Structural Integration in Language and Music: Evidence for a Shared System.

Citation	Fedorenko, Evelina et al. "Structural Integration in Language and Music: Evidence for a Shared System." Memory & Cognition 37.1 (2009) : 1-9.
As Published	http://dx.doi.org/10.3758/mc.37.1.1
Publisher	Springer New York
Version	Author's final manuscript
Citable link	http://hdl.handle.net/1721.1/64631
Terms of Use	Creative Commons Attribution-Noncommercial-Share Alike 3.0
Detailed Terms	http://creativecommons.org/licenses/by-nc-sa/3.0/

Structural integration in language and music:

Evidence for a shared system

Evelina Fedorenko[1], Aniruddh Patel[2], Daniel Casasanto[3],

Jonathan Winawer[3], and Edward Gibson[1]

[1]Massachusetts Institute of Technology, [2]The Neurosciences Institute,

[3]Stanford University

Memory and Cognition, In Press

Send correspondence to the first author:

Evelina (Ev) Fedorenko, MIT 46-3037F, Cambridge, MA 02139, USA

Email: evelina9@mit.edu

Abstract

This paper investigates whether language and music share cognitive resources for structural processing. We report an experiment which used sung materials and manipulated linguistic complexity (subject-extracted relative clauses, object-extracted relative clauses) and musical complexity (short / long harmonic distance between the critical note and the preceding tonal context, auditory oddball involving loudness increase on the critical note relative to the preceding context). The auditory oddball manipulation was included to test whether the difference between short- and long-harmonic-distance conditions may be due to any salient or 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 long-harmonic-distance conditions, compared to the short-harmonic-distance and the auditory-oddball conditions. These results provide evidence for an overlap in structural processing between language and music.

Introduction

The domains of language and music have been argued to 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 where a limited number of basic elements (e.g., words in language, tones and chords in music) can be combined into an infinite number of higher-order structures (e.g., sentences in language, harmonic sequences in music) (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 paper is concerned with the relationship between language processing and music processing at the structural level. Several approaches have been used in the past 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 et al., 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 et al., 1994; Peretz & Coltheart, 2003). 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 et al. (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. concluded that the P600 component indexes the difficulty of structural integration in language 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 et al., 1996), and its right hemisphere homolog (Maess et al., 2001; Koelsch et al., 2002; Levitin & Menon, 2003, 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 (roughly corresponding to the notion of long-term memory) and a system for integrating elements with one

---

[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 a high degree of anatomical and functional variability has been reported, especially in the frontal lobes (e.g., Amunts et al., 1999; Juch et al., 2005; Fischl et al., 2007).

another in the course of on-line processing (roughly corresponding to the notion of working memory) (for an alternative view, see MacDonald & Christensen, 2002, who hypothesize that no distinction exists between representational and processing networks). Patel 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 and tones/chords in music) into evolving structures (sentences in language, harmonic sequences in music). Specifically, it was proposed that structural integration involves rapid and 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 representation reside. This idea can be conceptually diagrammed as shown in Figure 1.

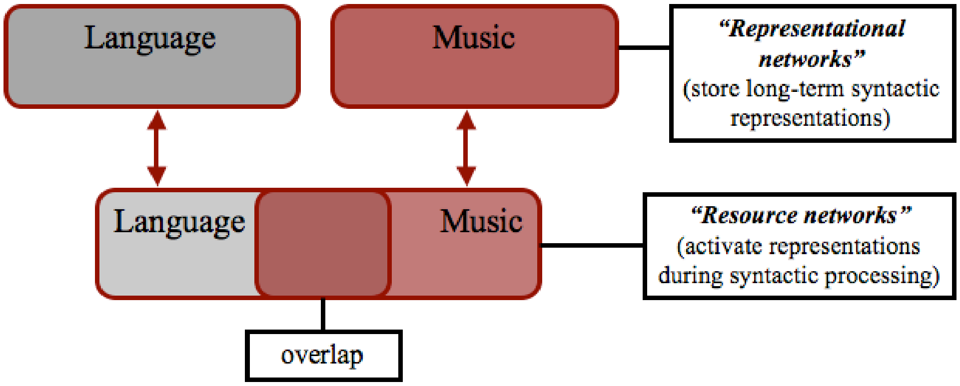

Figure 1. Schematic diagram of the functional relationship between linguistic and musical syntactic processing (adapted from Patel, 2008).

