

Structural integration in language and music: Evidence for a shared system

EVELINA FEDORENKO

Massachusetts Institute of Technology, Cambridge, Massachusetts

ANIRUDDH PATEL

Neurosciences Institute, La Jolla, California

DANIEL CASASANTO AND JONATHAN WINAWER

Stanford University, Stanford, California

AND

EDWARD GIBSON

Massachusetts Institute of Technology, Cambridge, Massachusetts

In this study, we investigate whether language and music share cognitive resources for structural processing. We report an experiment that used sung materials and manipulated linguistic complexity (subject-extracted relative clauses, object-extracted relative clauses) and musical complexity (in-key critical note, out-of-key critical note, auditory anomaly on the critical note involving a loudness increase). The auditory-anomaly manipulation was included in order to test whether the difference between in-key and out-of-key conditions might be due to any salient, unexpected acoustic event. The critical dependent measure involved comprehension accuracies to questions about the propositional content of the sentences asked at the end of each trial. The results revealed an interaction between linguistic and musical complexity such that the difference between the subject- and object-extracted relative clause conditions was larger in the out-of-key condition than in the in-key and auditory-anomaly conditions. These results provide evidence for an overlap in structural processing between language and music.

It has been argued that the domains of language and music share a number of similarities at the sound level, at the structure level, and in terms of general domain properties. First, both language and music involve temporally unfolding sequences of sounds with a salient rhythmic and melodic structure (Handel, 1989; Patel, 2008). Second, both language and music are rule-based systems in which a limited number of basic elements (e.g., words in language, single tones and chords in music) can be combined to generate an infinite number of higher order structures (e.g., sentences in language, harmonic sequences in music) (see, e.g., Bernstein, 1976; Lerdahl & Jackendoff, 1983). Finally, both appear to be universal human cognitive abilities, and both have been argued to be unique to our species (see McDermott & Hauser, 2005, for a recent review of the literature).

This article is concerned with the relationship between language processing and music processing at the structural level. Past studies have taken several approaches to investigate whether the two domains share psychological and neural mechanisms. Neuropsychological investigations of patients with selective brain damage have revealed cases of double dissociations between language and music. In particular, there have been reports of patients who suffer

from a deficit in linguistic abilities without an accompanying deficit in musical abilities (e.g., Luria, Tsvetkova, & Futer, 1965; but cf. Patel, Iversen, Wassenaar, & Hagoort, 2008), and conversely, there have been reports of patients who suffer from a deficit in musical abilities without an accompanying linguistic deficit (e.g., Peretz, 1993; Peretz & Coltheart, 2003; Peretz et al., 1994). These case studies have been interpreted as evidence for the functional independence of language and music.

In contrast, studies using event-related potentials (ERPs), magneto-encephalography (MEG), and functional magnetic resonance imaging (fMRI) have revealed patterns of results inconsistent with the strong domain-specific view. The earliest evidence of this kind comes from Patel, Gibson, Ratner, Besson, and Holcomb (1998; see also Besson & Faïta, 1995; Janata, 1995), who presented participants with two types of stimuli—sentences and chord progressions—and varied the difficulty of structural integration in both. It was demonstrated that difficult integrations in both language and music were associated with a similar ERP component (the P600) with a similar scalp distribution. Patel et al. (1998) concluded that the P600 component indexes the difficulty of structural integration in language

E. Fedorenko, evelina9@mit.edu

and music. There have been several subsequent functional neuroimaging studies showing that structural manipulations in music appear to activate cortical regions in and around Broca's area, which has long been implicated in structural processing in language (e.g., Stromswold, Caplan, Alpert, & Rauch, 1996), and in its right-hemisphere homologue (Koelsch et al., 2002; Levitin & Menon, 2003; Maess, Koelsch, Gunter, & Friederici, 2001; Tillmann, Janata, & Bharucha, 2003).¹ In summary, the results from the neuropsychological case studies and the neuroimaging studies appear to be inconsistent with regard to the extent of domain specificity of language and music.

Attempting to reconcile the neuropsychological and neuroimaging data, Patel (2003) proposed that, in examining the relationship between language and music, it is important to distinguish between long-term structural knowledge (corresponding roughly to the notion of long-term memory) and a system for integrating elements with one another in the course of online processing (corresponding roughly to the notion of working memory) (for an alternative view, see MacDonald & Christiansen, 2002, who hypothesized that no distinction exists between representational and processing networks). Patel (2003) argued that, whereas the linguistic and musical *knowledge systems* may be independent, the system used for *online structural integration* may be shared between language and music (the shared syntactic integration resource hypothesis, SSIRH). This non-domain-specific working memory system was argued to be involved in integrating incoming elements (words in language, tones/chords in music) into evolving structures (sentences in language, harmonic sequences in music). Specifically, it was proposed that structural integration involves rapid, selective activation of items in associative networks and that language and music share the neural resources that provide this activation to the networks where domain-specific representations reside.

Figure 1 illustrates the hypothesis that linguistic and musical syntactic representations are stored in distinct brain networks (and hence can be damaged selectively), whereas there is overlap in the networks that provide neural resources for the activation of stored syntactic representations. Arrows indicate functional connections between networks. Note that the boxes do not necessarily imply focal brain regions. For example, linguistic and musical representation networks could extend across a number of

brain regions or could exist as functionally segregated networks within the same brain regions.

