Perception of Emotion in Sounded and Imagined Music

Brian J. Lucas
Bucknell University

Emery Schubert
University of New South Wales, Sydney, Australia

Andrea R. Halpern
Bucknell University

We studied the emotional responses by musicians to familiar classical music excerpts both when the music was sounded, and when it was imagined. We used continuous response methodology to record response profiles for the dimensions of *valence* and *arousal* simultaneously and then on the single dimension of *emotionality*. The response profiles were compared using cross-correlation analysis, and an analysis of responses to musical feature turning points, which isolate instances of change in musical features thought to influence valence and arousal responses. We found strong similarity between the use of an emotionality arousal scale across the stimuli, regardless of condition (imagined or sounded). A majority of participants were able to create emotional response profiles while imagining the music, which were similar in timing to the response profiles created while listening to the sounded music. We conclude that similar mechanisms may be involved in the processing of emotion in music when the music is sounded and when imagined.

Received November 11, 2008, accepted February 25, 2010.

Key words: musical imagery, emotion, continuous response, cross-correlation, arousal

Mental imagery has intrigued cognitive psychologists for several reasons. On the one hand, most people experience these quasi-perceptual representations and report them as being very vivid and powerful. For example, among Gabrielsson's (2010) analysis of reports of people's strong experiences in music, he found that some:

> . . . may be elicited by 'inner music'; that is, imagined music . . . Music is 'heard', it just comes for no obvious reason and 'sounds' as clear and distinct as live music. Some respondents realized that it was 'only' imagined music, but for others the experience was so vivid that they were surprised to learn that there was in fact no sounding music present at all. (Gabrielsson, 2010, p. 559)

On the other hand, a mental process that in some ways mimics a perceptual process is resource intensive and thus not very efficient. However, despite this inefficiency, there may be some kinds of information that are ideally suited for this kind of representation. Kozhevnikov, Kosslyn, and Shephard (2005) found that scientists, who often have to think about how variables are related, were particularly good at a task of spatial imagery. The same tradeoff of efficiency for accuracy obtains for auditory imagery, especially for music.

From a myriad of composers who insist that music is composed in the mind before it is ever sounded, to the average person who cannot get a melody line out of his or her head, it is likely that most musicians and nonmusicians alike have at one time or another experienced such music images. A number of studies have shown that these representations maintain features of sounded music, such as pitch, tempo, melody, mode, and loudness. For instance, Halpern (1988a) asked participants to identify whether or not two lyrics were part of the same song and observed that reaction times increased as the distance, in beats, between the lyrics in the sounded tune increased. These results suggested that participants temporally scanned the musical image to locate the lyrics and this demonstrates that at least one temporal aspect of music is maintained in imagery. Halpern, Zatorre, Bouffard, and Johnson (2004) showed that similarity ratings of instrument sounds were nearly equivalent whether the sounds were perceived or imagined (as were several areas of neural activation), suggesting that another musical feature, timbre, is represented in musical images.

Although it has been established that these musical features can be represented in musical imagery, it is less clear whether the emotion in the music also is represented. We know very little about this, although we may get a hint from self-report studies of persistent musical images ("earworms") that are often of preferred music (Bailes, 2007). To the extent that anyone voluntarily "replays" a tune in his or her head, it is a reasonable assumption that this is partly because the tune is conveying some affect

Music Perception volume 27, issue 5, pp. 399–412, issn 0730-7829, electronic issn 1533-8312 DOI:10.1525/mp.2010.27.5.399

desired at the moment. And musicians who engage in mental practice may perhaps be doing so partly to review the emotional expression they wish to convey in actual performance. However, with one exception noted below, no one has examined this question under controlled conditions.

Furthermore, we know that some structural features communicate emotion. Hevner (1935) presented participants with an adjective circle that clustered together similar adjectives that represented a general emotion category. Hevner found, for example, that the major mode generally is associated with positive descriptors and the minor mode with negative descriptors. Using a similar checklist methodology, Motte-Haber (as cited in Gabrielsson & Lindstrom, 2001) found that music with faster tempi and higher event densities are rated as happier than music with slower tempi and lower event densities. Other musical features that have been linked to emotional descriptors include pitch and loudness. Higher pitches have been associated with happiness and lower pitches with sadness. Louder music and faster tempo have been associated with excitement and softer music with sleepiness (Gabrielsson & Lindstrom, 2001, Schubert, 2004). Other researchers have combined these basic features. For instance, Gagnon and Peretz (2003) found that both (major) mode and (faster) tempo elicited "happy" judgments and vice versa for "sad," with tempo being the more salient influence.

If at least some of these features can be represented in imagery, we propose that related emotional judgments may also be represented. However, measuring emotional judgment of imagined music presents challenges. Mental imagery is a cognitively demanding process, as it requires considerable working memory capacity. Making emotional judgments during this retrieval might be difficult. Thus, if an emotional judgment is elicited after the retrieval, the respondent could be basing the judgment on semantic associations to the music, or perhaps reporting socially desirable responses. It is difficult to verify the accuracy of the memory representation of the piece, and the retrospective judgment could be affected by that accuracy (or lack thereof).

It seems more desirable to assess emotional judgments of music as the experience unfolds (we want to clarify here that we are asking people to judge what the music is communicating, not how the music makes one *feel*; see Evans & Schubert, 2008). This approach allows a more temporally fine-grained analysis of emotional response, and also allows a mapping of emotional response to particular structural features in the music. Continuous response (CR) methodology (Schubert, 2001) allows emotional response to be measured in one or two dimensions as the music unfolds. *Valence* and *arousal* have been recognized by several researchers (e.g., Nagel, Kopiez, Grewe, & Altenmüller, 2007) as adequately capturing the gamut of emotional response. Valence is an emotional dimension that ranges from negative emotions, such as sad and angry, to positive emotions, such as happy and calm. Arousal is a dimension that ranges from excited/ energized to sleepy/bored. Arousal can differentiate between emotions with similar levels of valence. For example, "sad" would be distinguished from "angry" by its low level of arousal compared to anger's high level of arousal (Russell, 1980).

In a typical CR task, a listener is trained to indicate with a computer mouse his or her response on the emotional dimension of interest, as the music is played. The mouse position in one- or two-dimensional space is recorded at frequent, regular time intervals. The data can later be correlated with events in the musical stream. Although the response is unlikely to be instantaneous (for example, Schubert & Dunsmuir, 1999, found a lag of about 2–3 s for emotional response to changes in loudness), these responses are systematically related to musical features, and can provide a fine-grained and real-time portrait of emotional response to even complex music.

With this paradigm, we are in a position to ask whether the type and time course of emotional judgments of music are similar in sounded and imagined situations. We can ask a participant first to listen to a piece of familiar music and respond with valence and arousal ratings (or emotionality ratings, as in our Experiment 2) as the music unfolds. Then the participant repeats the task while imagining the music. The response profile to the sounded music and to the imagined music can then be compared. Schubert, Evans, and Rink (2006) reported a case study of one individual in this task. A professional pianist listened to his own recording of a Chopin *Nocturne* while making valence and arousal judgments. He then repeated the procedure while imagining his recording. The authors found that the continuous response profile of the sounded music strongly resembled that of the imagined music, although they also found that the pianist gradually slowed his responses in the imagined condition, perhaps due to the cognitive load of retrieving the imagined music (Sweller, 1988, 2006). However, we do not know if this response was idiosyncratic, confined to a professional musician, or confined to imagining one's own performance.