The diagram in Figure 1 represents the hypothesis that linguistic and musical syntactic representations are stored in distinct brain networks (and hence can be selectively damaged), whereas there is overlap in the networks which 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 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 super-additive processing difficulty because of competition for limited resources. A recent study evaluated this prediction empirically and found support for the SSIRH. In particular, Koelsch et al. (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 either be the tonic chord of that key or an unexpected out-of-key chord from a distant key. Previous research on language or music alone had shown that structures with gender agreement errors, like the ones used by Koelsch et al. (2005), elicit a left anterior negativity (LAN), while the musical incongruities elicit an early right anterior negativity (ERAN) (Gunter et al., 2000; Koelsch et al., 2000; Friederici, 2002)[2]. During

---

[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 ms) and a later positivity (P600, latency ~600 ms). The ERAN may reflect the brain's response to the violation of a structural prediction in music, while the P600 may index processes of structural integration of the unexpected element into the unfolding

the critical condition where 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 oddball manipulation involving physically deviant tones on the last word in a sentence. Thus the results of Koelsch et al.'s study provided support for the Shared Syntactic Integration Resource hypothesis.

The experiment reported here is aimed at further evaluating the SSIRH and differs from the experiments of Koelsch et al. in two ways. First, the experiment manipulates linguistic complexity via the use of well-formed sentences (with subject- vs. object-extracted relative clauses). To test claims about overlap in 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 ill-formed in some way. This is because the processing of ill-formed sentences may involve additional cognitive operations (such as error detection and attempts at reanalysis / revision), making the interpretation of language-music interactions more difficult (e.g., Caplan, 2007). Another difference between our experiment and that of Koelsch et al. is the use of sung materials, in which words and music are integrated into a single stream. Song is an ecologically natural stimulus for humans which has been used by other researchers to investigate the relationship between musical processing and linguistic *semantic*

---

sequence.

processing (e.g., Bonnel et al., 2001). To our knowledge, the current study is the first to use song to investigate the relationship between musical processing and linguistic *syntactic* processing.

Experiment

In the experiment described here 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 Shared Syntactic Integration Resource hypothesis is as follows: the condition where both linguistic and musical structural integrations are difficult should be more difficult to process than would be expected if the two effects – the syntactic complexity effect and the musical complexity effect – were independent. This prediction follows from the additive factors logic (Sternberg, 1969; see Fedorenko et al. (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 will therefore only present and discuss 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. The highly rhythmic nature of the materials (see Methods) may have led participants to pace themselves in a way that would allocate the same amount of time to each fragment regardless of condition-type, 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 of structural integration in language has to do with the need to retrieve the structural dependent(s) of an incoming element from memory in cases of non-local structural dependencies. Retrieval difficulty has been hypothesized to depend on the linear distance between the two elements (e.g., Gibson, 1998). We here compared structures containing local vs. non-local dependencies. In particular, we compared sentences containing subject- and object-extracted relative clauses (RCs), as shown in (1).

(1a) Subject-extracted RC: The boy who helped the girl got an "A" on the test.
(1b) Object-extracted RC: The boy who 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 "who helped the girl" contains only local dependencies (between the relative pronoun "who" co-indexed with the head noun "the boy" and the verb "helped", and between the verb "helped" and its direct object "the girl"), while in (1b) the RC "who the girl helped" contains a non-local dependency between the verb "helped" and the pronoun "who". The processing difficulty difference between subject- and object-extracted RCs is therefore plausibly 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., King & Just, 1991; Gibson, 1998, 2000; Gordon et al., 2001; Grodner & Gibson, 2005; Lewis & Vasishth, 2005).

