5-2011

An Experimental Analysis of the Role of Harmony in Musical Memory and the Categorization of Genre

David H. Witkowsky
College of William and Mary

Running Head: Harmony, Memory, and Genre

An Experimental Analysis of the Role of Harmony in Musical Memory and the Categorization of Genre

A thesis submitted in partial fulfillment of the requirement
for the degree of Bachelors of Arts in Music from
The College of William and Mary

by

David Witkowsky

Accepted for High Honors
(Honors, High Honors, Highest Honors)

Greg Bowers, Director

Gayle Murchison

Peter Vishton

Williamsburg, VA

May 2, 2011

Abstract

Expectations play a significant role in the way a listener experiences a piece of music. These expectations have been thought to generate through the probabilistic learning of harmonic structures by exposure to music. They make up what is called schematic memory. Through this understanding of musical memory, researchers have found that different schematic frameworks exist in participants from different cultures. This study has two primary goals. The first is to isolate harmony as a key element of schematic memory for music. The second is to consider the possibility that different genres within Western music may have their own schematic frameworks, as represented in musical memory, in a way similar to music from different cultures. While harmony was found to play a significant role in schematic memory, it remains unclear whether or not different genres within Western culture are categorized independently due to the differences in their harmonic frameworks.

An Experimental Analysis of the Role of Harmony in Musical Memory and the Categorization of Genre

Music and Memory

The focus on memory as a method for studying music first arose when Leonard Meyer claimed in his book, *Emotion and Meaning in Music*, that a person's reaction to a piece of music was ultimately guided by his or her expectations (1956). Meyer concerned himself with emotional response, as the title of his book indicates, but not without implicating the role of learning (1956, p. 43), and thus memory.

For Meyer, emotional responses are elicited by the way in which a piece of music conforms to or deviates from a listener's expectations. The essence of this claim is understood in his analogy:

> The sensation of falling through space, unconditioned by any belief or knowledge as to the ultimate outcome, will, for instance, arouse highly unpleasant emotions, [while] a similar fall experienced as a parachute jump in an amusement park may, because of our belief in the presence of control and in the nature of the resolution, prove most pleasing. (p. 20)

Applied to music, the musical moment's impact on the listener is dictated by the expectations the listener has at that moment. These expectations, for Meyer, are based in the listener's understanding of musical styles. Therefore the listener learns, i.e. establishes a memory, for music that informs his or her listening experience (1956).

It is important to clarify Meyer's use of the word "styles" when referring to music. For Meyer, a musical "style" could pertain to an individual composer, such as Franz Joseph Haydn. Haydn's music, in turn, belonged to a "style system" to which many composers belonged. In this example, the "style system" would be Western common practice music. "Common practice" refers to music in which harmony is emphasized and guided by conventions that are the foundation of music by Western European composers such as Bach and Mozart. For the purposes of this paper, the term "genre" will approximate what Meyer termed "style system," although his definition of both style and style system were more flexible (1956, p. 64).

Meyer used the analogy of music as language in his discussion of musical style. While the analogy makes sense anecdotally, linguists Ray Jackendoff and Fred Lerdahl further developed the concept into a theory of music that attempted to define a "grammar" of music. In contrast with Schenkerian analysis of harmony, they derived structure by referring to rules such as "proximity" and "similarity," terms that were derived from Gestalt psychology, though they were applied to musical concepts such as register and dynamics (1981, p. 58, 62). Ultimately, their analysis included a wide variety of musical elements, and ways in which these elements can be used to determine grammatical units or groupings in a piece of music. These rules for groupings make up the structure that Jackendoff and Lerdahl referred to as musical grammar (1981).

Meyer's method for defining musical styles, and consequently expectations, was rooted in the use of probabilities as a method for defining structure in music (1956). His approach to probabilistic structure in music specifically emphasized harmony as the key feature of music that informs expectations. The most general example of this is the probability that a tonic chord (I) will be followed by a pre-dominant chord (ii/IV), which will be followed by the dominant chord (V), which returns to the tonic chord (I). As a result, much of the research that exists today treats harmony as the fundamental element guiding music memory, and specifically expectations (1956).

Schematic Expectations

The assumption that harmony defines and explains listeners' expectations has led to the development of a theoretical schema, a harmonic framework that explains those expectations. The general principle is that, through experience, the listener gains knowledge of the underlying harmonic scheme of the musical style he or she is listening to. Ultimately, the listener develops an internalized framework that has taken into account the probability of musical events, specifically harmony, and can then base his or her expectations on this framework (Bharucha, 1994). For example, after having listened to Mozart enough to internalize expectations for the probable events in his music, these expectations would serve as the reference while listening to a piece of his music.

One basis for this schema lies in what Krumhansl and Shepard call "tonal hierarchies" (1979). In their experiment, Krumhansl and Shepard used a "probe tone technique," in which musical sequence was presented and then followed by this "probe tone" (Bharucha 1994, p.223). The musical sequence presented a context and the participants were asked to judge how well the probe tone completed that sequence on a scale of 1 to 7, on which 7 meant that it completed the sequence the best. The result of their work was a visualized map of tonal hierarchies, demonstrating which pitches were selected as best fits for given sequences (see Figure 1). The musical sequences were meant to establish a clear key. This meant that the tonal hierarchies demonstrated what participants believed the most probable next pitch should be within the context of the key.

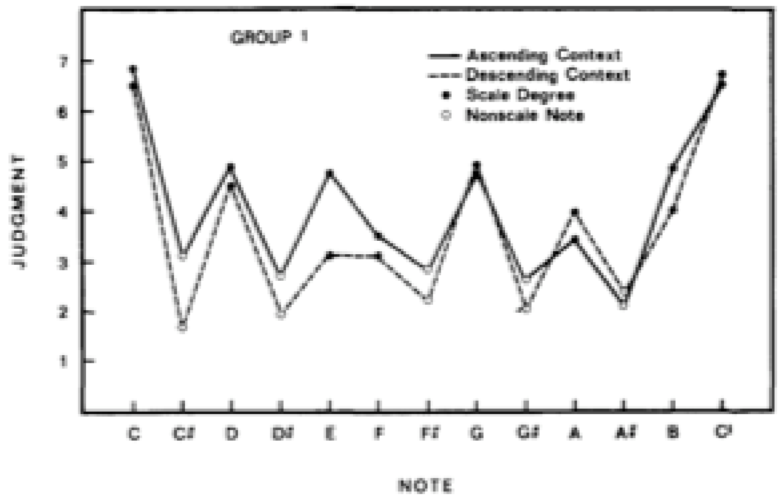


Figure 1. The judgment profiles for the 12 chromatic pitches, given a sequence in the key of C (Krumhansl & Shepard, 1979).

Similarly, Diana Deutsch conducted an experiment in which participants were asked to recall a 12-note sequence that had been presented to them. If the sequences followed the rules of common practice theory, the participants were more likely to correctly recall the sequence of notes (1980). This study is crucial to the literature for its demonstration that schematic memory enhances the ability to encode memory for novel pieces of music.

The assumption that common practice conventions are representative of musical schemata, however, is a relatively common one in the existing literature. Bharucha has explained the acquisition of the schematic framework using a neural networks model that incorporates common practice conventions as the de facto framework. The neural networks model begins with all pitches sharing an equal probabilistic weight, on one level. On another level, representations of chords are developed when certain pitches occur together frequently. Eventually, this second level represents all possible chords. Within each representation of a chord there is the probabilistic information regarding which pitches are more likely to occur. At a third layer, the concept of key develops as certain chords occur in various relationships with each other. At this level, entities representing each possible key are established, taking into account the probabilities of the chords within that key, and by extension pitches too. This model explains how, for example, listeners exposed to common practice conventions can internalize the conventions as the set of expectations that make up the schema. Ultimately, this understanding of the schema supports the legitimacy of using common practice conventions as the assumed schematic framework in research.

