

Musicae Scientiae

<http://msx.sagepub.com/>

The first three seconds : Listener knowledge gained from brief musical excerpts

Joseph Plazak and David Huron

Musicae Scientiae 2011 15: 29

DOI: 10.1177/1029864910391455

The online version of this article can be found at:

<http://msx.sagepub.com/content/15/1/29>

Published by:



<http://www.sagepublications.com>

On behalf of:

European
Society for the
Cognitive Sciences
Of
Music

[European Society for the Cognitive Sciences of Music](http://www.sagepub.com)

Additional services and information for *Musicae Scientiae* can be found at:

Email Alerts: <http://msx.sagepub.com/cgi/alerts>

Subscriptions: <http://msx.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>

Permissions: <http://www.sagepub.com/journalsPermissions.nav>

Citations: <http://msx.sagepub.com/content/15/1/29.refs.html>

The first three seconds: Listener knowledge gained from brief musical excerpts

Musicae Scientiae

15(1) 29–44

© The Author(s) 2011

Reprints and permission: sagepub.co.uk/journalsPermissions.nav

DOI: 10.1177/1029864910391455

msx.sagepub.com

msx.sagepub.com



Joseph Plazak and David Huron

Ohio State University, USA

Abstract

The human auditory system can rapidly process musical information such as, for example: the recognition and identification of sound sources, the deciphering of meter, tempo, mode, and texture, the processing of lyrics and dynamics, the identification of musical style and genre, the perception of performance nuance, and the apprehension of emotional character. Two empirical studies are reported that attempt to chronicle when such information is processed. In the first exploratory study, a diverse set of musical excerpts was selected and trimmed to various durations, ranging from 50 ms to 3000 ms. These samples, beginning with the shortest and ending with the longest, were presented to participants, who were then asked to free associate and talk about any observations that came to mind. Based on these results, a second main study was carried out using a betting paradigm to determine the amount of exposure needed for listeners to feel confident about acquired musical information. The results suggest a rapid unfolding of cognitive processes within a 3-second listening span.

Keywords

betting paradigm; brief musical excerpts; chronology; listener knowledge

All mental tasks, including music listening, take place in real time. Some tasks take longer to perform than others. When a person is listening to sounds, the amount of time needed to process certain kinds of information may reflect the complexity of the stimulus, the expertise of the listener, the length of the associated neural pathways, or the listener's physiological state. Additionally, the time required to perform a task may also reflect the importance of the information to the listener. The purpose of this study is to attempt to chronicle a possible timeline of individual mental processes involved in music listening.

The human auditory system is able to rapidly process many types of musical information: the recognition and identification of sound sources (e.g., Patterson et al., 2008); the deciphering of meter, tempo, and form (e.g., Lerdahl & Jackendoff, 1983; London, 2004); the identification of musical style and genre (e.g., Gjerdingen & Perrott, 2008); the processing of pitch and pitch patterns (e.g., Krumhansl, 1990); the perception of performance nuance (e.g., Johnson, 1996); the apprehension of emotional character (e.g., Huron, 2006); and so on. With the exception of a

Corresponding author:

Joseph Plazak, School of Music, 1866 College Rd., Columbus, OH 43210, USA

Email: plazak.1@osu.edu

few empirical findings (discussed below), the amount of time needed to acquire various types of musical information is unknown. In attempting to decipher the broad outline of music perception it may be useful to determine the chronology by which different perceptual processes unfold.

Most of the research on the timeline of auditory processing comes from studies in psychoacoustics. Psychoacoustic research exemplifies a bottom-up approach to auditory perception by studying relatively basic sound properties. For example, the perception of pitch can occur with remarkable quickness. Under ideal circumstances, trained listeners have been able to resolve the pitch of an isolated sound in as little as 10 ms (Warren et al., 1991). Likewise, Robinson and Patterson (1995) found that only 1–4 cycles of a sound are required in order to extract timbre information from complex tones, whereas pitch information requires 8–10 cycles.

A few studies have investigated the chronology of acquiring higher-level auditory information. Gjerdingen and Perrott (2008) and Schellenberg et al. (1999) both focused on the time required to identify musical style. Gjerdingen and Perrott likened the task of style identification to “scanning the radio dial,” where a listener decides whether to remain at the current station or move to the next station in a very short period of time. In a formal study, they found that listeners were capable of deciphering musical style within 250 ms following the onset of a sound excerpt, and that listeners were remarkably accurate in identifying specific styles within this period. Schellenberg had participants identify popular tunes by name from short sound excerpts that lasted either 100 ms or 200 ms. Results from this study found that listeners performed the task well with only 200 ms stimuli, and that they performed above chance levels when the excerpts were just 100 ms. This latter study also found that playing the samples backwards impaired participants’ ability to complete the task, which suggests that temporal (or time-varying) elements of music are important even for durations as short as 100 ms. London (2004) reviewed evidence that the shortest perceptual beat or tactus in a piece occurs with an interonset interval between 200 and 250 ms. Therefore, London further deduced that the shortest possible subdivisions would have a lower interonset interval limit of approximately 100 ms, depending on whether the subdivisions are simple or complex.

In the present research, many types of musically-pertinent information are chronicled in addition to the identification of style. Two related studies aimed to identify a possible chronology or order for different music-related processes. The first study was exploratory in nature. The goal was to identify some of the musical characteristics that listeners draw out of a brief listening experience. These data formed the basis for the ensuing main study in which the researchers probed *when* in the listening experience a particular type of musically-pertinent information becomes available to the listener. To anticipate our results, it will be shown that there is a rapid unfolding of cognitive processes within a 3-second listening span. We will offer a tentative chronology of information processing, including such acquired musical knowledge as instrument identification, genre, mode, mood, form, density, pleasantness, meter, and tempo.