The difficulty of structural integration in music was manipulated by varying the harmonic distance between an incoming tone and the key of the melody, as shown in

Figure 2. Harmonically distant (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 et al., 2006; Huron, 2006). Crucially, the linguistic and the musical manipulations were aligned: the musical manipulation occurred on the last word of the relative clause. This is the point (a) where the structural dependencies in the relative clause have been processed, and (b) which is the locus of processing difficulty in the object-extracted relative clauses 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.

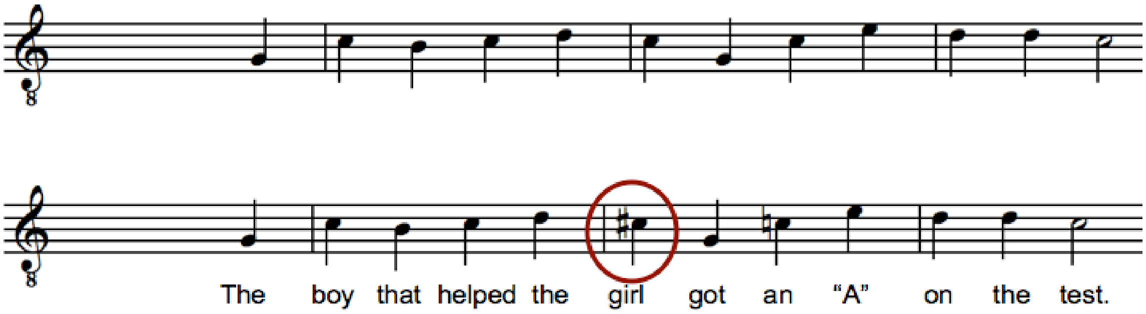

Figure 2. A sample melody (in the key of C major) with a version where all the notes are in-key (top) and a version where the note at the critical position is out-of-key (bottom; the out-of-key note is circled).

As stated above, under concurrent linguistic and musical processing conditions, the SSIRH predicts that linguistic integrations should interact with musical integrations, such that when both types of integrations are difficult, super-additive 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 non-specific acoustically unexpected event. To evaluate this possibility, we added a condition where the melodies had a perceptually salient increase in intensity (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.

Methods

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.

Design and materials The experiment had a 2 x 3 design, manipulating syntactic complexity (subject-extracted RCs, object-extracted RCs) and musical complexity (short harmonic distance between the critical note and the preceding tonal context, long harmonic distance between the critical note and the preceding tonal context, loudness increase on the critical note relative to the preceding context (auditory oddball)).

The language materials consisted of 36 sets of sentences, with two versions as shown in (2). Each sentence consisted of 12 mostly monosyllabic words[3], so that each word corresponded to one note in a melody, and was divided into four regions for the

---

[3] The first nine words of each sentence (which include the subject noun phrase, the relative clause, and the main verb phrase) were always monosyllabic and were sung syllabically (one 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 one a single trisyllabic word ("yesterday"). For the 16 items where 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.

purposes of recording and presentation, as indicated by slashes in (2a)-(2b): (1) a subject noun phrase, (2) an RC (subject-/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 different versions of a melody (short harmonic distance, long harmonic distance, auditory oddball), differing only in the pitch (between short- and long-harmonic-distance conditions) and only in the loudness (between short-harmonic-distance and auditory-oddball conditions) of the note corresponding to the last word of the relative clause (underlined in (2a)-(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 in length, and, like the experimental items, they consisted mostly of monosyllabic words, so that each word corresponded to one note in a melody. The words in the filler sentences were grouped into regions (with each sentence consisting of 3-5 regions and each region consisting of 1-6 words) to resemble the experimental items, which always consisted of 4 regions, as described above.

