DID YOU HEAR THE VOCALIST? DIFFERENCES IN PROCESSING BETWEEN SHORT SEGMENTS OF FAMILIAR AND UNFAMILIAR MUSIC

SHANNON L. LAYMAN
The University of Texas at Arlington

W. JAY DOWLING
The University of Texas at Dallas

PREVIOUS RESEARCH INDICATES THAT PEOPLE GAIN musical information from short (250 ms) segments of music. This study extended previous findings by presenting shorter (100 ms, 150 ms, and 200 ms) segments of Western popular music in rapid temporal arrays; similar to scanning through music listening options. The question remains, is there a critical feature, such as the song's vocalist, that listeners used when processing the complex timbral arrangements of Western popular music? Participants were presented with familiar and unfamiliar music segments, four segments in succession. Each trial contained a female or a male vocalist, or was purely instrumental. Participants were asked whether they heard a vocalist (Experiment 1) or a female vocalist (Experiment 2) in one of the four music segments. Vocalist detection in Experiment 1 was well above chance for the shortest stimuli (100 ms), and performance was better in the familiar trials than the unfamiliar. When instructed in Experiment 2 to detect a female vocalist, however, participants performed better with the unfamiliar trials than the familiar trials. Together, these findings suggest that the vocalist and vocalist gender may be stored as separate features and their utility differs based on one's familiarity with the musical stimulus.

Received: November 9, 2017, accepted February 2, 2018.

Key words: popular music, music memory, familiarity, featural processing, cognition

PROCESSING SHORT SEGMENTS OF MUSIC IS, for most people, a daily occurrence; for example, when they judiciously scan through music options such as listening to Pandora, Spotify, iTunes, or the radio. Even passively watching television and movies provides chances for daily music listening experiences and has, in part, continued to make us expert music listeners. While the literature of music perception with extremely short segments has a rather recent beginning, findings provide converging evidence that music, like other auditory stimuli, is something people are quite adept in processing to address the source and the importance of the sounds, even when those sounds are presented in very brief samples (Gjerdingen & Perrot, 2008; Grosjean, 1980; Mace, Wagoner, Teachout, & Hodges, 2011; Orgs, Lange, Dombrowski, & Heil, 2006; Schweinberger, Herholz, & Sommer, 1997; Tekman & Bharucha, 1998). Much of the previous literature involving music perception, however, did not scrutinize the musical segments to address the importance and utilization of particular features when processing musical information. The current study focused on the utility of the vocalist as a feature in processing short segments of music.

In one of the initial behavioral studies involving extremely short segments of music, Schellenberg, Iverson, and McKinnon (1999) showed that individuals were able to identify recordings as brief as 100 ms. Their study suffered from methodological limitations, however, in that participants heard only five songs total and the researchers provided an answer sheet with a list of the song titles and artists. With such a short list of musical stimuli, participants could have used more of a deductive reasoning strategy for their answers as opposed to pure recognition. The authors concluded that the unique instrumental and vocal make-up specific to each song provided more useful information for song identification than the absolute pitches for each song.

Subsequent studies suggest that it takes at least a 250 ms segment of musical information for the listener to make basic emotional judgments, classify genre, and discern familiar music from unfamiliar music (Bigand, Gérard, and Molin, 2009; Filipic, Tillmann, & Bigand, 2010; Gjerdingen & Perrott, 2008, Plazak & Huron, 2011). One such result showed that when scrambled segments of familiar and unfamiliar classical music were presented via a gating paradigm that either increased in duration (the "bottom-up group") or decreased in duration (the "top-down group"), participants were able to differentiate between the unfamiliar and familiar music at 250 ms and, surprisingly, the top-

Music Perception, VOLUME 35, ISSUE 5, PP. 607–621, ISSN 0730-7829, ELECTRONIC ISSN 1533-8312. DOI: https://doi.org/10.1525/MP.2018.35.5.607

down group, who heard the longest stimulus length first, did not differ in performance compared to the bottom-up group (Bigand et al., 2009). As in Schellenberg et al. (1999), the overall conclusions were that participants used local timbre-specific features or salient features of the melodic segments that usually contained a distinctive instrumental voice or timbre, to discern between familiar and unfamiliar music segments (Bigand et al., 2009; Filipic et al., 2010; Gjerdingen and Perrott, 2008).

Probably the most extensive study that addressed how people process short segments of popular music was by Krumhansl (2010). Krumhansl used 300 and 400 ms segments of 28 popular songs from the 1960's through the 2000's. Participants identified title and artist of the song, how confident they were about their choice, decade of release, emotion conveyed by the song, and style. Results showed that over 95% of participants correctly identified both the title and artist of songs with which they were familiar, and provided the highest confidence ratings, indicating that this information is strongly linked to memory for the music. When identifying title and artist, participant performance dropped from the 400 ms segments compared to the 300 ms segments. Even if participants did not identify the song segment, they were still able to provide accurate information concerning song emotion and song style, with consistent performance between the 400 ms and 300 ms music segments. Krumhansl concluded that people were able to make intuitive judgments based on short or degraded amounts of information.

The previous studies show that people are able to glean musical information concerning genre, style, perceived emotional content, and decade of release, after hearing segments as short as 250 to 300 ms, even if they cannot correctly identify it. In most cases identification of the song requires 400 or 500 ms. The literature provided variable estimates of exactly how much musical information is needed to support judgments of familiarity, though the range is somewhere between 250 and 500 ms (Bigand et al., 2009; Filipic et al., 2010; Krumhansl, 2010). Due to this variability we investigated results in the electrophysiological realm in order to further dissect the cognitive processes occurring during music processing.

In most electrophysiological research involving processing short segments of music, researchers used Western classical music as the stimuli, and the component of interest was the N400, a negative component initially occurring around 200 ms and peaking around 400 ms after the onset of a stimulus. In some instances, the latency peaks around 500 ms, and the potential is termed the N400/N5. In the non-music literature, the general conclusion is that this component is involved in higher-order cognitive processes including attention, memory, sequential processing, speed of processing, and is an overall index of semantic processing (Barrett, Rugg, & Perrett, 1988; Kutas & Federmeier, 2000; Kutas & Hillyard, 1980). The conclusion from the most recent research involving music and the N400 is that this potential is representative of abstract ideas and general notions—termed conceptual information or concepts—that music naturally conveys as music has the power to trigger concepts stored in long-term memory (Daltrozzo & Schön, 2008; Daltrozzo, Tillmann, Platel, & Schön, 2009; Koelsch et al., 2004; Miranda & Ullman, 2007). Additionally, results show that music, like language, conveys conceptual and sematic information that influences subsequent auditory stimuli (Daltrozzo & Schön, 2008; Koelsch et al., 2004; Steinbeis & Koelsch, 2008).

Differing from the goals of previous literature, Daltrozzo et al. (2009) explored two characterizations of music: 1) that the feeling of familiarity with melodies increases over time, and 2) that with this increased familiarity with the music, conceptual information such as emotional context also more readily comes to mind. The authors specifically define concepts as emotional or associative aspects that were conveyed either by the melody or by the memory of the melody. Eighty melodies differing in degree of familiarity were presented via a gating paradigm. The gating paradigm started with the first three notes of a melody, with each successive gate adding one note. After each gate, participants judged whether that gate was familiar or not familiar. This decision point is termed by the authors to be the Familiarity Emergence Point (FEP), and it varied among stimuli and across participants. Results showed that it took an average of five tones to reach the FEP, the highly familiar melodies were judged familiar more quickly than the moderately familiar melodies. The ERPs to the FEP tones had a significantly larger negativity for the highly familiar melodies compared with the moderately familiar melodies. These results confirm the authors' hypothesis that the melodies with a higher familiarity conveyed more conceptual information because by virtue of being familiar, those melodies have more concepts to convey. With a greater N400 for highly familiar melodies compared to moderately familiar melodies, the authors concluded that music did indeed have the power to trigger concepts stored in long-term memory. This conclusion confirms Miranda and Ullman's (2007) notion that the N400 elicited by their stimuli probably corresponded with information stored in long-term memory. In conclusion, there is temporal consistency between the results of both behavioral and electrophysiological studies regarding the type of melodic information processed and the timeline of that processing. While there are no current electrophysiological studies using Western popular music, it is probable that experiments would show that the same N400 would be elicited.