Exploratory study

In our initial exploratory study, a small group of 7 listeners heard excerpts of Western music and were simply invited to comment on each excerpt. Transcripts of the comments were then analyzed for content. Each participant was tested individually.

Stimuli

A diverse set of 70 musical samples was harvested from 9 genre-specific radio stations available via XM-Sirius satellite radio. These genres included: classical, rock, jazz, blues, reggae,

electronic, rap, gospel, and country. In order to expand the sample further, listeners heard either an unusual amateur recording (*Neapolitan Nights*, Brown, 2005) or an intentionally composed “bad” piece (*The Most Unwanted Song*, Soldier, Komar, & Melamid, 1997). Each participant was exposed to 10 stimuli representing each of the 10 genres. In addition, these 10 excerpts were trimmed so as to represent each of the 10 temporal epochs: 50 ms, 100 ms, 200 ms, 400 ms, 800 ms, 1200 ms, 1600 ms, 2000 ms, 2500 ms, and 3000 ms. Musical samples were chosen randomly and cut to these durations using sound editing software. These short clips of music were assembled into playlists, each of which contained 10 independent samples and represented all 10 genres and all 10 durations.

Participants

Seven listeners were recruited from the Ohio State University School of Music subject pool. This study was one of several that could be selected by participants in order to receive course credit. Participants were primarily music majors. Each participant was tested individually in an Industrial Acoustics Corporation sound isolation room.

Participants heard 10 short musical excerpts while the experimenter remained present. After each excerpt the participants’ task was simply to talk about any aspect of the sound. In order to encourage the participants to speak at length about their perceptions, the experimenter transcribed their remarks. After each remark, the experimenter prompted the participant for further observations regarding any aspect of their experience with the sound. That is, there was constant pressure for participants to come up with additional thoughts or observations. It was expected that having the participants self-record monologues or type their own responses would have led to fewer remarks. The experimenter read the directions aloud while the participant read along.

Instructions: The purpose of this experiment is to gather information about how people experience sound. At the end of the experiment I’ll say more about our specific goals. But for now it is best that you understand that we’re simply interested in your reactions.

After each sound example, I want you to talk about your experience. Say as much as you can about the sound and about how you experience the sound. I will be transcribing your remarks, so I might ask you to slow down or repeat what you said.

I may ask you some questions, but the purpose of the questions is simply to get you to talk about what you hear. Ideally, I wouldn’t ask you any questions at all. Feel free to talk about anything at all. You can talk about any aspect of the sound; whatever catches your attention, whatever you think, whatever it reminds you of. My preference is for you to simply talk about what you experience without my prompting.

Do you have any questions about this?

With one exception (explained below), sound samples were presented from the shortest to the longest. That is, participants heard a 50 ms excerpt, followed by an independent 100 ms excerpt, and so on, up to an independent 3000 ms excerpt. The total duration of the study was roughly 30 minutes.

Results

In total, the participants provided 424 discrete comments about the sound examples. Figure 1 graphs the number of comments according to the durations of the different samples. In general, the number of comments appears positively correlated with sample duration ($r = +.91$). That is, listeners provided more (prompted) comments for the longer sound segments. Having observed this relationship, the researchers became concerned that the increased number of comments

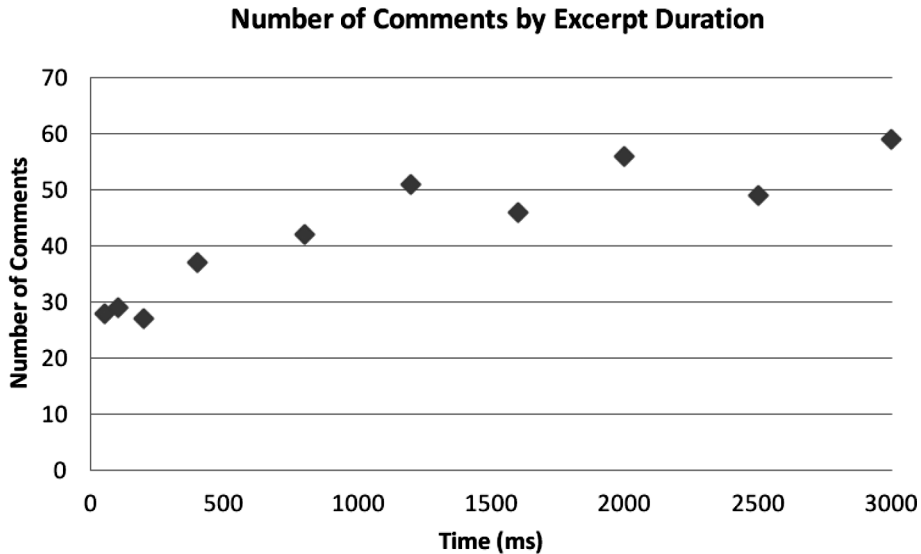


Figure 1. Number of spontaneous prompted comments evoked by musical excerpts of different lengths. The results suggest that the amount of musically related observation is positively correlated with the duration of exposure.

might be an artifact arising from participants growing accustomed to the research procedure. That is, as participants gained familiarity with the informal experimental situation, they may have become more eager to talk. As an informal test of this potential confound, one participant completed the study in reverse order (with the longest stimulus occurring first). In this single case, it was nevertheless observed that the number of comments remained greatest for the longest stimulus – suggesting that familiarity or comfort with the research procedure might not be a factor. A transcript for one of the participants' remarks is provided in Appendix 1. The experimenter's probing remarks were limited by the instructions provided above.

