



Prosodic Structure as a Parallel to Musical Structure

Christopher C. Heffner^{1,2,3*} and L. Robert Slevc^{1,4}

¹ Program in Neuroscience and Cognitive Science, University of Maryland, College Park, MD, USA, ² Department of Linguistics, University of Maryland, College Park, MD, USA, ³ Department of Hearing and Speech Sciences, University of Maryland, College Park, MD, USA, ⁴ Department of Psychology, University of Maryland, College Park, MD, USA

What structural properties do language and music share? Although early speculation identified a wide variety of possibilities, the literature has largely focused on the parallels between musical structure and syntactic structure. Here, we argue that parallels between musical structure and prosodic structure deserve more attention. We review the evidence for a link between musical and prosodic structure and find it to be strong. In fact, certain elements of prosodic structure may provide a parsimonious comparison with musical structure without sacrificing empirical findings related to the parallels between language and music. We then develop several predictions related to such a hypothesis.

OPEN ACCESS

Edited by:

Corianne Rogalsky,
Arizona State University, USA

Reviewed by:

Mireille Besson,
Centre National de la Recherche
Scientifique – Institut
de Neurosciences Cognitives de la
Méditerranée, France
Erin E. Hannon,
University of Nevada, Las Vegas, USA
Diana Omigie,
Max Planck Institute for Empirical
Aesthetics, Germany

*Correspondence:

Christopher C. Heffner
heffner@umd.edu

Specialty section:

This article was submitted to
Auditory Cognitive Neuroscience,
a section of the journal
Frontiers in Psychology

Received: 17 August 2015

Accepted: 07 December 2015

Published: 22 December 2015

Citation:

Heffner CC and Slevc LR (2015)
Prosodic Structure as a Parallel
to Musical Structure.
Front. Psychol. 6:1962.
doi: 10.3389/fpsyg.2015.01962

Keywords: musical structure, prosody, prosodic structure, music perception, speech perception

INTRODUCTION

Uncovering structural similarities between language and music has been an objective that has tantalized cognitive scientists for years. Music and spoken language both consist of complex auditory signals that are organized in line with an underlying structure. But which linguistic structures are musical structures parallel to? In their seminal work proposing a generative account of musical structure, Lerdahl and Jackendoff (1983, p. 330) concluded that “the similarity between prosodic and music structure can be used as a point of triangulation for approving an account of other temporally structured cognitive capacities.” In other words, they argued that musical structure and prosodic structure (patterns of rhythm, pitch, and tempo in speech) have convincing parallels, worthy of study and comparison to each other. Although it is certainly the case that subsequent investigations of the links between phonetic processing and musical structure have often relied on the idea of a music/prosody connection (Schön et al., 2004; Moreno et al., 2009; Besson et al., 2011), most subsequent work comparing musical structure to language has focused on comparisons with linguistic syntax (i.e., the combination of words into sentences) and not prosodic structure.

Here, we revisit Lerdahl and Jackendoff’s (1983) speculations about the similarities between the structure of music and the structure of prosody. We first review the literature connecting music and linguistic syntax, finding that evidence for certain parallels between the two modalities (especially with regard to recursion) surprisingly sparse. We suggest that this opens up opportunities to examine prosodic structure as another fitting analog to musical structure. We describe prosodic structure, highlighting the ways that it differs from syntactic structure, as well as its many similarities. We suggest that comparing musical structure to prosody does not require discarding syntactic parallels, and highlight some comparative insights that could be gained by additionally examining prosody. Taking prosodic structure seriously, we argue, also helps tie together a few disparate streams within the experimental literature that are currently unrelated. Finally, we make

some predictions for future studies that would be in line with a prosodic account of musical structure.

Although the purpose here is to accentuate the parallels between musical structure and prosodic structure, we do not mean to imply that musical processing does not have multiple parallels to language processing. It is certainly the case that comparisons between music and other linguistic systems are useful. However, in this review, we do seek to point out some of the ways that it is unfortunate that the use of prosodic structure to understand musical structure has fallen by the wayside. Studies that have analogized musical structure to syntactic structure often serve as a useful counterpoint in these discussions, particularly concerning the questions that are left unresolved or incomplete by a syntactic analysis of musical structure.

MUSIC AND SYNTAX

Evidence for Syntactic Parallels

Most of the discussion surrounding structural parallels between music and language has focused on using theories of syntactic structure to bridge the two fields. This may have something to do with the prominence of syntax within the language sciences, where it is undoubtedly accorded a place of honor. Indeed, although the idea is not shared by every language scientist, the recursive structure of syntax has been argued to be the attribute that separates human language from other communication systems (Hauser et al., 2002). Syntactic recursion allows syntactic structures to be embedded inside other syntactic structures, which can, in turn, be embedded inside others. This embedding process creates a hierarchical structure where certain parts of the structure are privileged over others and sometimes reduplicated at higher levels. Syntactic recursion is a powerful property because it allows speakers to potentially utter an infinite number of sentences.

Both music theorists and language scientists who have music-related interests agree that aspects of music can be described in terms of recursive formalisms similar to syntactic structure (although we note that this discussion has centered almost entirely on Western tonal music). These syntactic structures can be used to assemble a generative grammar of music. Such generative grammars for musical sequences have been proposed for jazz chord sequences (Steedman, 1984), tonal harmony (Lerdahl, 2001b; Rohrmeier, 2011), and rhythmic structure (Longuet-Higgins, 1976), and have often reflected contemporary conceptions of linguistic syntax (Katz and Pesetsky, 2009). Such ideas may have emerged in part from a long tradition of recursive structural analyses of music and from the influence of recursion on musical composition (Forte, 1959; Larson, 1998). This historical framework may similarly have influenced the argument that syntactic processes play a central role in the cognitive processing of music (e.g., Patel, 2003, 2008; Koelsch, 2005).

Indeed, there is evidence for music/syntax relationships in the experimental literature. For example, rhythm perception can predict syntactic skills among typically developing 6-year-olds (Gordon et al., 2015; see also Jentschke and Koelsch,

2009). Other work suggests similarities between the neural correlates of musical and linguistic structural manipulations; that is, event-related potential (ERP) components associated with syntactic phrase violations have often been uncovered in electroencephalography (EEG) studies that have investigated musical phrase structure (Patel et al., 1998a; Koelsch, 2005).

This is true not just for short-distance musical dependencies, but also ones that are spread across multiple musical phrases. For example, Koelsch et al. (2013) found EEG evidence for listeners' sensitivity to hierarchical structure by manipulating passages taken from two Bach chorales. Each chorale originally had an ABA structure, with information in the last phrase resolving musical structures left incomplete by the initial "A" phrase. In their experiment, Koelsch et al. (2013) contrasted the original chorales with modified versions with a CBA structure; that is the first "A" phrase was pitch-modified such that the final passage did not resolve the musical structures left incomplete in the initial passage. Because the pitch manipulations involved were performed on the initial phrase, not on the final chord, any ERP response differences to the final chord could not have been the result of an acoustic difference between the two stimuli. There were indeed ERP components that differentiated the two conditions, which they interpreted as evidence for sensitivity to hierarchical syntactic structure in music.

Other work localizes the neural correlates of musical structural violations to regions often implicated in syntactic structural violations, such as Broca's area, although such parallels may depend on the task demands of the experiments in question (LaCroix et al., 2015). Nevertheless, neural correlates of musical structural violations have also been found to resemble those associated with syntactic structural violations in both magnetoencephalography (MEG; Maess et al., 2001; Koelsch and Friederici, 2003) and functional magnetic resonance imaging (fMRI) studies (Tillmann et al., 2006; Oechslin et al., 2013; Seger et al., 2013).

Challenges for Syntactic Parallels

While this work supports a relationship between musical structure and linguistic syntax, there are both theoretical and empirical reasons to question the idea that musical structure and syntactic structure in language are fundamentally similar. First, some empirical work has failed to find overlapping regions associated with structural manipulations in music and language within individual participants (Fedorenko et al., 2011, 2012; Rogalsky et al., 2011). Fedorenko et al. (2011), for example, presented participants with sentences and lists of non-words (both presented word-by-word) and identified regions within individual participants that selectively responded more to sentences than to non-word lists. Many of these regions were in left frontal and temporal cortices (e.g., left inferior frontal gyrus, left anterior temporal cortex). They then had participants listen to music and scrambled music, and found that the regions that showed significant difference in activation between sentences and non-word lists were not consistently more activated by music more than scrambled music. Fedorenko et al. (2012) performed essentially the converse experiment. First, they identified regions within individual participants that were sensitive to music over

scrambled music; for most participants, this included bilateral temporal cortex and supplementary motor area. They then had participants view sentences, word lists and non-word lists. They found that the regions that were more activated by music than scrambled music did not show differential activation to each individual kind of sentence stimulus.

More generally, many fMRI studies of syntax/music connections suffer from challenges in differentiating shared patterns of fMRI activation and shared underlying circuitry (see Peretz et al., 2015, for discussion). Furthermore, the evidence for interactive effects between musical and syntactic manipulations (Carrus et al., 2013; Kunert and Slevc, 2015) is complicated by evidence for interactions between musical structure and non-syntactic linguistic variables (Perruchet and Poulin-Charronnat, 2013) and even between musical structure and non-linguistic processes (e.g., Escoffier and Tillmann, 2008).

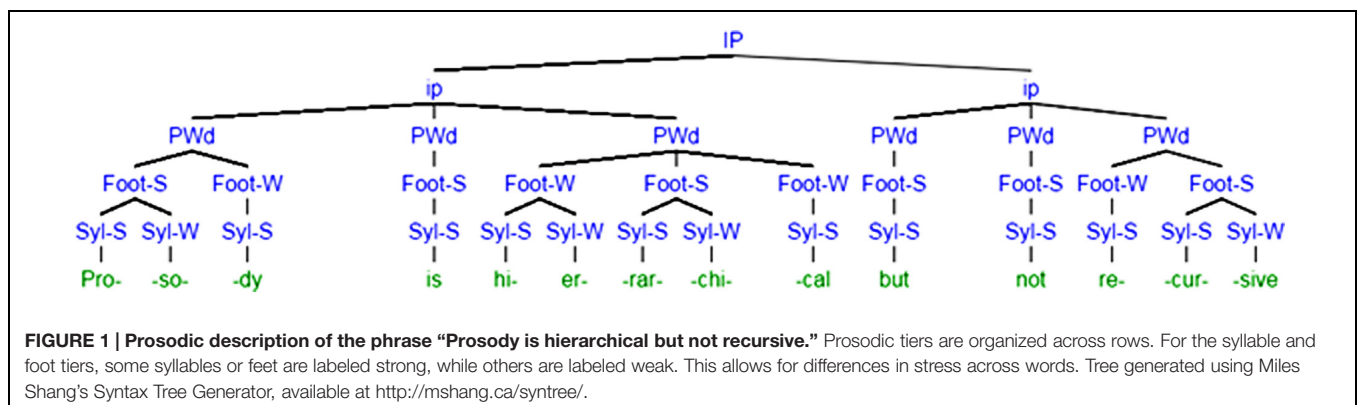
Finally, and perhaps more importantly, we argue that the evidence in favor of recursion being used in musical processing is scant. On the face of it, this may seem like a surprising claim. Composers, after all, do often include higher-level structure in their pieces; consider the rondo in many concerti (with an ABACABA structure) or variations on verse-and-chorus structure in popular music. Yet the fact that these structures exist does not necessarily mean that listeners are sensitive to them in day-to-day listening situations. Studies that have carefully investigated the perception of this long-distance structure have indicated that listeners, even ones with musical training, may simply ignore long-distance structural dependencies (Cook, 1987). For example, when listeners were asked complete a “musical puzzle”—that is, to organize parts of a minuet into a single, coherent piece—listeners often violated the typical structural relationships between each component part when constructing a piece (Tillmann et al., 1998). Although findings have shown evidence for some implicit learning of structural properties of language within a short time-window, the evidence for long-distance processing of dependencies is much weaker (Tillmann and Bigand, 2004). Studies that have assessed such processing have generally not used implicit measures that might indicate implicit learning of those sequences, however. Such findings suggest that listeners do not routinely extract recursive structure from a musical signal without explicit, post facto scrutiny.