The purpose of the present study was to address whether vocalist information was an important feature used in processing brief musical segments of Western popular music. In other words, would either the vocalist or vocalist gender be a useful feature for listeners when discerning between familiar music, which might possibly be music with which they intend to listen, from unfamiliar music, a task in which many music listeners experience. While previous studies indicated that participants used timbre-specific salient features of music to discern between familiar and unfamiliar music segments (Bigand et al., 2009; Filipic et al., 2010; Gjerdingen & Perrott, 2008), these studies did not actually break down the features of the music segments to address which specific feature participants were perhaps using during processing.

Based on findings from Krumhansl (2010), the vocalist stood out as an important feature when making a judgment of confidence in identification. As mentioned previously, Krumhansl found that in both 300 and 400 ms music clips, participants' confidence in song clip identification was bimodal with participants rating strongly that they either "don't recognize [the clip] at all" or that they "can name both artist and title" (Krumhansl, 2010, p. 341). Of the participants that were confident in naming artist and title, 95.2% correctly named both with the 400 ms clips, with performance dropping to 90.3% in the 300 ms clips. For participants that only identified either the artist or the title, they identified the artist most often. These results suggest that when a song becomes familiar, information regarding both title and artist, perhaps especially the artist, become a stable memory store and thus may be a useful feature when processing music. An additional reason for choosing the vocalist as the feature of interest was due to the usual musical make-up of Western popular music. The vocalist/instrumentals dichotomy is a widely used arrangement in Western popular music (Starr & Waterman, 2003). While there are many other usual instrumental voices in this type of music, such as the guitar, their presence was not guaranteed.

The present study theorizes that listeners using the vocalist features for processing music would be like an auditory analog of the recognition by components (RBC) model for visual object identification. The RBC model uses geometric shapes called geons as the basic building blocks of objects we identify in our environment (Hummel, 2013; Hummel & Biederman, 1992). The RBC model suggests that geons are combined to make-up the complex objects present in our environment and that a hierarchy of feature detectors in this model serve the purpose to not only identify object features at the lowest-level, but also the complex geon arrangements at the highest level. Another advantage of the RBC model is that the geons are viewpoint-independent in that you can recognize an object regardless of your perspective. Viewpoint-independence with the RBC model allows for quicker processing than other theories based on templates, which suggest that people have a template for every possible orientation of a stimulus in long-term memory. Comparing a current stimulus with the appropriate template would take longer processing time, which is not evident in the quick processing of short music clips shown in previous literature.

How the RBC model coincides with our proposal of the utility of the vocalist as a feature in processing music is that one artist's voice would be an "auditory geon" in that you can use this memory of the voice to not only assist you with identifying songs with which you are already familiar, but also hearing a new song and being able to recognize that song with a familiar artist as the vocalist. Vocalist features from familiar music would be more easily extracted than from unfamiliar music. Other auditory geons could potentially be that artist's usual instrumental accompaniment that makes up their usual songs and/or their musical genre. Using instrumentation or even specific instruments as auditory geons, however, could be potentially misleading when an artist transitions to a new genre, such as Taylor Swift's transition from country to pop music with her recent album, "1989" (Eells, 2014). Swift's songs on this recent album have an instrumental make-up that one would expect in pop music, which produces quite a different sound from the instrumental make-up that one would see in country music (Starr & Waterman, 2003). Thus, having auditory geons for instruments of a specific genre would be helpful in song identification up to a point, but we would argue that vocalist features are most useful for processing music and ultimately familiarity judgments.

In addition to focusing on the vocalist as a feature used in processing music, the present study used a more extensive list of song segments that were shorter in duration than that used in previous literature involving familiar songs (save for the 100 ms song stimuli from Schellenberg et al. 1999, which used a stimulus list of only five songs). Recall that previous studies indicated that at least 250 ms is sufficient to support judgments of familiarity (Bigand et al., 2009; Filipic et al., 2010; Krumhansl, 2010). The questions remain, are participants implicitly influenced by familiarity before they are explicitly able to make familiarity judgments? Is the vocalist a feature that assists in processing music that is so short in duration to eliminate explicit judgments based on content? Are participants utilizing the vocalist as a critical feature in the initial processing of such short segments?

In the present study, each trial contained short music segments in groups of four using an auditory version of the rapid serial visual presentation (RSVP) paradigm (Potter, 1976). Presenting the musical stimuli in quick succession is similar to how people may hear music in real life, as in scanning through musical options to find something to which they would like to listen. Each trial contained either familiar or unfamiliar music segments, with 80% of the trials containing the one target vocalist segment and three instrumental segments, and 20% of the trials containing four instrumentalist segments only. For each trial in Experiment 1, participants judged whether they heard a vocalist, the temporal position of the vocalist, and the vocalist gender. For each trial in Experiment 2, participants judged whether they heard a female vocalist and the temporal position of the female vocalist. Participants either heard 100 ms, 150 ms, or 200 ms versions of the trials. We varied the amount of silence between segments to either a constant 100 ms or to have the entire trial add up to 1 s. The purpose of this variance was to control for the effect of temporal context during encoding.

As previous literature suggests that some features of familiar music are already bound in one's memory, the hypothesis follows that in both experiments, participants will perform better at detecting the vocalist (Experiment 1) or the female vocalist (Experiment 2) in familiar music than in the unfamiliar music, indicating that these features may be used when processing and encoding musical information. As this study does not require music identification and the segments of music used were shorter in duration than necessary for identification, results from both experiments would suggest that people are influenced by music with which they are familiar with shorter segments of music information than utilized in the previous literature.

Method

PARTICIPANTS

One hundred graduate and undergraduate students from The University of Texas at Dallas (43 males, 57 females, $M_{\text{age}} = 21.14$ years) were recruited as participants in Experiment 1. Sixty additional graduate and undergraduate students from The University of Texas at Dallas (21 males, 39 females, $M_{\text{age}} = 25.10$ years) that had not participated in Experiment 1 were recruited as participants in Experiment 2. All participants were screened for normal hearing and indicated via questionnaire that they listened to popular music on a daily basis. Participants had on average 3.92 years of music experience ($SD = 3.64$), acquired mostly during music programs (orchestra, band, or choir) offered through middle school and high school. Informed consent was obtained before the beginning of the experiment.

STIMULI

To obtain the familiar songs, we distributed a survey of 200 familiar Western popular songs to a representative group of undergraduates ($n = 46$) at The University of Texas at Dallas, who did not participate in either of the present experiments. The songs chosen for this survey came from Billboard Music Top Charts such as “Hot 100” and “Greatest of All Time.”

To better ensure that the unfamiliar songs in our present study were truly unfamiliar as opposed to songs that were on the Billboard top charts, but considered less or unfamiliar to our sample, the unfamiliar songs used in our unfamiliar stimuli came from Rentfrow, Levitin, and Goldberg (2011). The songs that the researchers classified as unfamiliar were recorded at a professional level. These unfamiliar songs, however, did not make it to any popular charts, and were thus validated as unfamiliar to participants. A 2 (familiarity: familiar, unfamiliar) x 2 (vocalist gender: male, female) between-subjects analysis of variance (ANOVA) yielded no significant effects on average song frequency. This result excluded a possible effect of vocalist gender as a systematic confounding factor between the familiar and unfamiliar song segments (i.e., familiar songs with vocalists having a higher tessitura and sounding more female than unfamiliar songs).