Content analysis

Following the data collection, an informal content analysis was carried out on the participants' transcriptions. All of the 424 comments were printed on slips of paper, and the comments were then manually sorted according to whatever categories seemed appropriate. The categorization was done twice, once by one of the experimenters, and a second time by an independent researcher not involved with the project. After sorting, both researchers provided descriptive labels for each of their categories.

On the basis of the two independent content analyses, an aggregate set of categories was created as shown in Table 1. These categories were an effort to provide a superset of the two content analyses. This aggregate set of categories was created informally rather than employing any formal method. On the basis of the first study, 20 statements were formulated that were deemed to represent 20 music-related knowledge domains. This information is presented in Table 1.

It should be recognized that the initial choice of sample durations was arbitrary. In our content analysis we found that the number of comment categories increased significantly between 400 and 800 ms. As a result, we added an additional epoch at 600 ms in the main study.

Table 1. Knowledge domains obtained from the exploratory study with illustrative statements

Knowledge domain	Illustrative statement
Dynamics	This excerpt has a loud dynamic level.
Changing dynamics	This excerpt is part of a crescendo.
Acoustic environment	This recording was made in a highly reverberant room.
Critical judgment	Most people would say this music is pleasant.
Form	This excerpt is from the "chorus" part of a song.
Culture	This excerpt is from a piece of music from Mexico.
Instrumentation	There are drums in this musical excerpt.
Lyrics	The word "love" is in the lyrics of this musical excerpt
Intonation	This excerpt is out of tune.
Rhythm (syncopation)	This is a syncopated excerpt.
Rhythm (meter)	This excerpt has a triple meter.
Rhythm (tempo)	This is a slow-tempo excerpt.
Mode	This excerpt is in the minor mode.
Vocal presence	There is a singer in this musical excerpt.
Vocal sex	There is a female singer in this excerpt.
Performance quality	This excerpt is performed by an amateur group.
Affect	The mood of this excerpt is sad.
Density	This excerpt contains four or more instruments.
Texture	This excerpt is polyphonic in texture.
Style	This excerpt is from a piece of classical music.

Main study

In our main study, the goal was to answer the following research question: "At what point in some sonic exposure do listeners perceive X about the sound?" In brief, 20 new participants, again music students from the Ohio State University School of Music subject pool, were exposed to musical excerpts ranging from 50 ms to 3000 ms. Subjects were tested independently. In order to establish whether listeners had acquired particular types of musically-pertinent information, the researchers made use of a betting paradigm as described below.

Stimuli

Stimuli were once again collected from genre-specific radio stations available via XM-Sirius Radio. Specifically, these new samples were derived from 10 channels classified by XM-Sirius as rock, classical, jazz, blues, reggae, country, Latin, world, rap, and electronic. Supplementary stimuli were added from amateur performances available on YouTube. In total, 119 excerpts of 30 seconds each were randomly collected.

Musical works were randomly assigned to one of 10 revised temporal epochs: 50 ms, 100 ms, 250 ms, 400 ms, 600 ms, 800 ms, 1000 ms, 1500 ms, 2000 ms, and 3000 ms. A random excerpt was selected and a stimulus of appropriate duration was created. These samples were trimmed using sound editing software with a graphical user interface. The graphical interface allowed the

researchers to avoid randomly selecting passages of silence. As in the exploratory study, each participant heard only one excerpt from each of the 119 musical passages. This approach better ensured data independence and prevented possible confounds due to familiarity.

Procedure

The procedure was designed to determine when a listener acquires some piece of musically relevant knowledge. Consider, for example, the problem of how we know when a listener has determined that a trumpet is present in a sound. Of course, we can simply invite a spontaneous report of such information from the listener (as was done in the exploratory study). Alternatively, we might prompt the listener for specific information by asking a pertinent question, such as "Does the following excerpt contain a trumpet?". Especially in the case of very short excerpts, listeners are apt to engage in a certain amount of guessing. We might ask listeners to report their confidence in providing a given response, but listeners may be either over-confident or overly pessimistic in providing a direct confidence measure.

A more serious problem is establishing the ground truth for particular items of musical information. If we ask a listener "Is the following excerpt in the major or minor mode?", we must confidently be able to determine the mode of the excerpt in order to judge whether the listener's response is accurate. However, establishing the ground truth is not always straightforward. In many cases, we can confidently claim that a particular musical work is in the major or minor mode. However, this confidence may evaporate for brief excerpts. For example, a short excerpt might contain predominantly minor sonorities, despite being in a major mode.

In order to circumvent the problem of establishing ground truth, a method was employed that avoided the question of whether a particular item of information was correct. A useful method for determining whether people hold knowledge is to give them the opportunity to place wagers related to that knowledge. Accordingly, our main study employed a betting paradigm. Prior to hearing a sound stimulus, participants viewed a computer screen containing a written assertion such as: "This passage has a loud dynamic level" or "This passage includes a voice." After hearing the stimulus, participants were invited to place bets on the displayed statement. For any given statement, the participant could place a bet of up to 100 (virtual) dollars – either pro or con. That is, a listener could bet up to \$100 that a statement such as, "This passage is in the minor mode" was true or false. Bets were indicated on a linear slider via a computer mouse. This included a \$0 point, so a participant might elect to bet nothing. The participants were reminded that they were betting with pretend, rather than actual, money.

