

THEORETICAL REVIEW Processing structure in language and music: a case for shared reliance on cognitive control L. Robert Slevc & Brooke M. Okada Published online: 5 August 2014 # Psychonomic Society, Inc. 2014 Abstract The relationship between structural processing in music and language has received increasing interest in the past several years, spurred by the influential Shared Syntactic Integration Resource Hypothesis (SSIRH; Patel, *Nature Neuroscience*, 6, 674–681, 2003). According to this resource-sharing framework, music and language rely on separable syntactic representations but recruit shared cognitive resources to integrate these representations into evolving structures. The SSIRH is supported by findings of interactions between structural manipulations in music and language. However, other recent evidence suggests that such interactions also can arise with nonstructural manipulations, and some recent neuroimaging studies report largely nonoverlapping neural regions involved in processing musical and linguistic structure. These conflicting results raise the question of exactly what shared (and distinct) resources underlie musical and linguistic structural processing. This paper suggests that one shared resource is prefrontal cortical mechanisms of cognitive control, which are recruited to detect and resolve conflict that occurs when expectations are violated and interpretations must be revised. By this account, musical processing involves not just the incremental processing and integration of musical elements as they occur, but also the incremental generation of musical predictions and expectations, which must sometimes be overridden and revised in light of evolving musical input. Keywords Language . Music . Syntax . Cognitive control . Musical ambiguity

The impressive human ability to process complex structure is perhaps most evident in language and music. The existence (or nonexistence) of a relationship between musical and linguistic structure (syntax) has received increasing interest over the past several years (for reviews, see Patel, 2008; Slevc, 2012; Tillmann, 2012), partially because this issue speaks to the broad question of modularity: do the complex cognitive systems supporting music and language rely on separable, modular processes (Peretz & Coltheart, 2003), or does syntactic processing in music and language rely, at least in part, on a common system (Patel, 2003)? The second possibility gains some indirect support from a number of parallels between linguistic and musical structure. Both music and language can be characterized as hierarchical rule-based systems, and similar theories can be used to describe structural organization in both domains. In an influential set of talks, Leonard Bernstein (1976) linked musical structure to generative linguistic theory, leading to the development of several explicit theories of musical structure that draw on linguistic formalisms. The most well-known theory of this type is Lerdahl and Jackendoff's (1983) generative theory of tonal music (see also Hamanaka, Hirata, & Tojo, 2006; Lerdahl, 2001), but other linguistically motivated analyses of musical structure have been proposed by Longuet-Higgins (1976), Katz and Pesetsky (2011), and Rohrmeier (2011). Generally speaking, these proposals link hierarchical organization of (Western tonal) music (motivated to some extent by Schenkerian analysis; Schenker, 1935/1979) to a linguistically inspired structure of rules and constraints, leading to a generative theory of harmonic structure. Of course, describing linguistic and musical structure with similar formalisms does not mean the processes themselves are related (Jackendoff, 2009; London, 2012b). Nevertheless, these formal similarities have inspired questions about relatedness between the processing of linguistic and musical structure. L. R. Slevc (*) : B. M. Okada Department of Psychology, University of Maryland, College Park, MD 20742, USA e-mail: slevc@umd.edu *Psychon Bull Rev* (2015) 22:637–652 DOI 10.3758/s13423-014-0712-4 Indeed, linguistic and musical structure are not only formally related, but also show developmental, neural, and behavioral similarities. Children implicitly learn the structure of their native language (e.g., Gómez & Gerken, 1999; Saffran, Aslin, & Newport, 2001) and their native musical system (e.g., Corrigall & Trainor, 2010; Hannon & Trainor, 2007) along similar developmental trajectories (Brandt, Gebrian, & Slevc, 2012;