The musical materials were created in two steps: (1) 36 target melodies (with two versions each) and 25 filler melodies were composed by a professional composer (Jason Rosenberg), and (2) the target and the filler items were recorded by one of the authors – a former opera singer – Daniel Casasanto.

Melody creation

Target melodies

All the 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 above). All the melodies were isochronous: all notes were quarter notes except for the last note, which was a half note. They were sung at a tempo of 120 beats per minute, i.e., each quarter note lasted 500 ms. The first five notes established a strong sense of key. Both the short- and the long-harmonic-distance versions of each melody were in the same key and differed by one note. The critical (6th) note – falling on the last word of the relative clause – was either in-key (short-harmonic-distance conditions) or out-of-key (long-harmonic-distance conditions). It always was on the downbeat of second full bar. When the note was out-of-key, it was one of the five possible non-diatonic notes (C#, D#, F#, G#, A# in C major). Sometimes out-of-key notes were only different by a semi-tone (e.g., C vs. C#).

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 out-of-key notes were not odd in terms of voice leading compared to the in-key notes. In particular, the mean and standard deviation for the size of pitch jumps leading to the critical note was 2.1 (1.9) semitones for the in-key melodies and 2.5 (1.7) semitones for the out-of-key melodies (Mann-Whitney U test, p=.36). The mean and standard deviation for the size of pitch jumps leading from the critical note was 3.4 (2.3) semitones for the in-key melodies and 4.0 (2.3) for the out-of-key melodies (Mann-Whitney U test, p=.24). Out-of-key notes were occasionally associated with tritone jumps, but for every occurrence of this kind there was another melody where the in-key note had a jump of a similar size.

All 12 major keys were used three times (12 x 3 = 36 melodies). The lowest pitch used was C#4 (277 Hz), and highest was F5 (698 Hz). The range was designed for a tenor.

Filler melodies

All the melodies consisted of 10-14 notes, were tonal and resembled the target melodies in style. 8 (roughly third) of the filler melodies contained an out-of-key note at some point in the melody, and 8 contained an intensity manipulation, to reflect the distribution of the out-of-key note and intensity 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 short-harmonic-distance subject-extracted condition were recorded first, with each region recorded separately. Then, recordings of Region 2 of the remaining three conditions were made[4]. (Regions 1, 3 and 4 were only recorded once, since they were identical across the six conditions of the experiment.) For each filler item, every region was also recorded separately. After the recording process was completed, all the recordings were normalized for intensity (loudness) levels. Finally, the auditory-oddball conditions were created using the recordings of the critical region of the short-harmonic-distance conditions. In particular, the intensity (loudness) level of the last word in the RC was increased by 10 dB based on neuroimaging research indicating that this amount of change in an auditory sequence elicits a mismatch negativity (Jacobsen et al., 2003; Näätänen et al., 2004).

Pilot work

Prior to conducting the critical experiment, we conducted a pilot study in which 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, compared to 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. And second, we discovered that comprehension accuracies were very high, close to ceiling. As a result, we decided to increase the processing demands in the critical experiment, in order to increase the variance in comprehension accuracies. We reasoned that increasing the processing demands would lead to overall lower accuracies, thereby increasing the range of values

---

[4] In Region 2 (the RC region) of the long-harmonic-distance conditions, our singer tried to avoid giving any prosodic cues to upcoming out-of-key notes. Since the materials were sung rather than spoken, the pitches prior to the critical note were determined by the music, which should help minimize such prosodic cues. In future studies, cross-splicing could be used to eliminate any chance of such cues.

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, every audio file was sped up by 50% without changing pitch, using an audio-file manipulation program Audacity (available at http://audacity.sourceforge.net/).

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 http://tedlab.mit.edu/~dr/Linger/). The stimuli were presented 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 pseudo-randomized separately for each participant.

Each trial began with a fixation cross. Participants pressed the spacebar 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 key-presses. 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 current experiment (sung sentences), it would be very difficult for participants to ignore the music, because the words and the music are merged into one auditory stream. We further assumed that musical structure would be processed automatically, based on 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 minutes 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.

