Form and Meaning in Music: Revisiting the Affective Character of the Major and Minor Modes

Timothy Justus, Laura Gabriel and Adela Pfaff

Psychology and Cognitive Science Field Groups, Pitzer College, Claremont, CA, USA

ABSTRACT
Musical systems develop associations over time between aspects of musical form and concepts from outside of the music. Experienced listeners internalize these connotations, such that the formal elements bring to mind their extra-musical meanings. An example of musical form-meaning mapping is the association that Western listeners have between the major and minor modes and happiness and sadness, respectively. We revisit the emotional semantics of musical mode in a study of 44 American participants (musicians and non-musicians) who each evaluated the relatedness of 96 melody-word pairs. Among the tonal melodies, we manipulated mode (major and minor) and timbre (clarinet and flute) while systematically controlling for other musical factors including pitch register and melodic contour. Similarly, among the English words, we manipulated word affect (happy and sad) while systematically controlling for other lexical factors including frequency and word length. Results demonstrated that participants provided a higher proportion of related responses for major melodies paired with happy words and minor melodies paired with sad words than for the reverse pairings. This interaction between mode and word affect was highly significant for both musicians and non-musicians, albeit with a larger effect for the former group. Further interactions with timbre suggested that while both clarinet and flute conveyed happiness when in the major mode, the clarinet was somewhat more successful than the flute at conveying sadness in the minor mode. Debriefing questionnaires suggested that the majority of the participants, including all of the non-musicians, had no awareness of the major-minor manipulation, and instead directed their attention to register and contour. We argue that the affective character of the major and minor modes is but one example of form-meaning mapping in music, and suggest further exploration of the roles of timbre, register, and contour in conveying musical emotions.

How do musical systems come to represent meaning? For an experienced listener, sound forms bring to mind associations from outside of the music, including mental images, words, actions, and emotions. Such associations between musical forms and meanings from outside of the music, i.e., referential or extra-musical meaning, can be
contrasted with associations between one formal element and another, i.e., absolute or intra-musical meaning (see Meyer, 1956; Juslin & Sloboda, 2001; Kivy, 2002; Patel, 2008; Koelsch, 2012; Justus, in press). The present work concerns music’s ability to express and evoke emotion, phenomena we refer to collectively as the emotional semantics of music.1

Specifically, we revisit the affective connotations or “affective character” (Hevner, 1935) of the major and minor modes in Western tonal-harmonic music; all else being equal, the major mode is associated with happiness, and the minor mode with sadness. In the 21st century, Western music listeners grow up with countless examples that reinforce the convention, but its history extends back hundreds of years. Parncutt (2014) proposed that the emotional connotations of major and minor emerged gradually between the fourteenth and the seventeenth centuries (spanning the late Medieval, Renaissance, and Baroque periods). These centuries also witnessed the gradual replacement of the medieval modal system with common-practice major and minor tonality, suggesting that the emotional meanings may be as old as the forms themselves.

The psychological reality of the emotional connotations of mode has been studied experimentally since the early twentieth century. Beginning with studies conducted by Heinlein (1928) on the player piano (see Crowder, 1984 for discussion), behavioral and physiological experiments have illustrated this type of extra-musical meaning using a variety of stimulus materials and tasks (Collier & Hubbard, 2001; Costa, 2013; Crowder, 1985; Gagnon & Peretz, 2003; Gomez & Danuser, 2007; Hevner, 1935, 1936; Hunter, Schellenberg, & Schimmack, 2010; Husain, Thompson, & Schellenberg, 2002; Peretz, Gagnon, & Bouchard, 1998; Rigg, 1940a; Steinbeis & Koelsch, 2011, Exp. 2; Straehley & Loebach, 2014; Temperley & Tan, 2013; Thompson, Schellenberg, & Husain, 2001; Wedin, 1972). Corresponding developmental work has suggested that the association emerges by seven to eight years of age in Canadian, American, and British children (Dalla Bella, Peretz, Rousseau, & Gosselin, 2001; Gerardi & Gerkin, 1995; Gregory, Worrall, & Sarge, 1996), and perhaps earlier (Kastner & Crowder, 1990).

In Hevner’s (1935) study, “The Affective Character of the Major and Minor Modes in Music,” American participants were presented with musical stimuli based on compositions by Bach, Beethoven, and several other composers. Each major and minor excerpt was also rewritten in the other mode to help control for additional musical parameters. After each excerpt, participants were asked to select from a list of English adjectives the words that best described the piece. The excerpts in the major mode were more likely to be paired with affectively positive words (e.g., “light”) whereas the excerpts in the minor mode were more likely to be paired with affectively negative words (e.g., “stormy”). Hevner further examined potential effects of musical training, finding no significant differences between musicians and non-musicians regarding these affective connotations. Thus, early empirical work demonstrated that even non-musician listeners who have not been explicitly taught about major and minor modes can acquire an implicit association between this aspect of musical form and an extra-musical, emotional meaning.

Because the referential meanings associated with musical mode are emotional in nature, one might ask to what extent the emotion is represented (or expressed) as opposed to being evoked, a distinction known as the cognitivist and emotivist positions in the philosophy of music (Kivy, 1990; Tan, Pfordresher, & Harré, 2010). In other
words, whereas some scholars are concerned with how a piece of music accomplishes the expression of an emotion (perhaps of an imagined persona; Levinson, 2006), others might be more interested in whether and how that piece of music makes a listener feel the emotion (e.g., Vuoskoski & Eerola, 2012, 2015). We are aware of one study that addresses the issue somewhat directly in the context of the major and minor modes. Hunter et al. (2010) asked participants to rate 30-second excerpts from Bach inventions with respect to how happy or sad the music sounded, as well as how the music made them feel. While both sets of responses were influenced by whether the music was major or minor, the effect of how the music made participants feel was mediated by the effect of how the music sounded, suggesting that modal associations with emotional concepts may be primary.