These findings that listeners are not typically aware of long-distance musical structure conflict with other evidence taken to show recursive processing of musical structure. For example, Koelsch et al. (2013) found that participants’ EEG responses differentiated regular from irregular resolutions of long-distance dependences in music (as described above). However, it may be that these data do not actually reflect recursive processing of musical structure, but instead reflect listeners’ expectations about the timing and pitch information found within a piece rather than something fundamental about musical structure. This is plausible because, in Koelsch et al.’s (2013) experiment, participants listened to regular and irregular versions of only two different Bach chorales (transposed into each major key) 60 times each, presented in a randomized order. By the end of the experiment, it was likely that the rhythmic and pitch properties of these pieces became very well learned. Participants could therefore use information (explicitly or implicitly) early on within the chorale to predict future musical patterns. Alternatively, the listeners might simply be using the time they had been spending listening to the renditions of each version of each chorale to think about the music in a more abstract (and potentially recursive) way. Thus, these results are not inconsistent with the idea that listeners do not normally engage in recursive processing of musical structure during typical listening.

PROSODIC PARALLELS

Structural Description

Given these challenges for syntactic parallels to musical structure, we argue here that prosodic structure may better parallel the structure of music. The first thing to observe about prosodic structure is that it, like syntactic structure, is hierarchical. Indeed, a reader merely glancing at the prosodic tree in **Figure 1** should be forgiven for thinking that it is a syntactic tree. However, rather than phrases being different parts of speech, each unit is instead a level on the prosodic hierarchy. From bottom up, syllables are combined to form feet, which are combined to form prosodic words, which in turn are combined to form minor and major intonational phrases. This particular tree is not meant to be reflective of any single school of thought related to the



organization of prosodic structure, of which there are many. A discussion of these differences is outside the scope of this paper, but almost all agree on the details of prosodic structure we bring up here. Rather, we intend the description below to be an approximate sketch of commonalities between schools of thought within prosody. Interested readers should refer to Shattuck-Hufnagel and Turk (1996), which remains the best explanation of the prosodic hierarchy to novices.

From the very base of the hierarchy (phonemes, or speech sounds), smaller units are organized into larger groupings. These groupings in turn form larger ones. And the smallest groupings can be said to “belong” to higher level groupings by way of those intermediate levels. Some of these levels might be familiar, such as syllables and prosodic words (which usually, though imperfectly, correspond to an intuitive notion of “word”). Others might be unfamiliar outside of the realm of poetry, like the “foot”, which syllables combine together to create. But crucially, each of these structures plays a role in organizing the timing, pitch, and volume characteristics of the speech signal. In many languages, for example, syllables gain stress through an increase in some combination of volume, pitch, and length on the stressed syllable (Gay, 1978). The ends of intonational phrases and prosodic words are characterized by the lengthening of phrase-final material (Wightman et al., 1992; Byrd et al., 2006). Even newborn infants seem capable of discriminating short utterances with prosodic boundaries from short utterances without them (Christophe et al., 2001).

Similarly, within music, notes can associate to form beats, measures, and phrases, each of which has consequences for the timing, pitch, and volume characteristics of the notes in question. The structure and expression of musical phrases is sometimes referred to as “musical prosody” (see, e.g., Palmer and Hutchins, 2006) and shows many parallels with linguistic prosody; for example, phrase-final lengthening is found in music as well as speech (Large and Palmer, 2002). These predictable regularities in pitch, loudness, and duration information are tracked in real-time, and have been associated with an ERP component known as the “closure positive shift” (Steinhauer et al., 1999; Steinhauer and Friederici, 2001). This prosodic boundary-making can even sometimes lead listeners to misparse the syntactic content of sentences, as revealed, for example, through ERP components that typically manifest in response to syntactic anomalies (Steinhauer et al., 1999), causing some to claim a central role for prosody in sentence comprehension (Frazier et al., 2006).

The example in **Figure 1** shows that, besides being in the onset of the syllable “hi,” the phoneme /h/ within the word “hierarchical” also belongs to the word-initial foot “hier,” the prosodic word “hierarchical,” the minor intonational phrase “Prosody is hierarchical,” and to the major intonational phrase “Prosody is hierarchical but not recursive.” There are many prosodic structures that a syllable can map onto, but the mapping is essentially exhaustive. Like in prosody, syntactic structure is often posited in tree-like terms, with units contained in other units. Like syntax, too, it is generally non-overlapping. The sound /h/ does not also, say, belong to the end of the previous syllable, /ɪz/. Nor do phrases within each level of the

hierarchy cross, with /h/ belonging to the “hi” syllable but not to the “is” foot.¹ Again, this echoes syntactic structure, which has similar constraints against units that straddle higher-level boundaries.

The primary difference in description of prosodic and syntactic structure is in terms of the non-recursive nature of prosodic constituents. Recursion in syntax is, without a doubt, important; to some, it is absolutely essential (Hauser et al., 2002). Within syntax, sentences can be nested inside of sentences, adjectives inside other adjective phrases. This allows for the great complexity of human speech; for example, an infinite number of utterances can be created and comprehended by listeners who are perfectly capable of understanding sentences that have embedding. Recursion requires comprehenders to keep track of a wide variety of current dependencies and possible future dependencies based on the signal that they have heard so far, and these dependencies might be of exactly identical style. Prosody, in contrast, has no such recursive elements to it. There is no sense in which a prosodic word can be embedded within another prosodic word, or where a syllable is separated from another syllable by a few levels within hierarchy; one level of the structure is not duplicated anywhere else. This does not necessarily make prosodic structure any less rich. It simply makes it less self-embedded.

Prosodic Correspondences with Musical Structure

Prosodic structure shares some useful commonalities with musical structure. Like prosody, it includes phrases across a variety of time scales, with shorter-duration units nested inside larger ones. These phrases have more or less prominent units, and within phrases there are often restrictions on which segments can co-occur. And, in both cases, the structures that are imposed on the signal are capable of certain prescribed exceptions.

Take meter as an example (Lerdahl, 2001a; Jackendoff, 2009). Within musical theory, meter refers to regularly recurring sequences of more or less prominent beats across time. Within prosodic theory, meter refers to more or less prominent (i.e., stressed or unstressed) syllables across time, usually in terms of the organization of syllables into feet and feet into words. In both modalities, there is a typical rhythm that manifests itself across time, either within a piece (for music) or within a language (for prosody). For example, in common time (or 4/4 time) in music, it is often the first and third beats within a single measure that are the strongest, with the first beat receiving the most intense accent. However, a composer or musician has the option of subverting the typical metrical properties of a piece by placing accents on individual notes besides the first and third beats, or by using syncopation. This can sometimes be exploited to better sync the metrical properties of music with the metrical properties of language, which in turn leads to certain favorable behavioral

¹“Ambisyllabic” consonants are a proposed exception to this rule (Schiller et al., 1997), and under some theories a lower-level item (such as a single phoneme) might not be assigned to an individual syllable. However, their psychological status remains unclear. It is worth noting that notes that seem to straddle two phrases are also found sporadically in music (Lerdahl and Jackendoff, 1983).

and electrophysiological outcomes with simple language tasks (Gordon et al., 2011). Similarly, typical metrical prominences and organizations vary in their manifestation across the world's musical traditions. Balinese gamelan, for example, keeps track of the metrical beat through the use of an instrument called the *kajar*, which unambiguously (or, at least, unambiguously to the other gamelan players) sets out the location of each beat (McGraw, 2008).

In prosody, meanwhile, languages differ in how stress is assigned, but almost all show regular patterns; for example, they might assign stress to every other syllable starting from the last syllable within a word. Many languages allow for exceptions in the typical assignment of stress to syllables. Although all languages share some underlying structural properties, languages also have quite a bit of leeway in the realizations of that structure. In Czech and Polish, stress assignment is without exception (word-initial for Czech, penultimate for Polish), just as the *kajar* unambiguously determines the beat in gamelan music; in German, stress depends on the “weight” of the syllable in question, but tends to fall within the last three syllables; in Russian, it can occur anywhere at all. Thus, the two systems evince very similar properties to each other, including flexibility in broader organizational principles across traditions.

There is evidence that musical composers, and singers, are implicitly aware of these organizational parallels. The study of textsetting (the process of assigning syllables to individual notes in sung music) provides a useful guide. Textsetting has been described using a generative model that attempted “to achieve the best possible fit between the abstract structures representing rhythmic and linguistic surfaces” (Halle and Lerdahl, 1993, p. 7). Musical compositions seem to respect the structures of prosody composition in determining which notes are assigned which durations or which levels of accent (Palmer and Kelly, 1992). And singers spontaneously asked to chant simple sentences seem to assign syllables to beats in a way that respects their prosodic structural properties (Hayes and Kaun, 1996).

Another analogy can be drawn between musical modes or key systems and certain prosodically constrained phonological patterns, including vowel harmony. Musical modes constrain the set of pitches that are considered musically valid within a single piece or even a single phrase; valid notes within the key signature of G major are G, A, B, C, D, E, and F-sharp. Similarly, under phonological rules such as vowel harmony, the set of possible vowels that could be considered linguistically valid within a certain word is reduced. New suffixes that are attached to a word must maintain certain qualities in line with previous vowels in the word. For Turkish, a prodigious user of vowel harmony, certain vowels must share lip rounding and tongue position features within a single word (for example, the second vowel of *gördün*, “you saw,” is *ü* rather than *u*, *i*, or *ı* due to vowel harmony). Thus, in both modalities, the licit presence of certain subunits (vowels for language, pitches for music) inside longer sequences (words for language, musical phrases for music) is conditioned by constraints on co-occurrence within a larger sequence.

Even languages that do not display evidence for vowel harmony show examples of these constraints on co-occurrence within the prosodic domain of the word. In English, for example, the plural suffix has three phonetic forms: [s] as in “cats”, [z] as in “dogs”, and [əz] as in “horses”. Similarly, the past tense in English has three parallel phonetic forms: [t] as in “liked”, [d] as in “loved”, and [əd] as in “wanted”. Which variant appears in each form depends crucially on a constraint against consonants with dissimilar voicing appearing adjacently within words, as well as a constraint against consonants with a similar place of articulation appearing adjacently. These strict limits on adjacent co-occurrence resemble strict constraints on adjacent co-occurrence for the notes in Indian *ragas*, where there are often limits on, for example, which notes are “valid” parts of ascending or descending melodies within a piece of music (Castellano et al., 1984).