The current study built on this preliminary work by asking a group of student musicians to make continuous emotional judgments for excerpts of three different orchestral pieces. We looked at the congruence of emotional judgments with musical features that we predicted *a priori* would engender specific responses, and also compared the response profiles in the sounded and imagined conditions. In Experiment 1, participants performed the continuous response task using the dimensions of valence and arousal. In Schubert's review of the emotion labels used for continuous response, he argued that "arousal seems to be similar to emotional strength" (2010, p. 241) and that more research was called for to examine this similarity. In Experiment 2, participants used the single dimension of emotionality, both to see if a single global dimension could be extracted from the music, and to allow us to compare this global attribute to the more specific emotional labels of valence and arousal. Experiment 2 also replicated the procedure of Experiment 1 for several returning participants, in an attempt to assess test-retest reliability of results after several months.

We predicted that the continuous response time-series profiles of the sounded and imagined conditions would be structurally similar, which would demonstrate a musician's ability to extract emotion from a musical image. The Schubert et al. study (2006) reported consistent temporal drift, or lag, in the imagined responses compared to the sounded responses, but with good structural similarity between the sounded and imagined time-series profiles of emotional response. After correcting for any lagging that may occur, we predicted that the imagined responses would be structurally consistent with the sounded responses for experienced musicians who were familiar with the music. An alternative prediction stems from the fact that the musicians tested here, while experienced, were not as intimately familiar with the pieces as was the pianist who had performed the piece in the prior study. For these participants, performing this somewhat novel task of tracking emotion might induce some additional cognitive load, which could induce less correspondence between sounded and imagined emotion profiles than Schubert and colleagues had found.

Experiment 1

The aim of Experiment 1 was to investigate whether emotional responses made continuously to music were similar when the music was sounded compared to when it was imagined. Participants first were given a brief tapping task to establish their ability to keep time. In the experimental task, continuous response methodology was used to track participants' emotional responses as they judged three familiar pieces of classical music on the dimensions of valence and arousal. This task included a sounded condition and an imagined condition.

Method

PARTICIPANTS

Participants were 22 Bucknell University undergraduate students (18–21 years old). Five individuals later were excluded (for reasons explained in the Results section); thus the study included data from 17 (13 female and 4 male) participants. The group consisted of students from the Introductory Psychology participant pool ($N = 18$) and music majors ($N = 4$) who were compensated with course credit and movie tickets, respectively. To increase the probability that participants would be adept in musical tasks and familiar with the classical music excerpts used here, they were required to have a minimum of 8 years of private instrumental lessons. Experience ranged from 8–17 yrs with an average of 10.81 years.

MATERIALS

The stimuli were excerpts of Romantic and late Classical Western music of less than 1 min in length. Classical music was chosen because of the large selection of wordless pieces that avoid the complex interaction between musical features and lyrics (Serafine, Crowder, & Repp, 1984). High levels of richness and musical feature variety within a short excerpt is a common feature of Romantic and late Classical Western music not ordinarily found in other styles of music likely to be familiar to the participants. In order to identify highly familiar pieces that contain musical variability (changes in tempo, melody, loudness), we developed a survey that included 16 classical pieces. The survey was administered to six Bucknell University students with musical backgrounds. The three pieces listed below were chosen for the experiment because of high scores on scales measuring familiarity and confidence in one's ability to imagine the piece, and musical feature variability.

1. *Allegro con brio* from Beethoven's *5th Symphony in C minor, Op. 67* (*Classical Hits*, Films for Humanities and Sciences, performed by the Royal Philharmonic; abbreviated B). Measures 1–63 (48 s) were expected to occupy the low valence/high arousal and low valence/low arousal quadrants of the two-dimensional emotion space (see Figure 1). Its minor mode was predicted to evoke various levels of negative valence and its extreme changes in loudness were expected to produce varying levels of arousal.

2. The *Allegro* from Mozart's *Serenade in G K525 (Eine Kleine Nachtmusik)* (Deutsche Grammophon/Polygram, performed by the Berlin Philharmonic, M). Measures 1–28 (53 s) were expected to occupy the high valence/high arousal and high valence/low arousal quadrants of the two-dimensional emotion space. Its mode is major, which was predicted to evoke positive valence, and its variations

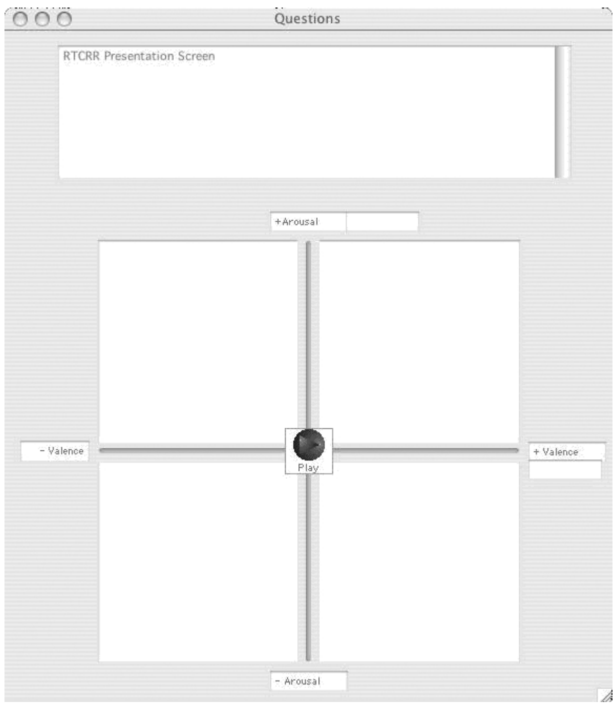

FIGURE 1. The RTCRR display.

in loudness and note density would produce various levels of arousal.

3. Tchaikovsky's *Waltz of the Flowers* from *Nutcracker Suite Op. 71* (*Classical Hits*, Films for Humanities and Sciences, performed by the Royal Philharmonic, T.) Measures 34–86 (51 s) were expected to occupy the high valence/low arousal, and high valence/high arousal quadrants of the two-dimensional emotion space. Its mode is major, which was predicted to evoke positive valence, and the soft passages were predicted to evoke low arousal. The soft and loud variation was predicted to produce varying low and high levels of arousal. This selection is henceforth referred to as 'T'.