McMullen & Saffran, 2004). Developmental deficits in linguistic syntax associated with specific language impairment also can affect structural processing in music (Jentschke, Koelsch, Sallat, & Friederici, 2008), supporting shared processing mechanisms. Both musical and linguistic structure are processed rapidly, and unexpected structural elements in music and in language are associated with similar electrophysiological responses (Koelsch, Gunter, Wittfoth, & Sammler, 2005b; Patel, Gibson, Ratner, Besson, & Holcomb, 1998; Sammler, Koelsch, & Friederici, 2011). In addition, manipulations of harmonic structure in fMRI paradigms show effects in brain areas typically associated with linguistic syntax including (most relevant to the following discussion) left inferior frontal regions, i.e., Broca's area (Janata, Tillmann, & Bharucha, 2002; Koelsch et al., 2002; Koelsch, Fritz, Schulze, Alsop, & Schlaug, 2005a; Minati et al., 2008; Oechslin, Van De Ville, Lazeyras, Hauert, & James, 2013; Tillmann, Janata, & Bharucha, 2003; Tillmann et al., 2006; Seger et al., 2013). These inferior frontal regions have also been implicated in the processing of rhythmic structure (Vuust, Roepstorff, Wallentin, Mouridsen, & Østergaard, 2006; Vuust, Wallentin, Mouridsen, Østergaard, & Roepstorff, 2011), and both frontal and temporal regions show equal sensitivity to temporal structure in music and speech (Abrams et al., 2011). Finally, there is a growing body of behavioral evidence linking the processing of musical and linguistic structure (e.g., Hoch, Poulin-Charronnat, & Tillmann, 2011; Fedorenko, Patel, Casasanto, Winawer, & Gibson, 2009; Slevc, Rosenberg, & Patel, 2009), as discussed below. Despite substantial evidence for similarities, it also is clear that musical and linguistic structure differ in many ways. For one, they serve quite different purposes. Linguistic structure represents propositional relationships between elements—i.e., who did what to whom. In contrast, musical structure does not reflect relational meaning but rather the relative stabilities of pitches in a tonal context and aesthetic/emotional patterns of tension and relaxation (for discussion, Jackendoff, 2009; London, 2012b). Empirically, distinct patterns of activation in recent functional neuroimaging studies of language and music (e.g., Rogalsky, Rong, Saberi, & Hickok, 2011) and double dissociations between musical and linguistic processing deficits (i.e., amusia and aphasia; see Peretz, 2006, for a review) suggest distinct neural systems underlying music and language. Although this work has not generally investigated structural processing per se, it does seem that deficits in musical structural processing can accompany preserved syntactic processing in language (Peretz, 1993) and that deficits in linguistic syntactic processing can accompany preserved processing of musical structure (Basso & Capitani, 1985).¹ Reconciling these differences with evidence for shared structural processing requires a more nuanced view of musical and linguistic structure that includes both shared and distinct elements of structure across domains. Music/language interactions and the shared syntactic integration resource hypothesis An influential reconciliation of this type is Patel's (2003; 2008; 2012) shared syntactic integration resource hypothesis (SSIRH), which claims that music and language rely on separable representations (e.g., nouns and verbs in language, tonal functions in music) but recruit a shared set of syntactic processing resources to integrate these separate representations into evolving sequences. The SSIRH is an appealing hypothesis because it can account both for similarities in the processing of musical and linguistic structure while also accounting for neuropsychological dissociations between processing of music and language. There is a growing body of evidence supporting the SSIRH, much of it relying on interference paradigms where participants are simultaneously presented with both musical and linguistic stimuli. In these paradigms, syntactic manipulations in both domains are crossed to look for interactive effects that indicate shared processing (in contrast to additive effects, which would indicate independent processes; Sternberg, 1969). For example, an electrophysiological effect characteristic of linguistic syntactic violations (the left anterior negativity, or LAN) is reduced when the linguistic manipulation is paired with a concurrent music-syntactic