Deutsch's results are also critical for understanding the role of the schema in musical memory. They suggest that having an internalized schema affects a participant's memory for music that he or she considers novel (1980). This finding has been instrumental in much of the research that aims to further understand the role of schematic expectations in memory. It also gives rise to the question: what is the relationship between the schema and memory for specific pieces of music? The distinction between the two types of memory is necessary for the concept of a schema. Otherwise, the emotional response towards a piece, in terms of Meyer's analysis, could only exist the first time a listener encountered it; after the first experience, the listener would be equipped with all the knowledge about how that piece unfolds. Instead, the listener's schema is violated in the same way every time, despite specific knowledge that should prevent such a sense of violated expectations.

Bharucha divides musical memory into two categories: schematic and veridical. The schema is the learned expectations gleaned from exposure to a musical style. Its expectations are based in the probability that something occurs within the context of the musical style, not the particular piece. Conversely, veridical memory encapsulates expectations for what should happen in a specific, familiar piece. As a result, an informed veridical memory for Beethoven's 5^{th} Symphony will not affect the schematic memory, insofar as deviations that are specific to that piece will not be incorporated into the schema. This explains why the same moments in a piece of music create a high level of affective response despite the large number of times a person might listen to it (1994).

Ultimately, this division of memory favors Deutsch's results (1980). In order to successfully encode a new piece of music into memory, it needs to be understood through the musical schema. In the case of Deutsch's study, the existence of a musical schema (using common practice conventions) aided in the encoding and subsequent recall of novel music sequences. Conversely, the sequences that did not follow the rules of the schema were harder to recall, due to the fact that they could not be interpreted through the schema.

This concept has guided a substantial amount of research on music and memory, especially as it concerns finding the boundaries of the schema's role in memory. Justus and Bharucha were interested in the interaction between schematic memory and veridical memory, by using "previews" prior to testing the participant. In the case of the first experiment, participants were tasked with judging intonation in a set of chords, one prime chord and one target chord. In half of the presentations the target chord that followed the prime chord was considered very unlikely given a common practice-based schema. In the other half, the target chord was very likely. In order to eliminate the difference this should make, the pairings were presented before judgments were made, so that the chord pairs were familiar, eliminating the uncertainty concerning whether the target chord would be close to or distant from the schematic expectations associated with the prime chord. Then they were presented a second time, and the participant had to decide whether or not the target chord was mistuned. In some cases, no preview was given (2001).

In the instances where there was no preview, schematically close targets resulted in faster processing if the targets were in tune. Likewise, schematically distant targets facilitated processing if the target was also mistuned. The latter situation was explained by "an observed bias to judge distantly related chords as mistuned" (Justus and Bharucha, 2001, p. 1005). When previews were given, overall processing improved, but the faster processing related to the relationship between schematic proximity and tuning was maintained. This indicated that although veridical memory based on the preview may affect performance, the role and effect of schematic information are not affected by the added veridical information.

This study raises yet another question due to the association between schematically probable chords and consonance and, conversely, schematically improbable chords and dissonance. Is the schema learned from musical exposure and probabilistic learning, or is it just due to natural, physical properties of sound? Is it possible that the physical nature of sound could have something to do with which chords or tones are expected, or desired, to follow after other chords or tones? Research suggests that this is not the case.

Bharucha points out that when given a choice to go from one chord to another, the greatest expectation is to go from one chord on the circle of fifths – a spatial representation of tonal hierarchies – to another, as opposed to what might make the most sense considering pitch frequencies (1994, p. 222). The relationship between the frequencies of the pitches in a chord generally dictates the degree to which that chord is perceived as consonant or dissonant. However, if a listener were to truly expect the most consonant chord to follow, he or she would expect a chord with the most common tones, so as to minimize conflicting frequencies. This chord relationship is different from that between adjacent chords on the circle of fifths, and it is the circle of fifths choice that is most expected.

A possible remedy for this contradiction becomes apparent when looking at the way in which the concept of a musical key ameliorates the problem. As seen in the neural networks model for the musical schema, the learned probabilities of certain chords following others are responsible for the construction of the concept of a key. That is, the key is more important than the pure relationships between chords; it is a higher-level concept. This is intuitively apparent in much of the research with pitch sequences in which the establishment of a key, rather than a sequence of chords, is used (Krumhansl and Shepard, 1979; Deutsch, 1980) to study schematic expectations. From a physical perspective, while the adjacent chord on the circle of fifths is not the most physically consonant choice, the key to which it corresponds is the most physically consonant choice. Each key along the circle of fifths differs by only a single pitch change when compared to its neighbor. In other words, the adjacent chord represents the maximum consonant overlap between two keys.

Even so, there is still evidence that schematic expectations are cognitively rather than physically based. Much of this evidence comes from research into how schematic expectations are learned during childhood.

Development of the Schema

Despite the "surprisingly sophisticated musical abilities" of musically untrained children and adults – such as being able to distinguish rhythmic patterns – a sense of key is not one of the structures young children appear to understand (Davidson, 1994, p.102; Hannon and Trehub, 2005). This implies that the acquisition of a sense of key, and likely therefore a musical schema, is something that is acquired during childhood. In particular, Davidson has codified the gradual development of a sense of key in young children.

Through Harvard's Project Zero, Davidson followed 9 children from age 1 to 7. In addition to those children, 70 others were observed at different points during those six years, for comparison with the behavior of the children who were observed throughout the course of the study. He found that children gradually arrived at a sense of key over the course of five discernable stages.

Davidson's findings were based on the concept of contour schemes, which contain three key elements. The first is the tonal frame, which describes the size of the interval the child uses in making music, specifically in singing. The second element is the level of pitch organization, which refers to the choice of pitch and its relationship to other pitches in the entire melody. If there were a sense of key, it would be manifested in the pitch organization, because there would be a central pitch, or tonic. The third element is the melodic motion, referring to a rise or fall in the melodic content of the child's song (1985).

The five stages outlined by Davidson track contour schemes, but most importantly focus on the interval content in the children's songsinging. He studied both the children's invented songs, as well as their renditions of standard songs such as "Row,

Row, Row Your Boat." In the first stage identified by Davidson, the children limit themselves to the use of an interval of a third. The most obvious part of this phenomenon is the fact that the standard melodies have their interval content condensed to include only the interval of a third, whether reached stepwise or as a leap. The second stage involves expansion to the interval of a fourth. The third stage is the interval of a fifth, and the fourth stage is the interval of a sixth. It is not until the fifth and final stage that the child incorporates the interval of an octave. This occurs around the age of 6, and is what Davidson claims is both responsible and necessary for the understanding of the concept of a tonal center, or key. In addition:

> The most striking phenomenon associated with reaching the fifth stage in contour scheme development is a striking change in behavior in which children become less inventive in their songsinging. In fact, at this age they seem to switch from spending most of their time inventing songs to spending more time listening to them (Davidson, 1994, p. 120-121).

This suggests an emphasis on gleaning a schematic framework by using their newfound sense of musical key.

Whereas Davidson's research ended at age 7, Krumhansl and Keil (1982) studied elementary school children in grades 1 through 6, as well as adults, in order to understand elements of tonal hierarchy acquisition. Their procedure included presenting a given sequence of tones and asking the participants to judge how good the sequence sounded as a potential melody. The researchers found that that the deciding factor in the judgments made by the 1st and 2nd grade students was whether or not the last two pitches pertained to the key, i.e. were diatonic tones. For 3rd and 4th graders, the deciding factor was not whether or not the last two tones were diatonic, but rather whether or not they were components of the tonic triad, even over other diatonic tones. This research indicates that schematic expectations are learned over a long period of time, and involve various levels of depth.

Cross-Cultural Considerations

The nature of developing a musical schema has led researchers to wonder if such schemata vary from culture to culture. Because different musical traditions around the world have their own musical conventions, researchers have correctly assumed that musical schemata should be different for people raised in different cultures. One of the first attempts to study the potentially different schematic expectations between different cultures was undertaken by Demorest, Morrison, Beken and Jungbluth (2008). They proposed that an enculturation effect, which they defined as “the process by which a person acquires the understandings and beliefs of a particular society from infancy without special training,” should account for differences in musical schemata between people raised in different musical cultures (p. 213).