In the first half of the study, participants heard 119 stimuli and were required to bet on a single statement for each stimulus. The specific statements used for betting were those listed in Table 1. The betting statements were displayed prior to hearing the stimulus. Participants themselves initiated the stimulus onset. This allowed participants mentally to prepare for the kind of information appropriate to the displayed statement. After the stimulus had sounded, the display was "unlocked" so that wagers for or against the displayed statement could be entered. Having placed their bet, the participant clicked on a "next" button which advanced the screen to the next trial. By way of example, a statement such as, "This is a syncopated excerpt" was displayed on the screen. After reading the statement, the participant then pressed a play button to hear a random musical sample. For example, the passage might be a 200 ms excerpt of country music. The participant would then bet on the truth or falsity of the statement for the given excerpt before pressing the "next" button.

After the main block of 119 stimuli, a dummy block of 20 "pseudo-stimuli" was introduced. Once again, participants were invited to place wagers on individual claims about a

sound excerpt. However, this time, they had to place their bets *without* hearing any sound. The aim was to establish a baseline or chance-level probability for each of the different music-related statements.

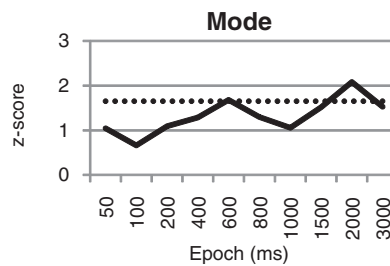
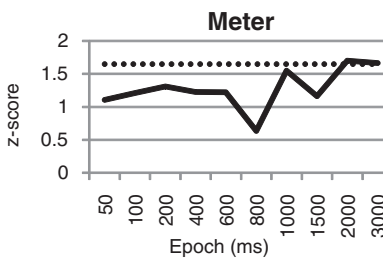
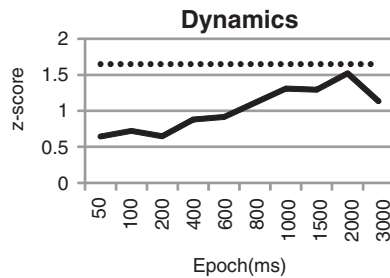
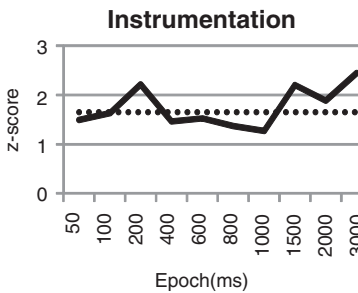
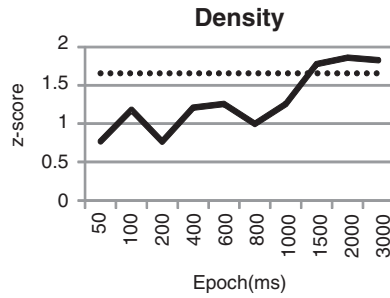
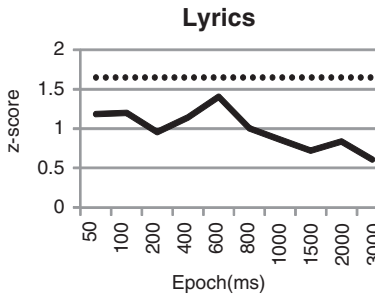
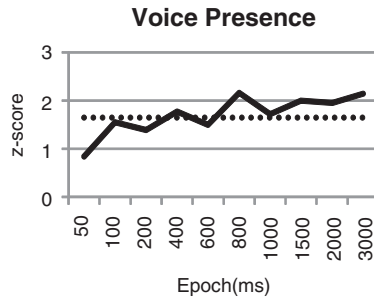
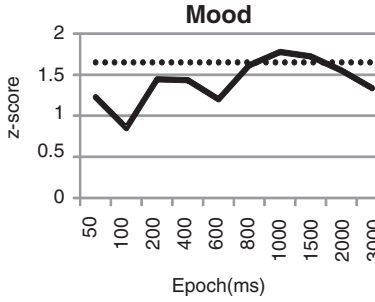
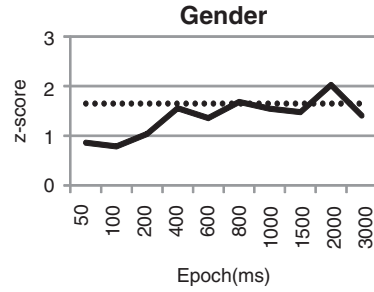
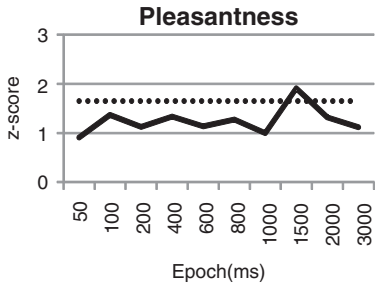
Notice that the data collected from the latter phase of this main study allows the calculation of a chance level for answering each knowledge assertion. Although this data was collected after the main part of the study, it represents an estimate of what statisticians would normally call “prior probability.” Using this probability data, it is possible to normalize the actual bets (related to a given knowledge assertion) placed during the main study. That is, the original wagers can be expressed as normalized *z*-scores. The assumptions underlying this procedure are further discussed below.

Analytic approach and results

Recall that the motivation for the main study was to determine when, in the listening experience, a listener has acquired a certain type of knowledge about a musical passage. For example, when does the listener know that an excerpt is “out of tune,” or when does a listener know that an excerpt exhibits a sad mood. Evidence of knowledge would be apparent when participants’ wagers deviate from a null or chance probability level. This chance level would be expected to differ depending on the type of question posed or statement made. The chance level probabilities were determined from the trailing block data (i.e., dummy questions) in which participants placed wagers without the presence of potentially informative sound stimuli.