One prediction of the SSIRH is that taxing the shared processing system with concurrent difficult linguistic and musical integrations should result in superadditive processing difficulty because of competition for limited resources. A recent study evaluated this prediction empirically and showed support for the SSIRH. In particular, Koelsch, Gunter, Wittfoth, and Sammler (2005; see also Steinbeis & Koelsch, 2008, for a replication) conducted an ERP study in which sentences were presented visually, word by word, simultaneously with musical chords, with one chord per word. In some sentences, the final word created a grammatical violation via gender disagreement (the experiment was conducted in German, in which nouns are marked for gender), thereby violating a syntactic expectation. The chord sequences were designed to strongly invoke a particular key, and the final chord could be either the tonic chord of that key or an unexpected chord from a distant key. Previous research on language or music alone had shown that structures with gender agreement errors, like those used by Koelsch et al. (2005), elicit a left anterior negativity (LAN), whereas the musical incongruities elicit an early right anterior negativity (ERAN) (Friederici, 2002; Gunter, Friederici, & Schriefers, 2000; Koelsch, Gunter, Friederici, & Schröger, 2000).² In Koelsch et al. (2005), during the critical condition, in which a sequence had simultaneous structural incongruities in language and music, an interaction was observed: The LAN to syntactically incongruous words was significantly smaller when these words were accompanied by an out-of-key chord, consistent with the possibility that the processes underlying the LAN and ERAN were competing for the same/shared neural resources. In a control experiment, it was demonstrated that this was not due to general attentional effects because the size of the LAN was not affected by a simple auditory anomaly manipulation that involved a physically deviant (e.g., ~4 dB louder than the context) chord on the last word in a sentence. Thus, the results of Koelsch et al.'s (2000) study provided support for the SSIRH.

The experiment reported here was aimed at further evaluating the SSIRH and differs from the experiments of Koelsch et al. (2000) in two ways. First, the experiment manipulates linguistic complexity via the use of well-formed sentences (with subject- vs. object-extracted relative clauses [RCs]). To test claims about overlap of

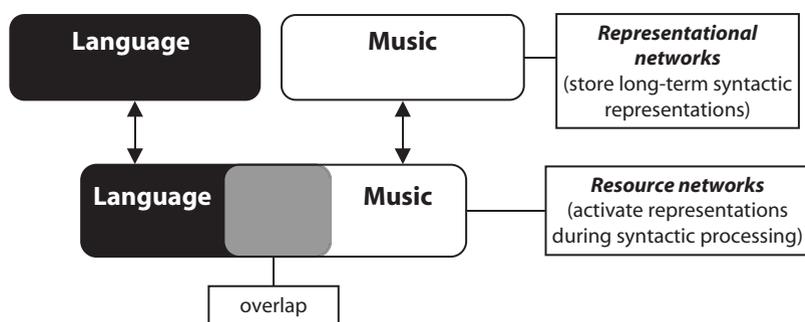


Figure 1. Schematic diagram of the functional relationship between linguistic and musical syntactic processing (adapted from Patel, 2008).

cognitive/neural resources for linguistic and musical processing, it is preferable to investigate structures that conform to the rules of the language rather than structures that are in some way ill formed. This is because the processing of ill-formed sentences may involve additional cognitive operations (such as error detection or attempts at reanalysis and revision), making the interpretation of language–music interactions more difficult (e.g., Caplan, 2007). Another difference is the use of sung materials, in which words and music are integrated into a single stream. For humans, song is an ecologically natural stimulus that researchers have used to investigate the relationship between musical processing and linguistic *semantic* processing (e.g., Bonnel, Faïta, Peretz, & Besson, 2001). To our knowledge, the present study is the first to use song to investigate the relationship between musical processing and linguistic *syntactic* processing.

The Present Experiment

We independently manipulated the difficulty of linguistic and musical structural integrations in a self-paced listening paradigm using sung stimuli to investigate the relationship between syntactic processing in language and music. The prediction of the SSIRH is as follows: The condition in which both linguistic and musical structural integrations are difficult should be more difficult to process than would be expected if the syntactic complexity and musical complexity effects were independent. This prediction follows from the additive factors logic (Sternberg, 1969; see Fedorenko, Gibson, & Rohde, 2007, pp. 248–249, for a summary of this reasoning and a discussion of its limitations).

We examined the effects of the manipulations of linguistic and musical complexity on two dependent measures: listening times and comprehension accuracies. However, only the comprehension accuracy data revealed interpretable results. We therefore present and discuss only the comprehension accuracy data. It is worth noting that the listening time data were not inconsistent with the SSIRH. In fact, there were some suggestions of the predicted patterns, but these effects were mostly not reliable. More generally, the listening time data were very noisy, as evidenced by high standard deviations (*SDs*). The highly rhythmic nature of the materials (see the Method section) may have led participants to pace themselves in a way that would allocate the same amount of time to each fragment, regardless of condition, thus possibly making effects difficult to observe. For purposes of completeness, we report the region-by-region listening times in Appendix A.

One source of difficulty for structural integration in language is related to the need, in cases of nonlocal structural dependencies, to retrieve the structural dependent(s) of an incoming element from memory. It has been hypothesized that retrieval difficulty depends on the linear distance between the two elements (e.g., Gibson, 1998). Here, we compared structures containing local versus nonlocal dependencies. In particular, we compared sentences containing subject- and object-extracted RCs, as shown in (1).

(1a) Subject-extracted RC: The boy that helped the girl got an “A” on the test.

(1b) Object-extracted RC: The boy that the girl helped got an “A” on the test.