In order to design their experiment, they relied on the fact that memory for novel musical stimuli should be enhanced if it adheres to the expectations inherent in the listener’s schema. They recruited native-born participants from the United States and Turkey to represent two different musical cultures. To draw on varied musical traditions, the stimuli were real musical examples taken from Western, Turkish, and Guangdong (Chinese) traditions. The procedure involved listening to 30-second clips followed by shorter clips. Target clips came from the original 30-second examples and foil clips came from the same pieces of music, but from different parts of the pieces. The participants were charged with correctly identifying the target and foil clips. The American participants were able to recall more correctly the target clips for the music of Western tradition, while the Turkish students were better at recalling the target clips from the Turkish tradition. Both groups performed poorly with the examples from the Guangdong tradition. These results indicate that there is no single universal musical schema (2008).

In 2009, Demorest, Morrison, Stambaugh, Beken, Richards, and Johnson repeated this experiment, though this time they also compared fMRI readings for participants listening to culturally familiar music with fMRI readings for participants listening to culturally unfamiliar music. They found that participants listening to culturally unfamiliar music experienced greater activation in the brain in areas associated with the processing of complex musical data – primarily the right frontal and parietal lobes.

In 2008, Morrison, Demorest, and Stambaugh repeated Demorest et al.'s 2008 study using 5^{th} grade elementary school children. The children were exclusively from the United States, but were tested using music from the Western tradition as culturally familiar music and music from the Turkish tradition as culturally unfamiliar music. As an additional variable, the experimenters varied the complexity of the musical examples used as stimuli, some of which were reduced to a single instrument playing a relatively more redundant version of the original melody. Regardless of these varying complexities, the results were the same for the children as they had been for the American students in Demorest et al.'s 2008 study. The culturally familiar stimuli were remembered better than the culturally unfamiliar ones during the testing phase. This study is important for two reasons. The first is that it supported the role of a schema in music memory at a young age. The second is that it implied that such a schema operates even on a basic, reduced level of complexity. This explains, in part, the relationship between melodic motives in research and the assumption that schematic expectations lie in harmony (Krumhansl and Shepard, 1979; Deutsch, 1980).

Curtis and Bharucha have also studied the influence of culture on musical expectation (2009). In their research, participants from the United States listened to a series of tones pertaining either to the culturally familiar major mode of the Western tradition, or to the culturally unfamiliar Indian *thaat* Bhairav. The key finding was that false alarm rates for what they called "congruous" tones were higher for the Western tradition tests than the Bhairav tests, and the false alarm rates for "incongruous" tones were lower for the Western tests than for the Bhairav tests (p. 371). A false alarm occurred when the participant believed that the test tone after the sequence had been part of the sequence, when it had not. "Congruous" tones were tones that pertained to the musical culture being presented but were not actually in the series of tones. "Incongruous" tones were tones that did not pertain to the musical culture being presented.

These findings indicate that as a result of their expectations, participants assumed that tones that belonged in the Western series had been presented when they actually had not, because the tone was expected. The number of false alarms for congruous tones in the unfamiliar culture was not as high, because there was no expectation leading the participants to identify the tone as having been in the series. Likewise, if a tone was incongruous with the familiar musical culture it would not be expected, and thus would lead to fewer false alarms than in the case of an unfamiliar culture in which no expectations can inform the listener's judgment.

Ultimately, the research suggests a clear division between different musical cultures and their respective schematic expectations. Wong, Roy, and Margulis have taken this research a step further and studied the possibility of having what they call "bimusicalism" (2009, p.85). For their study, they replicated Demorest et al.'s design (2008) but included a third participant group that included students who had been raised in a bicultural setting. The participants then engaged in a recall task. The American participants showed successful performance in recalling the examples of Western music; the Indian participants showed successful performance in recalling the examples of Indian music; and most importantly, the bicultural participants showed successful performance in recalling the examples of both musical traditions (2009).

Unique to this experiment, participants were also asked to assign a rating of perceived tension in musical examples from either Western or Indian culture. This reflects Meyer's original claim that the way music conforms to or deviates from our expectations affects a listener's emotional response to that music. Indeed, the results indicated that the American participants reported higher levels of tension in response to the Indian music, and the Indian participants reported higher levels of tension in response to the Western music. The implication is that deviation from a listener's expectations manifests itself as tension. This implication was strengthened by the fact that the bicultural participants did not report more tension for one musical tradition over the other (Wong et al., 2009). However, there are some problems with this assessment, in that ratings of tension are inherently subjective, and cannot be fully analyzed across participants. The only way in which these results are valuable is in the way that comparison within subjects was possible.

A similar approach involving a sort of bimusicalism found that it can be acquired, or at least begin to exist, in children as young as 12 months. Hannon and Trehub found that early exposure to foreign rhythm patterns enabled children to distinguish rhythm patterns from both their native culture and the introduced foreign culture. This does not involve harmony, but it does indicate the ease with which musical enculturation can occur, if introduced at a young age (2005).

Genre

A substantial amount of research has been dedicated to how genre is identified, but it has implicated qualities of music other than harmony. Schellenberg, Iverson, and McKinnon asked their participants to name the song and artist for short clips totaling 100msec or 200msec. With popular songs form the *Billboard* "Hot 100" as their stimuli, their participants were able to correctly identify the song or artist at a rate well above chance in clips that were 200msec in length, and still somewhat above chance in clips that were 100 msec. These results implicate timbre as the primary quality in the stimuli that allowed for proper identification, because of the relatively short amount of time required to assess the timbre of a given sound stimulus (1999).

Schellenberg et al. used familiar songs in their recognition test, thus not truly testing the impact of schematic expectations on encoding novel examples of music. However, other researchers have directly addressed the problem of recognizing the genre of an unknown piece of music, based on the implications of Schellenberg et al.'s findings.

Gjerdingen and Perrott asked their participants to identify the genre of short clips (250, 325, 400, 475, or 3000 msec) by selecting one of ten possible categories (blues, classical, country, dance, jazz, Latin, pop, R&B, rap, or rock). They found that the participants agreed with the category chosen by record labels well above chance in clips as short as 250 msec, though they saw statistically significant improvement up through 400 msec. The researchers explicitly address the fact that this does not allow enough time for any significant cues about harmony, melody, rhythm, or any other musical structure to register. In that length of time, no relationship between multiple notes or chords can be expressed. As a cue, only timbre, can explain their data, though they concede that the large variation between correct responses for the different genre categories implies something more complicated than merely using timbral cues to identify a genre. For example, at 250 msec 70% of the responses for identifying classical music were correct, compared with 27% of responses in the 250-475 msec range that were correct for blues music (2008).

Plazack and Huron found similar results, in a study in which a wide variety of musical characteristics were tested. Their design attempted to discover how much time it took a participant to make an accurate judgment about certain elements in music. They included aspects such as meter, the presence of vocals, and tempo in their analysis. By using clips of music of varying length, they found that a correct judgment regarding genre could be made at a frequency well above chance in clips as short as 400 msec (2010).

The study of genre has also focused on harmony. De Clerq and Temperley analyzed the harmonic content of 100 songs from *Rolling Stone Magazine's* "500

Greatest Songs of All Time," selecting the top 20 songs from each represented decade (the 1950s, '60s, '70s, '80s, and '90s). By aggregating their analyses and running a statistical analysis, they found trends that differed from common practice conventions, such as a heavy emphasis on the IV chord (2011).

In the field of Music Information Retrieval, Anglade, Ramirez, and Dixon were able to teach a computer, through exposure to examples, to identify the genre of musical examples, given the three categories of "rock," "classical," and "jazz" (2009). Nor are they the only ones who have achieved this. Pérez-Sancho, Rizo, Kersten, and Ramirez were similarly able to train a computer on the same task (2010). Both groups of researchers based their work on the differences in harmonic content between the different genres.

Objectives

The first objective of this paper is to address the assumption that harmony is a fundamental element of schematic expectations. While it has served to explain the phenomena encountered in research, how does it explain the enculturation effect, when not all musical traditions rely on harmony in the way that Western common practice music does? At the very least, this paper aims to discover whether or not harmony alone is enough to account for the phenomena in the literature concerning Western music, considering the confounding nature of other musical characteristics such as rhythm, timbre, or meter. The first part of this experiment engaged the participants in a recall task similar to that of Demorest et al. (2008). For the stimuli, they were presented with short chord progression piano reductions derived from real pieces of music. Since expectations aid in the encoding of novel music (Curtis and Bharucha, 2009; Demorest et al., 2008, 2009; Deutsch, 1980; Morrison et al., 2008; Wong et al., 2009), these reductions should be easy to recall. Evidence suggests that this should be the case, since Morrison et al. were able to produce these results using reduced versions of real music as well, though they were not explicitly using chord progressions (2008).