For each question or statement, a chance distribution was determined by calculating the means and standard deviations for the “uninformed” wagers. Ideally, this would be done within subjects, since the estimated chance level might differ between subjects. For example, one participant might suppose that major and minor modes are equally likely within the study and so provide a 50/50 guess wager for the major/minor question. Another participant, however, might recognize that the major mode is generally more prevalent, and so provide (say) a 30/70 guess wager for whether an unknown passage is in the minor mode. However, in order to calculate a chance distribution for each subject, multiple responses would be required for each dummy or chance question. Due to time constraints, this was deemed impractical. Only a single response was collected from each participant, and this single response was insufficient for calculating a chance-level distribution. Consequently, the chance distribution was calculated by amalgamating the dummy responses for all 20 participants. The resulting distribution therefore almost certainly exhibits a larger standard deviation than would occur for a within-subject distribution. In calculating normalized scores (see below), the use of a between-subject chance-level distribution therefore has a tendency to underestimate the true within-subject *z*-scores. Hence, our results will tend to underestimate the actual knowledge of individual participants.

Having calculated the distribution of guess wagers for each question, the participants’ responses to actual sound stimuli from the first part of the study were normalized and expressed as *z*-scores. These absolute *z*-score values were averaged across all listeners and then plotted as a function of epoch, as shown in Figure 2. The means and standard deviations which define the statement-related distributions are listed in Table 2. Those *z*-scores (both positive and negative) that differ significantly from zero may be interpreted as evidence of knowledge. That is, exposure to the sound stimulus encouraged the participant to place a wager that differed substantially from the wager they might have made if they had not heard the sound. Values close to the mean ($z = 0$) suggest that the listener has little or no musically pertinent knowledge to inform the betting – at least compared to the chance-level distribution. Conversely, large negative or positive *z*-scores suggest that the listener has obtained some knowledge from the excerpt and so



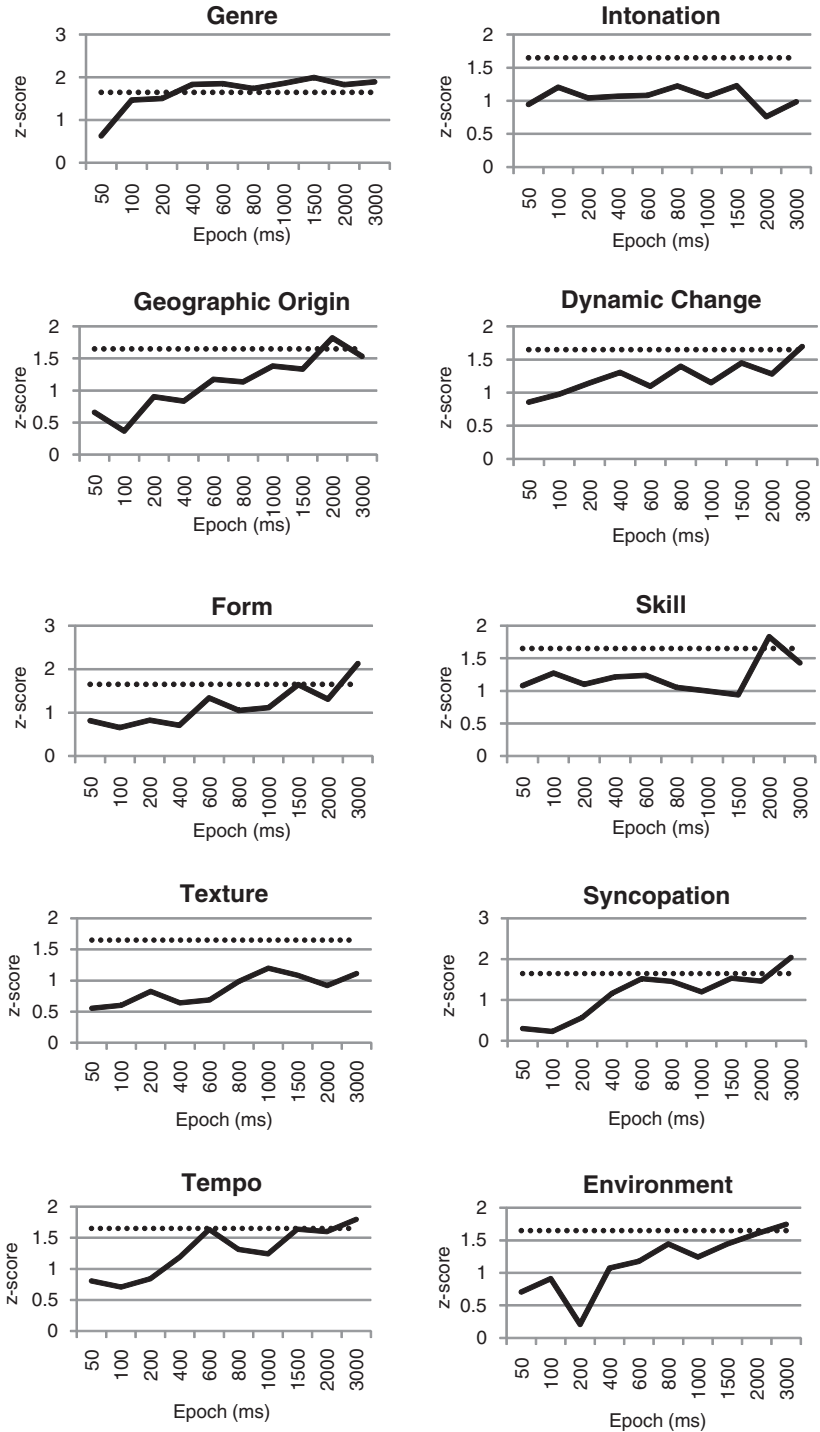


Figure 2. Normalized bets (expressed as absolute z-scores) plotted as a function of knowledge domain and epoch. The dotted line indicates $p = .05$. Graphs may be interpreted as indicating the amount of acquired knowledge with increasing excerpt duration according to knowledge type.