irregularity (Koelsch et al., 2005b). Similarly, facilitation for syntactically expected words in a lexical decision task is reduced when paired with harmonically unexpected chords (Hoch et al., 2011), and comprehension of sung complex sentences (object relative clauses) is worse when the critical regions are sung out-of-key (Fedorenko et al., 2009; cf. Fiveash & Pammer, 2014). Slevc et al. (2009) relied on temporary syntactic ambiguities (garden path sentences), where readers are slower to comprehend the disambiguating word was in a sentence, such as “It is worth noting that, while Basso and Capitani’s (1985) patient NS did show preserved harmonic processing despite quite severe global aphasia, it is not actually clear whether his ability to process linguistic structure was deficient because his severe anomia and apraxia make it difficult to evaluate his syntactic processing abilities per se. In fact, we know of no unambiguous reports of agrammatic individuals who show preserved harmonic processing in music. In addition, there is at least some evidence that agrammatism is associated with harmonic processing deficits in online tasks (Patel et al., 2008). *Psychon Bull Rev* (2015) 22:637–652 as “The scientist proved the hypothesis was false” compared to an unambiguous context, such as “The scientist proved that the hypothesis was false.” This slowed processing presumably reflects the need to revise an initial syntactic interpretation where “the hypothesis” was interpreted as the direct object of the verb proved rather than as the subject of an embedded sentence complement (Pickering & van Gompel, 2006, for review). This garden path effect was more pronounced when the disambiguating word (was) was accompanied by a harmonically unexpected chord (but not when accompanied by a chord of unexpected timbre). Importantly, there was no such interaction between harmonic unexpectedness and semantic unexpectedness in language. That is, while reading was slowed for semantically unexpected words, such as pigs, in the sentence, “The boss warned the mailman to watch for angry pigs when delivering the mail” (compared to the expected dogs), this effect did not differ as a function of the harmonic expectancy of the chord accompanying the critical (semantically surprising) word. This suggests that the interactive effects between musical structure and language are specific to syntax. However, a more recent finding casts doubt on this last conclusion: the same harmonic manipulations used by Slevc et al. (2009) did lead to interactive effects when paired with sentences containing “semantic garden paths” (Perruchet & Poulin-Charronnat, 2013). These were sentences, such as “When the exterminator found the bug, he quickly unplugged the spy equipment from the wall,” where the reader presumably interprets the semantically ambiguous word bug as referring to an insect until encountering the disambiguating information unplugged the spy equipment. This type of sentence is analogous to a syntactic garden path in the sense that a previous interpretation must be revised (as bug actually turns out to be referring to eavesdropping equipment); however, it differs in that this revision is—critically—not structural in nature. This interaction between a harmonic manipulation and a nonstructural manipulation in language suggests that shared integration resources between music and language are not limited to syntax per se (see also Poulin-Charronnat, Bigand, Madurell, & Peereman, 2005; Steinbeis & Koelsch, 2008). One might then imagine that what drives interactions between musical and linguistic structural processing is simply sensory attention (Poulin-Charronnat et al., 2005). This account is supported by demonstrations that the effects of many types of harmonic structural manipulations can be explained in terms of plausible sensory mechanisms (Collins, Tillmann, Barrett, Delbé, & Janata, 2014) and that harmonic manipulations can influence the attention devoted to concurrent nonmusical (and non-linguistic) tasks (e.g., Escoffier & Tillmann, 2008). However, it seems unlikely that the interactions between harmonic and linguistic structure described above are due entirely to shared reliance on attentional resources for two reasons. First, nonstructural musical manipulations of timbre or amplitude—investigated as controls for attentional capture—do not interact with linguistic syntactic or semantic manipulations (Fedorenko et al., 2009;

Fiveash & Pammer, 2014; Koelsch et al., 2005b; Slevc et al., 2009). Second, although semantically surprising words presumably also capture attention, manipulations of harmonic structure have generally not been found to interact with semantic unexpectedness (Besson, Faïta, Peretz, Bonnel, & Requin, 1998; Bonnel, Faïta, Peretz, & Besson, 2001; Hoch et al., 2011; Koelsch et al., 2005b; Perruchet & Poulin-Charronnat, 2013; Slevc et al., 2009; but see Poulin-Charronnat et al., 2005; Steinbeis & Koelsch, 2008). Thus, neither processes specific to syntactic processing nor general attentional mechanisms seem to adequately predict when musical and linguistic parsing do and do not interact. Neuroimaging evidence is similarly mixed. Although musical manipulations do activate “language regions” in frontal cortex (e.g., Koelsch et al., 2005b; Minati et al., 2008; Seger et al., 2013; Tillmann et al., 2006; Vuust et al., 2011), these fMRI studies have not examined musical and linguistic manipulations in the same participants, and thus do not necessarily show