If mode is capable of priming the concepts of happiness and sadness, one might predict that major and minor chords would facilitate the processing of happy and sad target words, respectively. Indeed, this was demonstrated by Costa (2013, Exp. 1), using an affective priming design. Specifically, major and minor triads in higher and lower registers of the piano served as primes for happy and sad target words. The task was explicit with respect to the target, but not the prime, i.e., participants were to rate the target as positive or negative, with the prime chord serving ostensibly as an alerting signal. Both mode (major, minor) and register (high, low) interacted with word affect (happy, sad) in the response times and errors, suggesting that the Italian participants, who were all non-musicians, had happier extra-musical associations with major chords and higher registers, and sadder extra-musical associations with minor chords and lower registers. The work complemented a previous finding that consonant and dissonant chords – specifically, chords containing perfect fifths compared to tritones – are associated with happy and sad words, respectively (Sollberger, Reber, & Ekstein, 2003).

Semantic associations between major and minor chords and happy and sad words have also been demonstrated electrophysiologically in German listeners. Steinbeis and Koelsch (2011, Exp. 2) demonstrated interactions between mode and word affect in both the behavioral data and the amplitude of the N400 event-related potential component, a neural index of semantic processing. While modulation of the N400 is classically associated with linguistic semantic manipulations, analogous effects have subsequently been demonstrated using other meaningful stimuli such as drawings, pictures, videos, and sounds (Kutas & Federmeier, 2011), including music (Daltrozzo & Schön, 2009a, 2009b; Koelsch et al., 2004). Specifically, Steinbeis and Koelsch (2011) found that response times and mean amplitudes of the N400 component were smaller for happy words when preceded by major triads, as opposed to minor triads, whereas the reverse pattern was observed for sad words. Like Hevner (1935), they compared musicians and non-musicians, finding no significant differences in emotional semantics, and, as with Costa (2013) and Sollberger et al. (2003), they also conducted analogous studies in which consonant and dissonant chords were found to prime happy and sad words, respectively (Steinbeis & Koelsch, 2008, 2011, Exp. 1).

The present study was designed to examine semantic associations between well controlled single-line melodies based on the major and minor modes and happy and sad words. A melody is a sequence of tones perceived as a gestalt; a single-line melody is unaccompanied by any additional chords or harmony. In contrast, the affective priming literature has tended to use major and minor triads (Costa, 2013; Steinbeis & Koelsch,
A triad is a chord comprising three tones separated by thirds; triads are the building blocks for harmony. Other designs employing subjective ratings of the music have tended to use real musical excerpts rich in both melody and harmony (Hevner, 1935; Hunter et al., 2010).

There are at least two important reasons why the extension from triads to melodies is not trivial. First, major and minor triads are not equivalent to major and minor tonality. Both types of triad appear within each mode, i.e., major tonality includes three major triads (I, IV, V), three minor triads (ii, iii, and vi), and one diminished triad (vii°), whereas minor tonality – if we take into account the raised seventh of the harmonic minor scale – includes two major triads (V, VI), two minor triads (i, iv), two diminished triads (ii°, vii°), and one augmented triad (III+). Second, any intrinsic psycho-physical differences between major versus minor triads that stem from the presentation of simultaneous complex tones would not apply to single-line melodies. Instead, the emotional connotations of major and minor melodies may stem from a learned association with their respective tonic triads (Parncutt, 2014), or with major and minor tonality more broadly.

Because several of the previous studies investigating the emotional associations of the major and minor modes included additional manipulations, such as faster and slower tempo (e.g., Hunter et al., 2010), and higher and lower register (e.g., Costa, 2013), typically finding that these variables modulated the effect of mode, we also included an additional exploratory factor in the present study, namely timbre. Timbre or tone color refers to the multidimensional attributes of sound that are distinct from pitch, loudness, spatial position, and duration (McAdams, 1999), as illustrated by the perceptual quality distinguishing two different musical instruments playing the same note. The distinct timbres of two complex tones result from differences in their harmonic spectra, including how the component frequencies change over time. As with mode, there are a variety of reasons to suspect that timbre, particularly that of a familiar musical instrument, is capable of conveying extra-musical, emotional meanings. These reasons may include inherent differences in consonance and dissonance – the pleasantness or unpleasantness with which tones combine – based on the timbre’s characteristic harmonic spectrum. In addition to such psychoacoustic factors, listeners may have learned cultural associations among the timbre, the instrument, and the musical settings in which they are characteristically found. For musicians, these include embodied, auditory-motor co-activations representing the physical affordances of familiar instruments (cf. De Souza, 2017).

