermans, D., Vansteenwegan, D., Crombez, G., Baeyens, F., & Eelen, P. (2002). Expectancy-learning and evaluative learning in human classical conditioning: affective priming as an indirect and unobtrusive measure of conditioned stimulus valence. *Behaviour Research and Therapy, 40*, 217-234.

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

Hevner, K. (1936). Experimental Studies of the Elements of Expression in Music. *The American Journal of Psychology, 48* (2), 246-268.

Hodges, D. A. (2010). Psychophysiological Measures. In P. N. Juslin, & J. A. Sloboda (Eds.), *Handbook of Music and Emotion: Theory, Research, Applications* (pp. 279-311). New York: Oxford University Press.

Huron, D. (2006). *Sweet Anticipation: Music and the Psychology of Expectation.* Cambridge: MIT Press.

Huron, D. (2008). A comparison of average pitch height and interval size in major- and minor-key themes: Evidence consistent with affect-related pitch prosody. *Empirical Musicology Review, 3* (2), 59-63.

Huron, D. (2012). Understanding Music-Related Emotion: Lessons from Ethology. In E. Cambouropolos, P. Mavromatis, & K. Pastiadis (Eds.), *Proceedings of the 12th International Conference on Music Perception and Cognition and the 8th Triennial Conference of the European Society for the Cognitive Sciences of Music*, (pp. 473-481). Thessaloniki.

Huron, D., & Davis, M. (2012). The harmonic minor scale provides an optimum way of reducing average melodic interval size, consistent with sad affect cues. *Empirical Musicology Review, 7* (3-4), 103-117.

Huron, D., Anderson, N., & Shanahan, D. (In press). You can't play sad music on a banjo: Acoustic factors in the judgment of instrument capacity to convey sadness. *Empirical Musicology Review*.

Huron, D., Kinney, D., & Precoda, K. (2006). Influence of pitch height on the perception of submissiveness and threat in musical passages. *Empirical Musicology Review, 1* (3), 170-172.

Huron, D., Yim, G., & Chordia, P. (2010). The effect of pitch exposure on sadness judgments: An association between sadness and lower than normal pitch. Paper presented at the International Conference on Music Perception and Cognition, Seattle, WA.

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. (2010). Introduction: Aims, Organization, and Terminology. In *Handbook of Music and Emotion: Theory, Research, Applications* (pp. 3-12). New York: Oxford University Press.

Kastner, M. P., & Crowder, R. G. (1990). Perception of the Major/Minor Distinction: IV. Emotional Connotations in Young Children. *Music Perception, 8* (2), 189-201.

Kivy, P. (1999). Feeling the musical emotions. *British Journal of Aesthetics , 39* (1), 1-13.

Klauer, K. C. (1997). Affective Priming. *European Review of Social Psychology, 8* (1), 67-103.

Klauer, K. C., & Musch, J. (2001). Does sunshine prime loyal? Affective priming in the naming task. *The Quarterly Journal of Experimental Psychology, 54A* (3), 727-751.

Kraepelin, E. (1899/1921). *Psychiatrie. Ein Lehrbuch für Studierende und Arzte, Klinische Psychiatrie. II.* Leipzig: Johann Ambrosius Barth, 1899. Trans. By R. M. Barclay as *Manic-depressive insanity and paranoia.* Edinburgh: E & S. Livingstone, 1921.

Logeswaran, N., & Bhattacharya, J. (2009). Crossmodal transfer of emotion by music. *Neuroscience Letters, 455*, 129-133.

Lövheim, H. (2012). A new three-dimensional model for emotions and monoamine neurotransmitters. *Medical Hypotheses, 78* (2), 341-348.

Matravers, D. (2001). *Art and emotion.* New York: Oxford University Press.

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: evidence of a dependence between retrieval operations. *Journal of Experimental Psychology, 90* (2), 227-234.

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

Morton, E. S. (1977). On the Occurrence and Significance of Motivation-Structural Rules in Some Bird and Mammal Sounds. *The American Naturalist, 111* (981), 855-869.

Muller, M., Klein, J., & Jacobsen, T. (2011). Beyond Demand: Investigating Spontaneous Evaluation of Chord Progressions with the Affective Priming Paradigm. *Music Perception, 29* (1), 93-108.

Murphy, S. T., & Zajonc, R. B. (1993). Affect, Cognition, and Awareness: Affective Priming With Optimal and Suboptimal Stimulus Exposures. *Journal of Personality and Social Psychology, 64* (5), 723-739.

Narmour, E. (1977). *Beyond Schenkerism: The Need for Alternatives in Music Analysis.* Chicago: University of Chicago Press.

Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner, & G. W. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (Vol. 11, pp. 264-336). Hillsdale: Lawrence Erlbaum Associates.