that the same neural regions are involved in the processing of musical and linguistic structure (cf. Fedorenko & Kanwisher, 2009). In fact, most of the few recent studies that have included within-subjects comparisons of linguistic and musical manipulations have not found substantial overlap between neural regions implicated in the processing of language and music (but see Abrams et al., 2011). For example, Fedorenko and colleagues (Fedorenko, Behr, & Kanwisher, 2011; Fedorenko, McDermott, Norman-Haignere, & Kanwisher, 2012) used a contrast between intact sentences and lists of unconnected words (visually presented word-by-word) to define a series of language-sensitive brain regions of interest (ROIs) for each participant, and then investigated whether a musical manipulation significantly engaged those same regions. The musical manipulation—a contrast between 24 second clips of rock/pop songs and pitch- and rhythm-scrambled versions of those same clips—did not lead to significant effects in the language-ROIs (frontal or otherwise), suggesting largely separable neural processes for language and music. But even these within-participant findings are equivocal; while comparing intact sentences versus nonword lists does broadly capture linguistic syntactic and semantic processing, it is less obvious that listening to pitch- and rhythm-scrambled music results in the absence of musical processing. In addition, these cross-modality comparisons—reading words vs. listening to music—may lead to increased separation. In a related paradigm, Rogalsky et al. (2011) found that listening to novel melodies (compared to silence) showed little or no overlap with a contrast between listening to intact “jabberwocky” sentences and scrambled sentences. However, neither the musical nor linguistic contrasts revealed prefrontal activation typically associated with syntactic processing (see Psychon Bull Rev (2015) 22:637–652 639 Friederici, 2011, for a review). Nevertheless, the point remains that there is little direct evidence for colocalization of structural processing in music and language. In sum, there is a growing body of evidence for shared processing of music and language, but also a growing body of work suggesting nonoverlapping processes. This motivates a reassessment of exactly what resources might be shared (and distinct) across domains. Cognitive control as a shared resource Resources that are shared between music and language must be those that link musical structural processing to some aspects of linguistic processing but not to other aspects. Specifically, musical structure processing seems to share resources involved in processing syntactic errors (Hoch et al., 2011; Koelsch et al., 2005b; Steinbeis & Koelsch, 2008), syntactic complexity (Fedorenko et al., 2009; Fiveash & Pammer, 2014), and both syntactic and semantic garden paths (Perruchet & Poulin-Charronnat, 2013; Slevc et al., 2009), but not resources involved in processing semantically surprising words (Hoch et al., 2011; Koelsch et al., 2005b; Perruchet & Poulin-Charronnat, 2013; Slevc et al., 2009) or related to the difference between intact and scrambled sentences (e.g., Fedorenko et al., 2012). One way to characterize this distinction is that the aspects of language processing that do interact with musical structure require not only the processing of an unexpected element, but also the revision or reinterpretation of a previous commitment to

a particular (syntactic or semantic) interpretation. Aspects of language processing that do not interact with musical manipulations, in contrast, may be those that do not require reinterpretation *per se*; for example, there is no obvious need to revise a previous interpretation when encountering a semantically surprising word or any clear way to revise the structural or semantic interpretation of a scrambled sentence. Revision or reinterpretation in these cases likely relies on the detection of conflict between new information and a current incrementally constructed interpretation, and also on the resolution of this conflict by biasing activation away from a current interpretation and toward a new one. This sort of conflict detection and resolution draws on processes of cognitive control that allow for the regulation of mental activity and the ability to adjust (on-the-fly) in the face of conflicting information (Botvinick, Braver, Barch, Carter, & Cohen, 2001; Miller & Cohen, 2001). This regulation of internal representations is distinct from mechanisms of perceptual (or “external”) attention (Elton & Gao, 2014; Chun, Golomb, & Turk-Browne, 2011; Lavie, Hirst, de Fockert, & Viding, 2004; Seeley et al., 2007) and is part of the flexible, goal-directed abilities associated with the prefrontal cortex (Miller & Cohen, 2001). There are two main components of cognitive