Nevertheless, there are exceptions that are tolerated in both modalities. Western music uses accidentals (sharps, flats, and naturals) in order to allow for a “banned” note, while almost every language with vowel harmony also has words that include exceptions to the general pattern of the language (the last vowel of the Turkish word *görüyor*, “you are seeing,” does not harmonize with the previous two vowels in the word because the present progressive suffix invariably uses the vowel *o*). And both music and language show variation in these properties. Languages that have vowel harmony may differ in which vowels undergo harmony, as well as what features of the vowel harmonize with previous ones. Likewise, musical structure varies from culture to culture, with patterns of greater typological frequency but not necessarily strict universals (see, e.g., Savage et al., 2015). All considered, it appears that prosody and music are well-aligned and, remarkably, even exhibit a similar flexibility within and across cultures.

Prosodic Structure and Syntactic Structure

Many previous findings arguing for a relationship between music and linguistic syntax can also be used to justify the idea of relationships between music and prosody. This is true because syntactic and prosodic structure are, in many respects, closely related. For example, patterns of pitch and syllable length (i.e., prosodic structure) aid participants in uncovering syntactic structure in an artificial language learning task (Langus et al., 2012). Indeed, prosodic boundaries are typically found near syntactic boundaries. Some theories even suggest that prosodic boundary placement is *dependent* on syntactic phrases, with prosodic phrase boundaries only being possible at the beginning or end of a syntactic phrase (Selkirk, 1984). Given the reciprocal nature of these two linguistic dimensions, it might be that many structural manipulations in studies of musical structure actually reflect changes to prosodic structure, not necessarily to syntactic structure.

The roles of syntactic and prosodic structure are likely indistinguishable in many previous studies, perhaps because of how closely the two types of structure are tied. Many studies that have revealed interactions of musical structure with syntactic

manipulations have also included manipulations of prosodic structure by dint of their auditory presentation (Fedorenko et al., 2009; Sammler et al., 2013). That is, because spoken sentences include patterns of pitch, rate, and volume that characterize prosody, manipulating the sentences in a way that affects syntax (by, say, scrambling them) generally also involves a concurrent manipulation of prosody (as re-ordering words implies also re-ordering the locations of pitch excursions, rate changes, and volume modulations). Similarly, studies of production that have found overlapping activation for music and language would by their nature have required the listener to create and deploy prosodic structure to produce sentences (Brown et al., 2006). Therefore, the erstwhile syntactic manipulations in such studies may actually just reflect the prosodic structures that are being built during the course of sentence production. Even self-paced reading studies (Slevc et al., 2009), which on the surface appear to divorce syntax from prosody, may not necessarily do so. Indeed, self-paced reading times correlate with prosodic structure (Koriat et al., 2002; Ashby, 2006) and self-paced reading elicits some of the very same prosody-related ERP components that show up in auditory speech perception (Steinhauer and Friederici, 2001).

How might these two similar structures be dissociated from one another? One possibility is to divorce syntactic cues from prosodic cues. Although some studies that have done this successfully in EEG have still exhibited significant effects of syntactic structure (discussed elsewhere in this paper), in many cases when using prosody-free presentation of language (e.g., isochronous visual word-by-word presentation), manipulations of syntactic and musical structure show non-overlapping patterns of activation in fMRI (Fedorenko et al., 2011, 2012). Similarly, studies that have crossed non-prosodic linguistic manipulations of semantic expectancy with musical harmonic violations in sung music have failed to find an interaction between each type of violation, suggesting that the linguistic semantic and musical harmonic systems are independent of each other (Besson et al., 1998; Bonnel et al., 2001 but see Poulin-Charronnat et al., 2005).

Even manipulations of long-distance musical information may fit into a prosodic framework. Prosodic structure also allows for the incorporation of long-distance information despite its non-recursive nature. Long-distance prosodic information has been shown to be informative to word segmentation, both in terms of long-distance pitch patterns (Dilley and McAuley, 2008) and long-distance rate information (Dilley and Pitt, 2010). Similarly, long-distance pitch and timing information can play a role in the processing of musical timing judgments (Povel and Essens, 1985; Boltz, 1993; McAuley and Kidd, 1998). However, critically, this long-distance information is not necessarily *recursive*. Incidental music in movie soundtracks, for example, quite often lacks any sort of long-distance structure or melody, but its ability to accentuate the action or mood on the screen makes it indispensable in modern movies (Boltz, 2004), while also being music that can be successfully marketed on its own. Listeners' sensitivity to long-distance information in music may instead reflect the gradient processing of relative timing or pitch properties

or the manipulation of expectations about rate or pitch information.

Emotional Aspects of Language and Music

Musical links to prosodic structure also allow for a variety of phenomena that are presently treated as unrelated curiosities to all be motivated by the same root causes. One notable example is the connection between the emotional content of speech and music. It is probable that most researchers reading this paper have largely seen prosody discussed in the context of emotion. Prosody is essential for the comprehension of emotion in speech, which fits nicely with the broader links between musical and prosodic structure we advocate here. There is no question that speech is capable of conveying rich emotional information, above and beyond the lexical content of the words in question, and that much of this information is conveyed through patterns of pitch, duration, and volume discussed in the context of prosody (Williams and Stevens, 1972; Murray and Arnott, 1993). The potential for the connection between these emotional aspects of language and those found in music has already been discussed extensively and convincingly elsewhere (e.g., Juslin and Laukka, 2003). As such, our focus here is on non-emotional aspects of language. However, it would be negligent on our part to avoid discussing emotional prosody entirely.

As an example of these connections, experience with music is tied to success in decoding prosodic aspects of language. There is a close correlation between prosody and emotional processing (Bänziger and Scherer, 2005; Grandjean et al., 2006), including an intriguing tendency for specific musical intervals (particularly the minor third) to emerge in emotional or stylized aspects of language (Curtis and Bharucha, 2010; Day-O'Connell, 2013). Prosody/emotion links suggest that advantages in prosodic processing may account for enhanced emotional processing abilities by musically trained adults (Thompson et al., 2004; Lima and Castro, 2011; Pinheiro et al., 2015; though cf. Trimmer and Cuddy, 2008). On the other end of the continuum of musical abilities, adults with congenital amusia exhibit a deficit in identifying emotional expressions in speech based on prosodic cues (Patel et al., 1998b, 2008; Jiang et al., 2012; Thompson et al., 2012).

More generally, it is often said to be the case that appreciation of music is derived in part from its emotional content (Juslin and Laukka, 2003; Juslin et al., 2008; Salimpoor et al., 2009). This emotional content has been argued to be universally accessible even across cultures (Fritz et al., 2009). Just as there are structural properties that appear to be shared across musical styles, there are also commonalities in how emotions are conveyed through music (Brown and Jordania, 2011). These bear many similarities to expressions of emotion in prosody (Juslin and Laukka, 2003). In one such parallel, infant-directed speech, also known as "motherese," has been claimed to primarily derive its status from uninhibited vocal emotion on the part of people addressing themselves to infants (Trainor et al., 2000). Likewise, infants and young children are also often sung to using unique vocal performance styles, such as the lullaby. Perhaps musical content gains its emotional valence in part through the same mechanisms that allow for the decoding of speech prosody.

Tying Up Loose Ends

It is certainly the case that the emotional links between music and prosody are important; however, as we have tried to spotlight, prosody is used for more in language than just emotion. Indeed, several aspects of prosody that are independent of emotion also gain greater salience under our present account. Although non-emotional aspects of prosodic perception have been less frequently studied than emotional prosody, Moreno et al. (2009) found that 8-year-olds with six months of musical training were more accurate than children without musical training at detecting small pitch excursions at the end of simple sentences; this was similar to previous results obtained for adults (Schön et al., 2004; Marques et al., 2007). This aligns with a version of the OPERA hypothesis (Patel, 2014), which is a framework that outlines how and why musical training might improve performance in other auditory domains. Under the OPERA hypothesis, speech processing is enhanced in musicians when speech and music both share certain cognitive resources, but music taxes those resources more than language does. Perhaps music is more demanding on prosodic structure-building than language. This in many ways resembles the perspective of Besson et al. (2011), who contrasted performance improvements on phonetic tasks that they claimed relied on shared auditory resources versus performance improvements that were the result of training transfer.

An additional example of a relevant non-emotion-related finding is the correlation between how variable the duration of a language's segments are (as measured using nPVI—the normalized Pairwise Variability Index—or other indices of vowel and consonant duration, such as the coefficient of variation) and how variable the length of different notes in that culture's music is. Patel and colleagues (Huron and Ollen, 2003; Patel and Daniele, 2003; Patel, 2005; Patel et al., 2006) found that, just as English speech has more variable segment durations than French speech, instrumental music from English composers has more variable note durations than music from French composers. Similar correlations exist among different dialects of English, as revealed from recordings made in the mid-twentieth century in the United States, Ireland, and Scotland (McGowan and Levitt, 2011). Under a prosodic account of musical structure, the reason for the correlation is quite clear: if musical structure and prosodic structure are closely linked, it makes sense for the statistical properties of the music within a culture to echo those of the prosodic structure.

Another prosodic process that might interact with musical experience is word segmentation, the process of locating word boundaries in fluent speech. Word segmentation is in part prosodic, as prosodic words are a part of the prosodic hierarchy. If prosody is linked to musical structure, it might then make sense for more musically experienced children to outperform their peers at tasks of word segmentation, as segmenting diverse types of musical streams might translate to the ability to segment auditory speech streams into prosodic words. And, indeed, this is exactly what was found for 8-year-old children on a test of their statistical learning proficiency (François et al., 2013).

Relatedly, modeling approaches that have made use of transitional probabilities have been successful in modeling

perceived musical expectedness ratings. In one experiment testing this, notes that had higher transitional probabilities with regard to their context were rated as more expected and elicited different electrophysiological responses from notes that had lower transitional probabilities (Pearce et al., 2010). Transitional probabilities are, by their nature, non-recursive. Although evidence for statistical learning has occasionally been demonstrated across long distances (Hsu et al., 2014; Kabdebon et al., 2015), this long-distance learning typically requires strong inter-unit similarity between the units that form the long-distance dependencies (Gebhart et al., 2009). These constraints on which units can show long-distance statistical learning are reminiscent of the co-occurrence restrictions in systems such as vowel harmony (e.g., vowel harmony applies in a long-distance fashion across adjacent vowels, “skipping over” consonants). Although it may be possible to incorporate ideas of statistical learning under the purview of syntax, the findings take on new significance under a prosodic account, as statistical learning has been clearly shown to be relevant for prosodic processes such as word segmentation (Saffran et al., 1996).