The quick auditory temporal array of short segments of music that were presented to participants consisted of four short segments of music in rapid succession. The length of the music segments varied between groups: one group of participants heard only 100 ms music segments, one group heard only 150 ms music segments, and one group heard only 200 ms music segments. There were two different conditions that altered the amount of silence between the successive music segments within a trial. The first was the constant interstimulus interval (ISI) condition, which had 100 ms of silence in between each pair of stimuli. The second was

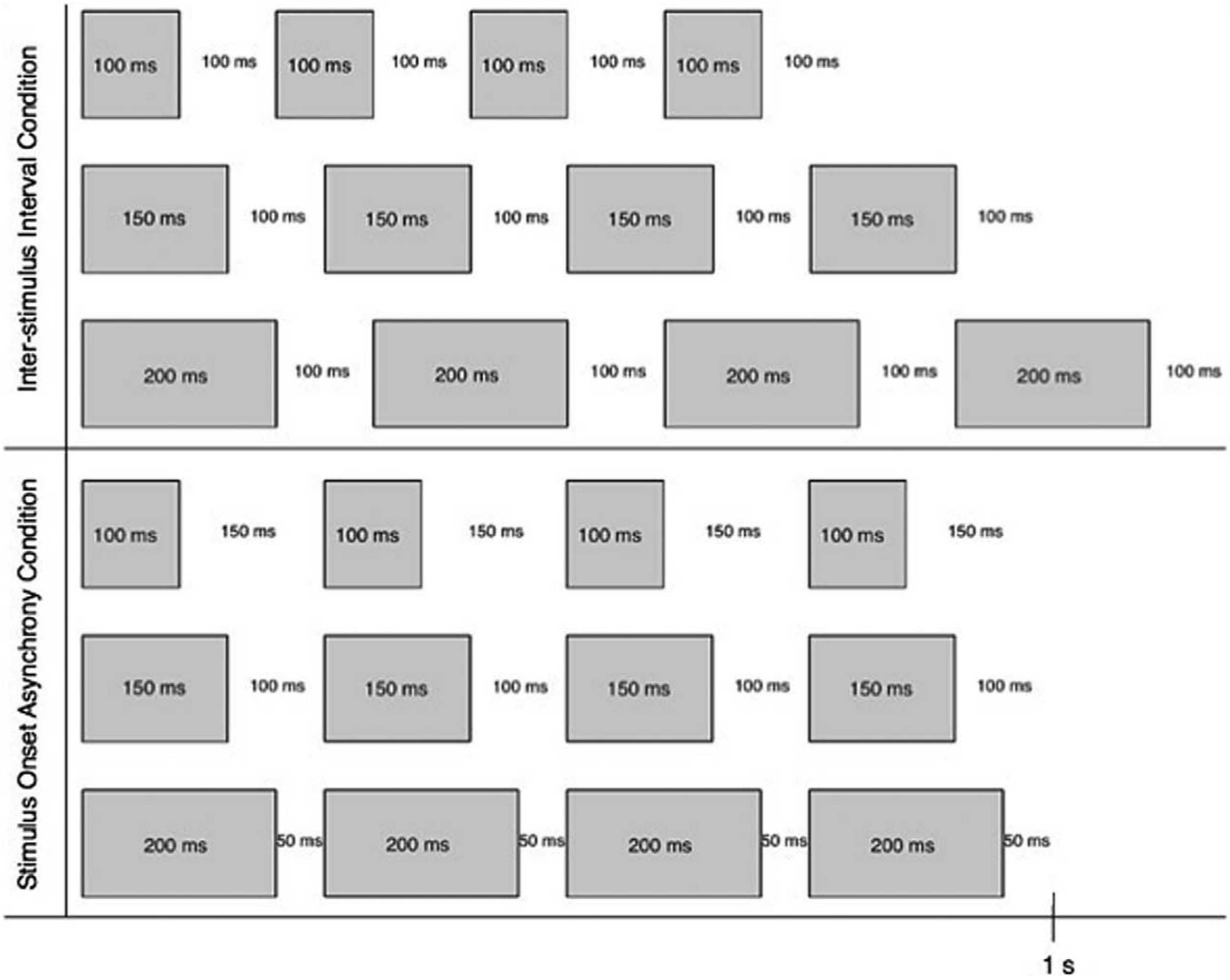

FIGURE 1. Examples of the different silence gaps for the three lengths of stimuli (100 ms, 150 ms, and 200 ms) in the ISI condition and the SOA condition.

the constant stimulus onset asynchrony (SOA) condition, in which each trial had a total duration of 1 s. Thus in the SOA condition the amount of silence between stimuli varied depending on stimulus length. If the stimuli were 100 ms, the amount of silence between each pair was 150 ms. If the stimuli were 150 ms; the amount of silence between each pair was 100 ms, which is the same amount of silence as the ISI condition. If the stimuli were 200 ms, the amount of silence between stimuli was 50 ms (see Figure 1).

The target in each trial was a music segment with a female or male vocalist. In the 80 trials there was one target present in 64 trials and no target present in 16 trials. The target was equally probable in each of the positions 1 through 4 within the trials. The target thus occurred in each of the positions for 16 trials out of the 64 target-present trials. For 32 trials (half of the total trials containing a target), the target was a part of a vocal line from a familiar song. Half of the familiar targets contained a female vocalist and the other half contained a male vocalist. Another 32 trials had vocals from unfamiliar music as the target. Half of the unfamiliar targets contained a female vocalist and the other half contained a male vocalist. Thus, there were 16 trials with a familiar female vocalist as the target, 16 trials with an unfamiliar female vocalist as the target, 16 trials with a familiar male vocalist as the target, and 16 trials with an unfamiliar male vocalist as the target. Familiar and unfamiliar music never occurred within the same trial. The targets were randomized in position (1 through 4), type (familiar or unfamiliar), and gender (male or female), with equal numbers of each type of trial in each condition. The three remaining segments of music within a trial were instrumental. Each instrumental excerpt only appeared once in the experiment. The instrumental music was either familiar or unfamiliar, depending on the corresponding target within a trial. If the target was a familiar vocalist, then the three instrumental segments were also from familiar music, and vice versa if the target was unfamiliar (see Figure 2).

The stimuli in Experiment 2 were the same as those used in Experiment 1 except for two differences. First, the only type of silence was the ISI condition. We excluded the SOA condition because results from Experiment 1 showed no difference in vocalist detection performance between the two silence conditions, indicating that the amount of silence between music segments, the temporal context, was not influential during encoding. Second, in Experiment 2, familiar and unfamiliar trials with male vocalists were now considered distracters instead of Targets (see Figure 2).

PROCEDURE

Prior to participation, participants filled out a consent form and a short music questionnaire. The purpose of the questionnaire was to ensure that they normally listened to the music included in the study (see Appendix A for the music questionnaire). After completion, a hearing screening test was administered to ensure participants had normal hearing before participating in the study. The test presented pure tones in order to examine hearing sensitivity on both ears at 20 dB HL at octave intervals from 500 to 8000 Hz. Participants who did not have normal hearing ($n = 15$) were thanked for their time and did not continue with the study.

The stimuli in both Experiments 1 and 2 were presented via Matlab on a Macbook Pro. Participants listened binaurally with Koss UR-20 headphones and were able to adjust the volume to a comfortable listening level in the neighborhood of 70 db sound pressure level (SPL).

All participants were tested individually. Seated comfortably in front of the computer, participants were guided through six example trials of the 150 ms ISI condition in order to orient them to the task. Participants were instructed to respond via keyboard as quickly and accurately as possible as to whether or not they heard a vocalist in each group of short segments. They responded on a 1 (*Very confident, NO there was not a vocalist*) to 6 (*Very confident, YES there was a vocalist*) confidence scale. If participants responded that they heard a vocalist, they were asked in which position, 1 through 4, they thought that they heard the vocalist. The final question posed to participants was whether or not the vocalist was male or female. Each participant took part in one experimental session, which consisted of only one length of music segment, 100 ms, 150 ms, or 200 ms, and one type of silent interval, ISI or SOA, with trial presentation randomized for each participant.

The procedure for Experiment 2 was the same as Experiment 1, except that participants were instead instructed to respond as to whether or not they heard a female vocalist only. If participants responded that they heard a female vocalist, they were asked in which position, 1 through 4, they thought that they heard the female vocalist.