control that are associated with distinct neural regions. Monitoring for and detecting conflict is primarily associated with the dorsal anterior cingulate cortex (dACC) (Botvinick et al., 2001; Shenhav, Botvinick, & Cohen, 2013; Yeung, Botvinick, & Cohen, 2004). Conflict detection then leads to regulatory activity in the lateral prefrontal cortex (e.g., Kerns, Cohen, MacDonald, Cho, Stenger, & Carter, 2004; Kounieher, Charron, & Koechlin, 2009), with increasingly more abstract forms of control recruiting increasingly more anterior/rostral regions (following a more general “gradient of abstractness” in the prefrontal cortex; Badre & D’Esposito, 2009; Koechlin & Summerfield, 2011). The resolution of relatively abstract representational conflict (versus response conflict) is assumed to rely importantly on the left inferior frontal gyrus (LIFG) including Broca’s area (e.g., Badre & Wagner, 2007; Miller & Cohen, 2001; Novick, Trueswell, & Thompson-Schill, 2005; 2010). Given that Broca’s area—a classical language region—is involved in cognitive control, it is perhaps unsurprising that the role of cognitive control in linguistic syntactic processing is part of a larger debate on the role of Broca’s area in language (see Rogalsky & Hickok, 2011, for discussion). While cognitive control is typically investigated using nonlinguistic tasks, such as the Stroop task (MacLeod, 1991; Stroop, 1935), or memory tasks that manipulate proactive interference (Jonides, Smith, Marshuetz, Koeppe, & ReuterLorenz, 1998), aspects of linguistic parsing have been argued to critically rely on cognitive control to detect and resolve conflict that occurs when expectations are violated and interpretations must be revised (Novick et al., 2005; 2010). Conflict resolution in language can be syntactic in nature; for example, LIFG-based cognitive control processes have been implicated in resolution of syntactic conflict in garden path sentences (January, Trueswell, & Thompson-Schill, 2009; Novick, Kan, Trueswell, & Thompson-Schill, 2009). Importantly, cognitive control also is recruited to resolve nonsyntactic conflicts; for example the LIFG is recruited when resolving conflict between semantic plausibility and thematic roles (Thothathiri, Kim, Trueswell, & ThompsonSchill, 2012; Ye & Zhou, 2008; 2009), resolving competition in lexical selection (Schnur et al., 2009), and resolving semantic ambiguities (Bedny, Hulbert, & Thompson-Schill, 2007; Rodd, Johnsrude, & Davis, 2010; Vuong & Martin, 2011). These findings map relatively straightforwardly onto the cases where linguistic manipulations interact with musical structure. In particular, garden path sentences (Slevc et al., 2009) and morpho-syntactic errors (Hoch et al., 2011; Koelsch et al., 2005b) involve reinterpretation of an incrementally constructed initial syntactic analysis based on latearriving syntactic information (cf. Novick et al., 2005). 2 See the Conclusions section below for discussion of some exceptions to this generalization. 640 *Psychon Bull Rev* (2015) 22:637–652 Syntactic complexity effects (Fedorenko et al., 2009; Fiveash & Pammer, 2014) involve resolving

temporary structural ambiguities and overcoming interference when establishing complex or long-distance dependencies (Fernandez-Duque, 2009; Lewis, Vasishth, & Van Dyke, 2006), and semantic garden paths (Perruchet & Poulin-Charronnat, 2013) involve reinterpretations based on incompatible semantic interpretations of homophones (Rodd et al., 2010). Thus, studies finding interactive effects between musical structure and language (be it linguistic syntax or non-syntactic situations that require resolution between conflicting representations like semantic garden paths) may be revealing simultaneous use of cognitive control resources. Because cognitive control is important primarily when there is a need to regulate mental activity, these relationships may be most evident when listeners are actively processing music and language. Indeed, one general distinction between studies of musical (and linguistic) processing that do and do not implicate prefrontal cortical regions associated with cognitive control is that frontal activation is found in studies employing active tasks (e.g., categorization or tapping tasks), whereas studies finding no frontal involvement typically employ passive listening (but see Abrams et al., 2011; Levitin & Menon, 2003). This suggests that active processing may be a prerequisite for the involvement of control processes (cf. effects of active processing tasks in other domains, such as vision (Beauchamp, Haxby, Jennings, & DeYoe, 1999)). If music/language interactions do reflect shared reliance on cognitive control, active musical syntactic processing as measured in the studies cited above also must rely on cognitive control mechanisms. Ambiguity and cognitive control in musical structure