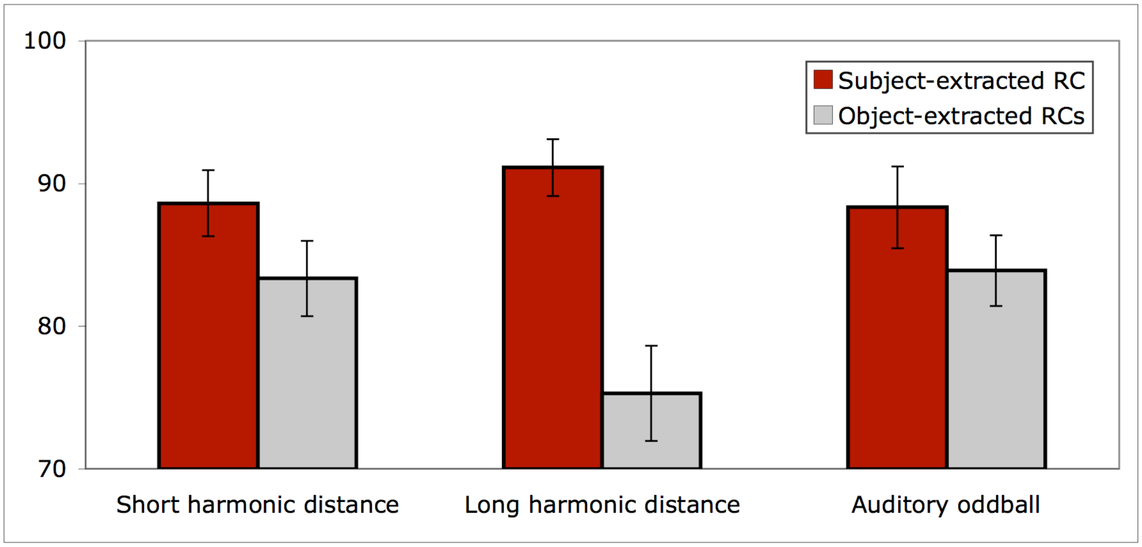

Figure 3. Comprehension accuracies in the six conditions of the experiment. Error bars represent standard errors of the mean.

A two-factor ANOVA with the factors (1) syntactic complexity (subject-extracted RCs, object-extracted RCs) and (2) musical complexity (short harmonic distance, long harmonic distance, auditory oddball) revealed a main effect of syntactic complexity and an interaction. First, participants were less accurate in the object-extracted conditions (80.8%), compared to the subject-extracted conditions (89.4%) (F1(1,59)=11.03; MSe=6531; p<.005; F2(1,35)=14.4; MSe=3919; p<.002). And second, the difference between the subject- and the object-extracted conditions was larger in the long-harmonic-distance conditions (15.8%), compared to the short-harmonic-distance conditions (5.3%) or the auditory-oddball conditions (4.4%) (F1(2,118)=6.62; MSe=1209; p<.005; F2(2,70)=7.83; MSe=725; p<.002).