ANALYSIS

The present study used IBM SPSS Statistics 21 statistical software program for data analysis. To assess the effect of condition on the behavioral measurements, a series of ANOVAs were completed on participants' hit rate, false alarm rate, and overall area scores. Area under the receiver operating characteristic (ROC) provides an unbiased estimate of proportion correct where chance = .50 (Swets, 1973). All models used area scores as the dependent variable. An alternative series of models with conditional probability as the dependent variable can be found in Appendix B. The within-subjects factors were 1) type of target with two levels: familiar vocalist and unfamiliar vocalist in Experiment 1, or familiar female vocalist and unfamiliar female vocalist in Experiment 2; 2) gender of the vocalist with two levels: male vocalists and female vocalists; and 3) position of the vocalist with four levels: positions 1, 2, 3, and 4. The between-subjects factors were: 1) length of music segment with three levels: 100 ms, 150 ms, and 200 ms; and 2) amount of silence with two levels: ISI and SOA. To lower the probability of making a Type I error, the Bonferroni correction for multiple comparisons (14 in total) yielded $\alpha = .0036$. Supplementary analyses can be found in Appendix B. Post hoc analyses used pairwise comparisons via t-tests using the Bonferroni correction with differences significant at $\alpha = .05$. Eta squared (η^2) was used as a measure of strength of the effect using the total sum of squares. These analyses were the same for Experiments 1 and 2.

Results

EXPERIMENT 1: DETECTING A VOCALIST

The purpose of the first series of ANOVAs was to assess the effect of amount of silence between music segments on vocalist detection performance in both the ISI condition (100 ms of silence between stimuli) and the SOA condition (the entire trial adding up to one second) (see Figure 1). Area scores were calculated on hits (correctly

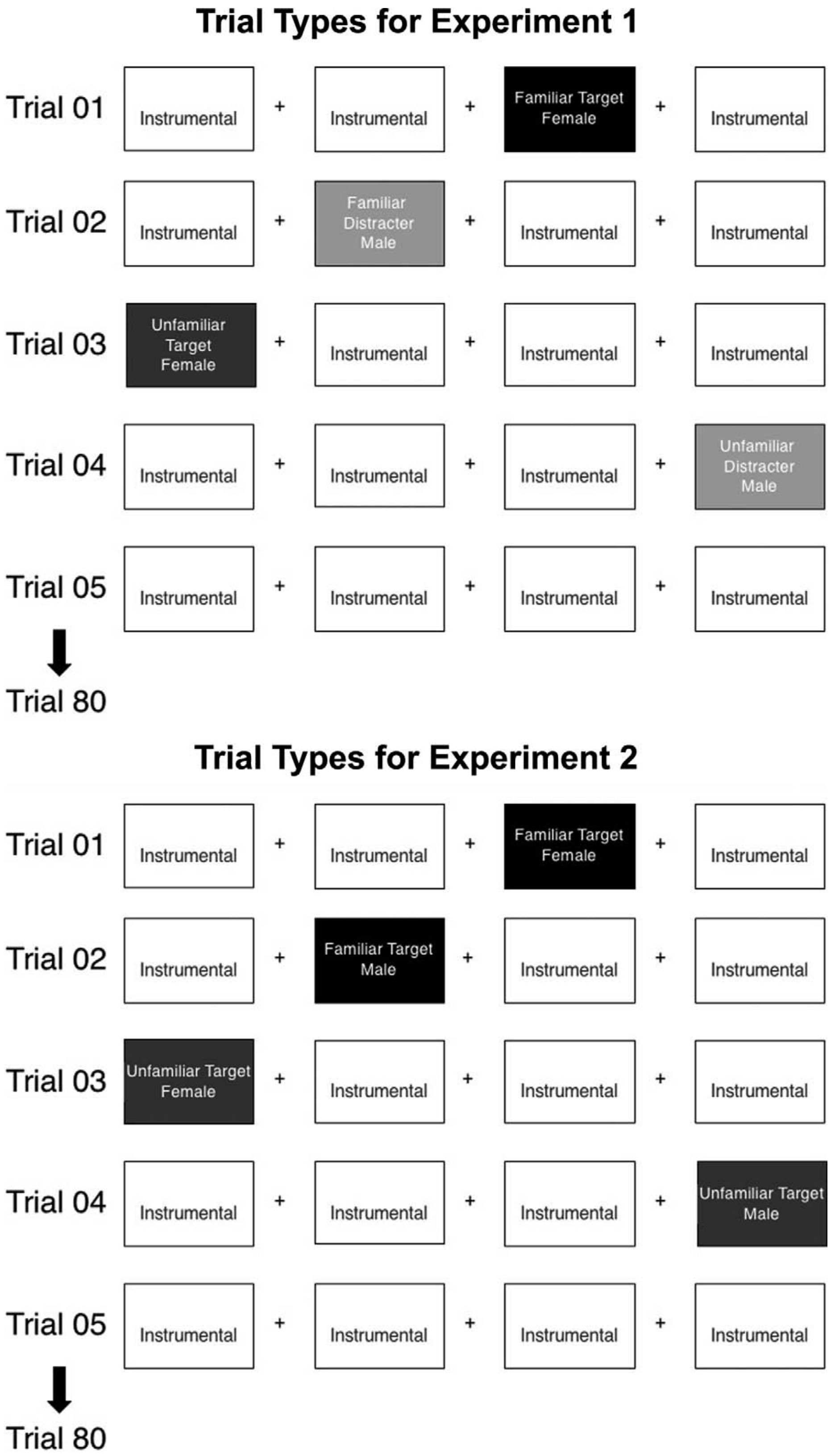

FIGURE 2. An example of the possible trial types in Experiments 1 and 2. For Experiment 1, familiar vocalist targets are black and unfamiliar vocalist targets are dark gray. For Experiment 2, familiar female vocalist targets are black and unfamiliar female vocalist targets are dark grey. Male vocalist distracters are light grey. Instrumental music segments are white for both experiments.

detecting a vocalist) versus false alarms (incorrectly hearing an instrumental music clip as a vocalist). We did not expect differences in vocalist detection performance between the ISI and the SOA conditions.

Two mixed-design ANOVAs were completed with target familiarity (familiar, unfamiliar) as the within-subjects variable and silence length between music segments (ISI, SOA) as the between-subjects variable, for both the 100 ms music segment condition and for the 200 ms segment condition. Results showed that with 100 ms music segments, participants were able to discriminate the presence of the vocalist in the familiar stimuli ($M = 90.98$, $SD = 5.97$) better than in the unfamiliar stimuli ($M = 82.73$, $SD = 8.62$), $F(1, 38) = 51.45$, $MSE = 26.46$, $p < .0001$, $\eta^2 = .24$. There was no significant difference in vocalist detection performance between the ISI condition and the SOA condition, $F(1, 38) = 0.06$, $MSE = 83.32$, $p = .81$. There was a weak interaction effect between target familiarity and silence length between music segments that yielded no significant differences among the individual means, $F(1, 38) = 4.17$, $MSE = 26.46$, $p = .048$, $\eta^2 = .02$, observed power = .51.

The pattern for the main effect for target familiarity remained the same with 200 ms music segments with participants having better vocalist detection performance in the familiar stimuli ($M = 95.80$, $SD = 4.21$) than in the unfamiliar stimuli ($M = 92.48$, $SD = 5.11$), $F(1, 38) = 11.67$, $MSE = 18.94$, $p = .002$, $\eta^2 = .11$. With the more stringent alpha level from the Bonferroni correction for multiple comparisons ($\alpha = .0036$) participants showed no difference in vocalist detection performance in the ISI condition ($M = 95.58$, $SD = 3.49$) compared to the SOA condition ($M = 92.70$, $SD = 5.75$), $F(1, 38) = 7.69$, $MSE = 21.49$, $p = .009$. The interaction was not significant, $F(1, 38) = 0.35$, $MSE = 18.94$, $p = .56$.

The next series of ANOVAs addressed whether there was a significant improvement in vocalist detection performance with an increase in music segment length between the 100 ms, 150 ms, and 200 ms conditions. As shown with the previous results and confirmed with subsequent analyses, temporal context was not important to encoding. Since the SOA condition produced quantitatively similar effects, all subsequent results will present findings for the ISI condition only.