The subject-extracted RC (1a) is easier to process than the object-extracted RC (1b), because in (1a) the RC “that helped the girl” contains only local dependencies (between the relativizer “that” co-indexed with the head noun “the boy” and the verb “helped,” and between the verb “helped” and its direct object “the girl”), whereas in (1b) the RC “that the girl helped” contains a nonlocal dependency between the verb “helped” and the relativizer “that.” Therefore, it is plausible that the difference in processing difficulty between subject- and object-extracted RCs is related to a larger amount of working memory resources required for processing object extractions and, in particular, for retrieving the object of the embedded verb from memory (e.g., Gibson, 1998, 2000; Gordon, Hendrick, & Johnson, 2001; Grodner & Gibson, 2005; King & Just, 1991; Lewis & Vasishth, 2005).

The difficulty of structural integration in music was manipulated by varying the harmonic distance between an incoming tone and the tonal center (tonic) of the melody, as shown in Figure 2.³ Out-of-key notes are known to increase the perceived structural complexity of a tonal melody and are associated with increased processing demands (e.g., Eerola, Himberg, Toivainen, & Louhivuori, 2006; Huron, 2006). Crucially, the linguistic and the musical manipulations were aligned: The musical manipulation occurred on the last word of the RC. This is the point (1) where the structural dependencies in the RC have been processed and (2) that is the locus of processing difficulty in the object-extracted RCs due to the long-distance dependency between the embedded verb and its object. This created simultaneous structural processing demands in language and music in the difficult (object-extracted harmonically distant) condition.

As stated above, under the conditions of concurrent linguistic and musical processing, the SSIRH predicts that linguistic integrations should interact with musical integrations, such that when both types of integrations are difficult, superadditive processing difficulty should ensue. However, if linguistic and musical processing were shown to interact superadditively, in order to argue that linguistic and musical integrations rely on the same/shared pool of resources, it would be important to rule out an explanation whereby the musical effect is driven by shifts of attention due to any nonspecific acoustically unexpected event. To evaluate this possibility, we added a condition in which the melodies had a perceptually salient increase in loudness instead of an out-of-key note at the critical position. The SSIRH predicts an interaction between linguistic and musical integrations for the structural manipulation in music, but not for this lower level acoustic manipulation.

METHOD

Participants

Sixty participants from MIT and the surrounding community were paid for their participation. All were native speakers of English and were naive as to the purposes of the study.

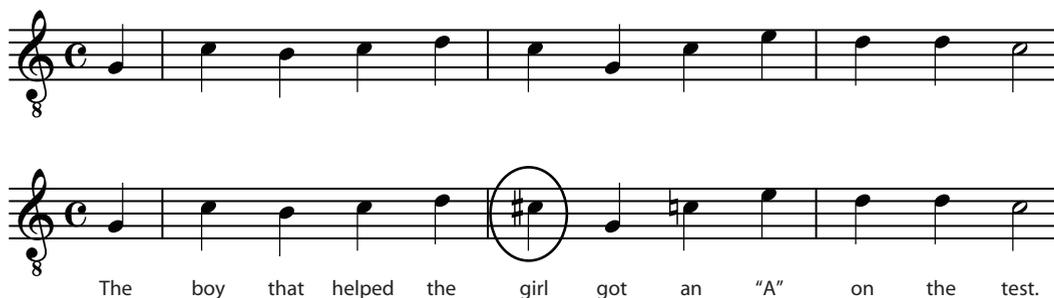


Figure 2. A sample melody (in the key of C major), with a version in which all the notes are in key (top) and a version in which the note at the critical position is out of key (bottom; the out-of-key note is circled).

Design and Materials

The experiment had a 2×3 design, manipulating syntactic complexity (subject-extracted RCs, object-extracted RCs) and musical complexity (critical note in the key of the preceding tonal context [*in key*], critical note not in the key of the preceding tonal context [*out of key*], electronically manipulated loudness increase on the critical note relative to the preceding context [*auditory anomaly*]).

The language materials consisted of 36 sets of sentences, with two versions as shown in (2). Each sentence consisted of 12 mostly monosyllabic words⁴ (so that each word corresponded to one note in a melody) and was divided into four regions for the purposes of recording and presentation, as indicated by slashes in (2a) and (2b): (1) a subject noun phrase, (2) an RC (*subject-extracted/object-extracted*), (3) a main verb with a direct object, and (4) an adjunct prepositional phrase. The reason we grouped words into regions, instead of recording and presenting the sentences word by word, was to preserve the rhythmic and melodic pattern.

(2a) Subject-extracted: The boy / that helped the girl / got an "A" / on the test.

(2b) Object-extracted: The boy / that the girl helped / got an "A" / on the test.

Each of these two versions was paired with three versions of a melody (in key, out of key, auditory anomaly), differing only in the pitch (between in-key and out-of-key conditions) and only in the loudness (between in-key and auditory-anomaly conditions) of the note corresponding to the last word of the RC [underlined in (2a) and (2b)].

In addition to the 36 experimental items, 25 filler sentences with a variety of syntactic structures were created. The filler sentences were 10–14 words long. Like the experimental items, they consisted mostly of monosyllabic words, so that each word corresponded to one note in a melody. To resemble the experimental items, which always consisted of four regions as described above, the words in the filler sentences were grouped into regions (with each sentence consisting of three to five regions and each region consisting of 1–6 words).

The 36 target melodies (with two versions each) and 25 filler melodies were composed by a professional composer (Jason Rosenberg), and the target and filler items were recorded by one of the authors, a former opera singer, Daniel Casasanto.