Table 2. Baseline or chance-level probabilities for each knowledge domain statement.

Statement	Mean	SD
This excerpt has a loud dynamic level.	11.2	50.7
This excerpt is part of a crescendo.	-34.6	37.1
This recording was made in a highly reverberant room.	-25.8	43.6
Most people would say this music is pleasant.	51.2	42.0
This excerpt is from the "chorus" part of a song.	-23.9	36.3
This excerpt is from a piece of music from Mexico.	-8.2	49.7
There are drums in this musical excerpt.	49.3	38.4
The word "love" is in the lyrics of this musical excerpt	-65.8	43.6
This excerpt is out of tune.	-50.1	44.6
This is a syncopated excerpt.	-15.2	43.7
This excerpt has a triple meter.	-41.2	33.8
This is a slow-tempo excerpt.	-28.6	42.8
This excerpt is in the minor mode.	-30.7	37.7
There is a singer in this musical excerpt.	30.8	47.7
There is a female singer in this excerpt.	-25.7	43.6
This excerpt is performed by an amateur group.	-55.0	40.4
The mood of this excerpt is sad.	-35.7	36.9
This excerpt contains four or more instruments.	31.5	46.5
This excerpt is polyphonic in texture.	13	70.6
This excerpt is from a piece of classical music.	-15.2	43.7

is betting differently from the chance-level distribution; only the absolute value of each z-score is pertinent. Therefore, these z-scores may be regarded as an index of "knowledgeableness."

Given the number of data points plotted in Figure 2, one might expect problems arising from multiple tests. For individual observations, a null distribution would lead one to expect a one in twenty chance of any given value reaching the $p = .05$ confidence level. However, it should be noted that each of the plotted data points represents the average absolute z-score across a number of listeners. This significantly reduces, although it does not eliminate, the problem of multiple tests. Given the complexity of the calculation, no effort will be made here to correct for multiple tests. Hence, the plotted $p = .05$ line in Figure 2 (dotted line) is uncorrected and should be taken as suggestive rather than literal.

Recall from our exploratory study that we observed a positive correlation between the epoch duration and the number of provided comments. In Figure 2, we can see a parallel phenomenon in which the absolute z-score values generally rise as the excerpt duration increases. This is consistent with the notion that listeners gain increasing music-related information with the longer exposure times. The results are consistent with our assumption that larger bets are symptomatic of increased knowledge.

Discussion

The purpose of this study was not to test one or more hypotheses, but rather to chronicle a possible timeline for music related cognitive processes. A tentative timeline is offered in Table 3.

Table 3. Tentative chronology of listener-acquired musical knowledge.

Music knowledge domain	First epoch of significance
Instrumentation	100 ms
Genre	400 ms
Voice	400 ms
Mode	600 ms
Gender	800 ms
Mood	1000 ms
Form	1500 ms
Density	1500 ms
Pleasantness	1500 ms
Meter	2000 ms
Geographic origin	2000 ms
Skill	2000 ms
Tempo	3000 ms
Syncopation	3000 ms
Changing dynamic	3000 ms
Environment	3000 ms

Notwithstanding our earlier caveat regarding multiple tests, we might deem statistical significance to be achieved when the absolute *z*-score value is greater than or equal to 1.65 (nominally $p \leq 0.05$). Of the 20 probing statements, 16 might be thought to have reached significance at a particular epoch. The entries in Table 3 correspond to the first epoch which showed data consistent with the significance criteria. However, Figure 2 shows oscillations above and below the significance level for several questions – variability suggesting that collecting more data would be warranted.

Limitations of the study

In an ideal study, the corpus of musical stimuli would be representative of all music. This would include every type and variety of music from around the world. Although an effort was made to include several selections of world music in the main study, the chosen musical corpus was highly biased towards Western-enculturated listeners.

Our main study employed just 119 excerpts. Although participants placed bets on different statements for the 119 stimuli, the 2,380 bets placed cannot be regarded as truly independent data. However, the alternative of providing 2,380 different musical stimuli was deemed impractical.

A single probing question was used to represent each music knowledge domain, despite that fact that a multitude of representative statements could have been generated. Using multiple probing statements might be expected to curtail some of the variability when a participant misinterprets a statement. In hindsight, the experimenters realized that this information could easily have been coded into the data set and would have eliminated the need to generalize the results of a single probe to an entire knowledge domain.

This study was also limited by employing a betting paradigm that did not provide feedback. Earlier it was rationalized that one method for determining whether or not a participant has gained knowledge is to ask them to place wagers. However, the wagers are subject to change based on the reward criteria. It seems highly plausible that participants would have bet differently if there had been an actual reward/punishment system providing continuous feedback. The lack of feedback could have caused the betting task to become tedious, and the lack of any real reward could have caused participants to bet recklessly.

Future research might consider employing the same paradigm used in this study to research various types of musical features that have been regarded as difficult to measure, including musical feeling, musical emotion, or virtuosity. Future studies might aim to refine the classification of music related knowledge domains and endeavor to increase the temporal resolution through a greater variety of window sizes. Response-time measures have long been known to correlate with the complexity of the neural processing. Fast response times suggest simpler neural pathways, whereas slow response times suggest more complex neural processes. Studies that refine the timeline tentatively offered here may ultimately prove helpful in deciphering music-related brain activity.