Steinbeis and Koelsch (2011, Exp. 3) manipulated timbre within the affective priming design previously described. A synthesized grand piano, which had previously served as the default timbre in their other experiments, was now considered the pleasant timbre, whereas a synthesized tin drum was considered the unpleasant timbre. Both always played a major chord. Analogous to their previous findings in which consonance and mode were manipulated, the authors demonstrated affective priming; mode and word affect interacted in both the behavioral data and the amplitude of the N400. A complementary study demonstrated similar effects of relatedness between musical timbres and words, using a range of semantic associations (Painter & Koelsch, 2011).
In a different experimental design, Lahdelma and Eerola (2016) explored the affective connotations of major, minor, diminished, and augmented triads (each in all three inversions), and major and minor seventh chords (in two of the four possible inversions). The participants, who were recruited online primarily from Finland, the United States, and New Zealand, rated the chords on nine emotional dimensions, each of which demonstrated an interaction between chord type and inversion (cf. Parncutt, 2014). Most relevant for the present purposes, these authors also included a manipulation of timbre (piano and strings), finding that the piano was rated more highly for happiness/joy, whereas strings were rated more highly for nostalgia/longing, melancholy/sadness, and tenderness. While this study understandably focused on only two timbres in the context of the chord manipulations, the same laboratory has employed a much larger range of timbres in studies using isolated tones (Eerola, Ferrer, & Alluri, 2012).

Other researchers have examined the emotional associations of instrumental timbres in the context of other experimental designs. In one study, British participants were presented with melodies designed to represent one of four emotions (happiness, sadness, fear, or anger) and were asked to identify these emotions using a four-alternative forced choice (Hailstone et al., 2009, Exp. 1). It is noteworthy that among other differences among the four sets of melodies in tempo, meter, and dynamics, the major mode was used to convey happiness whereas the minor mode was used for sadness, fear, and anger. The critical, additional manipulation was that the melodies were played by one of four instruments (piano, violin, trumpet, or electric synthesizer) to assess potential effects of timbre. The results of the study suggested first that participants identified the intended emotion more frequently for the happy and sad melodies compared to the fearful and angry ones. Further, this factor interacted significantly with timbre. For example, a melody intended to represent happiness was less likely to be labeled as such if played by a violin, whereas a melody intended to represent sadness was less likely to be labeled as such if played by a synthesizer.

Of particular relevance to the timbres chosen for the present experiment – clarinet and flute – Wu, Horner, and Lee (2014) presented Hong Kong participants with resynthesized sustained tones representing eight instruments: bassoon, clarinet, flute, horn, oboe, saxophone, violin and trumpet, all at a pitch of approximately E-flat 4. Participants indicated which of two instruments sounded more like a provided emotional word, including positive words such as happy, joyful, heroic, and comic, and negative words such as sad and depressed. Participant ratings in the first experiment suggested that the happier words tended to be best represented by the clarinet, trumpet, and violin, whereas the sad words tended to be best represented by the flute and horn. This observation is consistent with other studies suggesting a relatively sad affect when comparing the flute to instrumental timbres other than the clarinet (Balkwill, Thompson, & Matsunaga, 2004), and a relatively happy affect when comparing the clarinet to the violin (Paquette, Peretz, & Belin, 2013).

Much of the response variance in Wu et al. (2014, Exp. 1) was explained by the average spectral centroid, the mean frequency of the spectrum weighted by amplitude. This acoustic factor is associated with the psychological experience of brightness in the timbre among many instruments, and considering the clarinet (6.34) and flute (3.45) specifically, would suggest that the clarinet should sound brighter, and therefore happier, than the flute. However, the second and third experiments of Wu et al.
in which average spectral centroid was equated for all the instruments, revealed the importance of other factors such as the relative power of the even and odd harmonics. A low even/odd ratio (i.e., stronger odd harmonics) is characteristic of the clarinet (.37), particularly in the lowest, chalumeau register. In contrast, the power of the even and odd harmonics is relatively equal for the flute (.85). Across various instruments, the even/odd ratio correlates positively with happiness and negatively with sadness, which suggests that on the basis of this factor alone, the clarinet should sound sadder than the flute. This observation is consistent with the results of Huron, Anderson, and Shanahan (2014), whose musician participants suggested that the clarinet is somewhat better suited than the flute for conveying sadness. Overall, the two factors of spectral centroid and even/odd ratio would appear to be working in opposition with respect to the affective connotations of the two instruments.

It is worth noting that the choice of one controlled concert pitch in Wu et al. (2014) and other important studies of the affective connotations of timbre (e.g., Eerola et al., 2012) places the tone in very different resonances for instruments of different ranges, making the tones less comparable in other ways. For the B-flat soprano clarinet, E-flat 4 (notated as F4) is a nearly open fingering at the top end of the lowest register, resulting in a weaker tone relative to the clarinet timbre as a whole, whereas the same pitch on the flute is a relatively closed fingering at the low end of the lowest register, resulting in a richer, fuller tone relative to the flute timbre as a whole. Thus, despite the experimental value in controlling for pitch height, a single pitch is not representative of the instrumental timbres over the full ranges of the instruments. This may be particularly relevant considering that the emotional meanings of the clarinet and flute may reflect a balance of acoustic factors (i.e., spectral centroid and even/odd ratio), which may have different relative importance depending on the register of the instruments.

In the present study we investigated the emotional semantics of musical mode by asking whether brief major and minor melodies – as opposed to the isolated major and minor triads used in previous work – reliably express happiness or sadness to American listeners. We asked participants to judge the relatedness of such precisely controlled melodies and a set of precisely controlled English words with happy and sad connotations. We chose a relatedness judgment for the task, rather than an explicit judgment of emotion, because we did not wish to presume that the basis of the participants’ responses would be emotional, nor did we wish the structure of the task to presume a categorical rather than continuous approach to emotion (cf. Eerola & Vuoskoski, 2011) or to prime the basic emotion categories in English (cf. Barrett, 2017).