Melody Creation

Target melodies. All melodies consisted of 12 notes, were tonal (using diatonic notes and implied harmonies that strongly indicated a particular key), and ended in a tonic note with an authentic cadence in the implied harmony (see Figure 2). All melodies were isochronous, consisting only of quarter notes except for the final half note. They were sung at a tempo of 120 quarter-note beats/min (i.e., each quarter note lasted 500 msec). The first five notes established a strong sense of key. Both the in-key and the out-of-key versions of each melody were in the same key and differed by one note. The critical (sixth) note—falling on the last word of the RC—was either in key or out of key. It was always on the downbeat of the second

full bar. When the note was out of key, it was one of the five possible nondiatonic tones (e.g., C \sharp , D \sharp , F \sharp , G \sharp , A \sharp in C major). Sometimes out-of-key notes were different from their corresponding in-key critical note by only a semitone (e.g., C vs. C \sharp).

The size of pitch jumps leading to and from the critical note was matched for the in-key and out-of-key conditions, so that the voice leading for the out-of-key notes closely matched that for the in-key notes. In particular, the mean size of the pitch jumps leading to the critical note was 2.1 ($SD = 1.9$) semitones for the in-key melodies and 2.5 ($SD = 1.7$) semitones for the out-of-key melodies (Mann–Whitney U test, $p = .36$). The mean size of the pitch jumps leading from the critical note was 3.4 ($SD = 2.3$) semitones for the in-key melodies and 4.0 ($SD = 2.3$) for the out-of-key melodies (Mann–Whitney U test, $p = .24$). Occasionally, out-of-key tones were associated with tritone jumps, but for every occurrence of this kind there was another melody in which the in-key tone had a jump of a similar size.

Each of the 12 major keys was used three times ($12 \times 3 = 36$ melodies). The tenor vocal range was selected, C $\sharp 4$ to F 5 (277–698 Hz).

Filler melodies. Each melody consisted of 10–14 notes, was tonal, and resembled the target melodies in style. Eight (roughly one third) of the filler melodies contained an out-of-key tone at some point, and eight contained a loudness manipulation (*auditory anomaly*) to reflect the distribution of the out-of-key tone and loudness increase occurrences in the target materials. The out-of-key/loud note occurred at least five notes into the melody. The pitch range used was the same as that used for creating the target melodies.

Recording the Stimuli

The target and the filler stimuli were recorded in a soundproof room at Stanford's Center for Computer Research in Music and Acoustics. For each experimental item, Regions 1–4 of the in-key subject-extracted condition were recorded first, with each region recorded separately. Then, recordings of Region 2 of the remaining three conditions were made.⁵ (Regions 1, 3, and 4 were recorded only once, because they were identical across all six conditions.) For each filler item, every region was recorded separately. After the recording process was completed, all the recordings were normalized for loudness levels. Finally, the auditory-anomaly conditions were created using the recordings of the critical region of the in-key conditions. In particular, we used the audio manipulation program Audacity (available at audacity.sourceforge.net) to increase the loudness of the last word in the RC by 10 dB, a value chosen on the basis of neuroimaging research indicating that this amount of change in an auditory sequence elicits a mismatch negativity (Jacobsen, Horenkamp, & Schröger, 2003; Näätänen, Pakarinen, Rinne, & Takegata, 2004).

Pilot Work

Before conducting the critical experiment, we tested several participants on the full set of materials. This pilot study was informative in two ways. First, we established that the standard RC extraction

effect (lower performance on object-extracted RCs, as compared with subject-extracted RCs [e.g., King & Just, 1991]) can be obtained in sung stimuli and can therefore be used to investigate the relationship between structural integration in language and music. Second, we discovered that comprehension accuracies were very high, approaching ceiling. As a result, we increased the processing demands in the critical experiment, in order to increase the variance in comprehension accuracies. We reasoned that increasing the processing demands would lower accuracies overall, thereby increasing the range of values and, hence, the sensitivity in this measure. In order to increase the processing demands, we increased the playback speed of the audio files. In particular, we sped up each file by 50%, without changing pitch, using Audacity.

Procedure

The task was self-paced, phrase-by-phrase listening. The experiment was run using the Linger 2.9 software by Doug Rohde (available at tedlab.mit.edu/~dr/Linger). The stimuli were played to the participants via headphones. Each participant heard only one version of each sentence, following a Latin square design (see Appendix B for a complete list of linguistic materials). The stimuli were pseudorandomized separately for each participant.

Each trial began with a fixation cross. Participants pressed the space bar to hear each phrase (region) of the sentence. The amount of time the participant spent listening to each region was recorded as the time between keypresses. A yes/no comprehension question about the propositional content of the sentence (i.e., who did what to whom) was presented visually after the last region of the sentence. Participants pressed one of two keys to respond “yes” or “no.” After an incorrect answer, the word “INCORRECT” flashed briefly on the screen. Participants were instructed to listen to the sentences carefully and to answer the questions as quickly and accurately as possible. They were told to take wrong answers as an indication to be more careful. Participants were not asked to do a musical task. We reasoned that, because of the nature of the materials in the present experiment (sung sentences), it would be very difficult for participants to ignore the music, because the words and the music merged into one auditory stream. We further assumed that musical structure would be processed automatically, on the basis of research showing that brain responses to out-of-key tones in musical sequences occur even when listeners are instructed to ignore music and attend to concurrently presented language (Koelsch et al., 2005). Participants took approximately 25 min to complete the experiment.

RESULTS

Participants answered the comprehension questions correctly 85.1% of the time. Figure 3 presents the mean accuracies across the six conditions.