We further conducted three additional (2 x 2) ANOVAs, using pairs of musical conditions (short- vs. long-harmonic-distance conditions, short-harmonic-distance vs. auditory-oddball conditions, and long-harmonic-distance vs. auditory-oddball conditions), in order to ensure that the interaction above is indeed due to the fact that the extraction effect was larger in the long-harmonic-distance conditions, compared to the short-harmonic-distance conditions and the auditory-oddball conditions. The ANOVA where the two levels of musical complexity were short vs. long harmonic distance revealed a main effect of syntactic complexity, such that participants were less accurate in the object-extracted conditions, compared to the subject-extracted conditions (F1(1,59)=14.96; MSe=6685; p<.001; F2(1,35)=21.6; MSe=4011; p<.001), and an interaction, such that the difference between the subject- and the object-extracted conditions was larger in the long-harmonic-distance conditions, compared to the short-harmonic-distance conditions (F1(1,59)=10.76; MSe=1671; p<.005; F2(1,35)=9.49; MSe=1003; p<.005). The ANOVA where the two levels of musical complexity were short harmonic distance vs. auditory oddball revealed a main effect of syntactic complexity – marginal in the participants analysis – such that participants were less accurate in the object-extracted conditions, compared to the subject-extracted conditions (F1(1,59)=3.25; MSe=1418; p=.077; F2(1,35)=4.43; MSe=851; p<.05). There were no other effects (Fs<1). Finally, the ANOVA where the two levels of musical complexity were long harmonic distance vs. auditory oddball revealed a main effect of syntactic complexity, such that participants were less accurate in the object-extracted conditions, compared to the subject-extracted conditions (F1(1,59)=12.9; MSe=6168; p<.002; F2(1,35)=14.2; MSe=3701; p<.002), a marginal effect of musical complexity, such that participants were less accurate in the long-harmonic-distance conditions, compared to the auditory-oddball conditions (F1(1,59)=3.019; MSe=510; p=.088; F2(1,35)=3.046; MSe=306; p=.09), and an interaction, such that the difference between the subject- and the object-extracted conditions was larger in the long-harmonic-distance conditions, compared to the auditory-oddball conditions (F1(1,59)=8.31; MSe=1946; p<.01; F2(1,35)=15.98; MSe=1167; p<.001).

This pattern of results – an interaction between syntactic and musical structural complexity, and a lack of a similar interaction between syntactic complexity and a musical manipulation involving a lower-level (not structural) manipulation – is as predicted by the Shared Syntactic Integration Resource hypothesis.

General 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 where musical integrations were difficult compared to the conditions where musical integrations were easy (long- vs. short-harmonic-distance conditions). The auditory-oddball condition further showed that this interaction was not due to a non-specific perceptual saliency effect in the musical conditions: in particular, the accuracies in this control condition exhibited the same pattern as the short-harmonic-distance 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 on-line 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 on-line listening time data are: (a) the highly rhythmic nature of the materials; (b) generally longer reaction times in self-paced-listening, compared to the self-paced-reading, which may reflect not only the initial cognitive processes, but also some later processes; and (c) 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 on-line process of constructing structural representations in language and music (although this would be inconsistent with some of the existing data (e.g., Patel et al., 1998; Koelsch et al., 2005)), 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.

Based on the current data, it is difficult to determine the exact nature of the overlap. However, given that there already exists some suggestive evidence for an overlap between structural processing in language and music during the on-line stage (e.g., Patel et al., 1998; Koelsch et al., 2005), it is unlikely that the overlap occurs *only* at the retrieval stage. Future work will be necessary to better understand the nature of the shared structural integration system, especially in the on-line processing. Evaluating materials like the ones used in the current experiment 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., Gordon et al., 2002; Fedorenko et al., 2006, 2007; c.f. Caplan & Waters, 1999).

In summary, the contributions of the current work are as follows. First, these results demonstrate that there are some aspects of structural integration in language and music that appear to be shared, providing further support for the Shared Syntactic Integration Resource hypothesis. 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.

References

Amunts, 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 Univ. Press.

Besson, M. & Faïta, F. (1995). An event-related potential (ERP) study of musical expectancy: Comparison of musicians with non-musicians. *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 and 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. *Behavorial and Brain Sciences,* 22, 77-94.

Eerola, T., Himberg, T., Toiviainen, P., & Louhivuori, J. (2006). Perceived complexity of western and African folk melodies by western and African listeners. *Psychology of Music,* 34(3).

Fedorenko, E., Gibson, E. & Rohde, D. (2006). The Nature of Working Memory Capacity in Sentence Comprehension: Evidence against Domain Specific Resources. *Journal of Memory and Language*, 54(4).