A final group of results that are parsimoniously explained by a prosodic account of musical structure are auditory illusions. Consider the speech-to-song illusion (Deutsch et al., 2011). In the speech-to-song illusion, certain phrases (in Deutsch et al., 2011, “sometimes behave so strangely”) when repeated numerous times suddenly begin to sound less like speech and more like music. This transformation is accompanied by increases in the salience of the pitch contour to listeners and, after being asked to repeat the phrase in question, by increasing faithfulness of such repetitions to musical intervals. It appears likely that the repetition of the prosodic information of the sentence are what allows it to begin to take on the structural characteristics of a song.

Another frequently discussed auditory illusion is the tritone paradox, where tritone pairs (i.e., tone pairs separated by a half-octave) are variably perceived as increasing or decreasing in pitch depending on, for example, the dialect of the listeners (Deutsch, 1997). Intriguingly, which pairs are perceived as increasing and which are perceived as decreasing is correlated with the pitch range of the perceiver's speech (Deutsch et al., 1990). Forging a direct link between a speaker's prosodic abilities and musical abilities makes the link clearer: the prosody of one's native dialect would help determine which pitch ranges and intervals a speaker was most used to, and would therefore affect their perception of those intervals outside of the domain of language.

THE NATURE OF A MUSIC/PROSODY CONNECTION

Having established, then, that music and prosody are connected, this leads to the question: *how* are they connected? We propose that there are four possibilities, although not all of these possibilities seem equally likely. First, musical structure and prosodic structure might be functionally independent but parallel. That is, the similarities between them might either be coincidental or based on some underlying structural organization

of the brain. Second, the use of musical structure and prosodic structure may involve similar underlying resources, but could differ in some aspects of their implementation. Third, certain aspects of musical structure might be “fundamental”, while prosodic structure depends on it. Finally, prosodic structure might be “fundamental”, while certain aspects of musical structure grew up around it.

First, let us address the idea that the structural properties of prosody and music are similar but related only coincidentally, not because of any particular shared mechanism. The first, “coincidental,” argument relies on the premise that prosody and music might have similar structural properties, but do not share underlying structural units. We believe that there are simply too many similarities between the processing of music and the processing of prosody for this fact to be simply an accident of mental organization. Why, under a “coincidental” account, should children or adults with musical training be better at detecting pitch excursions or segmenting words in speech (Schön et al., 2004; Moreno et al., 2009; François et al., 2013; Patel, 2013; Tillmann, 2014) or should adults with congenital amusia be worse at detecting emotional prosody in speech (Thompson et al., 2012)?

The second, “passive,” argument would be based on the idea that musical structure and prosodic structure do have similarities, but that those similarities are entirely due to a third factor. It behooves someone selecting this option, then, to find the third factor that explains the similarities between the modes of processing, such as cognitive control (see Slevc and Okada, 2015). Unfortunately, though, prosody is not a frequent object of study with regard to more domain-general cognitive functions, so it seems that this argument awaits further data.

The latter two other hypotheses represent more direct (and more radical) ways that prosody and music might be interrelated. It will likely be exceedingly difficult to resolve between these two hypotheses, given how difficult music and language are to analyze in the fossil and evolutionary anthropological record (Fitch, 2000; Hauser et al., 2002; Hauser and McDermott, 2003); however, this has not prevented many researchers from trying (Wallin et al., 2000). Nevertheless, if either prosody underlies music or music underlies prosody, then we can develop numerous predictions for the parallels between musical structure and prosodic structure, as the development and deployment of one ability marches in tandem with the other. Under the third argument outlined above, prosodic structure follows from musical structure. Musical structure forms the underlying basis of the organization that we observe within prosodic units. This bears some similarities to, say, evolutionary accounts of language that put music as the origin of linguistic abilities (Mithen, 2007). This may explain, for example, why talkers’ utterances trade off at consonant intervals during conversations with perfectly shared information but trade off at dissonant intervals during conversations with imperfectly shared information (Okada et al., 2012), or why the “stylized interjection” (i.e., the “calling contour”) is often associated with a minor third interval (Day-O’Connell, 2013).

The opposite notion is equally persuasive. It might be the case that the development of musical structure followed from prosodic structure. Linguistic prosody bears many similarities to the sorts

of vocal signals that other animals produce emotionally (Juslin and Laukka, 2003). And prosodic structure seems to be closely linked to syntactic structure (Selkirk, 1984; Langus et al., 2012), which many theorists take as the crux of what a human language is (Hauser et al., 2002). Thus, while it is almost impossible to say for certain whether music or prosody preceded the other, in light of how closely related syntax and prosody are and accordingly how central and unique recursive (syntactic) thought is to human nature, we err on the side of the notion that prosody arrived on the scene first, providing a framework and fertile ground upon which music could grow and organize itself, perhaps (though not necessarily) capitalizing on the emotional or expressive aspects of prosodic structure (Panksepp, 2009). Under this account, then, music still continues, with some variation, to make use of that structure.

CONNECTIONS

Roles for Both Prosody and Phonetics

Prosody, of course, is not entirely separable from other domains of language perception. Consider the perception of segmental phonetics, that is, speech when considered at the level of individual sounds. Musical pitches, like speech sounds, show evidence of categorical perception in trained listeners. Musicians (and, to a lesser extent, non-musicians) show sharp identification curves when asked to characterize musical intervals, with very good discriminability across category boundaries but only middling discriminability within single categories (Burns and Ward, 1978; Zatorre and Halpern, 1979). Categorical perception of musical intervals has been localized to a similar (although right-lateralized rather than left-lateralized) brain region as the categorical perception of speech sounds (Klein and Zatorre, 2011). This pattern was not found for a pitch discrimination task that did not entail categorical perception (Klein and Zatorre, 2015).

Trained musicians perform broadly better than non-musicians at phonetic perception tasks across a variety of measures. For example, musicians show stronger categorical perception for speech sounds than non-musicians do, with the strength of that categorical perception corresponding to higher-fidelity subcortical encoding of speech sounds (Bidelman et al., 2014). In one comprehensive test of second language abilities comparing groups of musicians to non-musicians, musically trained second language learners of English scored higher on measures of phonetic and phonological learning—but not any higher-level linguistic abilities, such as syntactic grammaticality judgments or vocabulary size—than non-musically trained learners (Slevc and Miyake, 2006). As such, it is plausible to suggest an analogy between music and sound-based linguistic systems in lieu of the syntactic explanation of musical structure. Yet while a parallel with phonetic processing may explain these findings and some of the interactions apparent in early music and language processing, a phonetic account is not sufficient to explain many of the findings of a higher-level relationships between music and language processing described earlier (e.g., Patel et al., 1998a; Koelsch, 2005; Koelsch et al., 2013).

An increased focus on the ties between musical structure and prosodic structure does not preclude the idea that musical processing aids the perception of individual segments as well. Perhaps one way in which these abilities tie in to the suprasegmental nature of prosody is related to the perception of lexical tone. Like prosody, lexical tone is suprasegmental, as it relies on pitch contours. However, like segmental information, it is realized on individual phonetic segments, such as vowels, and is related to the expression of basic lexical distinctions. Musicians who speak a non-tonal language are better than their non-musician peers at detecting tonal variation in languages that use lexical tone (Wong et al., 2007; Delogu et al., 2010), in a way reminiscent of tonal language speakers' increased proficiency with musical pitch discrimination (Bidelman et al., 2013). Exploring segmental contrasts that are similar to prosody is one way in which the relationship between prosody, phonetics, and music may be related.

Roles for Both Prosody and Syntax

Despite reasons to posit links between musical and prosodic structure, there is still evidence that music can interact with linguistic syntactic manipulations that are divorced from prosody, and that prosody may not overlap with music. For example, Rogalsky et al. (2011) found that comparing (auditory) scrambled sentences (presumably with disrupted prosody and syntax) to intact sentences reveals brain regions that largely do not correspond to those that are activated by music. Those regions that do respond to both music and prosody seem to be largely related to lower-level amplitude modulation, not to higher-level structural processes. This suggests non-overlap between prosodic brain regions and musical brain regions, at least in regions involved in this particular comparison of music to language.

Other evidence for musical relationships with syntax that are not attributable to prosody come from studies finding music/syntax interactions using isochronous word-by-word (and chord-by-chord) presentation (Koelsch et al., 2005; Steinbeis and Koelsch, 2008; Carrus et al., 2013), as this removes the timing, loudness, and pitch cues necessary to produce prosodic structure (though note that these prosodic cues do seem to be better exploited by musicians than non-musicians; Besson et al., 2011). These studies rely on the interpretation of particular ERP components that are traditionally associated with syntax, especially the LAN and P600. Although the syntactic natures of both the LAN (Steinhauer and Drury, 2012) and the P600 (Van Herten et al., 2005; Martín-Loeches et al., 2006) have been brought into question, this still provides support for a tie between music and syntax.

Therefore, we suggest that the increased focus on prosodic structure that we recommend in this paper does not necessitate an abandonment of syntactic structure as a potential analog for at least some components of music. Music is a complex system that may have multiple parallels in the similarly complex structures of language. And, given the many parallels between the structures of prosody and those of syntax, it is likely that finding ways to isolate the two will require careful study. However, we entreat researchers studying the ties between music and language to more

carefully consider the possibility that it is in prosody, not simply in syntax, where parallels lie.

PREDICTIONS

Encouraging renewed interest in prosodic structure as an analog to musical structure does more than just suggest a convenient way to tie together diverse studies that have examined both music and language. It also allows for a rich set of predictions that could be useful to further work in elucidating parallels between these domains.

For example, it should be possible to show further enhancements for musically trained participants in their perception of prosodic information, echoing similar advantages associated with phonetic processing. Examples of the latter include MEG evidence that French speaking musicians showed stronger mismatch negativity (MMN) responses to duration-related mismatches in non-linguistic stimuli than non-musician French speakers and even than non-musician Finnish speakers, who use contrastive consonant and vowel information in their own language (Marie et al., 2012). These advantages may be related to differences in the left planum temporale (PT) that are evident when comparing musicians to non-musicians (Elmer et al., 2012, 2013).

Given these advantages in phonetic processing, along with the prosody/music parallels discussed so far, we would also expect musicians to have improved prosodic processing abilities. There is indeed considerable evidence for such a relationship with emotional prosody (Thompson et al., 2004; Trimmer and Cuddy, 2008; Lima and Castro, 2011; Pinheiro et al., 2015). However, prosodic perception is more than just emotional processing. A prosody/music link predicts that musicians should also show advantages in prosodic abilities unrelated to emotional contexts. There is already some evidence for non-emotional prosodic enhancements in musicians; for example, musically trained children find it easier than non-musically trained children to extract statistical regularities they can use to determine word boundaries (François et al., 2013). Musically trained participants have more veridical neurophysiological representations of speech sounds for both children (Chobert et al., 2014) and adults (Wong et al., 2007). Musicians are also more sensitive to subtle fundamental frequency (F0) changes in speech when compared to non-musicians (Schön et al., 2004) which could be the result of more robust fundamental frequency tracking in their auditory brainstem responses (Musacchia et al., 2007).