The second objective of this paper is, to use the analogy of language, to address whether different genres within Western culture are their own languages, or dialects of the same language. Analysis has shown that in examples such as electronic dance music, the harmonic progressions do not follow the rules of common practice western music. There is no prolongation or emphasis of a tonic chord, and there is no cadential pattern derived from common practice convention. For example, in the song “Nil by Mouth,” the chord progression repeated throughout the song is: i – v – IV – VII^9. Experimentally, this difference has been reflected in differences in recall task performance for Western participants listening to stimuli from classical and electronic dance music. The participants gave more correct responses for the classical stimuli than they did for the electronic dance stimuli (Witkowsky, Bowers, and Stefanucci, 2010).

Demorest et al. also suggest the possible application of their research to the comparison of genres such as jazz to common practice music (2008, p. 220). Thus, the second part of this experiment will include a human analog to Anglade et al. (2009) and Pérez-Sancho et al.’s (2010) work, by asking participants if they can identify the genre of chord progression piano reductions of songs taken from iTunes, in the categories of “rock,” “classical,” and “jazz.” This design is also similar to Gjerdingen and Perrott’s research (2008) in what it will be asking of the participants. If each genre has its own musical schema, then an effect similar to Wong et al.’s bimusicalism – in which performance in categorizing clips should be the same across genres – could occur. Likewise, there would be no difference in recall ability for the various reductions during the first part of the experiment, in which all three genres will be present. However, the hypothesis of this paper is that participants will demonstrate inferior performance in the two popular genre choices, “rock” and “jazz,” based on the assumption that exposure to these styles of music occurs later in life, and thus a schema for these genres cannot be as fully developed. This assumption is related to Hannon and Trehub’s research comparing 12-month-old children and adults, in which children responded more readily to the music from a foreign culture (2005).

Method

Participants

58 (39 female, 19 male) students (age *M* = 18.86 years, range = 18-22 years) participated in the experiment, and received credit as part of an undergraduate introductory psychology course. The participants’ prior musical training was measured using the Ollen Musical Sophistication Index (OMSI) survey (Ollen, 2006). Using their answers on the survey, a score was obtained for each of the participants. This score allowed for analysis along a continuous scale of musical expertise, rather than dividing the participants into musically trained or untrained groups. The mean score on the OMSI was 0.2474 (range = 0.0247-0.8138). These values indicate “respondent’s predicted probability of being more musically sophisticated” (Ollen, 2006, p.120).

Stimuli

The stimuli were chosen by selecting from the top 40 downloads on iTunes on October 26, 2010, for the categories of jazz and rock. For the category of classical music, the top 40 downloads were not offered, but the top downloaded album was “The 50 Most Essential Pieces of Classical Music,” providing a substitute for the use of top downloads. 10 pieces of music were chosen from each genre to use as stimuli. After each stimulus was selected, a musical phrase within the piece was chosen and reduced to a simple chord progression (see Appendix A). Since harmony can only occur in time, a removal of harmonic rhythm would not eliminate it, but rather arbitrarily normalize it across examples. This was considered problematic, and thus harmonic rhythm was maintained as a necessary element of harmony. These chord progressions were rendered in audio files using GarageBand, using the “Grand Piano” voice. The average length of the classical clips was 9.68 seconds (range = 2.55-16.22), the average length of the jazz clips was 10.03 seconds (range = 4.63-15.49), and the average length of the rock clips was 9.69 seconds (range = 4.74-15.3).

Design

The experiment employed a repeated-measures, within-participants design. In the first part of the experiment, participants listened to a set of fifteen clips, 5 from each genre, and were then asked to perform a recall task from the total set of 30 clips, 10 from each genre. This set of 30 clips was divided into two sets, A and B, each containing 5 examples of each genre. Clips were assigned randomly to the A or B set. The first fifteen clips in this part of the experiment came exclusively from either the A set or the B set.

In the second part of the experiment participants listened to a set of fifteen clips, either the A set or the B set, and were asked to identify the genre of each clip. Because there were two parts to the experiment, the possible combinations of the clip sets used were AA, AB, BA, and BB. All four combinations were employed equally in the experiment: AA (N=14), AB (N=15), BA (N=15), and BB (N=14).

Procedure

For the first part of the experiment, the participants listened to each clip in either the A set or the B set, depending on which group they were in. Each clip was prompted individually before being presented, and the participants were instructed to pay attention closely, since they would be asked about the clips later in the experiment. After all 15 clips had been presented, the participants were given a sheet of paper on which to mark their answers for the next section of this part of the experiment. They were instructed both in writing and verbally that they were to listen to a larger set of clips and decide whether each clip had been presented in the earlier part of the study or not; They were to select “Old” for previously heard clips and “New” for previously unheard clips. For each selection they were also instructed to a mark a confidence rating on a scale of 1-6, on which 6 indicated the highest degree of confidence.

After the first part of the experiment was completed, the participants were given the OMSI Survey to complete (Ollen, 2006). They were told to answer the questions as truthfully as possible, as there were no right or wrong answers to the survey.

For the second part of the experiment, participants listened to either the A set or B set of clips, depending on which group they were in. Each clip was prompted before being presented. The participants were given a sheet of paper on which to mark their answers, and instructed, both in writing and verbally, that after each clip was played, they were to circle “C” if they believed the clip was from a classical piece, “J” if from a jazz piece, or “R” if from a rock piece. The were also asked to provide a confidence rating for each judgment, on a scale of 1-6, on which 6 indicated the highest degree of confidence.

Results

Coding and Analysis

Participants’ responses were coded as either correct or incorrect. Correct responses included both correctly identifying a clip as one that had been previously heard, and correctly identifying a clip as one that had not been previously heard. Incorrect responses were those that identified incorrectly either clips that had been previously heard or clips that had not.

Mauchly's test of sphericity was used to assess this assumption of the repeated-measures ANOVA. When a significant violation was identified, the *p* values reflect the Greenhouse-Geisser degrees of freedom correction.

Recall Task

The mean number of correct responses given in this task was 21.81 (SD = 2.652, range = 16-28). For the classical clips the mean was 7.41 (SD = 1.522, range = 5-10); for the jazz clips the mean was 7.29 (SD = 1.124, range = 5-10); for the rock clips the mean was 7.14 (SD = 1.627, range = 2-10).

A one-sample t-test was run for each genre on the number of correct responses given, and for all three genres the participants performed well above chance, $t(57) = 62.631$, $p < .0005$.

Next, a repeated-measures analysis of variance (ANOVA) was run using the number of correct responses for each genre as a within-participants independent variable.

The results indicated no significant difference in the number of correct responses given for each genre, $F(2, 114) = 0.572$, $p = .566$.

A two-way repeated-measures ANOVA was run, adding a between-subjects variable which took into account whether the participant was given the A set of clips or the B set of clips as the target clips. A significant difference was found for the number of correct responses given for each genre, depending on whether the participants were given the A set or the B set as the target set for the recall task, $F(2, 112) = 3.538$, $p = .032$.

A third repeated-measures ANOVA was run, using the participants' scores on the OMSI as a covariate, to see if musical expertise affected performance on the recall task. It did not, $F(2, 112) = 0.587$, $p = .558$.

Genre Identification

The mean number of correct responses given in this task was 8.41 ($SD = 2.256$, range = 2-12). For the classical clips the mean was 3.21 ($SD = .951$, range = 1-5); for the jazz clips the mean was 2.22 ($SD = .974$, range = 0-4); for the rock clips the mean was 2.98 ($SD = 1.357$, range = 0-5).

Participants' answers were coded as correct, for this part of the experiment, if they were able to correctly identify the genre of the clip provided, from one of three categories. A One-Sample *t*-test concluded that the participants' number of correct responses was well above chance $t(57) = 28.405$, $p < .0005$.