A two-factor ANOVA with the factors syntactic complexity (subject-extracted RCs, object-extracted RCs) and musical complexity (in key, out of key, auditory anomaly) revealed a main effect of syntactic complexity and an interaction. First, participants were less accurate in the object-extracted conditions (80.8%) than in the subject-extracted conditions (89.4%) [$F_1(1,59) = 11.03$, $MS_e = 6,531$, $p < .005$; $F_2(1,35) = 14.4$, $MS_e = 3,919$, $p < .002$]. Second, the difference between the subject- and the object-extracted conditions was larger in the out-of-key conditions (15.8%) than in the in-key conditions (5.3%) or the auditory-anomaly conditions (4.4%) [$F_1(2,118) = 6.62$, $MS_e = 1,209$, $p < .005$; $F_2(2,70) = 7.83$, $MS_e = 725$, $p < .002$].

We conducted three additional (2×2) ANOVAs, using pairs of musical conditions (in key vs. out of key, in key vs.

auditory anomaly, and out of key vs. auditory anomaly), in order to ensure that the interaction above was indeed due to the fact that the extraction effect was larger in the out-of-key conditions than in the in-key conditions and the auditory-anomaly conditions. The ANOVA in which the two levels of musical complexity were in key versus out of key revealed a main effect of syntactic complexity, such that participants were less accurate in the object- than in the subject-extracted conditions [$F_1(1,59) = 14.96$, $MS_e = 6,685$, $p < .001$; $F_2(1,35) = 21.6$, $MS_e = 4,011$, $p < .001$], and an interaction, such that the difference between the subject- and the object-extracted conditions was larger in the out-of-key versus the in-key conditions [$F_1(1,59) = 10.76$, $MS_e = 1,671$, $p < .005$; $F_2(1,35) = 9.49$, $MS_e = 1,003$, $p < .005$]. The ANOVA in which the two levels of musical complexity were in key versus auditory anomaly revealed a main effect of syntactic complexity—marginal in the participants analysis—such that participants were less accurate in the object- than in the subject-extracted conditions [$F_1(1,59) = 3.25$, $MS_e = 1,418$, $p = .077$; $F_2(1,35) = 4.43$, $MS_e = 851$, $p < .05$]. There were no other effects ($F_s < 1$). Finally, the ANOVA where the two levels of musical complexity were out of key versus auditory anomaly revealed a main effect of syntactic complexity, such that participants were less accurate in the object- than in the subject-extracted conditions [$F_1(1,59) = 12.9$, $MS_e = 6,168$, $p < .002$; $F_2(1,35) = 14.2$, $MS_e = 3,701$, $p < .002$], a marginal effect of musical complexity, such that participants were less accurate in the out-of-key conditions than in the auditory-anomaly conditions [$F_1(1,59) = 3.019$, $MS_e = 510$, $p = .088$; $F_2(1,35) = 3.046$, $MS_e = 306$, $p = .09$], and an interaction, such that the difference between the subject- and the object-extracted conditions was larger in the out-of-key conditions than in the auditory-anomaly conditions [$F_1(1,59) = 8.31$, $MS_e = 1,946$, $p < .01$; $F_2(1,35) = 15.98$, $MS_e = 1,167$, $p < .001$].

This pattern of results is as predicted by the SSIRH: an interaction between syntactic and musical structural complexity and a lack of a similar interaction between

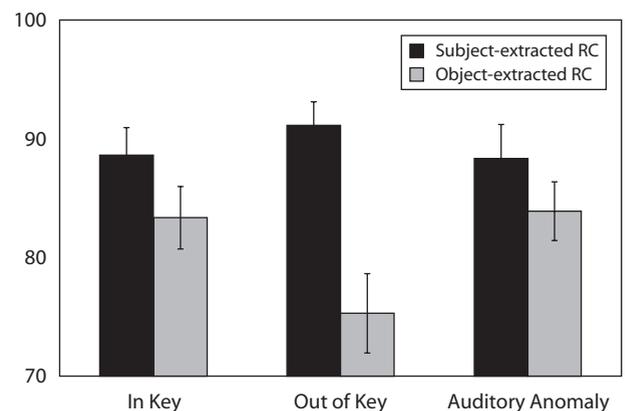


Figure 3. Comprehension accuracies in the six conditions of the experiment. Error bars represent standard errors of the mean.

syntactic complexity and a musical manipulation involving a lower level (not structural) manipulation.

DISCUSSION

We reported an experiment in which participants listened to sung sentences with varying levels of linguistic and musical structural integration complexity. We observed a pattern of results where the difference between the subject- and the object-extracted conditions was larger in the conditions in which musical integrations were difficult than in the conditions in which musical integrations were easy (out of key vs. in key). The auditory-anomaly condition further showed that this interaction was not due to a nonspecific perceptual salience effect in the musical conditions: In particular, the accuracies in this control condition exhibited the same pattern as that in the in-key conditions.

This pattern of results is consistent with at least two interpretations. First, it is possible to interpret these data in terms of an overlap between linguistic and musical integrations in online processing. In particular, it is possible that (1) building more complex structural linguistic representations requires more resources, and (2) a complex structural integration in music interferes with this process due to some overlap in the underlying resource pools. Three possible reasons for not obtaining interpretable effects in the online listening time data are (1) the highly rhythmic nature of the materials; (2) generally longer reaction times in self-paced listening than in the self-paced reading, which may reflect not only the initial cognitive processes, but also some later processes; and (3) the phrase-by-phrase presentation, which does not have very high temporal resolution. Therefore, the online measure used in the experiments reported here may not have been sensitive enough to investigate the relationship between linguistic and musical integrations online.