This account predicts that musicians' advantages should also extend to more subtle aspects of prosodic processing. Dilley and Heffner (2013) moved F0 peaks through utterances in 30 ms increments in a way that they hoped would evoke prosodic boundaries. They had participants perform a variety of tasks, including discriminating between two versions of the same utterance, repeating different versions of each utterance, and explicitly judging word stress, all meant to assess categorical perception of the location of prosodic prominences within the sentences. In general, they found that listeners perceived the differences between the locations of possible prominence

categorically. That is, just like what has been abundantly demonstrated in the perception of individual speech sounds (Liberman et al., 1957), listeners' judgments of prosodic contrasts showed strong evidence for the influence of prosodic categories on the basic perception of those contrasts. Put another way, participants judged changes in the linear continuum non-linearly. However, there were strong individual differences in how categorical (i.e., how non-linear) participants' judgments actually were. Perhaps participants' musical training is one factor to explain the variation between individuals. Given the importance of prosody to determinations of foreign accent (Munro and Derwing, 2001; Trofimovich and Baker, 2006), musicians may be better able to replicate the prosodic properties of a second language that they are trying to learn, and might therefore sound less accented to an L1 speaker (Slevc and Miyake, 2006).

This hypothesis could additionally relate to the development of musical and linguistic skills. Given the conjecture that prosodic structure and musical structure have shared underpinnings, prosodic abilities should correlate with musical abilities across development. Indeed, that is what is generally found; both abilities develop quite early in life and show similar developmental trajectories (Brandt et al., 2012). Infants develop a very early preference, at as young as one month of age, for infant-directed speech, with accompanying pitch, volume, and duration cues (Fernald and Kuhl, 1987; Cooper and Aslin, 1990). Similarly, infants also prefer infant-directed singing over adult-directed singing (Trainor, 1996). Children as young as six to nine months prefer the prosody of their native language (Jusczyk et al., 1993; Höhle et al., 2009), and can productively use that prosodic information to help them segment words and phrases (Soderstrom et al., 2003). At around the same time period that they are learning the prosodic structures of their native language, they also begin to show evidence for a preference for native musical structures. Infants of a similar age show sensitivity to the musical structures used around them (Krumhansl and Jusczyk, 1990; Jusczyk and Krumhansl, 1993), prefer the musical structures of familiar musical styles to those of unfamiliar ones (Soley and Hannon, 2010), and can recognize brief, familiar musical passages (Saffran et al., 2000).

Clinical populations might also help test these hypotheses. Dysprosody, the clinical loss of prosodic characteristics of speech, is an infrequent condition that is typically acquired at some point during adulthood (Whitty, 1964; Samuel et al., 1998; Sidtis and Van Lancker Sidtis, 2003). However, dysprosody in language production is often associated with neurodegenerative diseases such as Parkinson's disease, suggesting the production or perception of prosody may be associated with the basal ganglia and dopaminergic pathways within the brain (Caekebeke et al., 1991; Van Lancker Sidtis et al., 2006; Skodda et al., 2009). Although the roots of dysprosody in this population may reside with difficulties in production, perceptual prosodic capacities among individuals with Parkinson's disease are currently understudied. One investigation found that deficits in the perception of the emotional content of music did not seem to be strongly

correlated with deficits in the perception of emotional speech prosody in Parkinson's patients (Lima et al., 2013); nevertheless this framework predicts that patients with perceptual dysprosody should also suffer from deficits in perception and production of musical structure.

Similarly, people with schizophrenia are often said to have problems with the comprehension of prosody. This is particularly true for the comprehension of affective prosody, which is very well-studied in people with schizophrenia (Murphy and Cutting, 1990; Mitchell et al., 2004; Hoekert et al., 2007; Roux et al., 2010), and is connected to concurrent emotional face processing deficits (Edwards et al., 2001, 2002). Fewer studies have investigated prosodic processing in schizophrenia beyond emotion; those that have, though, have also found deficits in more general prosodic processing (Leitman et al., 2005; Matsumoto et al., 2006). Given these results, it may also be the case that people with schizophrenia show impaired perception of musical structure. So too with people with depression, who display differences from typical controls in the production (Garcia-Toro et al., 2000) and perception (Uekermann et al., 2008) of prosody. Intriguingly, they also show differences in activation within brain regions typically associated with reward from control participants when listening to their favorite songs (Osuch et al., 2009).

CONCLUSION

Much of the literature on the structural parallels between language and music has focused on syntax. Although this is understandable in light of the focus within the language sciences on syntactic structure, it leaves relatively understudied another important aspect of language: the prosodic patterns of loudness, pitch, and timing that make up the rhythms and melodies of speech. It is possible that, when creating the *Hallelujah* chorus, Handel wept for joy solely because of the pleasant syntactic structure of the music he was creating. But, if it is the case that music and language are entwined closely enough to share crucial structural ingredients, it seems more likely (at least to us) that the pleasure of music might be enabled by the acoustic characteristics shared between music and prosody. It might be the case that the speech-to-song illusion (Deutsch et al., 2011) occurs because of the repetition of the syntactic structure in the simple utterances that characterize the illusion; however, it at least appears to be more direct to see it as such because of the repeated prosodic patterns of pitch, timing, and volume present in every utterance.

Prosody has a structure of its own, independent from (but interwoven with) the syntactic structure of language. Recalling the initial speculation found in Lerdahl and Jackendoff (1983), we argue here that the relationship of musical structure to language may gain much from discussion in terms of prosodic structure. This helps tie together several previous strands of research related to language/music parallels, and allows for several novel predictions about the relationship between prosodic processing in language and music processing. Syntax is, without a doubt, a critical part of language processing. And to consider prosody as a potential analog of musical structure does not require

abandoning the syntactic parallels that have been described. But ignoring the rest of language when seeking the most analogous processes to music within language is an approach that, we believe, has proven misguided.

FUNDING

CH was supported by a National Science Foundation Graduate Research Fellowship (GRF) and LRS was supported in part by a grant from the GRAMMY Foundation. Funding for Open Access provided by the UMD Libraries Open Access Publishing Fund.

REFERENCES

- Ashby, J. (2006). Prosody in skilled silent reading: evidence from eye movements. *J. Res. Read.* 29, 318–333. doi: 10.1111/j.1467-9817.2006.00311.x
- Bänziger, T., and Scherer, K. R. (2005). The role of intonation in emotional expressions. *Speech Commun.* 46, 252–267. doi: 10.1016/j.specom.2005.02.016
- Besson, M., Chobert, J., and Marie, C. (2011). Transfer of training between music and speech: common processing, attention, and memory. *Front. Psychol.* 2:94. doi: 10.3389/fpsyg.2011.00094
- Besson, M., Faïta, F., Peretz, I., Bonnel, A.-M., and Requin, J. (1998). Singing in the brain: independence of lyrics and tunes. *Psychol. Sci.* 9, 494–498. doi: 10.1111/1467-9280.00091
- Bidelman, G. M., Hutka, S., and Moreno, S. (2013). Tone language speakers and musicians share enhanced perceptual and cognitive abilities for musical pitch: evidence for bidirectionality between the domains of language and music. *PLoS ONE* 8:e60676. doi: 10.1371/journal.pone.0060676
- Bidelman, G. M., Weiss, M. W., Moreno, S., and Alain, C. (2014). Coordinated plasticity in brainstem and auditory cortex contributes to enhanced categorical speech perception in musicians. *Eur. J. Neurosci.* 40, 2662–2673. doi: 10.1111/ejn.12627
- Boltz, M. G. (1993). The generation of temporal and melodic expectancies during musical listening. *Percept. Psychophys.* 53, 585–600. doi: 10.3758/BF03211736
- Boltz, M. G. (2004). The cognitive processing of film and musical soundtracks. *Mem. Cogn.* 32, 1194–1205. doi: 10.3758/BF03196892
- Bonnel, A.-M., Faïta, F., Peretz, I., and Besson, M. (2001). Divided attention between lyrics and tunes of operatic songs: evidence for independent processing. *Percept. Psychophys.* 63, 1201–1213. doi: 10.3758/BF03194534
- Brandt, A., Gebrian, M., and Slevc, L. R. (2012). Music and early language acquisition. *Front. Psychol.* 3:327. doi: 10.3389/fpsyg.2012.00327
- Brown, S., and Jordania, J. (2011). Universals in the world's musics. *Psychol. Music* 41, 229–248. doi: 10.1177/0305735611425896
- Brown, S., Martinez, M. J., and Parsons, L. M. (2006). Music and language side by side in the brain: a PET study of the generation of melodies and sentences. *Eur. J. Neurosci.* 23, 2791–2803. doi: 10.1111/j.1460-9568.2006.04785.x
- Burns, E. M., and Ward, W. D. (1978). Categorical perception—phenomenon or epiphenomenon: evidence from experiments in the perception of melodic musical intervals. *J. Acoust. Soc. Am.* 63, 456–468. doi: 10.1121/1.381737
- Byrd, D., Krivokapic, J., and Lee, S. (2006). How far, how long: on the temporal scope of prosodic boundary effects. *J. Acoust. Soc. Am.* 120, 1589–1599. doi: 10.1121/1.2217135
- Caekebeke, J. F., Jennekens-Schinkel, A., van der Linden, M. E., Buruma, O. J., and Roos, R. A. (1991). The interpretation of dysprosody in patients with Parkinson's disease. *J. Neurol. Neurosurg. Psychiatry* 54, 145–148. doi: 10.1136/jnnp.54.2.145
- Carrus, E., Pearce, M. T., and Bhattacharya, J. (2013). Melodic pitch expectation interacts with neural responses to syntactic but not semantic violations. *Cortex* 49, 2186–2200. doi: 10.1016/j.cortex.2012.08.024
- Castellano, M. A., Bharucha, J. J., and Krumhansl, C. L. (1984). Tonal hierarchies in the music of north India. *J. Exp. Psychol. Gen.* 113, 394–412. doi: 10.1037/0096-3445.113.3.394

ACKNOWLEDGMENTS

This paper is based on a draft originally written for a seminar taught by LRS; we would like to thank the members of that seminar (Shota Momma, Mattson Ogg, Amritha Mallikarjun, Alix Kowalski, Brooke Okada, Alia Lancaster) for their valuable feedback during the initial formulations of this idea. We would also like to thank Bill Idsardi and Rochelle Newman for additional comments and suggestions on early manuscript drafts, and Mattson Ogg in particular for insights on prosody-emotion connections.