We predicted that an emotional association would be observed, such that mode and word affect would interact in determining participant responses. In addition to manipulating mode (major and minor), we also manipulated timbre, presenting half of the melodies in a clarinet timbre and half in a flute timbre, both within a balanced pitch range spanning nearly two octaves. Although we expected emotional semantics to be driven primarily by mode, we were intrigued as to whether this would be modulated by timbre, given prior work suggesting different affective connotations for the two instruments. We also included equal numbers of musician and non-musician participants in the study, with the expectation that both groups would demonstrate the predicted associations.
Method

Participants

Forty-four undergraduate students at an American college participated in the study, 22 of whom were categorized as musicians with at least five years of musical training (age: $M = 20$ years, $SD = 2$; education: $M = 14$ years, $SD = 1$; musical training: $M = 11$ years, $SD = 4$; 12 women, 9 men, 1 non-binary), and 22 as non-musicians (age: $M = 19$ years, $SD = 1$; education: $M = 14$ years, $SD = 1$; musical training: $M = 2$ years, $SD = 2$; 14 women, 8 men). Thus, the musicians were moderately trained amateurs rather than professionals. All participants reported being fluent and proficient in English; all were either native English speakers or had acquired English by the age of seven. All participants reported normal or corrected-to-normal vision, and all but one participant reported normal hearing. A hearing threshold was obtained for that participant, which confirmed normal hearing in the good ear. The instruments that musicians reported studying for at least five years included piano ($n = 12$), guitar (9), voice (6), saxophone (2), clarinet (1), flute (1), violin (1), cello (1), pipa (1), and/or drums (1).

Stimuli and Design

Ninety-six seven-tone melodies represented a 2 mode (major, minor) $\times$ 2 timbre (clarinet, flute) design. The melodies were 4200 ms in duration, with each of the seven tones lasting 600 ms, corresponding to 100 beats per minute. Thus, all melodies were isochronous, with every tone of the same duration, and all were equal in tempo.

All melodies were derived from 12 distinct phrases that were initially composed in C major (Figure 1, upper example). To evoke a strong sense of tonality, the melodies always contained the tonic ($do$) twice, once low and once high, with one instance always serving as the final tone (six each). The melodies also always included the dominant ($sol$) and subdominant ($fa$), one of which, or the other instance of the tonic, was always the first tone (four each). To permit the manipulation of major and (harmonic) minor, the melodies also always contained the mediant ($mi$) and submediant ($la$). Finally, the remaining tone in the melody was either the supertonic ($re$) or leading tone ($ti$) (six

![Figure 1](image.png)

Figure 1. Experimental design. 96 seven-tone melodies, which manipulated the factors of mode (major, minor) and timbre (clarinet, flute), were followed by 96 written words, which manipulated the factor of word affect (happy, sad). Participants indicated the perceived relatedness of the pair.
The 12 minor melodies were created from the 12 major by lowering both mi and la by a half step, consistent with the harmonic minor scale (Figure 1, lower example). Thus, there were no confounds between the manipulation of mode and any other aspect of melody, including ascending versus descending movement, or the implied harmony.

These 24 major and minor melodies were then each transposed four times among the 24 major and minor keys to generate 96 experimental melodies. Thus, in the main experiment, each melodic pattern was heard in four different keys and each key was heard with four melodic patterns. An additional eight melodic transpositions were generated as practice stimuli. The pitch range of each melody was an octave (tonic to tonic), and given the transpositions, the entire stimulus set represented a span of nearly two octaves.

The 96 melodies were generated using clarinet and flute timbres (48 each), balanced for all factors above, using the music notation software Sibelius 7.5. The Sibelius timbres for clarinet and flute succeed in capturing the critical differences in average central spectroid (higher for the clarinet) and even/odd harmonics ratio (lower for the clarinet). The pitch height chosen for the transpositions to the 24 major and minor keys was identical for both clarinet and flute (C4 to B5 concert pitch), with fundamental frequencies ranging approximately from 262 Hz to 988 Hz, which is within the overlap in range for the two instruments (the soprano clarinet in B-flat and the concert flute in C). Thus, there was no confound between timbre and register, a factor that has been previously studied in affective priming designs (Costa, 2013), but unlike prior studies that restricted stimuli to a single pitch (Eerola et al., 2012; Wu et al., 2014), the melodies sampled from a two-octave range for both instruments.

Ninety-six written English words served as the target stimuli, representing the third factor in the design: word affect (happy, sad). These were selected using the Affective Norms for English Words, a set of over 1000 words previously rated on a five-point scale for happiness, sadness, and other basic emotions by nearly 300 American participants (Stevenson, Mikels, & James, 2007). As shown in Table 1, the 48 happy words (e.g., success) and 48 sad words (e.g., despair) were given comparably high ratings for the intended emotion and comparably low ratings for the other emotion. The happy and sad items were also matched as best as possible for a variety of lexical factors including the number of letters, number of syllables, frequency of the lemma, and part of speech, using the CELEX Lexical Database (Baayen, Piepenbrock, & Gulikers, 1995), given that these factors all influence the dynamics of visual word recognition. An additional eight words were selected as practice stimuli.