Second, it is possible to interpret these data in terms of an overlap at the retrieval stage of language processing. In particular, it is possible that (1) there is no competition for resources in the online process of constructing structural representations in language and music (although this would be inconsistent with some of the existing data; see, e.g., Koelsch et al., 2005; Patel et al., 1998), but (2) at the stage of retrieving the linguistic representation from memory, the presence of a complex structural integration in the accompanying musical stimulus makes the process of reconstructing the syntactic dependency structure more difficult.

On the basis of the present data, it is difficult to determine the exact nature of the overlap. However, given that there already exists some evidence suggestive of an overlap between structural processing in language and music during the online stage (e.g., Koelsch et al., 2005; Patel et al., 1998), it is unlikely that the overlap occurs *only* at the retrieval stage. Future work will be necessary to increase our understanding of the nature of the shared structural integration system, especially in the online processing. Evaluating materials like the ones in the present experiment by using temporally fine-grained measures, such as ERPs, is likely to provide valuable insights.

In addition to providing support for the idea of a shared system underlying structural processing in language and music, the results reported here are consistent with several recent studies demonstrating that the working memory system underlying sentence comprehension is not domain specific (e.g., Fedorenko, Gibson, & Rohde, 2006, 2007; Gordon, Hendrick, & Levine, 2002; cf. Caplan & Waters, 1999).

In summary, the contributions of the present work are as follows. First, these results demonstrate that some aspects of structural integration in language and music appear to be shared, providing further support for the SSIRH. Second, this is the first demonstration of an interaction between linguistic and musical structural complexity for well-formed (grammatical) sentences. Third, this work demonstrates that sung materials—ecologically valid stimuli in which music and language are integrated into a single auditory stream—can be used for investigating questions related to the architecture of structural processing in language and music. And fourth, this work provides additional evidence against the claim that linguistic processing relies on an independent working memory system.

AUTHOR NOTE

We thank the members of TedLab, Bob Slevc, and the audiences at the CUNY 2007 conference, the Language and Music As Cognitive Systems 2007 conference, and the CNS 2008 conference for helpful comments on this work. This research was supported in part by the Neurosciences Research Foundation as part of its program on music and the brain at The Neurosciences Institute, where A.P. is the Esther J. Burnham Senior Fellow. We are especially grateful to Jason Rosenberg for composing the melodies and to Stanford's Center for Computer Research in Music and Acoustics for allowing us to record the materials using their equipment and space. Finally, we are grateful to three anonymous reviewers. Send correspondence to E. Fedorenko, MIT 46-3037F, Cambridge, MA 02139 (e-mail: evelina9@mit.edu).

REFERENCES

- AMUNTS, K., SCHLEICHER, A., BURGEL, U., MOHLBERG, H., UYLINGS, H. B. M., & ZILLES, K. (1999). Broca's region revisited: Cytoarchitecture and inter-subject variability. *Journal of Comparative Neurology*, **412**, 319-341.
- BERNSTEIN, L. (1976). *The unanswered question*. Cambridge, MA: Harvard University Press.
- BESSON, M., & FAÏTA, F. (1995). An event-related potential (ERP) study of musical expectancy: Comparison of musicians with nonmusicians. *Journal of Experimental Psychology: Human Perception & Performance*, **21**, 1278-1296.
- BONNEL, A.-M., FAÏTA, F., PERETZ, I., & BESSON, M. (2001). Divided attention between lyrics and tunes of operatic songs: Evidence for independent processing. *Perception & Psychophysics*, **63**, 1201-1213.
- CAPLAN, D. (2007). Experimental design and interpretation of functional neuroimaging studies of cognitive processes. *Human Brain Mapping*. Unpublished appendix, "Special cases: The use of ill-formed stimuli."
- CAPLAN, D., & WATERS, G. S. (1999). Verbal working memory and sentence comprehension. *Behavioral & Brain Sciences*, **22**, 77-126.
- EEROLA, T., HIMBERG, T., TOIVAINEN, P., & LOUHIVUORI, J. (2006). Perceived complexity of western and African folk melodies by western and African listeners. *Psychology of Music*, **34**, 337-371.
- FEDORENKO, E., GIBSON, E., & ROHDE, D. (2006). The nature of working memory capacity in sentence comprehension: Evidence against domain-specific resources. *Journal of Memory & Language*, **54**, 541-553.
- FEDORENKO, E., GIBSON, E., & ROHDE, D. (2007). The nature of working memory in linguistic, arithmetic and spatial integration processes. *Journal of Memory & Language*, **56**, 246-269.