References

- Brown, T. (2005). *Neapolitan nights* [mp3 recording].
- Gjerdingen, R. O., & Perrott, D. (2008). Scanning the dial: The rapid recognition of music genres. *Journal of New Music Research*, 37(2), 93–100.
- Huron, D. (2006). *Sweet anticipation: Music and the psychology of expectation*. Cambridge, MA: MIT Press.
- Johnson, C. M. (1996). Musicians' and nonmusicians' assessment of perceived rubato in musical performance. *Journal of Research in Music Education*, 44(1), 84–96.
- Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music*. Cambridge, MA: MIT Press.
- London, J. (2004). *Hearing in time*. New York: Oxford University Press.
- Krumhansl, C. L. (1990). *Cognitive foundations of musical pitch*. New York: Oxford University Press.
- Patterson, R. D., Smith, D., Dinter, R., & Walters, T. (2008). Size information in the production and perception of communication sounds. In W. A. Yost, A. N. Popper, & R. R. Fay (Eds.), *Springer handbook of auditory research: Vol. 29. Auditory perception of sound sources* (pp. 43–75). New York: Springer.
- Robinson, K., & Patterson, R. D. (1995). The duration required to identify an instrument, the octave, or the pitch chroma of a musical note. *Music Perception*, 13, 1–15.
- 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.
- Soldier, D., & Komar & Melamid (1997). *The Most Unwanted Song* [CD]. Dia Art Foundation.
- Warren, R. M., Gardner, D. A., Brubaker, B. S., & Bashford, J. A. (1991). Melodic and nonmelodic sequences of tones: Effects of duration on perception. *Music Perception*, 8, 277–290.

Appendix I:

Sample transcript from the exploratory study

1. 50 ms Rap
 - “that was it, very short”
 - “two pitches, loud, feedback, rock concert-ish”
 - “amplified”
 - “overtones”
 - “splice of something”
 - “processed voice, probably male”

2. 100 ms Rock
 - “similar to the first one, the pitch is different than the first one”
 - “didn’t hear any voices, instrumental”
 - “sounded like a plucking, or a string instrument”
 - “I hear the attack of the instrument”
 - “rock concert-ish”
 - “able to hum the pitch”
3. 200 ms Gospel
 - “two pitches, first one lower than the second, half or whole step”
 - “group of horns, brassy woodwinds, sounded like more than one”
 - “less amplified, sounded more like studio recording as opposed to live”
 - “I can sing back the pitches”
 - “sounds like it is harmonized (sax and trumpet combination)”
 - “maybe big band timbres or Chicago band”
4. 400 ms Reggae
 - “rhythmic background, accompanying a singer (bass voice)”
 - “sounded pop-ish, R&B, sounded English”
 - “drum track, definitely a rhythm section (electric bass, guitar)”
 - “not bluesy, reminds me of late Motown, R&B, crossover into Gospel”
 - “maybe a little country – a bit of twangy-ness”
 - “heard a bit more of the guitar”
5. 800 ms Electronic
 - “ok techno”
 - “bass rhythmic idea”
 - “sequenced, in a drum machine”
 - “3 pitches – bass, two techno frequencies”
 - “reminds me of rave music, or the Soup on E!”
 - “European influence, Germany, rather than American pop music”
6. 1200 ms Country
 - “female vocalist”
 - “more pop-ish”
 - “clear drum track (probably a live drummer – bass/HH/snare)”
 - “vocal range (mid-alto)”
 - “some sort of chordal instrument”
 - “Maybe Sheryl Crow (mid 90’s 1992–1996)”
 - “bass line”
 - “reminds me of a Grammy nominee CD I have from 1997”
 - “sounds like ‘Every Day is a Winding Road’.”
 - “4th grade-ish”
7. 1600 ms Random
 - “choral thing”
 - “very percussive, maybe a marimba or a piano”
 - “has a glassy sound”
 - “sounds like a live recording in a hall, there is a little bit of reverb”
 - “a chorus may have been coming in at the last second”

- “I heard multiple parts (soprano voice) sounded like an ensemble”
- “more vibes than marimba (sounded like metal not wood)”

8. 2000 ms Classical

- “definitely a symphony”
- “could hear the brass coming in at the end”
- “bass trombone”
- “strings, maybe tremolo in the violins”
- “sounded like Wagner; perhaps late Beethoven”
- “sounded European”
- “not early classical”
- “multiple parts; lot of doubling of parts (sounded minor)”
- “relatively simple in structure”
- “large range, both on the high side and low side”
- “sounded like a fairly typical orchestral arrangement”
- “string bass included with the trombones”
- “sounded fairly aggressive, made me think of Beethoven”
- “not light and fluffy”

9. 2500 ms Blues

- “seemed a lot longer than previous”
- “walking bass, with a drummer”
- “sounds like a tenor or bari sax (scooping the note; going for the bend)”
- “bluesy aspect to it”
- “sounded like up a m3 from the bass”
- “sustained pitch, a wind instrument”
- “there was attack, no decay”
- “club scene; jazz club; background music; Park Street Tavern”
- “doesn’t sound like Pete Milnes”

10. 3000 ms Jazz

- “smooth jazz”
- “piano playing blue notes (flat 3 to major 3)”
- “soprano sax at the end”
- “reminds me of Kenny G”
- “a lot vibrato than the average player” [sic]
- “bluesy, dorian mode-like”
- “not reaching for whole tone or diminished”
- “bass player & drummer & acoustic piano”
- “longer than the first samples”
- “could be Kenny G; stuff you hear on 103.5”
- “cross-over between jazz and pop”
- “could have been someone I saw at the jazz festival this year”
- “in the soprano sax’s normal range (middle to low), not crazy squeaking”
- “definitely a jazz sax (as opposed to classical)”
- “less airy”