To assess the effect of target familiarity and music segment length on vocalist detection performance, a 2 (target familiarity: familiar, unfamiliar) x 3 (music segment length: 100 ms, 150 ms, 200 ms) mixed-design ANOVA was completed with music segment length as the between-subjects variable. The results followed the prediction that participants would detect the presence of the vocalist in the familiar stimuli more accurately than in the unfamiliar stimuli, $F(1, 57) = 57.31$, $MSE = 12.23$, $p < .0001$, $\eta^2 = .13$. Vocalist detection improved with an increase of music segment length, with a significant difference in vocalist detection performance between the 100 ms condition and the 200 ms condition, $F(2, 57) = 18.47$, $MSE = 43.18$, $p < .0001$, $\eta^2 = .29$. The interaction was trending toward significance, $F(2, 57) = 2.66$, $MSE = 12.23$, $p = .08$, $\eta^2 = .01$. In the unfamiliar stimuli, participant performance improved with an increase in music segment length. In the familiar stimuli, however, performance showed a ceiling effect from the 150 ms condition to the 200 ms condition indicating that 150 ms provided sufficient information for this vocalist detection task (see Figure 3).

The final analysis addressed performance in detecting vocalist gender with a 3 (music segment length: 100 ms, 150 ms, 200 ms) x 2 (target familiarity: familiar, unfamiliar) x 2 (vocalist gender: male, female) mixed-design ANOVA. We predicted that performance would improve with an increase in music segment length. We also predicted that participants would have better gender detection in the familiar stimuli than the unfamiliar stimuli. No effect of gender was anticipated.

There was a main effect of familiarity, $F(1, 57) = 48.79$, $MSE = 72.13$, $p < .001$, $\eta^2 = .11$, with more accurate participant performance in correctly detecting vocalist gender in the unfamiliar stimuli than in the familiar stimuli. There was an interaction between familiarity and vocalist gender, $F(1, 57) = 14.24$, $MSE = 56.38$, $p = .001$, $\eta^2 = .03$, observed power = .96. For the male vocalists, participants had much higher performance in the unfamiliar stimuli than the familiar stimuli. The pattern continued in the female vocalists, though performance became more similar between the unfamiliar and familiar stimuli (see Figure 4).

EXPERIMENT 2: DETECTING A FEMALE VOCALIST

Experiment 2 served the purpose to address further the importance of the vocalist as a feature used in processing music. Experiment 2 focused on participant performance when discerning the vocalist gender by having participants detect the female vocalist amongst not only instrumental distractors as in Experiment 1, but male vocalist distractors as well. This experiment served to provide evidence of the importance of the feature of vocalist gender when processing music.

While the participant task changed from distinguishing vocalists from instrumentalists in Experiment 1, to distinguishing female vocalists from both male vocalists and instrumentalists in Experiment 2, the prediction

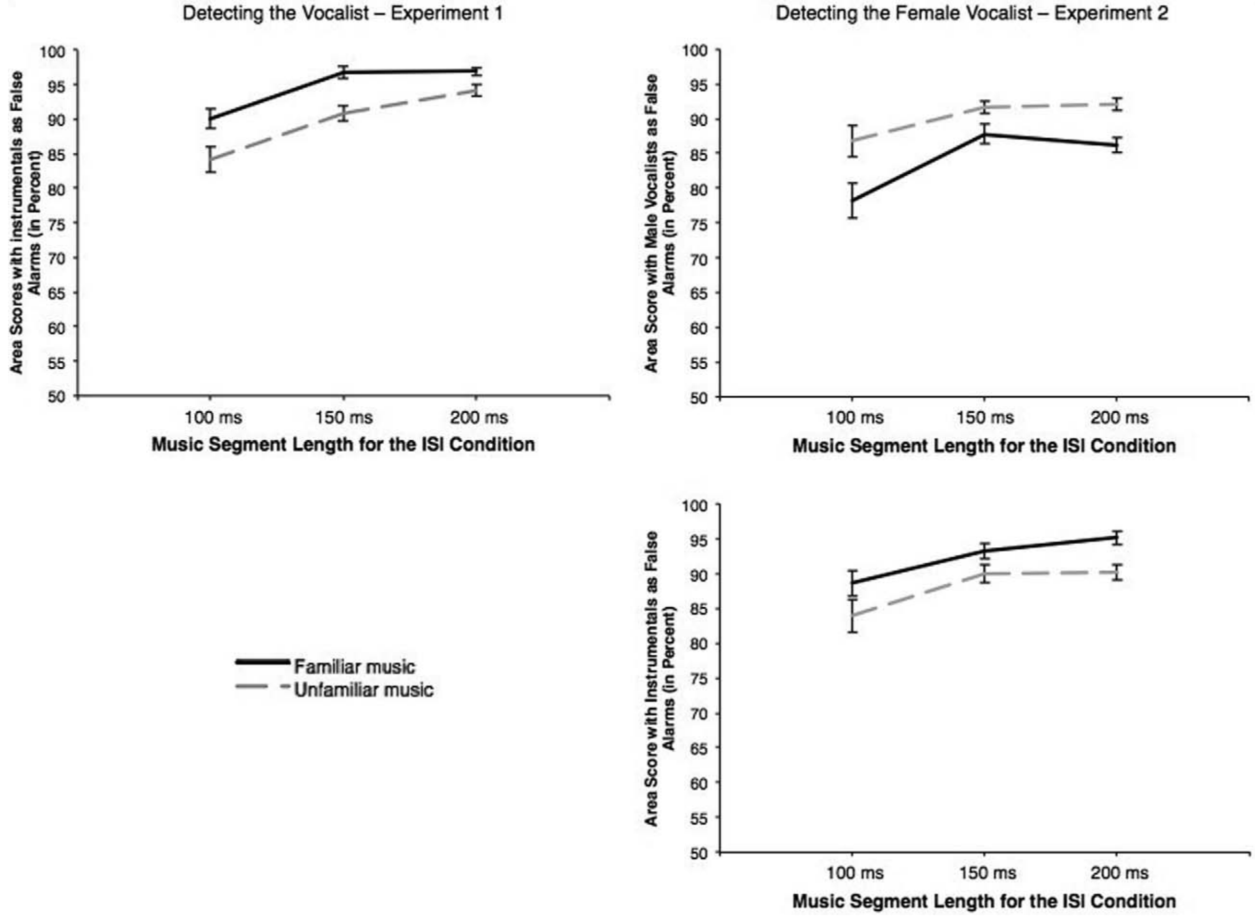

FIGURE 3. Correct detection of the vocalist (Experiment 1) or female vocalist (Experiment 2) as a function of target familiarity with familiar (solid line) and unfamiliar (dashed line) conditions and music segment length with chance level performance at 50%, in the ISI condition. Error bars represent standard error of the mean.

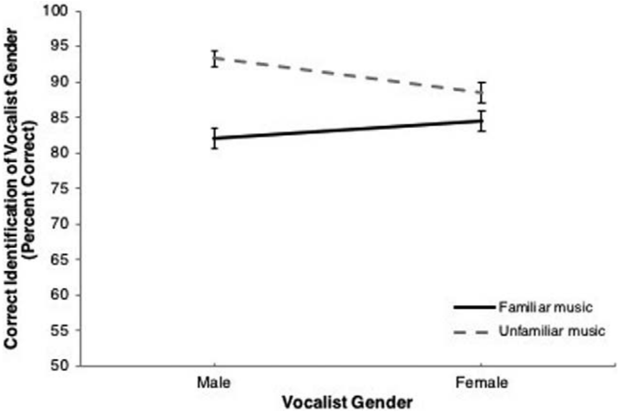

FIGURE 4. Correct identification of vocalist gender in Experiment 1 as a function of target familiarity with familiar (solid line) and unfamiliar (dashed line) conditions and vocalist gender in the ISI condition. Error bars represent standard error of the mean.

that participants would better detect female targets in the familiar stimuli than in the unfamiliar stimuli remained. To assess an improvement in performance in detecting a female vocalist with an increase in the length of the music segment length, a 2 (female target familiarity: familiar female vocalist, unfamiliar female vocalist) x 3 (music segment length: 100 ms, 150 ms, 200 ms) mixed-design ANOVA was used. The dependent variable was area score, which consisted of a hit as correctly detecting a female vocalist and a false alarm as identifying a male vocalist as a female vocalist.

Results showed a main effect of familiarity, $F(1, 57) = 39.50$, $MSE = 28.41$, $p < .001$, $\eta^2 = .13$. Participants were able to discriminate the presence of a female vocalist in the unfamiliar stimuli more accurately than in the familiar stimuli. These results neither matched the prediction that participants would perform better at correctly detecting female vocalists with the familiar stimuli than in the unfamiliar stimuli, nor the corresponding results from Experiment 1. Results also showed a main effect of music segment length, $F(2, 57) = 7.99$, $MSE = 80.11$, $p < .001$, $\eta^2 = .13$. Consistent with findings from Experiment 1, participant performance was most accurate in the 200 ms condition that only differed significantly from the 100 ms condition (see Figure 3).