- FISCHL, B., RAJENDRAN, N., BUSA, E., AUGUSTINACK, J., HINDS, O., YEO, B. T. T., ET AL. (2007). Cortical folding patterns and predicting cytoarchitecture. *Cerebral Cortex*, **18**, 1973-1980.
- FRIEDERICI, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences*, **6**, 78-84.
- GIBSON, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, **68**, 1-76.
- GIBSON, E. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. In A. Marantz, Y. Miyashita, & W. O'Neil (Eds.), *Image, language, brain* (pp. 95-126). Cambridge, MA: MIT Press.
- GORDON, P. C., HENDRICK, R., & JOHNSON, M. (2001). Memory interference during language processing. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **27**, 1411-1423.
- GORDON, P. C., HENDRICK, R., & LEVINE, W. H. (2002). Memory-load interference in syntactic processing. *Psychological Science*, **13**, 425-430.
- GRODNER, D., & GIBSON, E. (2005). Consequences of the serial nature of linguistic input for senten[t]ial complexity. *Cognitive Science*, **29**, 261-290.
- GUNTER, T. C., FRIEDERICI, A. D., & SCHRIEFERS, H. (2000). Syntactic gender and semantic expectancy: ERPs reveal early autonomy and late interaction. *Journal of Cognitive Neuroscience*, **12**, 556-568.
- HANDEL, S. (1989). *Listening: An introduction to the perception of auditory events*. Cambridge, MA: MIT Press.
- HURON, D. (2006). *Sweet anticipation: Music and the psychology of expectation*. Cambridge, MA: MIT Press.
- JACOBSEN, T., HORENKAMP, T., & SCHRÖGER, E. (2003). Preattentive memory-based comparison of sound intensity. *Audiology & Neurotology*, **8**, 338-346.
- JANATA, P. (1995). ERP measures assay the degree of expectancy violation of harmonic contexts in music. *Journal of Cognitive Neuroscience*, **7**, 153-164.
- JUCH, H., ZIMINE, I., SEGHER, M. L., LAZEYRAS, F., & FASEL, J. H. D. (2005). Anatomical variability of the lateral frontal lobe surface: Implication for intersubject variability in language neuroimaging. *NeuroImage*, **24**, 504-514.
- KING, J., & JUST, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory & Language*, **30**, 580-602.
- KOELSCH, S., GUNTER, T. C., FRIEDERICI, A. D., & SCHRÖGER, E. (2000). Brain indices of music processing: Nonmusicians are musical. *Journal of Cognitive Neuroscience*, **12**, 520-541.
- KOELSCH, S., GUNTER, T. C., VON CRAMON, D. Y., ZYSSET, S., LOHMANN, G., & FRIEDERICI, A. D. (2002). Bach speaks: A cortical "language-network" serves the processing of music. *NeuroImage*, **17**, 956-966.
- KOELSCH, S., GUNTER, T. C., WITTFOTH, M., & SAMMLER, D. (2005). Interaction between syntax processing in language and in music: An ERP study. *Journal of Cognitive Neuroscience*, **17**, 1565-1577.
- KRUMHANSL, C. L. (1979). The psychological representation of musical pitch in a tonal context. *Cognitive Psychology*, **11**, 346-374.
- LERDAHL, F., & JACKENDOFF, R. (1983). *A generative theory of tonal music*. Cambridge, MA: MIT Press.
- LEVITIN, D. J., & MENON, V. (2003). Musical structure is processed in "language" areas of the brain: A possible role for Brodmann area 47 in temporal coherence. *NeuroImage*, **20**, 2142-2152.
- LEWIS, R. L., & VASISHTH, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, **29**, 375-419.
- LURIA, A. R., TSVETKOVA, L., & FUTER, D. S. (1965). Aphasia in a composer. *Journal of Neurological Science*, **1**, 288-292.
- MACDONALD, M. C., & CHRISTIANSEN, M. H. (2002). Reassessing working memory: Comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review*, **109**, 35-54.
- MAESS, B., KOELSCH, S., GUNTER, T. C., & FRIEDERICI, A. D. (2001). Musical syntax is processed in Broca's area: An MEG study. *Nature Neuroscience*, **4**, 540-545.
- MCDERMOTT, J., & HAUSER, M. D. (2005). The origins of music: Innateness, uniqueness, and evolution. *Music Perception*, **23**, 29-59.
- NÄÄTÄNEN, R., PAKARINEN, S., RINNE, T., & TAKEGATA, R. (2004). The mismatch negativity (MMN): Towards the optimal paradigm. *Clinical Neurophysiology*, **115**, 140-144.
- PATEL, A. D. (2003). Language, music, syntax, and the brain. *Nature Neuroscience*, **6**, 674-681.
- PATEL, A. D. (2008). *Music, language, and the brain*. New York: Oxford University Press.
- PATEL, A. D., GIBSON, E., RATNER, J., BESSON, M., & HOLCOMB, P. J. (1998). Processing syntactic relations in language and music: An event-related potential study. *Journal of Cognitive Neuroscience*, **10**, 717-733.
- PATEL, A. D., IVERSEN, J. R., WASSENAAR, M., & HAGOORT, P. (2008). Musical syntactic processing in agrammatic Broca's aphasia. *Aphasiology*, **22**, 776-789.
- PERETZ, I. (1993). Auditory atonalia for melodies. *Cognitive Neuropsychology*, **10**, 21-56.
- PERETZ, I., & COLTHEART, M. (2003). Modularity of music processing. *Nature Neuroscience*, **6**, 688-691.
- PERETZ, I., KOLINSKY, R., TRAMO, M., LABRECQUE, R., HUBLET, C., DEMEURISSE, G., & BELLEVILLE, S. (1994). Functional dissociations following bilateral lesions of auditory cortex. *Brain*, **117**, 1283-1301.
- STEINBEIS, N., & KOELSCH, S. (2008). Shared neural resources between music and language indicate semantic processing of musical tension-resolution patterns. *Cerebral Cortex*, **18**, 1169-1178.
- STERNBERG, S. (1969). The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica*, **30**, 276-315.
- STROMSWOLD, K., CAPLAN, D., ALPERT, N., & RAUCH, S. (1996). Localization of syntactic comprehension by positron emission tomography. *Brain & Language*, **52**, 452-473.
- TILLMANN, B., JANATA, P., & BHARUCHA, J. J. (2003). Activation of the inferior frontal cortex in musical priming. *Cognitive Brain Research*, **16**, 145-161.