A repeated-measures ANOVA was run using the number of correct responses for each genre as a within-participants independent variable. The results indicated a significant difference in the number of correct responses given for each genre $F(1.787,$

101.847) = 15.399, $p < .0005$. The number of correct responses for the jazz clips was significantly below that of both the classical and the rock clips.

A two-way repeated-measures ANOVA was run, adding a between-subjects variable which took into account whether the participant was given the A set of clips or the B set of clips to identify. A significant difference was found for the number of correct responses given for each genre, depending on whether the participants were given the A set or the B set as the target set for the recall task, $F(1.851, 103.676) = 5.848$, $p = .005$.

Another two-way repeated-measures ANOVA was run to see if the participants who had gotten the same set of clips for both parts of the experiment (AA or BB group) performed better at this task than those who were given different clips (AB or BA group), due to a possible priming effect ($N = 28$). There was not a significant priming effect, $F(2, 112) = 0.089$, $p = .915032$.

A final repeated-measures ANOVA was run, using the participants' scores on the OMSI as a covariate, to see if musical expertise affected performance on the genre identification task. It did not, $F(1.783, 99.845) = 0.162$, $p = 0.826$.

Discussion

Research Findings

The design employed in this study maximized the isolation of harmony as the only independent variable in the stimuli presented to the participants. The fact that the participants were able to perform the tasks in a way predicted by theories of schematic memory strongly supports the explanation of schematic memory in terms of harmony. Being able to pinpoint a single musical element as being significantly involved in music memory processes, and by extension musical expectations, is critical for fully understanding the way in which these processes work. While focus on harmony has been implied in past research, this is the first time it has been truly isolated as a variable, minimizing the potential confound from elements such as timbre, melody, or rhythm that play some role in musical memory, though not necessarily schematic memory (Gjerdingen and Perrott, 2008; Plazack and Huron, 2011; Schellenberg et al., 1999).

This research has not conclusively shown whether or not different genres within Western culture merit independent categorization, as far as schematic memory is concerned. While no differences between genres were found in the participants' performance of the recall task, differences emerged between the classical and rock clips and the jazz clips in their performance of the genre identification task. This incomplete differentiation of genres may be due to the fact that participants reached a ceiling of performance, getting all the possible choices correct in the case of both classical and rock music during the genre identification task. This is problematic because it limits the possible validity of the data collected for those categories, since participants may have been capable of correctly identifying more clips, on average, in those two categories. If performance had not hit a ceiling, a difference between the rock and classical clips could have been more effectively analyzed.

This problem of achieving maximum performance in a single genre also occurred at least once for each of the three genres in the recall task. Ideally, participants should not achieve a perfect performance on any measure, because the possibility of a higher range of scores limits the validity of analyzing their scores. These results indicate that the task may in fact have been too easy for the participants. Since the point of the stimuli was to have chord reductions, increasing the complexity of the stimuli is a difficult option. However, increasing the total number of stimuli presented should solve this problem, if future research incorporates the design employed in this study.

What was also of interest in the results of this study was the apparent variability of participants' performance between clips. When looking directly at the data, it becomes apparent that there were observable differences in the performance of the participants on different clips. For example, for clip C7, participants were divided evenly among the three choices in the genre identification task (see Appendix B). Another striking example is that no one gave an incorrect response for clip C1 in set A in the recall task, yet 11 participants who got set B missed that same clip. The opposite phenomenon occurred with clips such as C4 (see Appendix C). In other cases, such as with clip J5, participants were evenly split between 2 out of the 3 choices in the genre identification task. Likewise, the fact that there were significant differences between performance with the A set or the B set in both tasks seems to suggest faulty stimuli, but in fact, when looking more closely, the differences between performance in each set seem to balance out (see Figure 2).

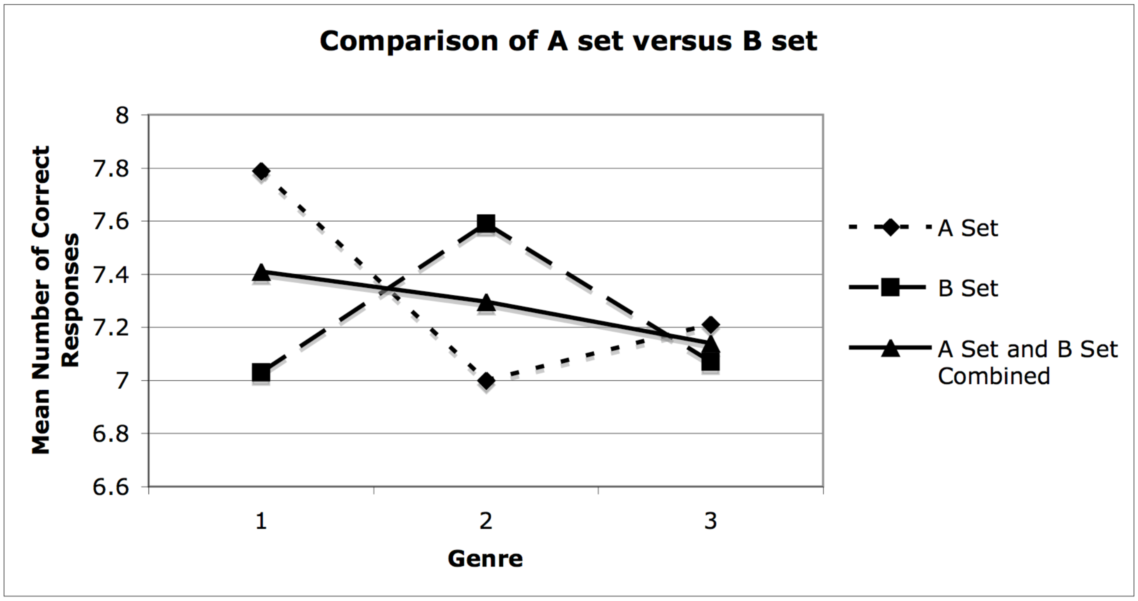


Figure 2. The mean number of correct responses for the recall task, comparing the difference between participants who had the A set or the B set as the target set. Genre “1” is classical music, genre “2” is jazz music, and genre “3” is rock music

The stimuli themselves were each constructed in an identical manner, and thus each deserves an equal degree of validity, as can be commanded by the above chance performance by the participants in both tasks. These results indicate, instead, the inherent variability in music. Each song is composed differently, and each harmony has its own nuance. In fact, as Meyer mentions, it is these differences that are responsible for affective responses to music (1956). This variability of performance between clips indicates that as long as real musical examples are used, there will be some differences in how well they allow the participant to perform the required tasks. In fact, even the computers that were trained to identify the genre of different pieces of music never performed anywhere near 100% accuracy (Anglade et al., 2009; Pérez-Sancho et al., 2010). For this reason a large set of stimuli should always be used. Looking at the data, no clip was always missed in the recall or identification tasks (see Appendix B and C), and average performance showed minimal variation across all the participants. This shows that variation in musical stimuli can be accounted for, just as easily as variation between participants is accounted for, by having a large number of pieces of music included in the stimuli used in an experiment.

The data also indicate that musical expertise did not influence participants' performance on either task of the experiment. This is consistent with the existing literature, and should be, since musical memory, as far as the schema is concerned, is a subconscious framework learned through exposure. It is not something mediated by expert knowledge.

Future Research

Since the results of this experiment indicate that harmony is a key contributor to schematic memory for music, it allows future research to rely on harmony as a method for investigating musical memory and expectations. It also allows for methodological streamlining of research on music memory, since chord reductions appear to have a great deal of validity as stimuli. Perhaps the approach here should be to discover other methods of testing musical memory using harmony, such as measuring reaction times for making recall judgments.