- Chobert, J., François, C., Velay, J.-L., and Besson, M. (2014). Twelve months of active musical training in 8- to 10-year-old children enhances the preattentive processing of syllabic duration and voice onset time. *Cereb. Cortex* 24, 956–967. doi: 10.1093/cercor/bhs377
- Christophe, A., Mehler, J., and Sebastián-Gallés, N. (2001). Perception of prosodic boundary correlates by newborn infants. *Infancy* 2, 385–394. doi: 10.1207/S15327078IN0203_6
- Cook, N. (1987). The perception of large-scale tonal closure. *Music Percept.* 5, 197–205. doi: 10.2307/40285392
- Cooper, R. P., and Aslin, R. N. (1990). Preference for infant-directed speech in the first month after birth. *Child Dev.* 61, 1584–1595. doi: 10.1111/j.1467-8624.1990.tb02885.x
- Curtis, M. E., and Bharucha, J. J. (2010). The minor third communicates sadness in speech, mirroring its use in music. *Emotion* 10, 335–348. doi: 10.1037/a0017928
- Day-O'Connell, J. (2013). Speech, song, and the minor third: an acoustic study of the stylized interjection. *Music Percept.* 30, 441–462. doi: 10.1525/MP.2013.30.5.441
- Delogu, F., Lampis, G., and Belardinelli, M. O. (2010). From melody to lexical tone: musical ability enhances specific aspects of foreign language perception. *Eur. J. Cogn. Psychol.* 22, 46–61. doi: 10.1080/09541440802708136
- Deutsch, D. (1997). The tritone paradox: a link between music and speech. *Curr. Dir. Psychol. Sci.* 6, 174–180. doi: 10.1111/1467-8721.ep10772951
- Deutsch, D., Henthorn, T., and Lapidis, R. (2011). Illusory transformation from speech to song. *J. Acoust. Soc. Am.* 129, 2245–2252. doi: 10.1121/1.3562174
- Deutsch, D., North, T., and Ray, L. (1990). The tritone paradox: correlate with the listener's vocal range for speech. *Music Percept.* 7, 371–384. doi: 10.2307/40285473
- Dilley, L. C., and Heffner, C. C. (2013). The role of F0 alignment in distinguishing intonation categories: evidence from American English. *J. Speech Sci.* 3, 3–67.
- Dilley, L. C., and McAuley, J. D. (2008). Distal prosodic context affects word segmentation and lexical processing. *J. Mem. Lang.* 59, 294–311. doi: 10.1016/j.jml.2008.06.006
- Dilley, L. C., and Pitt, M. A. (2010). Altering context speech rate can cause words to appear or disappear. *Psychol. Sci.* 21, 1664–1670. doi: 10.1177/0956797610384743
- Edwards, J., Jackson, H. J., and Pattison, P. E. (2002). Emotion recognition via facial expression and affective prosody in schizophrenia: a methodological review. *Clin. Psychol. Rev.* 22, 1267–1285. doi: 10.1016/S0272-7358(02)00162-9
- Edwards, J., Pattison, P. E., Jackson, H. J., and Wales, R. J. (2001). Facial affect and affective prosody recognition in first-episode schizophrenia. *Schizophr. Res.* 48, 235–253. doi: 10.1016/S0920-9964(00)00099-2
- Elmer, S., Hänggi, J., Meyer, M., and Jäncke, L. (2013). Increased cortical surface area of the left planum temporale in musicians facilitates the categorization of phonetic and temporal speech sounds. *Cortex* 49, 2812–2821. doi: 10.1016/j.cortex.2013.03.007
- Elmer, S., Meyer, M., and Jancke, L. (2012). Neurofunctional and behavioral correlates of phonetic and temporal categorization in musically trained and untrained subjects. *Cereb. Cortex* 22, 650–658. doi: 10.1093/cercor/bhr142
- Escoffier, N., and Tillmann, B. (2008). The tonal function of a task-irrelevant chord modulates speed of visual processing. *Cognition* 107, 1070–1083. doi: 10.1016/j.cognition.2007.10.007

- Fedorenko, E., Behr, M. K., and Kanwisher, N. (2011). Functional specificity for high-level linguistic processing in the human brain. *Proc. Natl. Acad. Sci. U.S.A.* 108, 16428–16433. doi: 10.1073/pnas.1112937108
- Fedorenko, E., McDermott, J. H., Norman-Haignere, S., and Kanwisher, N. (2012). Sensitivity to musical structure in the human brain. *J. Neurophysiol.* 108, 3289–3300. doi: 10.1152/jn.00209.2012
- Fedorenko, E., Patel, A., Casasanto, D., Winawer, J., and Gibson, E. (2009). Structural integration in language and music: evidence for a shared system. *Mem. Cogn.* 37, 1–9. doi: 10.3758/MC.37.1.1
- Fernald, A., and Kuhl, P. (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behav. Dev.* 10, 279–293. doi: 10.1016/0163-6383(87)90017-8
- Fitch, W. T. (2000). The evolution of speech: a comparative review. *Trends Cogn. Sci.* 4, 258–267. doi: 10.1016/S1364-6613(00)01494-7
- Forste, A. (1959). Schenker's conception of musical structure. *J. Music Theory* 3, 1–30. doi: 10.2307/842996
- François, C., Chobert, J., Besson, M., and Schön, D. (2013). Music training for the development of speech segmentation. *Cereb. Cortex* 23, 2038–2043. doi: 10.1093/cercor/bhs180
- Frazier, L., Carlson, K., and Clifton, C. (2006). Prosodic phrasing is central to language comprehension. *Trends Cogn. Sci.* 10, 244–249. doi: 10.1016/j.tics.2006.04.002
- Fritz, T., Jentschke, S., Gosselin, N., Sammler, D., Peretz, I., Turner, R., et al. (2009). Universal recognition of three basic emotions in music. *Curr. Biol.* 19, 573–576. doi: 10.1016/j.cub.2009.02.058
- García-Toro, M., Talavera, J. A., Saiz-Ruiz, J., and Gonzalez, A. (2000). Prosody impairment in depression measured through acoustic analyses. *J. Nerv. Ment. Dis.* 188, 824–829. doi: 10.1097/00005053-200012000-00006
- Gay, T. (1978). Physiological and acoustic correlates of perceived stress. *Lang. Speech* 21, 347–353. doi: 10.1177/002383097802100409
- Gebhart, A. L., Newport, E. L., and Aslin, R. N. (2009). Statistical learning of adjacent and nonadjacent dependencies among nonlinguistic sounds. *Psychon. Bull. Rev.* 16, 486–490. doi: 10.3758/PBR.16.3.486
- Gordon, R. L., Magne, C. L., and Large, E. W. (2011). EEG correlates of song prosody: a new look at the relationship between linguistic and musical rhythm. *Front. Psychol.* 2:352. doi: 10.3389/fpsyg.2011.00352
- Gordon, R. L., Shivers, C. M., Wieland, E. A., Kotz, S. A., Yoder, P. J., and McAuley, J. D. (2015). Musical rhythm discrimination explains individual differences in grammar skills in children. *Dev. Sci.* 18, 635–644. doi: 10.1111/desc.12230
- Grandjean, D., Bänziger, T., and Scherer, K. R. (2006). Intonation as an interface between language and affect. *Prog. Brain Res.* 156, 235–247. doi: 10.1016/S0079-6123(06)56012-1
- Halle, J., and Lerdahl, F. (1993). A generative textsetting model. *Curr. Musicol.* 55, 3–23.
- Hauser, M. D., Chomsky, N., and Fitch, W. T. (2002). The faculty of language: what is it, who has it, and how did it evolve? *Science* 298, 1569–1579. doi: 10.1126/science.298.5598.1569
- Hauser, M. D., and McDermott, J. (2003). The evolution of the music faculty: a comparative perspective. *Nat. Neurosci.* 6, 663–668. doi: 10.1038/nn1080
- Hayes, B., and Kaun, A. (1996). The role of phonological phrasing in sung and chanted verse. *Linguist. Rev.* 13, 1–60. doi: 10.1515/tlir.1996.13.3-4.243
- Hoekert, M., Kahn, R. S., Pijnenborg, M., and Aleman, A. (2007). Impaired recognition and expression of emotional prosody in schizophrenia: review and meta-analysis. *Schizophr. Res.* 96, 135–145. doi: 10.1016/j.schres.2007.07.023
- Höhle, B., Bijeljac-Babic, R., Herold, B., Weissenborn, J., and Nazzi, T. (2009). Language specific prosodic preferences during the first half year of life: evidence from German and French infants. *Infant Behav. Dev.* 32, 262–274. doi: 10.1016/j.infbeh.2009.03.004
- Hsu, H. J., Tomblin, J. B., and Christiansen, M. H. (2014). Impaired statistical learning of non-adjacent dependencies in adolescents with specific language impairment. *Front. Psychol.* 5:175. doi: 10.3389/fpsyg.2014.00175
- Huron, D., and Ollen, J. (2003). Agogic contrast in French and English themes: further support for Patel and Daniele (2003). *Music Percept.* 21, 695–701. doi: 10.1525/mp.2003.21.2.267
- Jackendoff, R. (2009). Parallels and nonparallels between language and music. *Music Percept.* 26, 195–204. doi: 10.1525/mp.2009.26.3.195
- Jentschke, S., and Koelsch, S. (2009). Musical training modulates the development of syntax processing in children. *Neuroimage* 47, 735–744. doi: 10.1016/j.neuroimage.2009.04.090
- Jiang, C., Hamm, J. P., Lim, V. K., Kirk, I. J., Chen, X., and Yang, Y. (2012). Amusia results in abnormal brain activity following inappropriate intonation during speech comprehension. *PLoS ONE* 7:e41411. doi: 10.1371/journal.pone.0041411
- Jusczyk, P. W., Cutler, A., and Redanz, N. J. (1993). Infants' preference for the predominant stress patterns of English words. *Child Dev.* 64, 675–687. doi: 10.1111/j.1467-8624.1993.tb02935.x
- Jusczyk, P. W., and Krumhansl, C. L. (1993). Pitch and rhythmic patterns affecting infants' sensitivity to musical phrase structure. *J. Exp. Psychol. Hum. Percept. Perform.* 19, 627–640. doi: 10.1037/0096-1523.19.3.627
- Juslin, P. N., and Laukka, P. (2003). Communication of emotions in vocal expression and music performance: different channels, same code? *Psychol. Bull.* 129, 770–814. doi: 10.1037/0033-2909.129.5.770
- Juslin, P. N., Liljeström, S., Västfjäll, D., Barradas, G., and Silva, A. (2008). An experience sampling study of emotional reactions to music: listener, music, and situation. *Emotion* 8, 668–683. doi: 10.1037/a0013505
- Kabdebon, C., Pena, M., Buiatti, M., and Dehaene-Lambertz, G. (2015). Electrophysiological evidence of statistical learning of long-distance dependencies in 8-month-old preterm and full-term infants. *Brain Lang.* 148, 25–36. doi: 10.1016/j.bandl.2015.03.005
- Katz, J., and Pesetsky, D. (2009). *The Identity Thesis for Language and Music*. Cambridge, MA: MIT Press.
- Klein, M. E., and Zatorre, R. J. (2011). A role for the right superior temporal sulcus in categorical perception of musical chords. *Neuropsychologia* 49, 878–887. doi: 10.1016/j.neuropsychologia.2011.01.008
- Klein, M. E., and Zatorre, R. J. (2015). Representations of invariant musical categories are decodable by pattern analysis of locally distributed BOLD responses in superior temporal and intraparietal sulci. *Cereb. Cortex* 25, 1947–1957. doi: 10.1093/cercor/bhu003
- Koelsch, S. (2005). Neural substrates of processing syntax and semantics in music. *Curr. Opin. Neurobiol.* 15, 207–212. doi: 10.1007/978-3-211-75121-3_9
- Koelsch, S., and Friederici, A. D. (2003). Toward the neural basis of processing structure in music: comparative results of different neuropsychological methods. *Ann. N. Y. Acad. Sci.* 999, 15–28. doi: 10.1196/annals.1284.002
- Koelsch, S., Gunter, T. C., Wittfoth, M., and Sammler, D. (2005). Interaction between syntax processing in language and in music: an ERP Study. *J. Cogn. Neurosci.* 17, 1565–1577. doi: 10.1162/089892905774597290
- Koelsch, S., Rohrmeier, M., Torrecuso, R., and Jentschke, S. (2013). Processing of hierarchical syntactic structure in music. *Proc. Natl. Acad. Sci. U.S.A.* 110, 15443–15448. doi: 10.1073/pnas.1300272110
- Koriat, A., Greenberg, S. N., and Kreiner, H. (2002). The extraction of structure during reading: evidence from reading prosody. *Mem. Cogn.* 30, 270–280. doi: 10.3758/BF03195288
- Krumhansl, C. L., and Jusczyk, P. W. (1990). Infants' perception of phrase structure in music. *Psychol. Sci.* 1, 70–73. doi: 10.1111/j.1467-9280.1990.tb00070.x
- Kunert, R., and Slevc, L. R. (2015). A Commentary on: “Neural overlap in processing music and speech.” *Front. Hum. Neurosci.* 9:330. doi: 10.3389/fnhum.2015.00330
- LaCroix, A. N., Diaz, A. F., and Rogalsky, C. (2015). The relationship between the neural computations for speech and music perception is context-dependent: an activation likelihood estimate study. *Front. Psychol.* 6:1138. doi: 10.3389/fpsyg.2015.01138
- Langus, A., Marchetto, E., Bion, R. A. H., and Nespore, M. (2012). Can prosody be used to discover hierarchical structure in continuous speech? *J. Mem. Lang.* 66, 285–306. doi: 10.1016/j.jml.2011.09.004
- Large, E. W., and Palmer, C. (2002). Perceiving temporal regularity in music. *Cogn. Sci.* 26, 1–37. doi: 10.1016/S0364-0213(01)00057-X
- Larson, S. (1998). Schenkerian analysis of modern jazz: questions about method. *Music Theory Spectr.* 20, 209–241. doi: 10.1525/mts.1998.20.2.02a00020
- Leitman, D. I., Foxe, J. J., Butler, P. D., Saperstein, A., Revheim, N., and Javitt, D. C. (2005). Sensory contributions to impaired prosodic processing in schizophrenia. *Biol. Psychiatry* 58, 56–61. doi: 10.1016/j.biopsych.2005.02.034
- Lerdahl, F. (2001a). The sounds of poetry viewed as music. *Ann. N. Y. Acad. Sci.* 930, 337–354. doi: 10.1111/j.1749-6632.2001.tb05743.x
- Lerdahl, F. (2001b). *Tonal Pitch Space*. New York, NY: Oxford University Press.