The claim that cognitive control is, in fact, a shared mechanism implies that aspects of music perception rely on cognitive control. Indeed, this is likely to be the case. Listening to music involves building up complex cognitive representations of musical structure over time. This involves not only the incremental processing and integration of musical elements as they occur, but also the incremental generation of musical predictions and expectations (for a recent discussion, see Rohrmeier & Koelsch, 2012). One hazard of this predictive processing is that new information can be inconsistent with one's prediction, thus harmonic processing requires both the ability to detect conflict between predicted and observed percepts and the ability to resolve this conflict by overriding and updating an evolving representation of musical structure. Conflict between musical percepts and predictions likely arises in many situations, not the least of which is cases of musical ambiguity (Bernstein, 1976; Jackendoff, 1991; Temperley, 2001; Thompson, 1983; see also Lewin, 1986). Structural ambiguity in music is common and occurs across diverse musical genres—not only in classical works (e.g., Smith, 2006; Temperley, 2001; Thompson, 1983), but also in jazz and blues (e.g., Blake, 1982; Ripani, 2006), rock music (e.g., McDonald, 2000; Hesselink, 2013), and electronic dance music (e.g., Butler, 2001; 2006). Of course, structural ambiguity is not limited to the Western musical tradition (e.g., Scherzinger, 2010; Stevens, 2012), but here we perpetuate a weakness of many cognitively oriented studies on musical structure by focusing on Western tonal music. Jackendoff (1991) distinguishes between two general accounts of how a listener could parse a musically ambiguous structure. One possibility is that parsing is serial: listeners commit to a single analysis at any point in time, choosing the most probable analysis in the face of ambiguity. When confronted with newly arriving information that is inconsistent with this parse, listeners would experience a “musical garden path” and have to revise their previous structural parse (alternatively, revision might not occur immediately, but only after sufficient evidence has accumulated). This serial parsing model is essentially analogous to the two-stage “garden path model” of sentence parsing (Frazier, 1987; Ferreira & Clifton, 1986), where the parser first forms a syntactic analysis based only on bottom-up information, then revises based on other available information (if necessary) in a second stage. Alternatively, musical parsing might be parallel, where multiple structural hypotheses are entertained at any given point, with more likely analyses (i.e., those that are better supported by any available data) given more weight. This

is analogous to interactive constraint-based (or constraint-satisfaction) models of sentence parsing (e.g., MacDonald, Pearlmutter, & Seidenberg, 1994; McClelland, St. John, & Taraban, 1989) where all possible sentence analyses are activated in parallel, to the extent that they are supported by all available sources of information.³ Of course, a third possibility is that listeners do not resolve musical ambiguity at all and simply do not assume structural coherence (cf. Cook, 1987; Tillmann, Bigand, & Madurell, 1998). Under either serial or parallel accounts of ambiguity resolution, when a musical piece provides new information that is inconsistent with a first or a dominant analysis, that primary analysis may need to be revised (or activation of alternative analyses adjusted) to incorporate this new information. The detection of conflict between these structural analyses and the revision of a previously formed musical interpretation in light of newly arriving information are exactly the sort of processes served by cognitive control. There are many types of musical 3 It seems unlikely that multiple musical (or linguistic) analyses are consciously available simultaneously; instead, musically ambiguous stimuli might be better construed as cases of multistability, such as the Necker cube, where only one interpretation can be experienced at a time (Repp, 2007). However, it remains possible that mechanisms of musical parsing construct and consider multiple analyses at some unconscious level of representation. *Psychon Bull Rev* (2015) 22:637–652 641 ambiguity that might draw on cognitive control mechanisms; we focus on ambiguity in meter, harmony, tonality, and contrapuntal structure (Temperley, 2001). Perhaps the most easily apparent form of musical ambiguity is metrical, when the apparent meter of a piece of music changes and must be reevaluated. Meter refers to the perceived organization of a series of beats, including both their cyclic pattern and additional higher