PROCEDURE

A tapping task was administered to establish beat-tracking ability. The task was designed to be a simplified parallel to the experimental task. It included a sounded and an imagined condition where participants were required to produce a steady tempo (160 bpm). In the sounded condition, which served as practice, participants pressed a keyboard key on a Yamaha PSR 500 digital keyboard to synchronize with a sounded metronome clicking quarter notes for 42 measures in standard 4/4 time. In the main imagined condition, participants were given a two-measure memory cue of the metronome tempo and then asked to continue tapping for the next 40 measures in the absence of the metronome clicks. Each trial consisted of approximately 60 s of tap production. Cakewalk Pro Audio 8 sound editing software was used to record participants' responses. Following the tapping task, participants filled out a questionnaire that recorded age, gender, and music training, as well as piece familiarity of the three stimuli used in the experiment on a scale of 1–7.

Participants then were introduced to the experimental task and the Two-Dimensional Emotion Space (2DES) software, a continuous response interface. We used the Real Time Cognitive Response Recorder (RTCRR) developed by Schubert (2007). The program was presented on a Macintosh G4 laptop computer. It allowed the user to respond to *x* and *y* dimension variables by moving the mouse along the continuum of the axes and was set to record responses in real time at 0.5 s intervals. Valence was presented on the *x*-axis and arousal on the *y*-axis. The four quadrants of the interface as presented to participants are illustrated in Figure 1: high valence/high arousal, low valence/high arousal, low valence/low arousal, and high valence/low arousal. The mouse began each trial centrally located in the middle of the emotion space. Valence was defined as the range of positive to negative emotions evoked by the music. Arousal was defined as the range of emotions covering excited to sleepy. The participants were then instructed as follows:

> You are going to listen to a piece of music and your task is to indicate the amount of valence and arousal expressed by the music as it unfolds. Move the mouse throughout the quadrants to indicate your response to emotional output. You are rating the emotion you believe the piece is trying to evoke.

Participants first executed practice trials with Vivaldi's *Four Seasons: Spring* (0:00–0:51). When participants showed proficiency with the software and confirmed that they understood the instructions they began the experimental task. First was the sounded condition where participants heard a musical excerpt and made valence and arousal responses as the music unfolded. Two trials of each of the three pieces were conducted in the sounded condition. Next, in three trials of the imagined condition, participants heard a 5–8 s memory cue of the recording and then responded as they imagined the rest of the excerpt. To help participants keep their place in the music, one-page reduced versions of the musical score were presented in both the sounded and imagined conditions with the instructions that the score was supplemental and did not need to be used if the participant thought it unnecessary. The musical score was placed on the keyboard (directly below the emotion space) so participants could see both the software interface and the music. Some participants chose to hold up the music so they were looking at it alongside the screen. During the task the experimenter was present in case any questions or problems arose.

Each trial of the sounded condition was conducted before the participants began the imagined condition, to allow the sounded condition to act as an additional training trial before the piece was to be imagined. For each piece, approximately 15 min elapsed before the imagined condition task was performed and participants were not told that they would be imagining the same pieces for the second task. This was done to reduce the likelihood of participants focusing on memorizing their responses. Order effects were addressed by counterbalancing the presentation order of the three pieces across all participants. Following the sounded condition, another questionnaire was given, collecting information on a 1–7 scale about the extent to which participants utilized the musical score and their perceived proficiency in imagining the piece. Ratings on score use indicated moderate score use in both Experiment 1 ($M = 4.69$) and Experiment 2 ($M = 4.73$). If time permitted, the experiment ended with a general discussion to obtain information that could be relevant in analysis. One participant, for example, discussed her role as a conductor and having performed the Mozart piece.

Results

INCLUDED AND EXCLUDED TRIALS

With each of the 22 participants rating three excerpts twice in the sounded condition and three times in the imagined condition, a total of 330 continuous response profiles were collected. In our analysis we used the second trial of the sounded condition and the third trial of the imagined condition. We chose to use the last trial of each condition rather than an average of all the trials because this method allowed the participants the most time to acquaint themselves with the task and the excerpts. Considering only the last trial of the sounded and the imagined conditions, 132 trials were eligible for analysis, or 66 within-subject sounded-to-imagined condition comparisons.

Next, we considered other factors that would require a trial or a participant to be excluded from analysis. We excluded response profiles from participants who scored below a 3 (1–7 scale) on familiarity with an excerpt, or who gave similarly low scores on the question, "How well could you imagine this piece?" Many of these respondents also showed incoherent, or "skywriting" patterns with the mouse. A third problem arose from technical difficulties or software error. These exclusions left a total of 80 trials, or 40 sounded-to-imagined condition comparisons, for inclusion in our analysis.

TAPPING TASK

As a measure of rhythmic consistency, we chose to analyze intertap intervals. The average intertap interval was 0.359 s ($SD = 0.019$), a 4.27% error (error range 4.00%–5.20%) from the actual interval of 0.375 s (at tempo 160 bpm). The consistency of tapping is in the range of typical synchronization performance summarized by Repp (2005), which indicates adequate rhythmic tapping ability from all participants.

COMPARISON OF MEAN SOUNDED AND IMAGINED SERIES FOR AROUSAL AND VALENCE

Comparisons of averaged sounded and imagined responses for each response dimension across the three works were calculated and plotted to provide initial visual comparison of the tasks. The plots are shown in Figure 2a for arousal and 2b for valence. In each case, a good similarity can be seen between the imagined and sounded responses.

As a simple first pass on analysis, we conducted a within-subject analysis by comparing the responses (mouse positions) in the sounded condition to the responses in the imagined condition of each participant. An average of 85% of all individual arousal sounded-to-imagined response profile correlations exceeded the critical value of r (as each of our music excerpts varied in time, and thus number of responses collected, so too did the degrees of freedom, and the corresponding critical values.) For valence, 66% exceeded the critical value. We save a more extensive and quantified analysis of these results after presenting results of Experiment 2.

MUSICAL FEATURES ELICITING RESPONSES

Given that our second goal was to look at specific sounded-imagined correspondences, we identified some musical feature turning points that we predicted would evoke an emotional response. They are characterized by salient increases or decreases in the amount of a musical feature (e.g., see Schubert, 2004), such as loudness, or a change in the nature of a musical feature, such as a major to minor change in mode. We then looked to see if imagined and sounded condition profiles would show similar changes at those points.