Similar to the previous ANOVA with the only difference being the make-up of the area scores, a 2 (female target familiarity: familiar female vocalist, unfamiliar female vocalist) x 3 (music segment length: 100 ms, 150 ms, 200 ms) mixed-design ANOVA was completed. In this ANOVA area scores were calculated as participant performance in distinguishing female vocalists from instrumentalists. Participants were able to discriminate the presence of a female vocalist in familiar stimuli better than in unfamiliar stimuli, $F(1, 57) = 14.72$, $MSE = 36.82$, $p < .001$, $\eta^2 = .08$. There was a main effect of music segment length, $F(2, 57) = 8.09$, $MSE = 57.91$, $p < .001$, $\eta^2 = .14$. Participants were most accurate in the 200 ms condition, followed by the 150 ms, and 100 ms conditions, with no significant difference in performance between the 200 ms and 150 ms conditions (see Figure 3).

Discussion

Results from both Experiment 1 and Experiment 2 indicated that participants were able to discern vocalists from instrumentalists in Western popular music well above chance, in as little as 100 ms music segments. While task performance significantly improved from 100 ms to 150 ms, no significant difference occurred between performance in the 150 ms and 200 ms conditions. This result indicates that 150 ms provides sufficient information needed to identify the vocalist (Experiment 1) or the female vocalist (Experiment 2), and discern these targets from their respective distractors. The addition of the extra 50 ms of auditory information did not significantly improve performance. While previous research indicates that people are able to discern spoken or sung voice segments from instrumental segments in less than 50 ms, these segments were controlled to contain only one type of timbre per segment (Bigand, Delbé, Gérard, & Tillmann, 2011; Suied, Agus, Thorpe, Mesgarani, & Pressnitzer, 2014), which is a stark contrast to the complex and dense timbral arrangements found in Western popular music (Starr & Waterman, 2003).

The previous study by Schellenberg et al. (1999) used segments of music in the duration of 100 ms and 200 ms, which were as short as the segments used in the present study. As mentioned previously, their methodologies were limited in that they only used five recordings of popular music and the participants matched the song segment to an alphabetical list of artists and song titles. The current study involved 320 separate music segments (160 familiar music segments) and provided no prompt for music identification assistance. Music segment length in the current study was shorter than the length deemed sufficient to gain information such as making basic emotional judgments, classifying genres, identifying title and artist and decade of release, and discerning familiar music from unfamiliar music, which ranged from 250 ms to 500 ms (Bigand et al., 2009; Filipic et al., 2010; Gjerdingen & Perrott, 2008, Krumhansl, 2010). Previous research also suggests that participants could not have used language semantics in the decision making process of discerning vocalists from instrumentalists as the amount of information provided is much shorter than 240 ms, the length that is considered sufficient for spoken word identification (Grosjean, 1980; Marslen-Wilson & Welsh, 1978; Miller & Grosjean, 1981).

In Experiment 1, participants identified the vocalist in the familiar stimuli more accurately than in the unfamiliar stimuli, though performance was well above chance in both conditions. The current study differs from previous studies regarding familiarity (Bigand et al., 2009; Filipic et al., 2010) in that participants were not explicitly informed that familiarity was a factor in either of the present experiments, nor did the task involve discerning familiar stimuli from unfamiliar stimuli. While the present study did not assess that all familiar songs were familiar to participants individually, results suggest that participant performance was influenced by the familiarity of the stimuli, perhaps outside of participant awareness.

When identifying vocalist gender, participant performance did not differ between identifying male and female vocalists. When coupled with target familiarity, however, there was an interaction between target familiarity and vocalist gender, with participants more accurately identifying vocalist gender in the unfamiliar stimuli compared to the familiar stimuli. In other words, when attention is directed at the more specific feature of vocalist gender, participants show better performance in gender identification in the unfamiliar stimuli compared to the familiar stimuli. Specifically, participants were better at identifying male vocalists in the unfamiliar stimuli than in the familiar stimuli. The same pattern emerged with female vocalists, though performance became more similar between familiarity types for female vocalists.

The only difference between Experiment 1 and Experiment 2 was the nature of the vocalist detection task. In Experiment 1, participants were instructed to detect vocalists, (regardless of gender). In Experiment 2, participants were instructed to detect female vocalists, a more specific feature than in Experiment 1. This change in task resulted in an interesting change in results from Experiment 2 in comparison to Experiment 1. In Experiment 2, results indicated that when using male vocalists as false alarms, participants correctly detected female vocalists more accurately in the unfamiliar stimuli than in the familiar stimuli. These results are the direct opposite of the results from Experiment 1. When using instrumentalists as false alarms (making the area scores parallel those of Experiment 1, also with instrumentalists as false alarms), performance converges with Experiment 1 and switches back to being more accurate in the familiar stimuli than in the unfamiliar stimuli. This switch in performance between experiments appears to hinge on the effect of target familiarity, though as mentioned previously, participants were unaware of this factor.

The relationship with a piece of music during the transition from the unfamiliar to the familiar ultimately has to do with repetition. In a striking finding from Margulis (2013), participants rated that they found more enjoyment with, were more interested in, and thought as more artistic excerpts of music that were modified to include repetition. The excerpts utilized is what makes this finding so interesting in that they were drawn from modernist music, music that is most notably devoid of repetition. The sections chosen for repetition were not even chosen for an appropriate continuation in the musical structure of the piece, but rather for convenience such as a break in the music. Findings from North and Hargreaves (1995) provided further evidence in the relationship between liking and familiarity in music. Their results show a positive relationship between liking and familiarity, suggesting that there is no limit to how much you can like a piece of music with which you become more and more familiar. The caveat is that our liking is tempered by complexity. People seem to have an optimal complexity for liking music; however, if music is either too simple or too complex, liking drops, making an inverted U relationship such as the Wundt curve. So, in terms of liking and complexity, we can also become too familiar with music, finding it no longer as stimulating and thus enjoyable (North & Hargreaves, 1995).

The underlying force influencing an increase in liking from the findings in both Margulis (2013) and North and Hargreaves (1995) was the mere exposure effect (Zajonc, 1968). Previous research on the mere exposure effect indicates that in the absence of extreme negative qualities, liking increases with the more exposure one has to a stimulus. Part of this liking occurs because over multiple exposures, people build a better understanding and thus predictability of their experience with the stimulus, which as such has become less complex. The nature of Western popular music is that it is known for much repetition, and as mentioned previously, timbrewise it is very dense. Within one piece the listener experiences the verse/chorus pattern upwards of three times, making it a type of music that is cleverly feeding the human desire for repetition and understanding of one's environment. Additionally, it may take time for the listener to become bored with a piece of music, as there are so many layers of sounds in which to attend. With more repetition comes an even more involved relationship with the music such that it becomes an enduring aspect of long-term memory, so much so that listeners can show absolute pitch and rhythm for their favorite pieces (Levitin, 1994; Levitin & Cook, 1996). When re-listening to music, previous electrophysiological results suggest that listeners may not be able to separate previous memories and emotions from past encounters with this piece, which may be related to using more attention towards the music shown by a larger N400 response (Daltrozzo & Schön, 2008; Koelsch et al., 2004; Miranda & Ullman, 2007; Steinbeis & Koelsch, 2008).

To explain the influence of familiarity on the switch in performance accuracy between Experiment 1 and Experiment 2, it is hypothesized that using vocalist gender may not be helpful and/or widely used as a feature in processing familiar music. Regarding our theorized auditory RBC model, present results suggest that the vocalist appears to be stored as a useful auditory geon in long-term memory, whereas vocalist gender does not. Perhaps storing both the vocalist and vocalist gender as separate features is redundant and unnecessary, as arguably if one can identify the vocalist, the detail of that vocalists' gender is automatically and simulataneously activated. Another reason that vocalist gender may not be stored as a feature when processing familiar music is the wide use of falsetto in popular music, especially in male vocalists (Ravens, 2014). If a male vocalists uses falsetto, then the discrimination of females using higher vocal registers than males may be blurred, thus rendering vocalist gender a possibly unreliable feature in processing music. If listening to an unfamiliar song, however, using vocalist gender might be helpful in not only understanding the song but also narrowing down possible artists.