levels of temporal structure. It is distinct from rhythmic grouping in that it relies on our endogenous perception of musical rhythm (as can be seen, for example, by our ability to synchronize to syncopated rhythms where the acoustic signal may not correspond to the beat). Meter perception may be driven by entrainment (Repp, 2007) and temporal expectancies (Large & Palmer, 2002; London, 2012a). Because of the predictive and entraining nature of metrical perception, listeners not only interpret incoming music in terms of a metrical structure, but form expectations and predictions about future metrical events. A melodic line is metrically ambiguous when it can be perceived in one of several possible meters (Fig. 1). In such cases, an ambiguous stimulus is presumably interpreted with the most plausible meter until later information conflicts with that first metrical interpretation (Jackendoff, 1991; Temperley, 2001). In order to form a coherent structure of the piece overall, the listener must resolve the conflict between the new musical information and the currently entrained/ predicted pattern; this detection and reconstruing of meter forms a type of “rhythmic garden path,” as illustrated in Fig. 2. 4 To our knowledge, there has been only one attempt to investigate whether listeners actually resolve a disambiguated metrical interpretation: Vazan and Schober (2004) asked listeners to tap along to a song where an ambiguous rhythm is strongly biased toward a triple meter but later resolves to a duple meter (“Murder by Numbers” by The Police). Over multiple rehearsals, only a few participants showed evidence of having reinterpreted the initial rhythmic structure (by tapping in duple meter from the beginning), suggesting that many listeners do not successfully revise metrical ambiguities, at least in this particular song (Vazan & Schober, 2004). Note, however, that metrical ambiguity is not always disambiguated or resolved; some types of music may actively engage listeners precisely because of long-lasting ambiguity in meter (e.g., Butler, 2006). Managing these multiple interpretations also is likely to draw on cognitive control mechanisms. Consistent with this claim, keeping a specific rhythm in a polyrhythmic context engages the LIFG, an area often associated with cognitive control (Vuust et al., 2006; 2011). In fact, Vuust and colleagues speculate that “the inferior frontal lobe is crucially involved in processing discrepancy or tension between the anticipatory neuronal model and relevant

features of the incoming stimuli, be it in language, music or other communicational systems.” (Vuust et al., 2011, p. 216). Musical ambiguity can occur in harmonic structure as well (cf. Lewin, 1986). Figure 3 shows an example of a chord that, heard in isolation, can be perceived as either a C Major chord or an A minor chord, because it only contains two pitches: C and E. The notes C-E-G would make a C Major chord and the notes A-C-E would make an A minor chord. However, these types of chords are rarely perceived as ambiguous, because they are usually interpreted within their surrounding harmonic context. In Fig. 3, the interpretation of this two-note chord is colored by the context. In the context of 3a, the chord is perceived as C Major, but the same chord, in the context of 3b, is perceived as A minor. A closely related form of ambiguity is tonal ambiguity. In contrast to harmonic ambiguity, which refers to individual ambiguous chords (Fig. 3), tonal ambiguity deals with a piece’s overall key. Just as listeners build up expectations of metrical structure, they also predict information about the tonal structure of an evolving musical piece. Changes in musical structure often occur with diatonic pivot chords, which are common to at least two different keys (and are thus harmonically ambiguous—when heard in isolation, they alone do not establish a key). Pivot chords can serve as a smooth transition between two keys, because they are harmonically appropriate in either key. For example, the circled chord in Fig. 4 acts as a minor six chord (vi6) in the key of C Major, but also as a minor two chord (ii6) in the new key of G Major. In the case of a pivot chord modulation (and most other types of modulation), the pivot chord (e.g., the A minor chord in Fig. 4) is initially interpreted as belonging to the original key. However, the following chords are unambiguously in another key, which may lead listeners to reinterpret the pivot chord and to revise their analysis of the musical key as the music continues. If listeners do, in fact, reinterpret both the pivot chord itself and the tonal center of the piece from the pivot chord onward, this can be characterized as a “tonal garden path,” which likely relies on the information recharacterization processes of cognitive