An example of a musical turning point is given in Figure 3, which shows a participant's time-series data in response to Beethoven. Emotional response value is on the y-axis and it can be tracked by time in seconds along the x-axis. One musical turning point we identified, at 25 s (measure 26), is characterized by a sudden decrease in loudness and is labeled in the figure as B2. We examined the 3 s window between B2 and B2', where the apostrophe

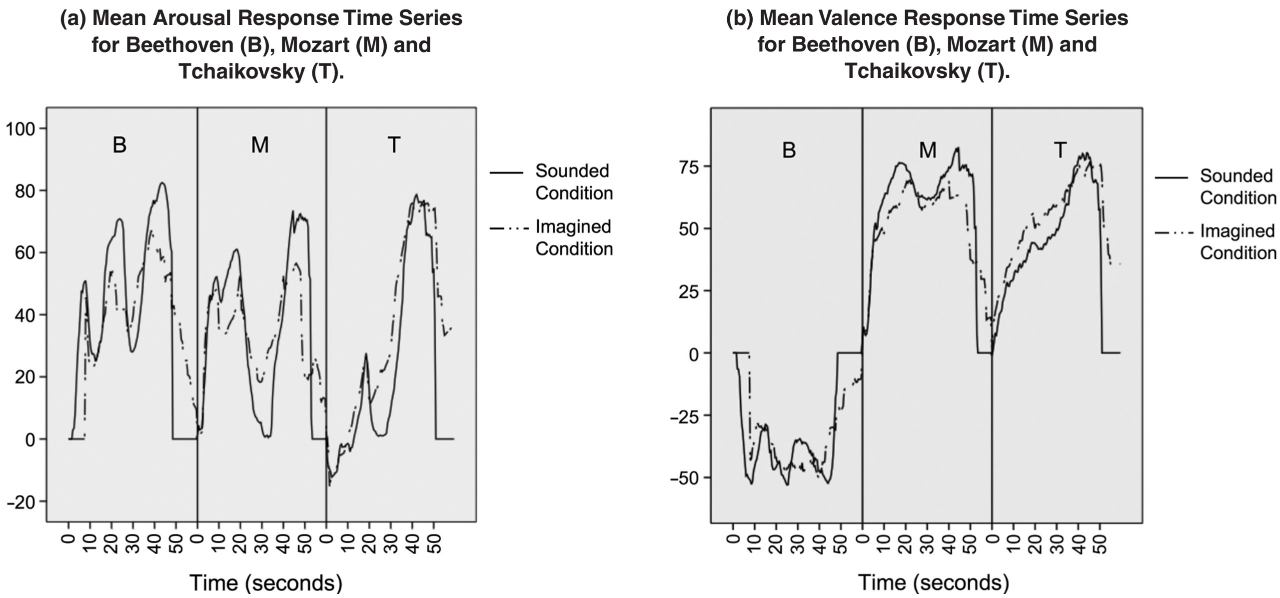

FIGURE 2. Mean sounded and imagined time series responses.

(B2′) indicates that the response was expected to be due to the feature change (B2). During the time between B2 and B2′, this participant responded with a clear decrease in arousal in the sounded condition. For the imagined condition the decrease in arousal from B2 to B2' in the time-series profile provides evidence that the turning point was imagined.

Eight musical feature turning points were identified for our analysis across all pieces. Seven of the events involved loudness, pitch, note density, or a combination of these features and were predicted to evoke changes in arousal. One event involved mode and was predicted to evoke a change in valence. Overall, participants responded to the musical feature turning points in the sounded music as expected. For all eight turning points, an average of 84% of the participants responded to the event in the way we had predicted in the sounded condition.

Experiment 2

Experiment 2 was designed to complement Experiment 1 in two main respects. First, we explored emotional labeling in Experiment 2. Although valence and arousal have been used extensively in continuous response experimentation, other labels, such as tension or pleasantness, have been used as well. There is some controversy over which labels are the most appropriate and provide the most clarity and accuracy in capturing emotional experience (Schubert, 2001, 2010; Sloboda & Lehmann, 2001).

We explored emotional labeling by adapting our software to test for the dimension of *emotionality*. Emotionality is measured in a single dimension and was defined by Sloboda and Lehmann (2001) as "the capacity of the performance at [any given] moment to suggest,

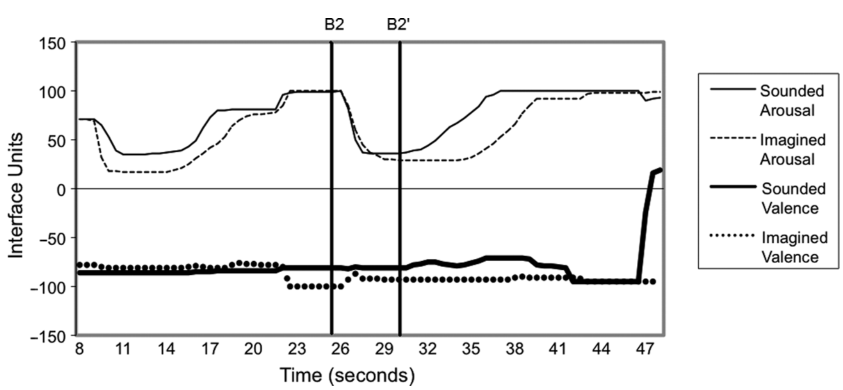

FIGURE 3. A depiction of a musical feature turning point measurement for Beethoven.

communicate, or evoke musically relevant emotion." Our first goal was to determine if emotionality could be used successfully in our continuous response paradigm. We were also interested in whether using one dimension might be easier for participants than tracking two dimensions, due to the reduction in number of decisions at each time point. Another hypothesis was that emotionality responses would resemble those for arousal more so than those for valence, as Schubert (2001) suggested.

Experiment 2 also allowed us to test whether or not participants could replicate their continuous response profiles (test-retest reliability). To do this, we retested three participants from Experiment 1 on the valence and arousal continuous response task. Successful replication from these repeat participants would increase the reliability of our Experiment 1 findings, augmenting the few studies on test-retest reliability in continuous response (Nagel et al., 2007; Schubert, 1999). Finally, we wanted to apply cross-correlation analysis to the combined results of both experiments, to capture the fine points of the relationship between profiles generated in the sounded and imagined conditions.

Method

PARTICIPANTS

The second experiment involved 11 Bucknell University undergraduate students (18–22 years old; 8 female, 3 male). Seven participants were previously untested students from the Introductory Psychology participant pool and four participants were retested from Experiment 1. All returning participants indicated on a questionnaire that recollections of the task in Experiment 1 (15–18 weeks prior) did not affect their responses in Experiment 2. Music experience ranged from 8–14 years with an average of 11.27 years.

MATERIALS

The materials used in the tapping task and the continuous response task were the same as in Experiment 1, with one exception. One change was made to the interface of the RTCRR for the continuous "emotionality" data collection: *Emotionality* replaced *valence* along the x-axis, and the y-axis was left unlabeled so that participants were presented with a single response axis.

STIMULI

The three musical excerpts used in Experiment 1 also were used in this experiment.

PROCEDURE

The first protocol was designed for repeat participants from Experiment 1. Their first task involved continuous response testing using the emotionality dimension. Consistent with the methods of Experiment 1, participants responded to each excerpt twice in the sounded condition and three times in the imagined condition. We then retested these participants on the continuous response task from Experiment 1 using the valence and arousal dimensions. One trial was taken in each of the sounded and imagined conditions in this task.

The second protocol was designed for previously untested participants. This procedure was the same as in Experiment 1 except that the experimental task measured emotionality in place of valence and arousal.

Results

The results of Experiment 2 are organized as follows. First, we report which trials were retained and which were excluded. This is followed by a pooling of data from the two experiments in which a test-retest comparison is reported, followed by a comparison of dimensions, and finally a comparison of mean responses and individual response lag structure between sounded and imagined conditions.