### Procedure

Participants were tested in a quiet, sound-attenuated room. Each trial began with a fixation cross for 1000 ms, followed by a 4200-ms melody, presented binaurally through closed

<table>
<thead>
<tr>
<th>Table 1. Descriptive statistics for word stimuli (means with SD, or Counts).</th>
</tr>
</thead>
<tbody>
<tr>
<td>Happiness Rating</td>
</tr>
<tr>
<td>Happy Words (N = 48)</td>
</tr>
<tr>
<td>Sad Words (N = 48)</td>
</tr>
</tbody>
</table>

* log lemma frequency per million.
Sennheiser headphones. After a 1000-ms interstimulus interval, the target word was visually presented for 1000 ms, followed by a brief reminder of the instructions, which remained on the screen until a response was provided. Participants were asked to indicate whether the melody and the word were related or unrelated. They were not informed about the manipulations of mode, timbre, or word affect, and were asked to make these judgments on any basis of their choosing, without time pressure. Participants recorded their responses by pressing one of two buttons on the keyboard with their index fingers, after which the 1000-ms fixation for the next trial immediately began. The response mappings were counterbalanced, with half of each participant group indicating related pairs with their right index finger and the other half indicating related pairs with their left index finger. The 96 trials were presented in a different pseudorandom order for each participant, with the condition that no more than two trials from a given experimental cell occur in succession, and with a different pseudorandom pairing of particular melodies and words for each participant. Trials were divided among four experimental blocks of 24 each, providing an opportunity for participants to rest. For 11 of the 44 participants, one response each was input too quickly to be recorded, and for 5 participants, two responses each, always from different conditions, were input too quickly. Therefore, the analysis of those participants was based on 94 or 95 trials each, rather than 96, with at least 11 observations per cell. The 21 excluded responses out of a grand total of 4224, represent less than one half of one percent of the group data.

Following the experiment, participants were interviewed regarding their musical and language backgrounds. Finally, the experimenter asked four questions regarding what participants had noticed about the experiment, in the following order: (1) How would you describe the melodies you heard in the experiment? (2) Did you notice anything about the scale or key of the melodies? (3) Did you notice anything about the musical instruments or timbres? (4) How would you describe the words presented in the experiment? The participants were then debriefed and received either a small payment or course credit for their participation. The Institutional Review Board of the college approved all procedures.

**Results**

For each participant, the proportion of stimulus pairs judged to be related was calculated for each of the eight experimental cells, with eleven or twelve observations per cell. A $2 \times 2 \times 2 \times 2$ mixed-design ANOVA with within-subjects factors mode (major, minor), timbre (clarinet, flute), and word affect (happy, sad), and between-subjects factor musical training (musician, non-musician) first revealed a main effect of mode, $F(1, 42) = 7.9, p = .007, \eta^2_p = .16$, such that the proportion of related responses was slightly higher, in general, for minor melodies ($M = 0.50$) compared to major melodies ($M = 0.45$).

More central to our hypothesis concerning the emotional connotations of the major and minor modes, the analysis also revealed a highly significant interaction between mode and word affect, $F(1, 42) = 69.0, p < .0001, \eta^2_p = .62$ (Figure 2). Specifically, the proportion of related pairs was higher in the conditions pairing major melodies with happy words ($M = 0.56$) and minor melodies with sad words ($M = 0.61$), than in the conditions pairing major melodies with sad words ($M = 0.34$) and minor melodies with happy words ($M = 0.38$).

Regarding effects of timbre, while a two-way interaction between timbre and word affect did not reach significance, $F(1, 42) = 3.4, p = .07, \eta^2_p = .07$, there was a small but
significant three-way interaction among mode, timbre, and word affect, $F(1, 42) = 4.1$, $p = .049$, $\eta^2_p = .09$, suggesting that the emotional semantics of musical mode were subtly different for the two timbres. To better understand this interaction, we conducted separate analyses for the trials containing major melodies and the trials containing minor melodies. When considering the major trials alone, there was no interaction between timbre and word affect, $p = .72$; significant and comparable effects of Word Type were observed for both clarinet, $t(43) = 4.3$, $p < .001$, and flute, $t(43) = 4.5$, $p < .001$. However, when considering the minor trials alone, there was a significant interaction between these variables, $F(1, 42) = 6.1$, $p = .02$, $\eta^2_p = .13$; significantly larger effects of Word Type were observed for clarinet, $t(43) = −6.2$, $p < .001$, compared to the flute, $t(43) = −3.8$, $p < .001$, though both were significant.

Regarding effects of musicianship, there was a three-way interaction among mode, word affect, and musical training, $F(1, 42) = 5.4$, $p = .03$, $\eta^2_p = .12$. Examination of the means in Figure 2 suggests that the musicians showed a larger two-way interaction between mode and word affect, i.e., stronger emotional semantics of musical mode, than did the non-musicians; nevertheless, the interaction was highly significant for both musicians, $F(1, 21) = 46.6$, $p < .0001$, $\eta^2_p = .68$, and non-musicians, $F(1, 21) = 24.3$, $p < .0001$, $\eta^2_p = .55$, when analyzed separately.

When asked the general question of how they would describe the melodies heard in the experiment, only five of the participants, all musicians, mentioned major and minor, and only two musicians mentioned the flute in this initial response. When asked more directly about the scale or key of the melodies, five additional musicians mentioned major and minor. When asked more directly about the musical instruments and timbres, 22 additional participants, mostly musicians, mentioned the flute by name, 8 of whom also identified the clarinet. And when asked what they noticed about the words, 31 of the participants, including both musicians and non-musicians, described them as happy and sad, positive and negative, or emotional. Throughout the
questionnaire, we noticed that 21 of the participants, including both musicians and non-musicians, volunteered something about pitch height (using words such as “high” and “low”) and/or contour (using words such as “rising” or “falling”) in their responses, typically indicating that they perceived the higher and/or ascending melodies as happier than the lower and/or descending melodies. While not the main focus of our study, pitch height and contour were carefully controlled during the construction of the stimuli and were orthogonal to our manipulations of mode and timbre.