An essential direction for future research will be to address how to treat genre. It is still unclear whether or not different genres should be considered as having their own respective schematic frameworks. This experiment, in addition to previous research (Anglade et al., 2009; Pérez-Sancho et al., 2010; Witkowsky et al., 2010), suggests that this is not outside the realm of possibility. What remains to be seen is a method for testing this hypothesis with certainty. The problem with the recall method employed in this experiment is that it requires differing levels of performance between genres in order to demonstrate different schematic frameworks. This requires not only that the genres have different schematic frameworks, but also that these frameworks vary in how engrained they are in the participant, as a results of earlier or later exposure in life. This is one assumption too many upon which to build a hypothesis. Indeed, even the genre identification task would not effectively confirm the hypothesis that different genres have their own schematic frameworks, because it also allows for equal performance across genres. On the other hand, if performance were *not* equal across genres, but above chance, then the hypothesis would certainly be supported.

The crucial element in research pursuing the potentially different schematic frameworks for different genres will be to eliminate a cue such as timbre, which has been shown to serve as an identifier of genre. There is a theoretical difference between something that can identify a genre and something that defines a genre. It is this distinction that should separate harmony from timbre in future research. Evidence of this separation can be found in Gjerdingen and Perrott's results, in which there was a significant variability between genres and the percentage of correct judgments made by the participants (2008). Perhaps certain timbral cues elicit information about genre, but that alone is not enough to identify or define genre, something that studying harmonic schemata suggests is possible.

Ultimately, the findings of this research have served to confirm and reinforce research based in the harmonic underpinnings of musical memory, and to inform an approach in researching the categorization of genre as it relates to schematic memory.

While the research was not conclusive on the latter question, it has opened the possibilities for future research, which should be better equipped to test the question. It has also given promising indications that a differentiation of genre-based schematic frameworks may exist. The only question raised without an answer is how researchers can approach schematic memory for musical traditions that are not based in harmony. While this research does not answer the question, it certainly suggests the need to find a way to do so.

References

Anglade, A., Ramirez, R., & Dixon, S. (October 2009). Genre Classification Using Harmony Rules Induced From Automatic Chord Transcriptions. Unpublished paper presented at the 10th International Society for Music Information Retrieval Conference, Kobe, Japan.

Bharucha, J. (1994). *Tonality and Expectation*. In Aiello, Rita and Sloboda, John A. (Eds.), Musical Perceptions (pp. 213-239). New York, NY US: Oxford University Press.

Curtis, M., & Bharucha, J. (2009). Memory and musical expectation for tones in cultural context. *Music Perception, 26(4)*, 365-375.

Davidson, L. (1985). Tonal structures of children's early songs*.* *Music Perception, 2(3)*, 361-373.

Davidson, L. (1994). *Songsinging by young and old: A developmental approach to music.* In Aiello, Rita and Sloboda, John A. (Eds.), Musical Perceptions (pp. 99-130). New York, NY US: Oxford University Press.

De Clercq, T., Temperley, D. (2011). A corpus analysis of rock harmony. *Popular Music 30(1),* 47-70.

Demorest, S. M., Morrison, S. J., Beken, M. N., & Jungbluth, D. (2008). Lost in Translation: An Enculturation Effect in Music Memory Performance. *Music Perception, 25(3),* 213-223.

Demorest, S. M., Morrison, S. J., Stambaugh, L. A., Beken, M., Richards, T. L., & Johnson, C. (2009). An fMRI investigation of the cultural specificity of music memory. *Social Cognitive and Affective Neuroscience 5(2-3)*, 282-291.

Deutsch, D. (1980). The processing of structured and unstructured tonal sequences. *Perception and Psychphysics, 28,* 381-389.

Gjerdingenm R. O., Perrott, D. (2008). Scanning the dial: The Rapid Recognition of Music Genres. *Journal of New Music Research, 37(2),* 93-100.

Hannon, E. E., & Trehub, S. E,. (2005). Tuning in to musical rhythms: Infants learn more readily than adults. *Proceedings of the National Academy of Sciences of the Unite States of America, 102(35),* 12639-12643.

Jackendoff, R., & Lerdahl, F. (1981). Generative music theory and its relation to psychology. *Journal of Music Theory 25(1),* 45-90.

Justus, T. & Bharucha, J. (2001). Modularity in musical processing: The automaticity of harmonic priming. *Journal of Experimental Psychology: Human Perception and Performance, 27(4)*, 1000-1011.

Krumhansl, C. L. & Keil, F. (1982). Acquisition of the hierarchy of tonal functions in music. *Memory & Cognition, 10(3)*, 243-251.

Krumhansl, C. L. & Shepard, R. N. (1979). Quantification of the hierarchy of tonal functions withing a diatonic context. *Journal of Experimental Psychology: Human Perception and Performance, 5(4),* 579-594.

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

Morrison, S. J., Demorest, S. M., & Stambaugh, L. A. (2008). Enculturation effects in music cognition: The role of age and music complexity. *Journal of Research in Music Education, 56(2)*, 118-129.

Ollen, J. E. (2006). *A criterion-related validity test of selected indicators of musical sophistication using expert ratings.* Unpublished doctoral dissertation, The Ohio State University, Columbus, Ohio.

Pérez-Sancho, C., Rizo, D., Kersten, S., & Ramirez, R. (2010). Genre classification of music by tonal harmony. *Intelligent Data Analysis, 14(5),* 533-545.

Plazack, J. & Huron, D. (2011). The first three seconds: Listener knowledge gained from brief musical excerpts. *Musicae Scientae, 15(1),* 29-44.

Schellenberg, E. G., Iverson, P., McKinnon, M. C. (1999). Name that tune: Identifying popular recordings from brief excerpts. *Psychonomic Bulletin & Review, 6(4),* 641-646.

Witkowsky, D., Bowers, G., & Stefanucci, J (2010). *The effect of genre on musical memory.* Unpublished manuscript.

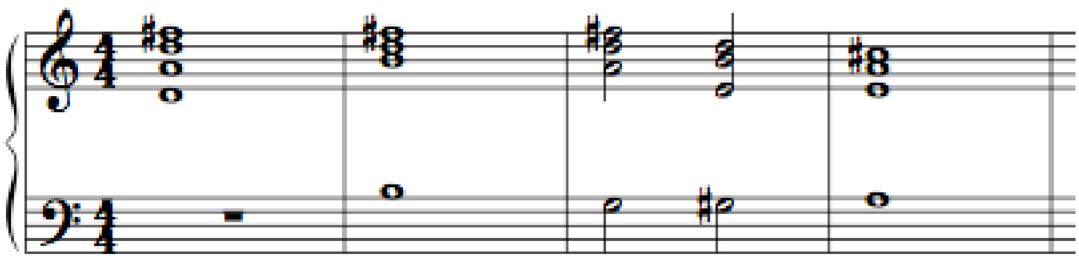
Wong, P. C. M., Roy, A. K., & Margulis, E. H. (2009). Bimusicalism: The implicit dual enculturation of cognitive and affective systems. *Music Perception 27(2)*, 81-88.

APPENDIX A

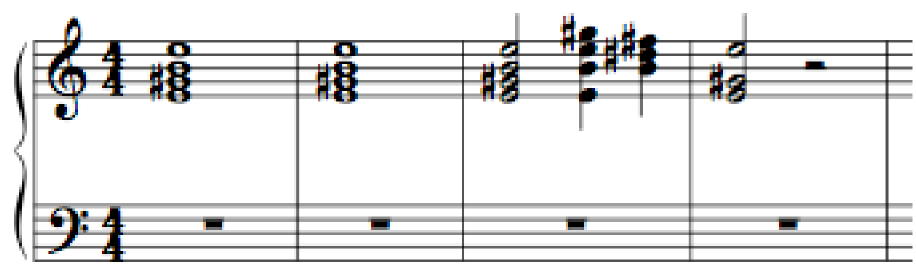
Clips Used as Stimuli and Their Chord Reductions

C1: Mozart's *The Magic Flute*, K. 620: Overture

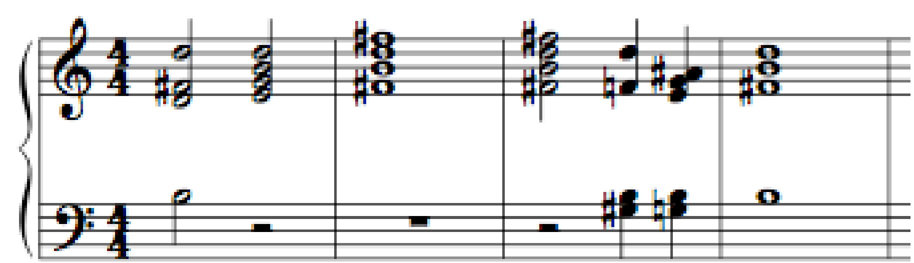
C2: Bach's *Suite for Orchestra No. 3 in D Major*, BWV 1068: II. Air