Los primeros tres segundos: Conocimiento del oyente obtenido a partir de breves extractos musicales

El sistema auditivo humano puede procesar rápidamente la información musical, como, por ejemplo, el reconocimiento y la identificación de fuentes sonoras, el desciframiento del metro, el *tempo*, el modo y la textura, el proceso de textos y dinámicas, la identificación de estilo y género musical, la percepción del matiz interpretativo, y la aprehensión del carácter emocional. Se comentan dos estudios empíricos que intentan hacer una crónica cuándo dicha es procesada información. En el primer estudio exploratorio, un conjunto diverso de fragmentos musicales fue seleccionado y recortado a duraciones diferentes, entre los 50 ms y los 3000 ms. Estas muestras, que comienzan con la más corta y terminan con la más larga, fueron presentadas a los participantes, a los que se pidió que respondiesen mediante asociación libre y comentasen cualquier asociación que les viniese a la mente. Basado en estos resultados, se realizó un segundo estudio principal utilizando un paradigma de apuestas para determinar la cantidad de exposición necesaria para que los oyentes se sintieran seguros acerca de la información musical adquirida. Los resultados sugieren un rápido despliegue de procesos cognitivos dentro de un periodo de escucha de tres segundos.

I primi tre secondi: Conoscenza acquisita dall'ascolto di brevi estratti musicali

Il sistema uditivo umano può elaborare rapidamente le informazioni musicali, come, ad esempio, il riconoscimento e l'identificazione delle fonti sonore, l'individuazione del metro, del tempo, della modalità e della tessitura, l'assimilazione dei testi e delle dinamiche, l'identificazione dello stile e del genere musicali, la percezione dell'andamento dell'esecuzione e la comprensione del carattere emozionale. Sono riportati due studi empirici che si prefiggono di individuare il momento cronologico in cui queste informazioni sono elaborate. Nel primo studio esplorativo è stata selezionata una sequenza diversificata di estratti musicali di durata variabile da 50 a 3000 secondi. Gli estratti, partendo dal più breve fino al più lungo, sono stati presentati ai partecipanti ai quali poi è stato richiesto di fare delle libere associazioni e di esporre qualsiasi osservazione fosse loro venuta in mente. Sulla base dei risultati di questo primo studio è stato condotto un altro studio utilizzando un paradigma per determinare il tempo necessario di esposizione dell'ascoltatore per potersi sentire sicuro delle informazioni musicali recepite. I risultati suggeriscono un rapido dispiegarsi dei processi cognitivi pari a tre secondi di ascolto.

Les trois premières secondes: la connaissance que l'auditeur retire de courts extraits musicaux

Le système auditif humain peut traiter rapidement l'information musicale, comme par exemple reconnaître et identifier les sources des sons, déchiffrer une mesure, un tempo, un mode, ou une texture, traiter les paroles et les nuances, identifier le style et le genre musical, percevoir les différences d'interprétations, et appréhender les caractères émotionnels. Deux études empiriques sont présentées, qui tentent de dresser une chronologie des moments où de telles informations sont traitées. Dans la première recherche - exploratoire - un ensemble de divers extraits musicaux a été sélectionné, et ces extraits coupés à différentes durées, allant de 50 à 3000 millisecondes. Ces échantillons, en commençant par le plus court pour finir par le plus long, ont été

présentés aux participants, à qui on a ensuite demandé de faire librement des associations et de dire toutes les observations qui leur venaient à l'esprit. Sur base des résultats, une l'étude principale fut menée en utilisant un paradigme de pari pour déterminer le nombre d'expositions nécessaires aux auditeurs pour qu'ils se sentent certains de l'information musicale acquise. Les résultats indiquent un déroulement rapide des processus cognitifs pour une audition d'une durée de 3 secondes.

Die ersten drei Sekunden: Hörererkenntnisse aus kurzen Musikausschnitten

Das menschliche auditorische System kann musikalische Informationen zügig verarbeiten. Dazu zählen beispielsweise das Erkennen und die Identifizierung von Klangquellen, die Entzifferung des Metrums, 'Tempo, Modus' und der Textur, die Verarbeitung des Texts und der Dynamik, die Identifikation des musikalischen Stils und Genres, die Wahrnehmung der Nuancen in der Interpretation und das Verstehen des emotionalen Charakters. In zwei empirischen Studien wird aufzuzeigen versucht, wann solche Informationen verarbeitet werden. In der ersten explorativen Studie wurden vielfältige Musikbeispiele ausgewählt und auf verschiedene Dauern von 50 ms bis 3000 ms geschnitten. Diese Beispiele wurden Versuchsteilnehmern vorgespielt, wobei die kürzesten Beispiele am Anfang und die längsten Beispiele am Ende erklangen. Die Teilnehmer sollten frei über alle Beobachtungen und Assoziationen reden, die ihnen in den Sinn kamen. Basierend auf diesen Ergebnissen wurde eine zweite Hauptstudie durchgeführt. Durch ein Wett-Paradigma sollte die Darbietungsdauer bestimmt werden, die Hörer benötigen, um sich hinsichtlich der erworbenen musikalischen Informationen sicher zu sein. Die Ergebnisse lassen auf einen zügigen Ablauf der kognitiven Prozesse innerhalb einer Hördauer von drei Sekunden schließen.