EXCLUDED TRIALS

In the emotionality task, each of 11 participants rated three excerpts twice in the sounded condition and three times in the imagined condition, for a total of 165 continuous response profiles. We used the second trial of the sounded condition and the third trial of the imagined condition in our analysis of the emotionality task. This left 66 trials eligible for analysis, or 33 within-subject sounded-to-imagined condition comparisons.

For each retest participant ($n = 3$; one person's retest data file had technical problems, so data for three participants remained) on the valence-arousal task, we collected one trial for each excerpt in both the sounded and imagined conditions. This yielded 18 trials for our analysis.

For the emotionality tasks, no participants were excluded from analysis and, other than the trials excluded because of our trial selection method, no individual trials were excluded. All participants scored a 3 or above (using a 1–7 scale) on excerpt familiarity and perceived proficiency in imagining the excerpt.

All participants completed the tapping task (either in Experiment 1 or 2) sufficiently well to be retained for the study, using similar criteria for new participants in Experiment 2 as were used in Experiment 1.

TEST-RETEST COMPARISON

Three participants completed the arousal and valence response tasks twice, allowing us to examine test-retest reliability in a small sample. Repeated measures were compared for each of the three participants by piece (3), dimension (arousal and valence) and condition (sounded and imagined) using a Pearson correlation analysis. As shown in Table 1, all participants' repeated responses most closely resembled the original for the arousal-sounded (AS) condition, regardless of the piece (all correlations were significant). Further, all participants' repeated responses were significantly correlated for the Tchaikovsky regardless of the condition. The Beethoven responses were least reliable for participant 1, as reflected by the negative correlations (which were due to the numerous changes in response that were "out of phase" when the valence-sounded (VS) condition in particular [$r = -0.50$] was repeated).

Test-retest analyses showed that valence responses had lower reliability than arousal responses, and the responses in the imagined condition were less reliable than the sounded conditions. The pauses and impulsive nature of the Beethoven example may have been responsible for the less consistent responses found compared to the other pieces. However, seven of the nine arousal-imagined (AI) correlations were significant, many impressively so—and six of the nine valence-imagined (VI) responses—suggesting that reliability of imagined responses, although lower than sounded responses, is still considerable.

VISUAL COMPARISON OF SOUNDED AND IMAGINED SERIES ON EMOTIONALITY JUDGMENTS

The plots of the superimposed time series comparing sounded and imagined conditions for the emotionality dimension is shown in Figure 4. Visual inspection of this figure shows similarity between sounded and imagined conditions at least as striking as we saw in Experiment 1. All three pieces showed quite a large range of emotionality judgments in both conditions, which is more similar to the arousal than valence profiles seen in Experiment 1. We quantify these relationships in the next sections.

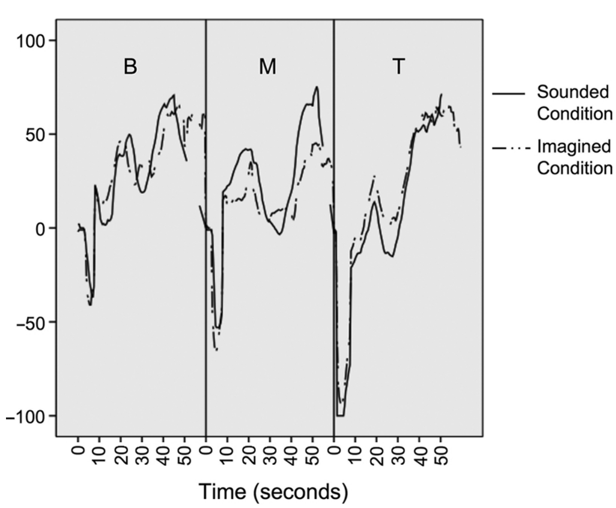

FIGURE 4. Mean emotionality response time series for Beethoven (B), Mozart (M), and Tchaikovsky (T).

ARE AROUSAL AND EMOTION DIMENSIONS RELATED IN SOUNDED MUSIC?

CCFs (Cross-Correlation Functions, see Campbell, Lo, & MacKinlay, 1997) were performed between pairs of emotion response dimension mean time series for the sounded conditions to determine whether any pairs had similarities. Cross-correlation takes two time series and performs several correlation analyses between them. Each correlation analysis is performed at different time lags between the two series. This allows identification of the lag at which the two series are maximally correlated. Further, similarity is indicated by large peaks in correlations for at least one lag in the CCF. The CCFs in Figure 5 show that arousal and emotion produce the strongest and most frequent correlation coefficients (correlation coefficients above the confidence interval), with the other two pairs (emotion with valence and arousal with valence) less so, confirming the prediction that emotion and arousal tap into a similar semantic dimension.

ANALYSIS OF LAG STRUCTURE BETWEEN IMAGINED AND SOUNDED CONDITIONS

In analyzing cross-correlations between imagined and sounded responses for each of the three tasks, two analyses were of interest: (1) The number of responses that produced significant lags for at least one lag of the CCF. A large number of responses producing at least some significant lag for a condition indicates that participants are able to reproduce the imagined condition in a way that resembles the sounded condition response. (2) Of these "significant peak lag" responses, the lag at which the peak correlation occurred within the comparison. This will provide information about the lag structure between imagined and sounded responses. For example, is there a greater delay in imagined responses when arousal is the dimension being responded to? If so, how long is this delay (lag)?

For these analyses, a difference transformation was made. This means that instead of examining the absolute values recorded by participants, the change in value from one sample to the next is computed. This approach has been shown to reduce the effects of serial correlation and thus provide more valid results (see Schubert, 2002). Further, only the 100 samples beginning from the 8th second were used in the analyses. This was to ensure that