Post, O., & Huron, D. (2009). Western Classical Music in the Minor Mode is Slower (Except in the Romantic Period). *Empirical Musicology Review, 4* (1), 2-10.

Ragozzine, F. (2011). Cross-Modal Affective Priming With Musical Stimuli: Effect of Major and Minor Triads on Word-Valence Categorization. *Ninad, 25*, 13-29.

Reber, R., Winkielman, P., & Schwarz, N. (1998). Effects of Perceptual Fluency on Affective Judgments. *Psychological Science, 9* (1), 45-48.
Rice, T. (2004). *Music in Bulgaria: Experiencing Music, Expressing Culture.* New York: Oxford University Press.
Riener, C. R., Stefanucci, J. K., Proffitt, D. R., & Clore, G. (2011). An effect of mood on the perception of geographical slant. *Cognition and Emotion, 25* (1), 174-182.
Roederer, J. G. (1984). The Search for a Survival Value of Music. *Music Perception, 1* (3), 350-356.
Rotteveel, M., Groot, P. D., Geutskens, A., & Phaf, R. H. (2001). Stronger Suboptimal Than Optimal Affective Priming? *Emotion, 1* (4), 346-364.
Rubin, M., Paolini, S., & Crisp, R. J. (2010). A processing fluency explanation of bias against migrants. *Journal of Experimental Social Psychology, 46* (1), 21-28.
Russell, J. A. (1980). A Circumplex Model of Affect. *Journal of Personality and Social Psychology, 39* (6), 1161-1178.
Schaffrath, H. (1995). The Essen folksong collection in the humdrum kern format. (D. Huron, Ed.) Menlo Park, CA: Center for Computer Assisted Research in the Humanities.
Scherer, K. R. (2005). What are emotions? And how can they be measured? *Social Science Information, 44* (4), 695-729.
Schutz, M., Huron, D., Keeton, K., & Loewer, G. (2008). The Happy Xylophone: Acoustic Affordances Restrict An Emotional Palate. *Empirical Musicology Review, 3* (3), 126-135.
Sollberger, B., Reber, R., & Eckstein, D. (2003). Musical Chords as Affective Priming Context in a Word-Evaluation Task. *Music Perception, 20* (3), 263-282.
Spruyt, A., Hermans, D., De Houwer, J., & Eelen, P. (2002). On the nature of the affective priming effect: affective priming of naming responses. *Social Cognition, 20* (3), 227-256.
Spruyt, A., Hermans, D., Pandelaere, M., De Houwer, J., & Eelen, P. (2004). On the replicability of the affective priming effect in the pronunciation task. *Experimental Psychology, 51* (2), 109-115.
Spruyt, A., De Houwer, J., Hermans, D., & Eelen, P. (2007). Affective Priming of Nonaffective Semantic Categorization Responses. *Experimental Psychology, 54* (1), 44-53.
Steinbeis, N., & Koelsch, S. (2010). Affective Priming Effects of Musical Sounds on the Processing of Word Meaning. *Journal of Cognitive Neuroscience, 23* (3), 604-621.
Storbeck, J., & Clore, G. L. (2008). The Affective Regulation of Cognitive Priming. *Emotion, 8* (2), 208-215.
Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of experimental psychology, 18* (6), 643-662.
Temperley, D. (2001). *The Cognition of Basic Musical Structures.* Cambridge: MIT Press.
Temperley, D., & Tan, D. (2013). Emotional connotations of diatonic modes. *Music Perception, 30* (3), 237-257.
Tomkins, S. S. (1962). *Affect, imagery, consciousness: Vol. I. The positive affects.* New York: Springer.

Turner, B., & Huron, D. (2008). A Comparison of Dynamics in Major- and Minor-key Works. *Empirical Musicology Review, 3* (2), 64-68.
Västfjäll, D. (2010). Indirect, perceptual, cognitive and behavioural measures. In P. N. Juslin, & J. A. Sloboda (Eds.), *Handbook of Music and Emotion: Theory, Research, Applications* (pp. 255-277). New York: Oxford University Press.
Vuoskoski, J. K., & Eerola, T. (2011). Measuring music-induced emotion: A comparison of emotion models, personality biases, and intensity of experiences. *Musicae Scientiae, 18* (1), 159-173.
Wachsmann, K. P. (1971). Universal Perspectives of Music. *Ethnomusicology, 15* (3), 381-384.
Winkielman, P., Zajonc, R., & Schwarz, N. (1997). Subliminal Affective Priming Resists Attributional Interventions. *Cognition & Emotion, 11* (4), 433-465.
Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist, 35* (2), 151-175.
Zentner, M., & Eerola, T. (2010). Self-Report Measures and Models. In P. N. Juslin, & J. A. Sloboda (Eds.), *Handbook of Music and Emotion: Theory, Research, Applications* (pp. 187-221). New York: Oxford University Press.