C3: Vivaldi's *The Four Seasons - Concerto for Violin in E Major*, RV 269, Op. 8:1, "Spring": I. Allegro



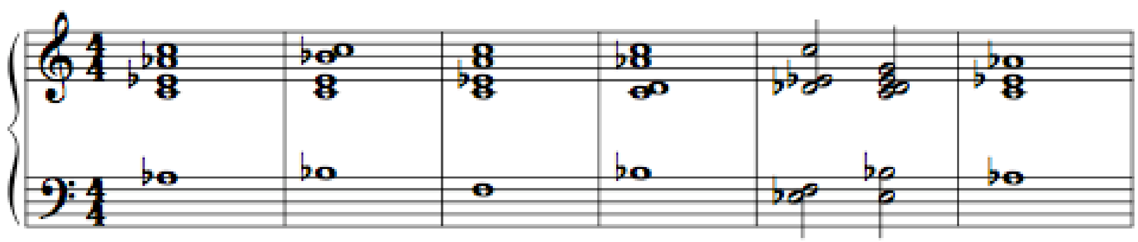
C4: Tchaikovsky's *Swan Lake Suite*, Op. 20: Scéne



C5: Grieg's *Peer Gynt Suite No. 1*, Op. 46: IV. In the Hall of the Mountain King

C6: Beethoven's *Symphony No. 5 in C Minor*, Op. 67 "Fate": I. Allegro con Brio

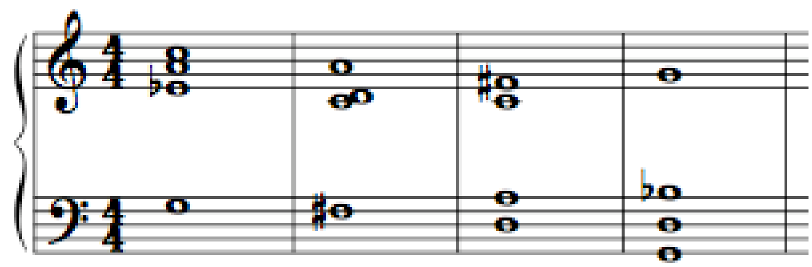
C7: Liszt's *Liebestraum No. 3 in A-Flat Major*, Op.62: "O Lieb So Lang' Du Lieben Kannst"



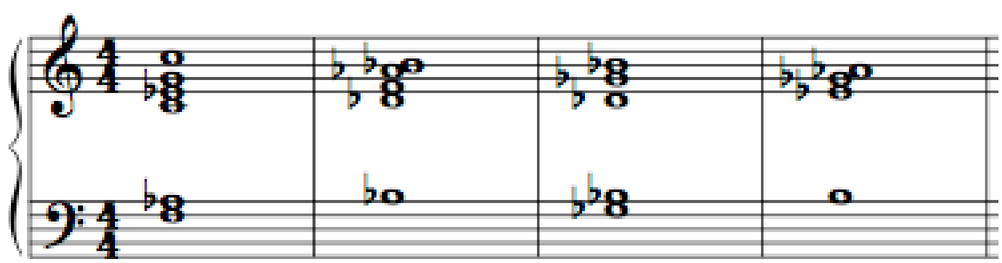
C8: Bach's *Brandenburg Concerto No.3 In G Major*, BWV 1048, 1. Allegro

C9: Mozart's *Concerto for Piano and Orchestra No. 21 in C Major*, K. 467: II. Andante

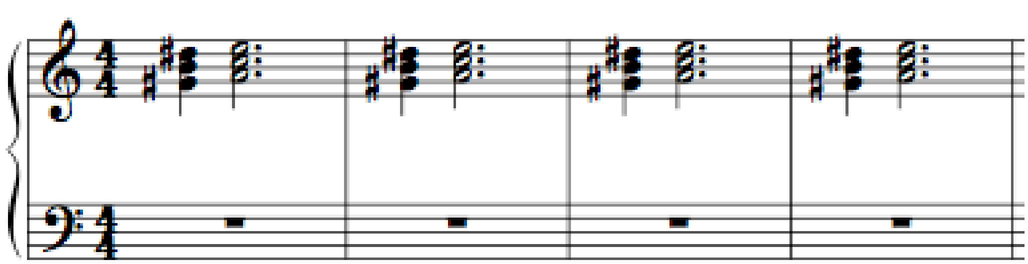
C10: Mozart's *Symphony No. 40 in G Minor*, KV 550: I. Allegro Molto



J1: "Rise" by Herb Alpert



J2: "Sing, Sing, Sing" by Benny Goodman



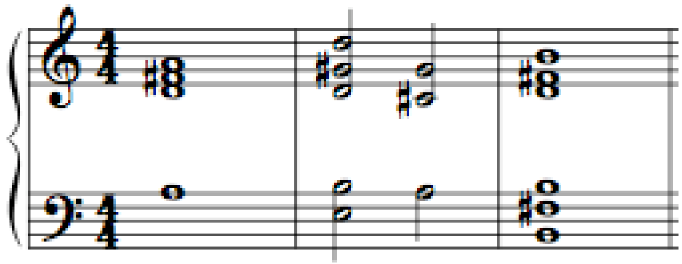
J3: "What a Wonderful World" by Louis Armstrong

J4: “ A Kiss to Build a Dream On” by Louis Armstrong

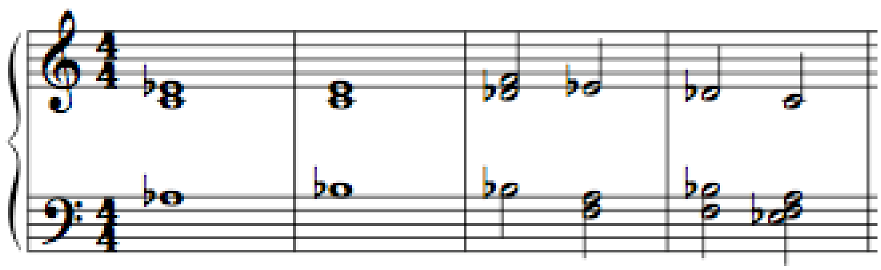
J5: “Dream a Little Dream of Me” by Ella Fitzgerald and Louis Armstrong