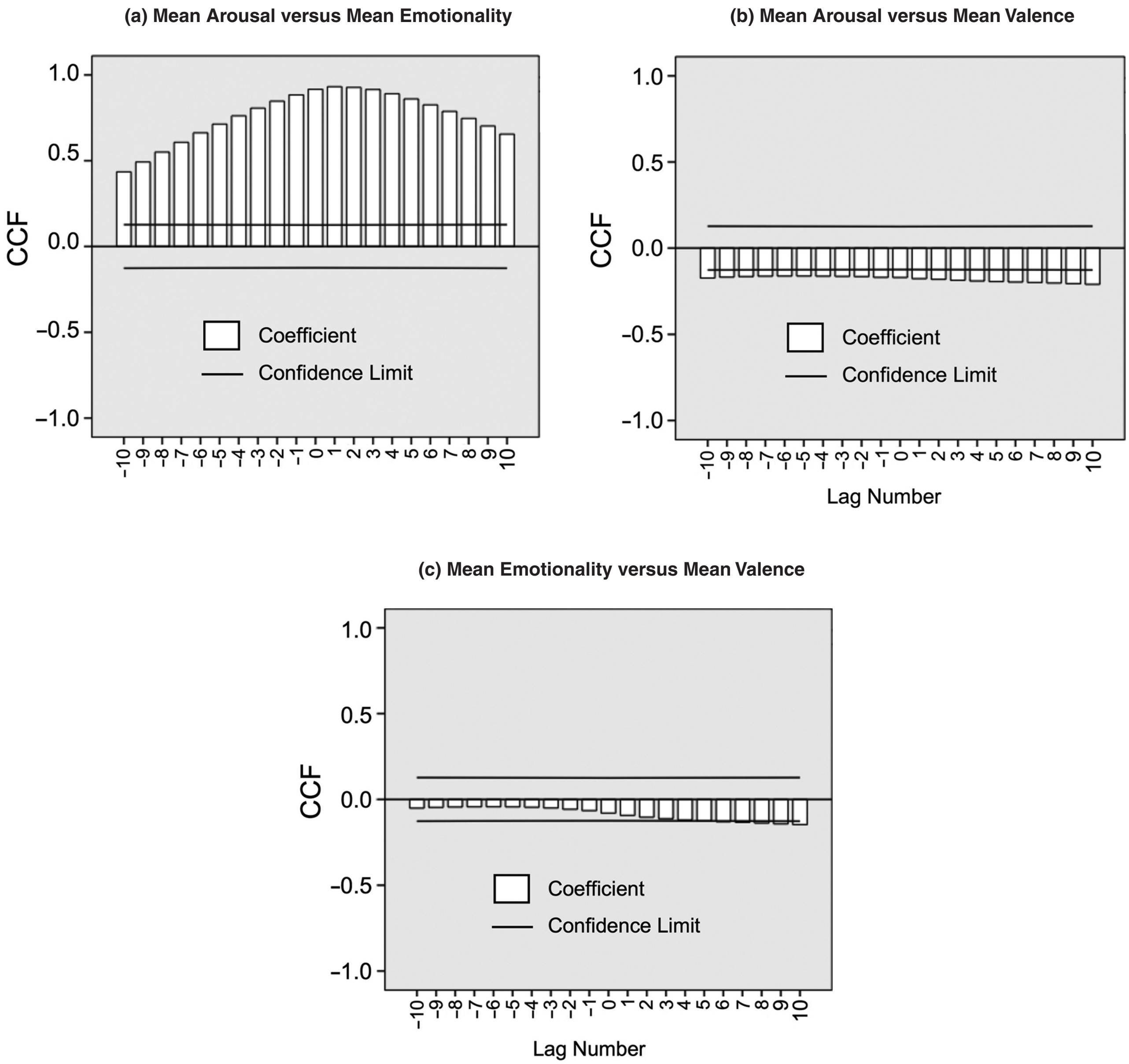

FIGURE 5. Analysis of lag structure between emotional dimensions in sounded music.

results were not biased by the presence of sounded music in the sounded condition and the imagined condition (in which the music was used for cueing the participant for the first 5 to 7 s). Sampling was made at 2Hz.

The CCF output was analyzed for presence of at least one significant correlation at any lag, and for these, the lag at which the peak significant correlation occurred. This was determined by categorizing a time series response pair (imagined versus sounded) as "significant" if the largest correlation coefficient was above the confidence interval, which in this case was set to one standard error (using the standard error of cross-correlation at each lag as reported by Box & Jenkins, 1976, p. 376). That is, if for an imagined-sounded comparison, no correlations were greater than 1 *SE* (at any lag in the CCF), then the pair was categorized as "not significant." Table 2 summarizes the results of the analyses. The mean lag between imagined and sounded conditions was less than 2 samples (1 s) for each cell (in Table 1), with the exception of emotionality response in Beethoven, in which the mean peak imagined response leads the sounded response by 2.73 (1.36 s). This means that imagined

TABLE 1. Pearson Correlation Coefficients of Responses For Repeat Participants in Experiments 1 and 2.

		Condition				
Stimulus	Participant	VS	VI	AS	AI	Mean
B	1	<u>−.50</u>	<u>−.11</u>	.39	<u>−.02</u>	−.05
	2	.86	<u>.02</u>	.78	.77	.61
	3	<u>−.38</u>	<u>−.59</u>	.78	.85	.17
	Mean	.00	−.23	**.65**	.55	
M	1	<u>−.12</u>	.41	.84	<u>−.10</u>	.34
	2	.59	.86	.96	.89	**.83**
	3	.77	.78	.92	.84	**.83**
	Mean	.49	**.69**	**.91**	.57	
T	1	.81	.83	.77	.82	**.81**
	2	.95	.92	.95	.98	**.95**
	3	.96	.88	.97	.89	**.93**
	Mean	**.91**	**.88**	**.90**	**.90**	

Note. Underlined correlation coefficients are either negative or non-significant (p = .05). Bold indicates all coefficients in the row or column were significant. *Df* varies for each cell, see text. Codes: V = Valence, A = Arousal, S = Sounded, I = Imagined, B = Beethoven, M = Mozart, T = Tchaikovsky.

responses tended to be slightly rushed with respect to the sounded condition when rating the emotion expressed for the Beethoven excerpt. However, it also should be noted that this mean peak lag returned a relatively large variability in location (second only to Mozart arousal), with peak lags distributed with a *SD* of 5.14 samples (2.57 s) across participants.

The most consistent responses (considering peak deviation) were for the valence response for Beethoven (*SD* = 1.54 samples), and the arousal response to Tchaikovsky (*SD* = 1.98 samples). The closest-to-instantaneous mean lag structure (i.e., imagined and sounded condition response timings best matched) occurred for arousal responses for all three pieces, in each case less than ±0.36 samples (±0.18 s). However, the variability of the peak lag for Beethoven and Mozart are relatively large (3.14 samples and 5.80 samples respectively). This suggests that participants had less idiosyncratic recall of imagined arousal responses for the Tchaikovsky. Tchaikovsky was also the only piece where the average peak lag was leading (imagined condition response occurring faster relative to the sounded condition—see negative coefficients in third column of Table 2). The results indicate that lag structure between sounded and imagined responses are quite reliably correlated, with 83% to 100% of participants producing at least one significant correlation (at some

TABLE 2. Peak Lag Statistics by Stimulus and Condition.

Stimulus	Task	# Responses	Mean Peak Lag (in samples)	SD Peak Lag (in samples)	% Significant (>1 *SE*)
B	A	16	−.03	3.14	94
	V	16	.03	1.54	94
	E	12	−2.73	5.14	92
M	A	12	.03	5.80	100
	V	12	2.00	4.13	83
	E	12	.55	3.15	92
T	A	15	−.04	1.98	93
	V	15	−.71	3.42	93
	E	12	1.36	3.49	92

Note. B = Beethoven, M = Mozart, T = Tchaikovsky, V = Valence, A = Arousal, E = Emotionality. % Significant refers to the percentage of participants for whom the correlation coefficient of the peak lag between imagined and sounded conditions was greater than the significance level of 1 *SE*.