NOTES

1. To the best of our knowledge, there have been no fMRI studies to date comparing structural processing in language and music within individual subjects. In order to claim that shared neural structures underlie linguistic and musical processing, within-individual comparisons are critical, because high degrees of anatomical and functional variability have been reported, especially in the frontal lobes (e.g., Amunts et al., 1999; Fischl et al., 2007; Juch, Zimine, Seghier, Lazeyras, & Fasel, 2005).

2. Unexpected, out-of-key chords in harmonic sequences have been shown to elicit a number of distinct ERP components, including an early negativity (ERAN; latency, ~200 msec) and a later positivity (P600; latency, ~600 msec). The ERAN may reflect the brain's response to the violation of a structural prediction in music, whereas the P600 may index processes of structural integration of the unexpected element into the unfolding sequence.

3. In tonal-harmonic music, out-of-key notes are perceived as psychologically distant from the prevailing tonal center; see Krumhansl (1979) for empirical data and Patel (2003) for a discussion of parallels between integrating cognitively distant elements in linguistic and musical sequences.

4. The first nine words of each sentence (which included the subject noun phrase, the RC, and the main verb phrase) were always monosyllabic and were sung syllabically (1 note per syllable). The last word of the fourth region ["test" in (2)] was monosyllabic in 19/36 items, bisyllabic in 15/36 items, and trisyllabic in 1/36 items. Furthermore, in one item, the fourth region consisted of a trisyllabic word ("yesterday"). For the 16 items in which the fourth region consisted of more than three syllables, the extra syllable(s) were sung on the last (12th) note of the melody, with the beat subdivided among the syllables.

5. In Region 2 (the RC region) of the out-of-key conditions, our singer tried to avoid giving any prosodic cues to upcoming out-of-key tones. Because the materials were sung rather than spoken, the pitches prior to the critical note were determined by the music, which should have helped minimize such cues. In future studies, cross-splicing could be used to eliminate any chance of such cues.

APPENDIX A
Listening Times

Table A1 presents region-by-region listening times (in seconds) in the six conditions. No trimming/outlier removal was performed on these data. (Item 28 contained a recording error, and therefore, it is absent from the listening time data.)

Table A1

In Key/Out of Key/Anomaly	Relative Clause	Region 1		Region 2		Region 3		Region 4	
		Time	<i>SE</i>	Time	<i>SE</i>	Time	<i>SE</i>	Time	<i>SE</i>
In key	Subject extracted	1,344	30	1,931	50	1,593	44	1,858	61
	Object extracted	1,370	32	1,871	38	1,570	31	1,886	62
Out of key	Subject extracted	1,344	29	1,922	41	1,531	33	1,848	59
	Object extracted	1,326	31	1,905	43	1,582	32	1,808	45
Anomaly	Subject extracted	1,369	33	1,899	46	1,533	26	1,991	147
	Object extracted	1,366	30	1,943	42	1,605	38	2,006	128

APPENDIX B
Language Materials

The subject-extracted version is shown below for each of the 36 items. The object-extracted version can be generated as exemplified in (1) below.

1. a. Subject-extracted, grammatical:

The boy that helped the girl got an "A" on the test.

b. Object-extracted, grammatical:

The boy that the girl helped got an "A" on the test.

2. The clerk that liked the boss had a desk by the window.
 3. The guest that kissed the host brought a cake to the party.
 4. The priest that thanked the nun left the church in a hurry.
 5. The thief that saw the guard had a gun in his holster.
 6. The crook that warned the thief fled the town the next morning.
 7. The knight that helped the king sent a gift from his castle.
 8. The cop that met the spy wrote a book about the case.
 9. The nurse that blamed the coach checked the file of the gymnast.
 10. The count that knew the queen owned a castle by the lake.
 11. The scout that punched the coach had a fight with a manager.
 12. The cat that fought the dog licked its wounds in the corner.
 13. The whale that bit the shark won the fight in the end.
 14. The maid that loved the chef quit the job at the house.
 15. The bum that scared the cop crossed the street at the light.
 16. The man that phoned the nurse left his pills at the office.
 17. The priest that paid the cook signed the check at the bank.
 18. The dean that heard the guard made a call about the matter.
 19. The friend that teased the bride told a joke about the past.
 20. The fox that chased the wolf hurt its paws on the way.
 21. The groom that charmed the aunt raised a toast to the parents.
 22. The nun that blessed the monk lit a candle on the table.
 23. The guy that thanked the judge left the room with a smile.
 24. The king that pleased the guest poured the wine from the jug.
 25. The girl that pushed the nerd broke the vase with the flowers.
 26. The owl that scared the bat made a loop in the air.
 27. The car that pulled the truck had a scratch on the door.
 28. The rod that bent the pipe had a hole in the middle.
 29. The hat that matched the skirt had a bow in the back.
 30. The niece that kissed the aunt sang a song for the guests.
 31. The boat that chased the yacht made a turn at the boathouse.
 32. The desk that scratched the bed was too old to be moved.
 33. The cook that hugged the maid had a son yesterday.
 34. The boss that mocked the clerk had a crush on the intern.
 35. The fruit that squashed the cake made a mess in the bag.
 36. The dean that called the boy had a voice full of anger.
-

(Manuscript received May 31, 2006;
revision accepted for publication August 28, 2008.)