Overall, the questionnaire data suggest that participants were typically aware of the manipulation of word affect, and sometimes aware of the manipulation of timbre. However, very few participants were aware of the manipulation of mode, and such awareness was limited to those with musical training. Instead, participants reported that their attention was more frequently directed to pitch height and contour, which cannot explain the observed interaction between mode and word affect.

Discussion

The purpose of the study was to determine whether listeners would associate seven-tone melodies composed in the major and minor modes with happy and sad English words, respectively, whether these associations would be affected by a further manipulation of clarinet and flute timbre, and whether such emotional associations would depend on musical training. The results illustrated a clear relationship between mode and word affect; participants indicated higher relatedness ratings for major melodies paired with happy words and for minor melodies paired with sad words, than for major melodies paired with sad words and minor melodies paired with happy words. This effect was observed for two instrumental timbres, but also subtly modulated by this factor, with the suggestion that in the minor mode, the clarinet represented sadness more effectively than did the flute. The interaction between mode and word affect was observed for both musicians and non-musicians, with musicians showing a larger effect.

This work complements prior studies in which the musical stimuli included major and minor triads rather than melodies (Costa, 2013; Steinbeis & Koelsch, 2011, Exp. 2; Lahdelma & Eerola, 2016). The use of major and minor single-line melodies rather than major and minor triads further suggests that emotional semantics can be driven by learned associations to major and minor tonality per se, rather than a response to psychoacoustic differences based on the simultaneous presentation of three complex tones. It should be noted that major and minor triads are both consonant chords; for example, the participants of McDermott, Lehr, and Oxenham (2010) provided pleasantness ratings that placed the minor triad in a range similar to that of the highly consonant perfect fifth and fourth. However, major and minor triads do not appear to be equally consonant; participants provided even higher pleasantness ratings for major triads in the same study. Major and minor triads may also differ in the salience of the root when considered over the possible range of inversions, registers, spacings, and doublings (Parncutt, 2011), an additional psychophysical basis for positive connotations with the more “certain” major triad over the more “ambiguous” minor (Parncutt, 2012, 2014; cf. Lahdelma & Eerola, 2016).

Single-line melodies drawn from the major and minor modes, on the other hand, do not vary on any psychoacoustic factors based on the simultaneous presentation of complex tones. They also need not differ with respect to the set of pairwise ascending
and descending intervals, although in practice these factors may co-vary (Huron & Davis, 2012). In Figure 1, for example, both sample melodies begin with descending perfect fourths (C down to G), followed by either an ascending major second (G to A) or an ascending minor second (G to A-flat). Both melodies continue with additional descending fourths (A to E, or A-flat to E-flat), followed by either an ascending minor second (E to F) or major second (E-flat to F). They then continue identically, with a descending minor third (F to D) and a descending major second (D to C). Thus, inherent differences in the consonance of pairwise intervals (e.g., McDermott et al., 2010) also cannot account for the affective connotations of the two melodies; both examples contain one minor second, two major seconds, one minor third, and two perfect fourths. The stimulus set was similarly balanced, with each major melody having a corresponding minor melody.

While the sensory consonance and salience of the melodies’ component tones and intervals did not differ, we suggest that they succeeded in priming associated chords and keys in listeners’ schematic knowledge of Western tonality as a whole. Connectionist models of such schemata would predict that the mental representation of the tonic chord, a major or minor triad, would be strongly activated (Tillmann, Bharucha, & Bigand, 2000). From a music theoretical perspective, melodies in the major or minor modes can be considered to be a prolongation (or elaboration) of the tonic triad (Parncutt, 2012; Schenker, 1921/2004). For example, the melodies in Figure 1 might be considered a prolongation of a descending arpeggio on the tonic chord in root position (C, G, E or E-flat, C), with the remaining diatonic tones serving as elaborations. Such learned associations between major and minor melodies and their corresponding major and minor tonic triads may mediate their association with happiness and sadness (Parncutt, 2014). Alternatively, rather than the tonic triads specifically, major and minor melodies may bring to mind major and minor tonalities more broadly (each including a mixture of major and minor triads as described earlier), which in turn are associated with positive and negative affect.

Rich associations exist in Western music between the major and minor modes and other factors with more iconic relationships to happiness and sadness. Classical music written in the major mode has tended to use higher pitch register (Huron, 2008), larger intervals (Huron, 2008), and faster tempi (Post & Huron, 2009) than music in the minor mode. These relationships are not intrinsic to the modes; they changed for example between the Classical and Romantic periods in Western music (Horn & Huron, 2015). But register, interval size, and tempo are likely associated with mode in the minds of Western listeners, consistent with how these variables interact in psychological experiments (Costa, 2013; Gagnon & Peretz, 2003; Hunter et al., 2010; Rigg, 1940b).