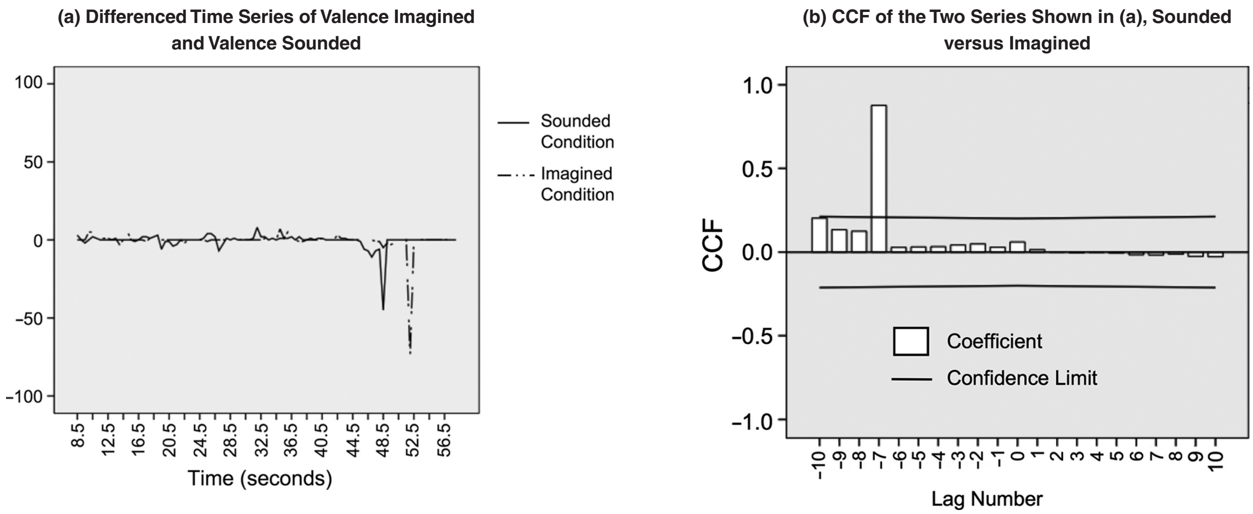

FIGURE 6. Time series and CCF for participant 21 valence response to Mozart.

lag) for each of the three pieces, and each of the three conditions.

To exemplify these results, consider the case of the valence response from participant 21 to the Mozart. Figure 6a demonstrates that the imagined valence response time series resembles the sounded valence series, but the imagined responses occur relatively earlier (most clearly depicted between the 44th and 53rd second of the response). This difference is quantified in the CCF plot (Figure 6b), where a peak at −7 (3.5 s) samples is clearly visible. It tells us that the imagined response is similar to the sounded response but leading (rushing or jumping ahead) by 3.5 s. It is these peak correlation coefficients that were collated and reported in Table 2.

General Discussion

This paper described two experiments that investigated how well the emotion associated with musical excerpts could be recalled. The novel aspects of the study were that emotional responses were recalled continuously as music unfolded mentally, and that three emotional dimensions were measured. Further, one-minute excerpts of late Classical and Romantic music were used to allow a variety of emotions within works that were familiar and did not refer to words.

At the most basic level, this study is one of a number of studies that report evidence of use of auditory imagery (Halpern, 1988a; Hubbard & Stoeckig, 1992; Intons-Peterson, 1992). The consistency of profiles over sounded and imagined conditions demonstrates that a majority of participants could successfully track emotion in a sounded excerpt of music, and that they could then extract emotion from their memory of that excerpt.

It is evident that familiarity with the piece is critical for this complex task. We screened all of our participants for music training, which made it likely they would have encountered famous pieces like these. Furthermore, we insured that they knew these particular pieces. However, we think it unlikely that our trained participants had already coded the emotional changes in each piece that we required them to externalize in the moment-to-moment paradigm of continuous response. Rather, we propose that participants recalled the requested piece and made the requested judgments *de novo,* utilizing the memory representations. Combined with similarity in response to features, this success suggests that the memory representation is quite detailed. The representation is also apparently quite stable, as evidenced by the similar responses over a lengthy period by our participants who took part in both experiments.

One argument against this notion is that emotional responses were consciously utilized during the sounded condition and that they were therefore readily accessible in the auditory image of the piece, within a session. However, we think this is unlikely, as the three excerpts were tested back-to-back in the sounded condition before being presented in the imagined condition. Participants were not aware that they would be responding to the same piece, and there was no reason to believe that they used conscious strategies to memorize responses. It also strains credulity that the precise judgments could be replicated on demand, at the sampling rate (2 Hz) used for continuous responses.

We think it particularly noteworthy that the emotions extracted from the musical image are similar to the emotion expressed by sounded music. This similarity allows us to propose that the underlying processing mechanisms for the two conditions are similar, as has been suggested for musical imagery tasks involving pitch, time, or timbre judgments (Halpern, 1988a, 1988b; Halpern et al., 2004). The retrieval of auditory images for music thus may not only be fairly easy for listeners, but also enjoyable (many people report being able to retrieve images of their favorite tunes, in whatever genres they are familiar with).

In this study, lag structure was more or less instantaneous (typically less than 0.5 s), meaning that participants did not fluctuate in time (slow down or speed up) when reporting emotion expressed from a musical image compared to a sounded recording. Thus, although imagery processes are typically slow compared to other kinds of memory retrievals (judgments are on the order of seconds in standard mental rotation or other imagery tasks, as compared to, for instance, a few hundred milliseconds in a lexical decision task), in this task, the emotional judgments apparently were available at about the same time as the musical information itself was retrieved (on the order of less than one second). It may be the case that using very familiar pieces allowed some anticipation of the next note or group of notes (Leaver, Van Lare, Zielinksi, Halpern, & Rauschecker, 2009). The only occasion when there was notable variability was for the piece that had several stop-starts (fermata), namely the opening of Beethoven's Fifth Symphony. This could be connected with the mismatch between perceptions of silences (Fabian & Schubert, 2008; ten Hoopen, et al., 2006) and the subsequent variability in response that this might have caused.

However, the most important evidence that emotional identification in sounded and imagined conditions was the same is through error analysis. If we set the baseline measurement error from the tapping task (which was, on average, 4.27%), then we would expect an error through tempo instability alone to produce a margin of 2.14 s (4.27% of the 50 s excerpt) or 4.28 samples. From Table 2 we can see that of none of the nine experimental cells (three pieces by three emotion dimension) had a mean lag structure greater than this value (greatest being −2.73 samples for emotionality identification in Beethoven). While the range of lag structure is fairly large for some pieces (in particular, Beethoven emotionality, as discussed), these findings nevertheless suggest that timing may be better overall when tracking emotion in familiar music than it is for a simple tapping task. The "high-level" cognitive task of tracking emotion in familiar music may facilitate low-level tempo tracking because tempo is integrated into the recall of the higher-level percept. This could also be indicative of the greater difficulty in the production task (tapping a tempo) versus a perception or emotion task. Further research on whether the nature of the task (emotion rating versus pure tempo tapping; recall versus production) may influence tempo is a corollary of our study.

The relationships among valence, arousal, and emotionality were consistent with that predicted by Schubert (2001, 2010), supporting the idea that emotionality and arousal may be semantic constructs with shared ontology. However, imagined emotionality did have larger mean lag than arousal (and valence, for that matter), and tended to rush ahead with respect to emotionality response in the sounded condition. This is in contrast to imagined arousal, which remained instantaneous with sounded arousal. It is, therefore, possible that "arousal" provides a more reliable and stable indication of emotional response than its related counterpart, emotionality. This paper does not aim to address the relationship and further research may explain the similarity in profiles between arousal and emotionality, but at the same time the difference in lag structure between the two.