One important contribution of the findings from this paper is that the current results provide converging evidence that people who do not necessarily have long-term or intensive formal music training, or are considered music experts in relation to music analysis or performance, are able to process such short segments of musical information. In the present study of 160 participants, all indicated that they listened to music every day and often for many hours a day. While it cannot be stated that the present participants were music experts per se, they certainly obtained quite a lot of practice listening to music, which may involve finding to what they want to listen, changing from song to song to avoid to what they do not want to listen, and so on. The ability to perceive, organize, comprehend, and—with familiar music—recall such a dense and complex array of sound should not be taken lightly. The ability to discern between familiar sounds and unfamiliar sounds with such a limited amount of information, perhaps so limited that one is not consciously aware of the difference, is also evidence towards the importance of assessing new sounds in one's environment. This contribution aids the changing perception of lay listeners as not being nonmusicians, but rather musically untrained listeners who are effective music processors (Bigand, 2003; Bigand & Poulin-Charronnat, 2006).

The current study provided converging evidence with the previous literature, concluding "Humans have exceptional capabilities when it comes to recognizing the source and significance of sounds" (Gjerdingen & Perrott, 2008, p. 93). Additionally, the current study showed that when people are familiar with music, they know it so well that even a tenth of a second of information influences their processing, a much shorter duration than was shown in previous literature. Finally, the current findings provide insight as to the importance of the vocalist as a feature for processing familiar songs, and vocalist gender as a feature perhaps used in processing unfamiliar songs, an area that was largely unexplored in the previous literature. To understand better how people process music, especially music that is familiar and more meaningful to them, is a line of research to be continued.

Author Note

The authors would like to acknowledge Dr. Peter F. Assmann, Dr. James F. Jerger, and Dr. Robert Xavier Rodriguez, for their assistance in critiquing the manuscript and their commitment and participation in the doctoral committee for Shannon L. Layman.

This manuscript is part of Shannon L. Layman's doctoral dissertation.

Correspondence concerning this article should be addressed to Shannon L. Layman, The University of Texas at Arlington, Department of Psychology Box 19528, Arlington, TX 76019. E-mail: shlayman@uta.edu

References

Barrett, S. E., Rugg, M. D., & Perrett, D. I. (1988). Event related potentials and the matching of familiar and unfamiliar faces. *Neuropsychologia, 26,* 105–117.

Bigand, E. (2003). More about the musical expertise of musically untrained listeners. *Annals of the New York Academy of Sciences, 999*, 304–312.

Bigand, E., Delbé, C., Gérard, Y., & Tillmann, B. (2011). Categorization of extremely brief auditory stimuli: Domain-specific or domain-general processes? *PLOS ONE, 6*(10), e27024.

Bigand, E., Gérard, Y., & Molin, P. (2009). The contribution of local features to familiarity judgments in music. *The Neurosciences and Music III – Disorders and Plasticity: Annals New York Academy of Sciences, 1169,* 234–244.

Bigand, E., & Poulin-Charronnat, B. (2006). Are we "experienced listeners"? A review of the musical capacities that do not depend on formal musical training. *Cognition, 100*, 100–130.

Daltrozzo, J., & Schön, D. (2008). Conceptual processing in music as revealed by N400 effects on words and musical targets. *Journal of Cognitive Neuroscience, 21,* 1882–1892.

Daltrozzo, J., Tillmann, B., Platel, H., & Schön, D. (2009). Temporal aspects of the feeling of familiarity for music and the emergence of conceptual processing. *Journal of Cognitive Neuroscience, 22*, 1754–1769.

Eells, J. (2014, September 8). *The reinvention of Taylor Swift.* Retrieved from http://www.rollingstone.com/music/features/taylor-swift-1989-cover-story-20140908

Filipic, S., Tillmann, B., & Bigand, E. (2010). Judging familiarity and emotion from very brief musical excerpts. *Psychonomic Bulletin and Review, 17,* 335–341.

Gjerdingen, R. O., & Perrott, D. (2008). Scanning the dial: The rapid recognition of music genres. *Journal of New Music Research, 37,* 93–100.

Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception and Psychophysics, 28,* 267–283.

Hummel, J. E. (2013). Object recognition. In D. Reisberg (Ed.), *The Oxford handbook of cognitive psychology* (pp. 32–45). New York: Oxford University Press.

Hummel, J. E., & Biederman, I. (1992). Dynamic binding in a neural network for shape recognition. *Psychological Review, 99,* 480–517.

Koelsch, S., Kasper, E., Sammler, D., Schulze, K., Gunter, T., & Friederici, A. D. (2004). Music, language, and meaning: Brain signatures of semantic processing. *Nature Neuroscience, 7,* 302–307.

Krumhansl, C. L. (2010). Plink: "Thin slices" of music. *Music Perception, 27,* 337–354.

Kutas, M., & Federmeier, K. (2000) Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences, 4,* 463–470.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 204*, 203–205.

Levitin, D. L. (1994). Absolute memory for musical pitch: Evidence from the production of learned melodies. *Perception and Psychophysics, 56,* 414–423.

Levitin, D. L., & Cook, P. R. (1996). Memory for musical tempo: Additional evidence that auditory memory is absolute. *Perception and Psychophysics, 58,* 927–935.

Mace, S. T., Wagoner, C. L, Hodges, D., & Teachout, D. J. (2011). Gender identification of very brief musical excerpts (2011). *Psychology of Music, 40,* 112–128.

Margulis, E. H. (2013). Aesthetic responses to repetition in unfamiliar music. *Empirical Studies of the Arts, 31*(1), 45–57.

Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*, *10*, 29–63.

Miranda, R. A., & Ullman, M. T. (2007). Double dissociation between rules and memory in music: An event-related potential study. *NeuroImage, 38,* 331–345.

Miller, J. L., & Grosjean, F. (1981). How the components of speaking rate influence perception of phonetic segments. *Journal of Experimental Psychology: Human Perception and Performance*, *7*, 208–215.

North, A. C., & Hargreaves, D. J. (1995). Subjective complexity, familiarity, and liking for popular music. *Psychomusicology, 14,* 77–93.

Orgs, G., Lange, K., Dombrowski, J., & Heil, M. (2006). Conceptual priming for environmental sounds and words: An ERP study. *Brain and Cognition, 62,* 267–272.

Plazak, J., & Huron, D. (2011). The first three seconds: Listener knowledge gained from brief musical excerpts. *Musicae Scientiae, 15*, 29–44.

Potter, M. C. (1976) Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory 2*, 509–522.

Ravens, S. (2014). *The supernatural voice: A history of high male singing.* Woodbridge, Suffolk: Boydell & Bewer Ltd.

Rentfrow, P. J., Levitin, D. J., & Goldberg, L.R. (2011). The structure of musical preferences: A five-factor model. *Journal of Personality and Social Psychology, 100,* 1139–1157.

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

Schweinberger, S. R., Herholz, A., & Sommer, W. (1997). Recognizing famous voices: Influence of stimulus duration and different types of retrieval cues. *Journal of Speech, Language and Hearing Research, 40,* 453–463.

Starr, L., & Waterman, C. (2003). *American popular music from minstrelsy to MTV.* New York: Oxford University Press.

Steinbeis, N., & Koelsch, S. (2008). Comparing the processing of music and language meaning using EER and fMRI provides evidence for similar and distinct neural representations. *PLOS ONE, 3,* e2226.

Suied, C., Agus, T. R., Thorpe, S. J., Mesgarani, N., & Pressnitzer, D. (2014). Auditory gist: Recognition of very short sounds from timbre cues. *Journal of the Acoustical Society of America, 135*, 1380–1391.

Swets, J. A. (1973). The relative operating characteristic in psychology. *Science, 182*, 990–1000.

Tekman, H. G., & Bharucha, J. J. (1998). Implicit knowledge versus psychoacoustic similarity in priming of chords. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 252–260.