J6: “Georgia on My Mind” by Ray Charles and the Count Basie Orchestra

J7: “Summertime” by Ella Fitzgerald and Louis Armstrong

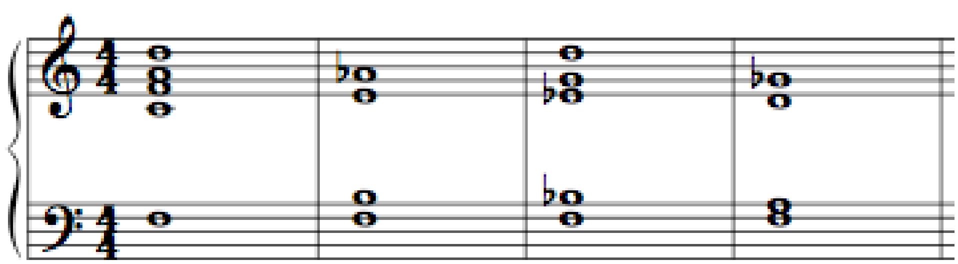


J8: "I'll Be Seeing You" by Billie Holiday

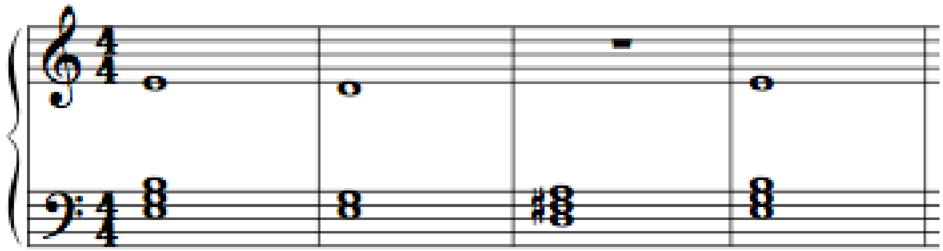


J9: "In a Sentimental Mood" by John Coltrane and Duke Ellington

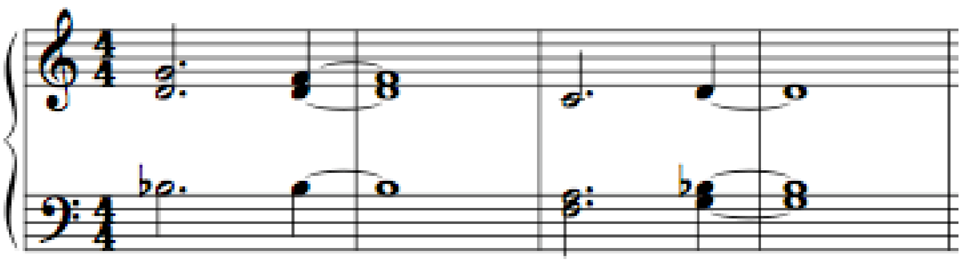
J10: "The Look of Love" by Diana Krall



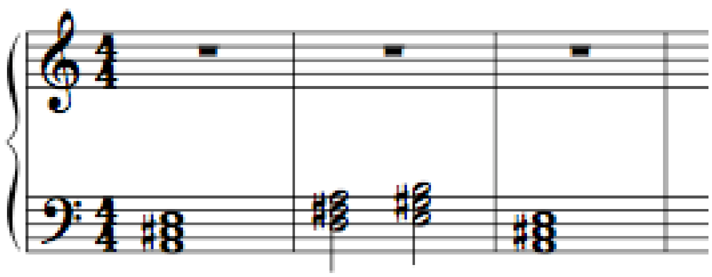
R1: "Bring Me Back to Life" by Evanescence



R2: “Breakeven (Falling to Pieces) by The Script



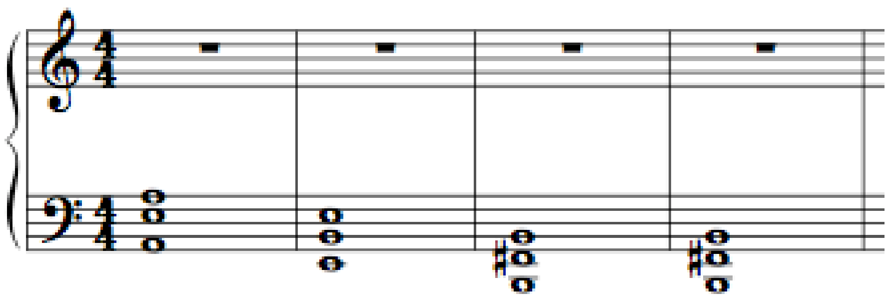
R3: “Season of the Witch” by Donovan



R4: “Half of My Heart” by John Mayer

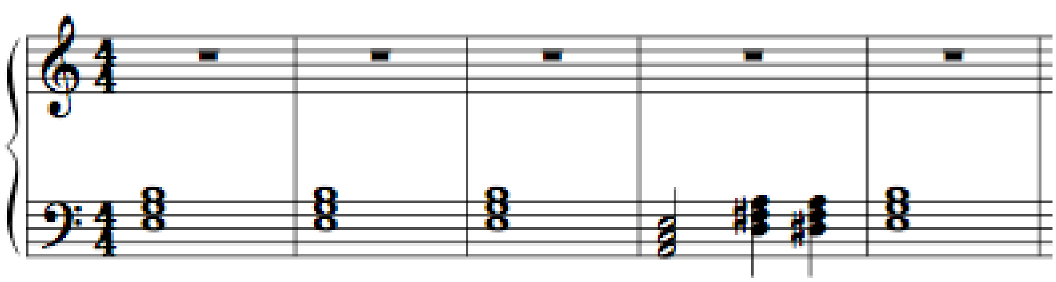
R5: “Brand New Key” by Melanie

R6: "Bad Company" by Five Finger Death Punch

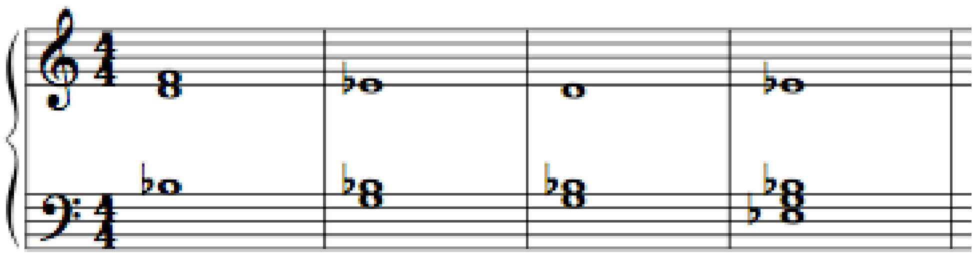


R7: "Don't Stop Believing" by Journey

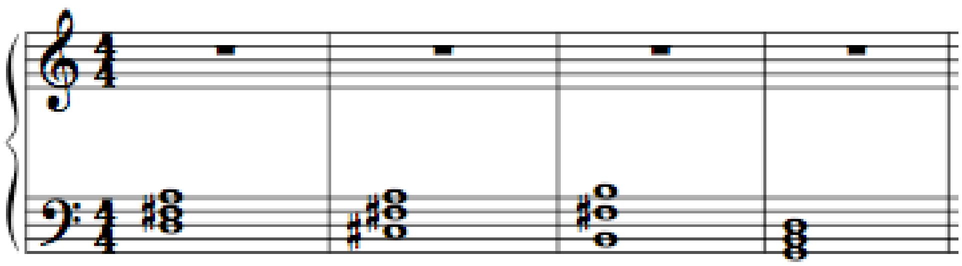
R8: "Porn Star Dancing" by My Darkest Days and Ludacris



R9: "September" by Daughtry



R10: “Secrets” by OneRepublic



APPENDIX B

Responses Given by Participants in the Genre Identification Task, for Each Clip

CLIP	GENRE	Responded "C"	Responded "J"	Responded "R"
C1	Classical	14	13	2
C2	Classical	17	2	10
C3	Classical	24	2	3
C4	Classical	14	9	6
C5	Classical	21	4	4
C6	Classical	23	3	3
C7	Classical	10	10	9
C8	Classical	22	5	2
C9	Classical	24	3	2
C10	Classical	17	8	4
J1	Jazz	5	16	8
J2	Jazz	5	8	16
J3	Jazz	13	5	11
J4	Jazz	2	16	11
J5	Jazz	14	13	2
J6	Jazz	6	15	8
J7	Jazz	12	9	8
J8	Jazz	17	9	3
J9	Jazz	9	15	5
J10	Jazz	4	23	2
R1	Rock	7	5	17
R2	Rock	2	2	25
R3	Rock	7	3	19
R4	Rock	11	0	18
R5	Rock	13	9	7
R6	Rock	5	2	22
R7	Rock	11	1	17
R8	Rock	9	4	16
R9	Rock	11	4	14
R10	Rock	8	3	18

APPENDIX C

Number of Times a Clip Was Incorrectly Judged in the Recall Task, for Each Clip

CLIP	Group A Inccorect	Group B Incorrect	Total Incorrect
C1	0	11	11
C2	3	9	12
C3	14	0	14
C4	13	0	13
C5	3	19	22
C6	2	22	24
C7	9	5	14
C8	4	10	14
C9	12	2	14
C10	4	9	13
J1	10	4	14
J2	0	6	6
J3	19	6	25
J4	11	8	19
J5	11	8	19
J6	6	4	10
J7	14	6	20
J8	5	9	14
J9	5	2	7
J10	5	17	22
R1	3	18	21
R2	9	6	15
R3	18	4	22
R4	9	13	22
R5	2	11	13
R6	3	4	7
R7	4	8	12
R8	3	1	4
R9	15	14	29
R10	13	7	20