Fedorenko, E., Gibson, E. & Rohde, D. (2007). The Nature of Working Memory in Linguistic, Arithmetic and Spatial Integration Processes. *Journal of Memory and Language,* 56(2).

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Science,* 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: Y. Miyashita, A. Marantz, & 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. *Cognitive Science*, Vol. 29 (2), 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.* MIT Press. Cambridge, MA.

Huron, D. (2006). *Sweet Anticipation: Music and the Psychology of Expectation.* Cambridge, MA: MIT Press.

Jacobsen, T., Horenkamp, T., & Erich Schroger, T. (2003). Preattentive Memory-Based Comparison of Sound Intensity. *Audiology & Neuro-Otology,* 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., Seghier, M.L., Lazeyras, F., & Fasel, J.H. (2005). Anatomical variability of the lateral front lobe surface: implication for intersubject variability in language neuroimaging. *NeuroImage,* 24, 504-514.

King & Just (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language,* 30, 580-602.

Koelsch, S., Gunter, T. C., Friederici, A. D., & Schröger, E. (2000). Brain indices of music processing: 'non-musicians' 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., Wittforth, M., & Sammler, D. (2005). Interaction between syntax processing in language and music: An ERP study. *Journal of Cognitive Neuroscience,*17:1565-1577.

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, 2142-52.

Lewis, R. & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*.

Luria, A, Tsvetkova, L., & Futer, J. (1965). Aphasia in a composer. *J. Neurol. Sci.* 2, 288-292.

MacDonald, M. & Christiansen, M. (2002). Reassessing Working Memory: Comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review,* 109 (1), 35-54.

Maess, B., Koelsch, S., Gunter, T., & 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. *Clin Neurophysiology,* 115:140-4.

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 Univ. Press.

Patel, A.D., Gibson, E., Ratner, J., Besson, M. & Holcomb, P. (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., Kolinsky, R., Tramo, M., Labrecque, R., Hublet, C., Demeurisse, G., & Belleville, S. (1994). Functional dissociations following bilateral lesions of auditory cortex. *Brain,* 117, 1283-1302 .

Peretz, I., & Coltheart, M. (2003) Modularity of music processing. *NatureNeuroscience,* 6, 688-691.

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. In Koster, W. G. (Ed.), *Attention and performance II.* Amsterdam: North-Holland (Reprinted from Acta Psychologica, 30, 276-315 (1969).)

Stromswold, K., Caplan, D., Alpert, N., Rauch, S. (1996). Localization of syntactic comprehension by positron emission tomography. *Brain & Lang*, 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.

Acknowledgments

We would like to thank the members of TedLab, Bob Slevc, and the audiences at the *CUNY 2007* conference, the *Language & Music as Cognitive Systems 2007* conference, and the *CNS 2008* conference for helpful comments on this work.

Supported in part by Neurosciences Research Foundation as part of its program on music and the brain at The Neurosciences Institute, where ADP is the Esther J. Burnham Senior Fellow.

We are also especially grateful to Jason Rosenberg for composing the melodies used in the two experiments, as well as 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.

Appendix A

Listening times

The table below presents region-by-region listening times in the six conditions (standard errors in parentheses). No trimming / outlier removal was performed on these data. [One item (#28) contained a recording error, and therefore, it is absent from the listening time data.]

	Reg 1	Reg 2	Reg 3	Reg 4
Short HD / Subj-extr. RC:	1344 (30)	1931 (50)	1593 (44)	1858 (61)
Short HD / Obj-extr. RC:	1370 (32)	1871 (38)	1570 (31)	1886 (62)
Long HD / Subj-extr. RC:	1344 (29)	1922 (41)	1531 (33)	1848 (59)
Long HD / Obj-extr. RC:	1326 (31)	1905 (43)	1582 (32)	1808 (45)
Oddball / Subj-extr. RC:	1369 (33)	1899 (46)	1533 (26)	1991 (147)
Oddball / Obj-extr. RC:	1366 (30)	1943 (42)	1605 (38)	2006 (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.