- Lerdahl, F., and Jackendoff, R. (1983). *A Generative Theory of Tonal Music*, 1st Edn. Cambridge, MA: MIT Press.
- Lieberman, A. M., Harris, K. S., Hoffman, H. S., and Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *J. Exp. Psychol.* 54, 358–368. doi: 10.1037/h0044417
- Lima, C. F., and Castro, S. L. (2011). Speaking to the trained ear: musical expertise enhances the recognition of emotions in speech prosody. *Emotion* 11, 1021–1031. doi: 10.1037/a0024521
- Lima, C. F., Garrett, C., and Castro, S. L. (2013). Not all sounds sound the same: Parkinson's disease affects differently emotion processing in music and in speech prosody. *J. Clin. Exp. Neuropsychol.* 35, 373–392. doi: 10.1080/13803395.2013.776518
- Longuet-Higgins, H. C. (1976). Perception of melodies. *Nature* 263, 646–653. doi: 10.1038/263646a0
- Maess, B., Koelsch, S., Gunter, T. C., and Friederici, A. D. (2001). Musical syntax is processed in Broca's area: an MEG study. *Nat. Neurosci.* 4, 540–545. doi: 10.1038/87502
- Marie, C., Kujala, T., and Besson, M. (2012). Musical and linguistic expertise influence pre-attentive and attentive processing of non-speech sounds. *Cortex* 48, 447–457. doi: 10.1016/j.cortex.2010.11.006
- Marques, C., Moreno, S., Castro, S. L., and Besson, M. (2007). Musicians detect pitch violation in a foreign language better than nonmusicians: behavioral and electrophysiological evidence. *J. Cogn. Neurosci.* 19, 1453–1463. doi: 10.1162/jocn.2007.19.9.1453
- Martín-Loeches, M., Casado, P., Gonzalo, R., De Heras, L., and Fernández-Frías, C. (2006). Brain potentials to mathematical syntax problems. *Psychophysiology* 43, 579–591. doi: 10.1111/j.1469-8986.2006.00463.x
- Matsumoto, K., Samson, G. T., O'Daly, O. D., Tracy, D. K., Patel, A. D., and Shergill, S. S. (2006). Prosodic discrimination in patients with schizophrenia. *Br. J. Psychiatry* 189, 180–181. doi: 10.1192/bjp.bp.105.009332
- McAuley, J. D., and Kidd, G. R. (1998). Effect of deviations from temporal expectations on tempo discrimination of isochronous tone sequences. *J. Exp. Psychol. Hum. Percept. Perform.* 24, 1786–1800. doi: 10.1121/1.417584
- McGowan, R. W., and Levitt, A. G. (2011). A comparison of rhythm in English dialects and music. *Music Percept.* 28, 307–314. doi: 10.1525/mp.2011.28.3.307
- McGraw, A. C. (2008). Different temporalities: the time of Balinese gamelan. *Yearbook Tradit. Music* 40, 136–162.
- Mitchell, R. L. C., Elliott, R., Barry, M., Cruttenden, A., and Woodruff, P. W. R. (2004). Neural response to emotional prosody in schizophrenia and in bipolar affective disorder. *Br. J. Psychiatry* 184, 223–230. doi: 10.1192/bjp.184.3.223
- Mithen, S. (2007). *The Singing Neanderthals: The Origins of Music, Language, Mind, and Body*. Cambridge, MA: Harvard University Press.
- Moreno, S., Marques, C., Santos, A., Santos, M., Castro, S. L., and Besson, M. (2009). Musical training influences linguistic abilities in 8-year-old children: more evidence for brain plasticity. *Cereb. Cortex* 19, 712–723. doi: 10.1093/cercor/bhn120
- Munro, M. J., and Derwing, T. M. (2001). Modeling perceptions of the accentedness and comprehensibility of L2 speech: the role of speaking rate. *Stud. Second Lang. Acquis.* 23, 451–468.
- Murphy, D., and Cutting, J. (1990). Prosodic comprehension and expression in schizophrenia. *J. Neurol. Neurosurg. Psychiatry* 53, 727–730. doi: 10.1136/jnnp.53.9.727
- Murray, I. R., and Arnott, J. L. (1993). Toward the simulation of emotion in synthetic speech: a review of the literature on human vocal emotion. *J. Acoust. Soc. Am.* 93, 1097–1108. doi: 10.1121/1.405558
- Musacchia, G., Sams, M., Skoe, E., and Kraus, N. (2007). Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proc. Natl. Acad. Sci. U.S.A.* 104, 15894–15898. doi: 10.1073/pnas.0701498104
- Oechslin, M. S., Van De Ville, D., Lazeyras, F., Hauert, C. A., and James, C. E. (2013). Degree of musical expertise modulates higher order brain functioning. *Cereb. Cortex* 23, 2213–2224. doi: 10.1093/cercor/bhs206
- Okada, B. M., Lachs, L., and Boone, B. (2012). Interpreting tone of voice: musical pitch relationships convey agreement in dyadic conversation. *J. Acoust. Soc. Am.* 132, EL208–EL214. doi: 10.1121/1.4742316
- Osuch, E. A., Bluhm, R. L., Williamson, P. C., Théberge, J., Densmore, M., and Neufeld, R. W. J. (2009). Brain activation to favorite music in healthy controls and depressed patients. *Neuroreport* 20, 1204–1208. doi: 10.1097/WNR.0b013e32832f4da3
- Palmer, C., and Hutchins, S. (2006). “What is musical prosody?,” in *Psychology of Learning and Motivation*, Vol. 46, ed. B. H. Ross (Oxford: Elsevier), 245–278. doi: 10.1016/S0079-7421(06)46007-2
- Palmer, C., and Kelly, M. H. (1992). Linguistic prosody and musical meter in song. *J. Mem. Lang.* 31, 525–542. doi: 10.1016/0749-596X(92)90027-U
- Panksepp, J. (2009). The emotional antecedents to the evolution of music and language. *Music. Sci.* 13(2 Suppl), 229–259. doi: 10.1177/1029864909013002111
- Patel, A. D. (2003). Language, music, syntax and the brain. *Nat. Neurosci.* 6, 674–681. doi: 10.1038/nn1082
- Patel, A. D. (2005). The relationship of music to the melody of speech and to syntactic processing disorders in aphasia. *Ann. N. Y. Acad. Sci.* 1060, 59–70. doi: 10.1196/annals.1360.005
- Patel, A. D. (2008). *Music, Language, and the Brain*. Oxford: Oxford University Press.
- Patel, A. D. (2013). “Sharing and nonsharing of brain resources for language and music,” in *Language, Music, and the Brain: A Mysterious Relationship*, ed. M. Arbib (Cambridge, MA: MIT Press), 329–355.
- Patel, A. D. (2014). Can nonlinguistic musical training change the way the brain processes speech? The expanded OPERA hypothesis. *Hear. Res.* 308, 98–108. doi: 10.1016/j.heares.2013.08.011
- Patel, A. D., and Daniele, J. R. (2003). An empirical comparison of rhythm in language and music. *Cognition* 87, B35–B45. doi: 10.1016/S0010-0277(02)00187-7
- Patel, A. D., Gibson, E., Ratner, J., Besson, M., and Holcomb, P. J. (1998a). Processing syntactic relations in language and music: an event-related potential study. *J. Cogn. Neurosci.* 10, 717–733. doi: 10.1162/089892998563121
- Patel, A. D., Peretz, I., Tramo, M., and Labreque, R. (1998b). Processing prosodic and musical patterns: a neuropsychological investigation. *Brain Lang.* 61, 123–144. doi: 10.1006/brln.1997.1862
- Patel, A. D., Iversen, J. R., and Rosenberg, J. C. (2006). Comparing the rhythm and melody of speech and music: the case of British English and French. *J. Acoust. Soc. Am.* 119(5 Pt 1), 3034–3047. doi: 10.1121/1.2179657
- Patel, A. D., Wong, M., Foxtton, J., Lochy, A., and Peretz, I. (2008). Speech intonation perception deficits in musical tone deafness (congenital amusia). *Music Percept.* 25, 357–368. doi: 10.1525/mp.2008.25.4.357
- Pearce, M. T., Ruiz, M. H., Kapasi, S., Wiggins, G. A., and Bhattacharya, J. (2010). Unsupervised statistical learning underpins computational, behavioural, and neural manifestations of musical expectation. *Neuroimage* 50, 302–313. doi: 10.1016/j.neuroimage.2009.12.019
- Peretz, I., Vuvan, D., Lagrois, M. -É, and Armony, J. L. (2015). Neural overlap in processing music and speech. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 370, 20140090. doi: 10.1098/rstb.2014.0090
- Perruchet, P., and Poulin-Charronnat, B. (2013). Challenging prior evidence for a shared syntactic processor for language and music. *Psychon. Bull. Rev.* 20, 310–317. doi: 10.3758/s13423-012-0344-5
- Pinheiro, A. P., Vasconcelos, M., Dias, M., Arrais, N., and Gonçalves, Ó. F. (2015). The music of language: an ERP investigation of the effects of musical training on emotional prosody processing. *Brain Lang.* 140, 24–34. doi: 10.1016/j.bandl.2014.10.009
- Poulin-Charronnat, B., Bigand, E., Madurell, F., and Peereman, R. (2005). Musical structure modulates semantic priming in vocal music. *Cognition* 94, B67–B78. doi: 10.1016/j.cognition.2004.05.003
- Povel, D.-J., and Essens, P. J. (1985). Perception of temporal patterns. *Music Percept.* 2, 411–440. doi: 10.2307/40285311
- Rogalsky, C., Rong, F., Saberi, K., and Hickok, G. (2011). Functional anatomy of language and music perception: temporal and structural factors investigated using functional magnetic resonance imaging. *J. Neurosci.* 31, 3843–3852. doi: 10.1523/JNEUROSCI.4515-10.2011
- Rohrmeier, M. (2011). Towards a generative syntax of tonal harmony. *J. Math. Music* 5, 35–53. doi: 10.1080/17459737.2011.573676
- Roux, P., Christophe, A., and Passerieux, C. (2010). The emotional paradox: dissociation between explicit and implicit processing of emotional prosody in schizophrenia. *Neuropsychologia* 48, 3642–3649. doi: 10.1016/j.neuropsychologia.2010.08.021
- Saffran, J. R., Loman, M. M., and Robertson, R. R. W. (2000). Infant memory for musical experiences. *Cognition* 77, 1–9. doi: 10.1016/S0010-0277(00)00095-0