In summary, it appears that despite the attention-demanding nature of musical imagery, people can reliably maintain that representation and extract novel information from it. We also note that the length of time in which we asked our participants to do these tasks was much longer than in most prior auditory imagery experiments, suggesting that these are not fleeting experiences. And we showed that the emotional information thus extracted is similar to the perceptual experience in important ways. In this context, a quote from a musician entombed for 18 hours in the rubble of the Haiti earthquake of 2010 is illuminating. He passed the time by imagining some of his familiar concertos: "For example, if I perform the Franck sonata, which is [sic] 35 minutes long in my honors recital at Juilliard, then I would bring myself to that time. That allows me not only to kill time, *but also to mentally take myself out of the space where I was*" (National Public Radio, 2010, italics added).

Author Note

Correspondence concerning this article should be addressed to Andrea Halpern, Psychology Department, Bucknell University, Lewisburg, PA 17837. E-MAIL: ahalpern@bucknell.edu

References

Bailes, F. (2007). The prevalence and nature of imagined music in the everyday lives of music students. *Psychology of Music*, *35*, 555–570.

Box, G. E. P., & G. M. Jenkins. (1976). *Time series analysis: Forecasting and control*. San Francisco, CA: Holden-Day.

Campbell, J. Y., Lo, A. W., & MacKinlay, A. C. (1997). *The econometrics of financial markets*. Princeton, NJ: Princeton University Press.

Evans, P., & Schubert, E. (2008). Relationships between expressed and felt emotions in music. *Musicae Scientiae, 12*, 75–99.

Fabian, D., & Schubert, E. (2008). Musical character and the performance and perception of dotting, articulation and tempo in 34 recordings of Variation 7 from JS Bach's Goldberg Variations (BWV 988). *Musicae Scientiae, 12*, 177–206.

Gabrielsson, A. (2010). Strong experiences with music. In P. N. Juslin & J. A. Sloboda (Eds.), *Handbook of music and emotion: Theory, research, applications.* (pp. 547–574). Oxford: Oxford University Press.

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

Gagnon, L., & Peretz, I. (2003). Mode and tempo relative contributions to "happy-sad" judgments in equitone melodies. *Cognition and Emotion, 17*, 25–40.

Halpern, A. R. (1988a). Mental scanning in auditory imagery for songs. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 434–443.

Halpern, A. R. (1988b). Perceived and imagined tempos of familiar songs. *Music Perception, 6*, 193–202.

Halpern, A. R., Zatorre, R., Bouffard, M., & Johnson, J. (2004). Behavioral and neural correlates of perceived and imagined musical timbre. *Neuropsychologia, 42*, 1281–1292.

Hevner, K. (1935). The affective character of the major and minor modes in music. *American Journal of Psychology, 47*, 103–118.

Hubbard, T. L., & Stoeckig, K. (1992). The representation of pitch in musical imagery. In D. Reisberg (Ed.), *Auditory imagery* (pp. 199-235). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Intons-Peterson, M. J. (1992). Components of auditory imagery. In D. Reisberg (Ed.), *Auditory imagery* (pp. 45–71). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Kozhevnikov, M., Kosslyn, S., & Shepard, J. (2005). Spatial versus object visualizers: A new characterization of visual cognitive style. *Memory and Cognitio*n, *33*, 710–726.

Leaver, A. M., Van Lare, J., Zielinski, B., Halpern, A. R., & Rauschecker, J. P. (2009). Brain activation during anticipation of sound sequences. *Journal of Neuroscience, 29*, 2477–2485.

Nagel, F., Kopiez, R., Grewe, O., & Altenmuller, E. (2007). EMuJoy: Software for continuous measurement of perceived emotions in music. *Behavior Research Methods, 39*, 283–290.

National Public Radio (2010, January 23). *Wife, school lost in quake, violinist vows to rebuild* [Radio program]. Retrieved from http://www.npr.org/templates/story/story.php?storyId=122900781&ft=1&f=122900781

Repp, B. H. (2005). Sensorimotor synchronization: A review of the tapping literature. *Psychonomic Bulletin and Review*, *12*, 969–992.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Social Psychology, 39*, 1161–1178.

Schubert, E. (1999). Measuring emotion continuously: Validity and reliability of the two-dimensional emotion-space. *Australian Journal of Psychology, 51*, 154–165.

Schubert, E. (2001). Continuous measurement of self-report emotional response in music. In P. N. Juslin & J. A. Sloboda (Eds.), *Music and emotion: Theory and research* (pp. 393–414). Oxford: Oxford University Press.

Schubert, E. (2002). Correlation analysis of continuous emotional response to music: Correcting for the effects of serial correlation. *Musicae Scientiae, Special Issue*, 213–236.

Schubert, E. (2004). Modeling perceived emotion with continuous musical features. *Music Perception*, *21*, 561–585.

Schubert, E. (2007). Real time cognitive response recording. *Proceedings of the Inaugural International Conference on Music Communication Science*, Sydney, Australia. Paper retrieved from http://marcs.uws.edu.au/links/ICoMusic/ArchiveCD/fullpaper.html

Schubert, E. (2010). Continuous self-report methods. In P. N. Juslin & J. A. Sloboda (Eds.), *Handbook of music and emotion: Theory, research, applications* (pp. 223–253). Oxford: Oxford University Press.

Schubert, E., & Dunsmuir, W. (1999). Regression modeling continuous data in music psychology. In S. W. Yi (Ed), *Music, mind, and science* (pp. 298–352). Seoul, South Korea: Seoul National University.

Schubert, E., Evans, P., & Rink, J. (2006). Emotion in real and imagined music: Same or different? In M. Baroni, A. R. Addessi, R. Caterina, & M. Costa (Eds.), *Proceedings of the Ninth International Conference on Music Perception and Cognition* (pp. 810–814). Bologna, Italy.

SERAFINE, M. L., CROWDER, R. G., & REPP, B. H. (1984). Integration of melody and text in memory for songs. *Cognition*, *16*, 285–303.

SLOBODA, J. A., & LEHMANN, A. C. (2001). Tracking performance correlates of changes in perceived intensity of emotion during different interpretations of a Chopin piano prelude. *Music Perception*, *19*, 87–120.

SWELLER, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*, 257–285.

SWELLER, J. (2006). Discussion of 'emerging topics in cognitive load research: Using learner and information characteristics in the design of powerful learning environments.' *Applied Cognitive Psychology, 20*, 353–357.

TEN HOOPEN, G., SASAKI, T., NAKAJIMA, Y., REMIJN, G., MASSIER, B., RHEBERGEN, K. S., & HOLLEMAN, W. (2006). Time-shrinking and categorical temporal ratio perception: Evidence for a 1:1 temporal category. *Music Perception, 24*, 1–22.