The present study also complements prior work concerning the emotional semantics of timbre (e.g., Steinbeis & Koelsch, 2011, Exp. 3). Unlike this prior study, we found only a subtle effect of timbre: a small but significant interaction among mode, timbre, and word affect. A likely reason for the subtlety of this effect is in the particular choice of timbres; Steinbeis and Koelsch contrasted a consonant grand piano with a dissonant tin drum, which they described as “considerably harsher and unpleasant” (p. 614). Therefore, their timbre manipulation was also one of consonance, which is well established as a basis for affective priming (Costa, 2013; Sollberger et al., 2003;
Steinbeis & Koelsch, 2011, Exp. 1). In contrast, the differences between clarinet and flute timbres are far subtler. The spectra of both instruments are harmonic, and therefore generally consonant, with important differences including the higher spectral centroid and the lower even/odd harmonic ratio of the clarinet compared to the flute (Wu et al., 2014). The former suggests a way in which the clarinet might be regarded as brighter, and therefore happier, than the flute, whereas the latter suggests an opposing effect that might contribute to the clarinet’s sadness, especially for its lowest, chalumeau register.

Another important difference is that we manipulated mode and timbre within the same experiment, whereas Steinbeis and Koelsch (2011) manipulated these two factors in separate experiments (Exp. 2 and 3). Previous work in which mode and tempo were varied in the same experiment has typically suggested that strong manipulations of tempo can trump mode, e.g., when an up-tempo piece sounds happy despite a minor tonality (Gagnon & Peretz, 2003; Hunter et al., 2010; Rigg, 1940b). Manipulations of register can similarly outweigh the effect of mode, depending upon the task (Costa, 2013). Therefore it is possible that the subtle differences in emotional valence between the clarinet and flute could be more easily observed when not simultaneously testing for effects of mode. Similarly, it would be interesting to see whether the effect of timbre/consonance observed by Steinbeis and Koelsch (2011, Exp. 3) would still be observed if mode were also manipulated within the same experiment, rather than a separate one (2011, Exp. 2).

We expected that effects of timbre might depend upon not only psychoacoustics, but also learned associations among these sound forms, the associated instruments, and prior listening experiences. Contrary to the Wu et al. (2014) study using isolated tones, which suggested that, all else being equal, the clarinet has happier connotations than the flute, other evidence aligns with the present study in suggesting the opposite, that the clarinet more readily connotes sadness. Huron et al. (2014) found that when American musicians were asked to rate on a 7-point scale the relative sadness of various Western instruments (by name rather than sound), the clarinet was judged to have a somewhat higher mean sadness compared to that of the flute (4.86 vs. 4.00) and a somewhat higher mean capacity for sadness (6.24 vs. 5.95). Huron et al. (2014) suggested that such differences stem from the instruments’ varying capacity to emulate the acoustics of an emotional vocal tract. In their second experiment, musicians who again rated various instruments estimated that the clarinet was better suited than the flute for playing with some, but not all, of the acoustic parameters associated with vocal sadness, e.g., the capacity to play quietly (4.90 vs. 4.50), slowly (6.00 vs. 5.50), and with a mumbling quality (5.00 vs. 4.00).

The ranges of the two instruments also suggest that, for experienced listeners, the clarinet timbre may be associated with somewhat lower pitches compared to that of the flute, even if in an experiment such as ours, the ranges were equated. More specifically, the range of the B-flat soprano clarinet (the most commonly encountered) extends down to D3 concert pitch (notated E3), whereas the Western concert flute cannot play lower than C4 (or sometimes B3). In our study, timbre and register were not confounded; the lowest pitch in either timbre was C4. However, to the extent that listeners have prior associations between instrumental timbres and registers, and further
associate higher registers with happiness and lower registers with sadness (cf. Costa, 2013), a learned, mediated association might easily develop between the clarinet timbre and sadness.

We note that our ability to observe differences between the clarinet and flute may have been limited by the use of synthesized timbres and possibly by limited knowledge of the associated musical instruments, especially among the non-musician group. When asked about the musical instruments and timbres, only 24 of the 44 participants identified the flute by name, and only eight participants identified the clarinet. The fact that fewer than half of the musicians mentioned the clarinet when specifically asked about the musical instruments suggests that the synthesized Sibelius timbre did not successfully evoke the instrument, or that undergraduates with modest musical training on other instruments are not necessarily familiar with the clarinet or its timbre. One might speculate that if participants found the clarinet timbre to be noticeably synthesized, this might have led to an unintended difference in subjective consonance. This in turn might have driven the interaction such that the clarinet timbre was perceived as sadder than the flute when combined with minor tonality. On the other hand, the emotional semantics of timbre may not depend upon explicit knowledge or conscious recognition of the instruments, just as the emotional semantics of musical mode do not seem to require musical training or awareness that a given melody is major or minor.

Our study also examined whether the emotional semantics of musical mode depend on musical training, by including equal numbers of musicians and non-musicians, and whether the observed interaction between mode and word affect depended on conscious awareness or strategies, by including a debriefing questionnaire. While musicians demonstrated a larger effect than did the non-musicians, as suggested by a three-way interaction among mode, word affect, and group, both musicians and non-musicians demonstrated highly significant mode by word affect interactions when analyzed separately. This suggests that passive exposure is sufficient to learn associations between musical form and emotional meaning via implicit learning. In this respect, the present study complements prior work demonstrating the emotional connotations of the major and minor modes in non-musician listeners (Hevner, 1935; Costa, 2013; Steinbeis & Koelsch, 2011, Exp. 2), and suggests that the implicit learning of tonality extends to meaning as well as form.