- Saffran, J. R., Newport, E. L., and Aslin, R. N. (1996). Word segmentation: the role of distributional cues. *J. Mem. Lang.* 35, 606–621. doi: 10.1006/jmla.1996.0032
- Salimpoor, V. N., Benovoy, M., Longo, G., Cooperstock, J. R., and Zatorre, R. J. (2009). The rewarding aspects of music listening are related to degree of emotional arousal. *PLoS ONE* 4:e7487. doi: 10.1371/journal.pone.0007487
- Sammler, D., Koelsch, S., Ball, T., Brandt, A., Grigutsch, M., Huppertz, H. J., et al. (2013). Co-localizing linguistic and musical syntax with intracranial EEG. *Neuroimage* 64, 134–146. doi: 10.1016/j.neuroimage.2012.09.035
- Samuel, C., Louis-Dreyfus, A., Couillet, J., Roubeau, B., Bakchine, S., Bussel, B., et al. (1998). Dysprosody after severe closed head injury: an acoustic analysis. *J. Neurol. Neurosurg. Psychiatry* 64, 482–485. doi: 10.1136/jnnp.64.4.482
- Savage, P. E., Brown, S., Sakai, E., and Currie, T. E. (2015). Statistical universals reveal the structures and functions of human music. *Proc. Natl. Acad. Sci. U.S.A.* 112, 8987–8992. doi: 10.1073/pnas.1414495112
- Schiller, N. O., Meyer, A. S., and Levelt, W. J. M. (1997). The syllabic structure of spoken words: evidence from the syllabification of intervocalic consonants. *Lang. Speech* 40, 103–140. doi: 10.1177/002383099704000202
- Schön, D., Magne, C., and Besson, M. (2004). The music of speech: music training facilitates pitch processing in both music and language. *Psychophysiology* 41, 341–349. doi: 10.1111/1469-8986.00172.x
- Seger, C. A., Spiering, B. J., Sares, A. G., Quraini, S. I., Alpetter, C., David, J., et al. (2013). Corticostriatal contributions to musical expectancy perception. *J. Cogn. Neurosci.* 25, 1062–1077. doi: 10.1162/jocn_a_00371
- Selkirk, E. (1984). *Phonology and Syntax: The Relation between Sound and Structure*. Cambridge, MA: MIT Press.
- Shattuck-Hufnagel, S., and Turk, A. E. (1996). A prosody tutorial for investigators of auditory sentence processing. *J. Psycholinguist. Res.* 25, 193–247. doi: 10.1007/BF01708572
- Sidtis, J. J., and Van Lancker Sidtis, D. (2003). A neurobehavioral approach to dysprosody. *Semin. Speech Lang.* 24, 93–105. doi: 10.1055/s-2003-38901
- Skodda, S., Rinsche, H., and Schlegel, U. (2009). Progression of dysprosody in Parkinson's disease over time—A longitudinal study. *Mov. Disord.* 24, 716–722. doi: 10.1002/mds.22430
- Slevc, L. R., and Miyake, A. (2006). Individual differences in second-language proficiency: does musical ability matter? *Psychol. Sci.* 17, 675–681. doi: 10.1111/j.1467-9280.2006.01765.x
- Slevc, L. R., and Okada, B. M. (2015). Processing structure in language and music: a case for shared reliance on cognitive control. *Psychon. Bull. Rev.* 22, 637–652. doi: 10.3758/s13423-014-0712-4
- Slevc, L. R., Rosenberg, J. C., and Patel, A. D. (2009). Making psycholinguistics musical: self-paced reading time evidence for shared processing of linguistic and musical syntax. *Psychon. Bull. Rev.* 16, 374–381. doi: 10.3758/16.2.374
- Soderstrom, M., Seidl, A., Kemler Nelson, D. G., and Jusczyk, P. W. (2003). The prosodic bootstrapping of phrases: evidence from prelinguistic infants. *J. Mem. Lang.* 49, 249–267. doi: 10.1016/S0749-596X(03)00024-X
- Soley, G., and Hannon, E. E. (2010). Infants prefer the musical meter of their own culture: a cross-cultural comparison. *Dev. Psychol.* 46, 286–292. doi: 10.1037/a0017555
- Steedman, M. J. (1984). A generative grammar for jazz chord sequences. *Music Percept.* 2, 52–77. doi: 10.2307/40285282
- Steinbeis, N., and Koelsch, S. (2008). Comparing the processing of music and language meaning using EEG and fMRI provides evidence for similar and distinct neural representations. *PLoS ONE* 3:e2226. doi: 10.1371/journal.pone.0002226
- Steinhauer, K., Alter, K., and Friederici, A. D. (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nat. Neurosci.* 2, 191–196. doi: 10.1038/5757
- Steinhauer, K., and Drury, J. E. (2012). On the early left-anterior negativity (ELAN) in syntax studies. *Brain Lang.* 120, 135–162. doi: 10.1016/j.bandl.2011.07.001
- Steinhauer, K., and Friederici, A. D. (2001). Prosodic boundaries, comma rules, and brain responses: the closure positive shift in ERPs as a universal marker for prosodic phrasing in listeners and readers. *J. Psycholinguist. Res.* 30, 267–295. doi: 10.1023/A:1010443001646
- Thompson, W. F., Marin, M. M., and Stewart, L. (2012). Reduced sensitivity to emotional prosody in congenital amusia rekindles the musical protolanguage hypothesis. *Proc. Natl. Acad. Sci. U.S.A.* 109, 19027–19032. doi: 10.1073/pnas.1210344109
- Thompson, W. F., Schellenberg, E. G., and Husain, G. (2004). Decoding speech prosody: do music lessons help? *Emotion* 4, 46–64. doi: 10.1037/1528-3542.4.1.46
- Tillmann, B. (2014). Pitch processing in music and speech. *Acoust. Aust.* 42, 124–130.
- Tillmann, B., and Bigand, E. (2004). The relative importance of local and global structures in music perception. *J. Aesthet. Art Criticism* 62, 211–222. doi: 10.1111/j.1540-594X.2004.00153.x
- Tillmann, B., Bigand, E., and Madurell, F. (1998). Local versus global processing of harmonic cadences in the solution of musical puzzles. *Psychol. Res.* 61, 157–174. doi: 10.1007/s004260050022
- Tillmann, B., Koelsch, S., Escoffier, N., Bigand, E., Lalitte, P., Friederici, A. D., et al. (2006). Cognitive priming in sung and instrumental music: activation of inferior frontal cortex. *Neuroimage* 31, 1771–1782. doi: 10.1016/j.neuroimage.2006.02.028
- Trainor, L. J. (1996). Infant preferences for infant-directed versus noninfant-directed playsongs and lullabies. *Infant Behav. Dev.* 19, 83–92. doi: 10.1016/S0163-6383(96)90046-6
- Trainor, L. J., Austin, C. M., and Desjardins, R. N. (2000). Is infant-directed speech prosody a result of the vocal expression of emotion? *Psychol. Sci.* 11, 188–195. doi: 10.1111/1467-9280.00240
- Trimmer, C. G., and Cuddy, L. L. (2008). Emotional intelligence, not music training, predicts recognition of emotional speech prosody. *Emotion* 8, 838–849. doi: 10.1037/a0014080
- Trofimovich, P., and Baker, W. (2006). Learning second language suprasegmentals: effect of L2 experience on prosody and fluency characteristics of L2 speech. *Stud. Second Lang. Acquis.* 28, 1–30. doi: 10.1017/S0272263106060013
- Uekermann, J., Abdel-Hamid, M., Lehmkämpfer, C., Vollmoeller, W., and Daum, I. (2008). Perception of affective prosody in major depression: a link to executive functions? *J. Int. Neuropsychol. Soc.* 14, 552–561. doi: 10.1017/S135561708080740
- Van Herten, M., Kolk, H. H. J., and Chwilla, D. J. (2005). An ERP study of P600 effects elicited by semantic anomalies. *Cogn. Brain Res.* 22, 241–255. doi: 10.1016/j.cogbrainres.2004.09.002
- Van Lancker Sidtis, D., Pachana, N., Cummings, J. L., and Sidtis, J. J. (2006). Dysprosodic speech following basal ganglia insult: toward a conceptual framework for the study of the cerebral representation of prosody. *Brain Lang.* 97, 135–153. doi: 10.1016/j.bandl.2005.09.001
- Wallin, N. L., Merker, B. H., and Brown, S. (eds) (2000). *The Origins of Music*. Cambridge, MA: MIT Press.
- Whitty, C. W. M. (1964). Cortical dysarthria and dysprosody of speech. *J. Neurol. Neurosurg. Psychiatry* 27, 507–510. doi: 10.1136/jnnp.27.6.507
- Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., and Price, P. J. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *J. Acoust. Soc. Am.* 91, 1707–1717. doi: 10.1121/1.402450
- Williams, C. E., and Stevens, K. N. (1972). Emotions and speech: some acoustical correlates. *J. Acoust. Soc. Am.* 52, 1238–1250. doi: 10.1121/1.1913238
- Wong, P. C. M., Skoe, E., Russo, N. M., Dees, T., and Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nat. Neurosci.* 10, 420–422. doi: 10.1038/nn1872
- Zatorre, R. J., and Halpern, A. R. (1979). Identification, discrimination, and selective adaptation of simultaneous musical intervals. *Percept. Psychophys.* 26, 384–395. doi: 10.3758/BF03204164

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Heffner and Slevc. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.