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology, 9*, 1–27.

Appendix A

MUSIC PREFERENCE QUESTIONNAIRE

Music Questionnaire

1. Have you ever had any music lessons/ music experience (choir, orchestra, band, etc.)?

2. If you have had music lessons/ music experience, when did this occur and for how long?

3. How often do you listen to music?

4. What types of music do you listen to?

5. Please describe any other music listening habits that the previous questions did not address:

Appendix B

SUPPLEMENTAL ANALYSES

Experiment 1: Detecting a Vocalist

In addition to examining area scores, a series of ANOVAs addressed the difference between vocalist detection performance—assessed by the probability of participants correctly determining the presence of a vocalist (hit rate)—the temporal position of the vocalist (position 1, 2, 3, or 4), vocalist gender (male vocalist, female vocalist), and as well the probability of false alarms.

Probabilities of false alarms and hits. Before assessing hit rates, false alarm rates were assessed to address the effect of target familiarity and music segment length with a 2 (target familiarity: familiar, unfamiliar) x 3 (music segment length: 100 ms, 150 ms, 200 ms) mixed-design ANOVA. As predicted, false alarm rates were uniformly low (< 4.00%) but after the Bonferroni correction for multiple comparisons, were not significantly different between familiar and unfamiliar targets $F(1, 57) = 4.24$, $MSE = 2.55$, $p = .04$. No other results for false alarms were significant. As false alarm rates were constant, the analysis of hit rates agreed qualitatively with area scores.

The next analysis addressed performance in vocalist attribute identification with a 3 (music segment length: 100 ms, 150 ms, 200 ms) x 2 (target familiarity: familiar, unfamiliar) x 3 (vocalist attributes: position of the vocalist, vocalist gender, and both vocalist attributes) mixed-design ANOVA. It was predicted that participants would have better performance with the familiar targets than with the unfamiliar targets. It was also predicted that participants would be able to identify the correct position of the vocalist and correct gender of the vocalist above the chance levels of 25.00% for position and 50.00% for gender. An increase in participant performance in vocalist attribute identification was expected as music segment length increased. No interaction effects were expected.

There was not a main effect of music segment length, $F(2, 57) = 4.40$, $MSE = 425.83$, $p = .02$. There was a main effect for target familiarity, $F(2, 57) = 20.54$, $MSE = 76.48$, $p < .001$, $\eta^2 = .03$, with higher accuracy in vocalist attribute identification in the unfamiliar stimuli than the familiar stimuli. There was a main effect of vocalist attribute identification with results indicating no difference in performance between identifying vocalist position or vocalist gender, and significantly lower performance overall in correctly identifying both attributes within the same trial, $F(2, 114) = 78.35$, $MSE = 60.85$, $p < .001$, $\eta^2 = .17$. There was an interaction between vocalist attributes and music segment length, $F(4, 114) = 2.87$, $MSE = 18.73$, $p < .001$, $\eta^2 = .02$. When identifying vocalist position alone and identifying both vocalist attributes within the same trial, participants were significantly better in the 200 ms condition than in the 100 ms condition. Music segment length had no effect on vocalist gender identification. Finally, there was an interaction between vocalist attributes and target familiarity, $F(2, 114) = 529.88$, $MSE = 18.73$, $p < .001$, $\eta^2 = .02$. Results showed no effect of target familiarity when identifying vocalist position; however, when identifying vocalist gender, participants were more accurate with the unfamiliar stimuli. Identification of both vocalist attributes within the same trial yielded more accurate performance with the unfamiliar stimuli.

The final analysis assessed performance in correctly detecting the different conditions within vocalist position and vocalist gender separately. The first analysis addressed performance in detecting the position of the vocalist with a 3 (music segment length: 100 ms, 150 ms, 200 ms) x 2 (target familiarity: familiar, unfamiliar) x 4 (vocalist position: 1–4) mixed-design ANOVA. We predicted an increase in performance with an increase in music segment length. No effects of vocalist position or vocalist familiarity were anticipated.

There was a main effect of music segment length, with the most accurate participant performance in the 200 ms condition ($M = 91.86$, $SD = 12.01$), differing significantly from the 100 ms condition ($M = 80.83$, $SD = 22.13$), $F(2, 57) = 3975.73$, $MSE = 893.80$, $p < .001$, $\eta^2 = .05$. There was not a main effect of position, $F(3, 171) = 4.06$, $MSE = 402.85$, $p = .008$, an interaction effect between vocalist familiarity and music segment length, $F(2, 57) = 3.17$, $MSE = 125.94$, $p = .05$, or an interaction between vocalist position and music segment, $F(6, 171) = 2.62$, $MSE = 402.85$, $p = .02$.

Experiment 2: Detecting a Female Vocalist

In addition to examining area scores, a series of ANOVAs addressed the difference between participant performance assessed by the probability of participants correctly determining the presence of a female vocalist (hit rate), the temporal position of the vocalist (position 1, 2, 3, or 4), and as well the probability of false alarms.

Probabilities of false alarms and hits. Before assessing hit rates, false alarm rates were assessed to address the effect of target familiarity on the music segment length. As Experiment 2 contained two types of false alarms—a false alarm to male vocalists and a false alarm to instrumentalists—two 2 (target familiarity: familiar, unfamiliar) x 3 (music segment length: 100 ms, 150 ms, 200 ms) mixed-design ANOVAs were used for each false alarm type. We predicted that false alarm rates would be low overall and that for both false alarm types; rates would decrease with an increase of music segment length.

Results for false alarms to trials with male vocalists indicated that participants had a higher false alarm rate in the familiar stimuli than in the unfamiliar stimuli, $F(1, 57) = 92.67$, $MSE = 18.78$, $p < .0001$, $\eta^2 = .26$. There was no main effect of music segment length and the interaction was not significant, $F(2, 57) = 3.34$, $MSE = 18.78$, $p < .05$, $\eta^2 = .01$. Results for false alarms to trials with instrumental clips only were parallel to the results for false alarms in Experiment 1, with higher false alarm rates for unfamiliar stimuli than for the familiar stimuli $F(1, 57) = 18.93$, $MSE = 30.46$, $p < .0001$, $\eta^2 = .08$. As false alarm rates were constant, the analysis of hit rates agreed qualitatively with area scores.

Probabilities of hits. The final analysis addressed performance in detecting the position of the vocalist with a 3 (music segment length: 100 ms, 150 ms, 200 ms) x 2 (target familiarity: familiar, unfamiliar) x 4 (vocalist position: 1–4) mixed-design ANOVA. We predicted an increase in performance with an increase in music segment length. No effects of target position or familiarity were anticipated.

There was a main effect of music segment length, $F(2, 57) = 8.64$, $MSE = 1574.49$, $p = .001$, $\eta^2 = .08$. Most accurate participant performance occurred in the 200 ms condition ($M = 79.38$, $SD = 23.21$), which differed significantly from the 100 ms condition ($M = 60.94$, $SD = 27.71$). There was also a main effect of position, $F(3, 171) = 7.54$, $MSE = 530.45$, $p < .001$, $\eta^2 = .03$. Participant performance in detecting the female vocalist in position 3 was significantly lower than performance in all other conditions. There was an interaction between familiarity type and position, $F(3, 171) = 34.42$, $MSE = 341.12$, $p < .001$, $\eta^2 = .10$. In the familiar stimuli, performance was the most accurate in position 3, which only differed significantly from position 2. In the unfamiliar stimuli, the effect was driven by low average female vocalist detection in position 3, which was significantly lower than all other positions.

Probabilities of false alarms. A similar ANOVA, but with false alarm rate as the dependent variable, addressed the drop in performance in position 3. The only significant effect was a main effect of position, $F(3, 171) = 8.18$, $MSE = .000$, $p < .001$, $\eta^2 = .06$, with significantly lower false alarms in position 1 ($M = 0.008$, $SD = 0.01$) in comparison to positions 2 ($M = 0.013$, $SD = 0.02$), 3 ($M = 0.021$, $SD = 0.02$), and 4 ($M = 0.016$, $SD = 0.02$). No other differences were significant, suggesting that the drop in hit rate performance in position 3 was not due to misidentifying the female vocalist as present in different position, but rather missing the female vocalist altogether.