The approach that participants took with the relatedness judgment task, as evidenced by the debriefing questionnaire, also supports the notion that the emotional connotations of the major and minor modes are largely implicit. Nothing was mentioned to the participants regarding the manipulation of mode, timbre, and word affect until after the study was completed. Instead, the instructions were to judge the relatedness of the melodies and words on any basis the participants saw fit. While the vast majority became aware of the manipulation of word affect, only ten musicians from among the 44 participants commented on the manipulation of mode, and five of these ten only did so when directly asked about scale and key. Instead, the responses of most participants suggested that they were explicitly evaluating other factors, such as whether a particular melody was presented in a higher or lower register (cf. Costa, 2013), given that the transpositions resulted in a range of pitch heights, or responding to whether the contour of the melody was ascending or descending (cf. Gerardi & Gerkin, 1995; Hevner, 1936), given that the final tonic tone could occur in either the lower or
upper octave. While not intended manipulations, these factors were perfectly balanced between the major and minor melodies, and therefore cannot explain the observed interaction between mode and word affect. Instead, the results suggest that the emotional semantics of musical mode are largely implicit and do not require the listener’s conscious attention. In this respect, the present study complements prior work demonstrating that music generates extra-musical associations, as evidenced by responses to target words, even when these associations are completely incidental to the task (Daltrozzo & Schön, 2009b). Nevertheless, the debriefing questionnaire should be interpreted with caution, given that participants with less musical training might have noticed the mode or timbre manipulations without having the vocabulary to express this awareness.

An acknowledged limitation of this study is that response times were not recorded, in contrast to earlier studies of affective priming between music and language (Costa, 2013; Steinbeis & Koelsch, 2011). There were two reasons for this. First, because we did not specify the basis on which participants should make their relatedness judgments, the experimental instructions did not emphasize speed. Second, the text reminding participants of how to respond, which reappeared after each target word was presented, had an unintended effect of slowing most responses to at least 1000 ms. In contrast, speeded responses to visually presented target words, as in for example a lexical decision task, are generally expected much sooner, around 600 to 700 ms post word onset. In these respects, our study parallels that of Painter and Koelsch (2011), who also used a relatedness judgment and did not analyze response times for similar reasons. Future studies will examine whether the emotional semantic effects observed in the present study extend to other tasks in which speeded judgments are more appropriate.

Our framework for the emotional semantics of music is influenced by those of Patel (2008) and Koelsch (2012). Both authors are interested in the extent of cognitive and neural association between music and language. Patel (2008) develops an inclusive view of musical meaning, drawing upon the work of Nattiez (1990), and suggests a taxonomy that includes both intra-musical and extra-musical sources, with the latter including various types of cross-modal imagery, semantic associations, and emotions both expressed and evoked. Koelsch (2012) takes a similarly inclusive view, drawing upon experimental work in which music has been shown to influence the processing of semantically related words (and vice versa), and outlines a theory of musical signification based on icon, index, and symbol (Peirce, 1931). Given the many parallels found between music and language in cognitive neuroscience studies of syntax and semantics, Koelsch (2012) suggests that the brain may not make a categorical distinction between the two domains (see Justus, 2014, for discussion).

In referring to musical forms, meanings, and the mappings between them throughout the present paper, we have borrowed psycholinguistics terminology to suggest an analogy with the mental lexicon. Form-meaning mappings in language were once thought to be entirely symbolic, with the “arbitrariness of the sign” considered a design feature that distinguished human speech from the communication systems of other animals (de Saussure, 1916/1959; Hockett, 1960). However, in recent years studies have suggested that the lexicon may have more iconicity than traditionally acknowledged (Monaghan, Shillcock, Christiansen, & Kirby, 2014; Perniss,
Thompson, & Vigliocco, 2010). That language and music both employ a balance of iconic and symbolic mapping resonates with theories suggesting that speech and song were less differentiated earlier in human evolutionary-cultural history (Brown, 2017).

The elements of emotion and meaning in music examined in the present study likely stem from an intricate interaction between the acoustics of complex tones and learned, cultural associations about what makes music happy or sad. Future research concerning the clarinet and flute timbres might take full advantage of the three distinct registers and timbres associated with each instrument, further clarifying the relationships among timbre, register, and contour in conveying meaning. Including the timbres of a wider selection of musical instruments would allow a richer understanding of this multidimensional perceptual attribute and its affective character. So would a consideration not only of the perceptual side of timbre, but also the motor affordances of the associated instruments. More broadly, a complete understanding of musical form-meaning mapping would include both a synchronic perspective concerning contemporary musical systems and the associated mental representations of today’s listeners, and a historical, diachronic perspective concerning how these musical systems and listeners’ knowledge thereof might have evolved through the centuries.

Notes

1. It is possible to object to the notion of musical semantics, if one believes meaning in music has to be denotational, i.e., referring to something specific in the world and asserting something about it. However, if we allow meaning in music to be connotational, i.e., bringing to mind something other than itself, more scholars will agree that passages of music have meaning. See Patel’s (2008) discussion of Kivy (2002) and Nattiez (1990), and Koelsch’s (2012) application of Peirce’s (1931) theory of signs.

2. A lemma is a mental representation of a word, and corresponding stage of lexical access, concerning the word’s semantic and grammatical properties, but not its morphological or phonological properties (which are instead aspects of the lexeme; see Levelt, 1989). In the present context, balancing by lemma frequency – rather than by surface frequency – simply means that inflections (e.g., success, successes) are counted as the same word.

Disclosure Statement

No potential conflict of interest was reported by the authors.

Funding

This project was supported by Pitzer College faculty research funding and the Helen R. Trilling P’16 Psychology Research Internship. Preliminary findings were presented at the annual meetings of the Psychonomic Society (Gabriel & Justus, 2015) and the Association for Psychological Science (Pfaff & Justus, 2018).
References