control. This sort of tonal garden path is likely not limited to diatonic pivot chords, but may instead result from any sort of harmonic change that requires reevaluation of a previous tonal analysis. The harmonic manipulations that lead to music/language interactions are of this sort: both relatively coarse manipulations of musical key (e.g., Koelsch et al., 2005b; Slevc et al., 2009) and more subtle manipulations of tonal function (e.g., Hoch et al., 2011) likely involve reinterpretation of a previously established harmonic context, and thus draw on cognitive control. Of course, many of the manipulations used in investigations of music/language interactions are not resolvable ambiguities and it is not obvious that a harmonic context can be reinterpreted based on a single chord from another key. 4 For additional examples and a taxonomy of different types of metrical ambiguity, see Justin London’s collected list of “metric fakeouts,” available from <http://people.carleton.edu/~jlondon/>. 642 *Psychon Bull Rev* (2015) 22:637–652 However, such an unexpected tonal event likely still elicits an attempt at reconciliation, even if it is eventually abandoned. This attempt may occur as an automatic consequence of ACC-mediated conflict detection that occurs when new information conflicts with an expected tonal event (e.g., a tonic at the end of a cadence) or set of expected possibilities (e.g., possible chords from a particular key), which automatically signals prefrontal conflict resolution mechanisms. Alternatively, reinterpretation may not be so automatic, in which case one might observe reduced harmonic unexpectedness effects over the course of an experiment as participants realize that out-of-key and unresolvable chords are relatively common (although this has not been directly investigated as far as we know, it seems plausible given, e.g., evidence that participants rapidly develop expectancies based on a new musical system; Loui, Wu, Wessel, & Knight, 2009). Another type of musical ambiguity concerns the number of voices in a melodic line. This is referred to as contrapuntal ambiguity (Temperley, 2001) and draws on theories of auditory scene analysis (Bregman, 1990; see Moore & Gockel, 2012, for a review). When listening to music, we hear it as

coming from one or more sources, or streams. Fission (stream segregation) describes perception of a sequence of sounds as two or more separate streams. Conversely, fusion describes perception of a sequence of sounds as a single stream. Differences in pitch, loudness, timing, and timbre all affect how one perceives auditory streams (e.g., Iverson, 1995; Micheyl, Hanson, Demany, Shamma, & Oxenham, 2013); for instance, listeners may perceptually group notes that are most proximal in pitch, thus more distant pitches tend to be heard as two segregated streams. An example of this is shown in Fig. 5a, where the music comes from a single source, but the differences in pitch induce the listener to segregate the sequence into two streams (for related examples, see Deutsch, 1987; 1999; Dowling, Lung, & Herrbold, 1987). This type of contrapuntal ambiguity can occur in fugues, which contain multiple voices. For instance, the subject in the first three measures of Fig. 5b could initially be perceived as two voices (notated in dark blue and light blue) until the arrival of the answer (in red) in measure four. At this point, the listener may revise this segregated perception of the first voice (the subject) into a single fused interpretation, with the new information in the answer now interpreted as a second voice. This revision of melodic voices into fused or segregated Fig. 1 (a) A melodic line that can be perceived with different metrical analyses. (b) Analysis of the melody in 4/4 time, with the strongest pulses on the first and third beats (the number of dots indicate the perceived strength of the pulses). (c) Analysis of the melody in 3/4 time with the strongest pulses on the downbeats of every measure. (d) Alternative analysis in 4/4 time with the first C treated as a pick up note instead of the downbeat Fig. 2 A “rhythmic garden path” in which the listener may initially perceive the ambiguous meter as 2/4 time (metric analysis a) or in 3/4 time (metric analysis b). However, upon reaching measure 4, in which the rhythm is most common to 3/4 time, one would need to reconcile the predicted metric interpretation to 3/4 time and potentially revise the interpretation of the preceding rhythm Psychon Bull Rev (2015) 22:637–652 643 sources is yet another instance that likely relies on the information recharacterization functions of cognitive control.

Evidence for a cognitive control/music link These situations of musical ambiguity and revision suggest an important role for cognitive control in musical processing; however, there is, as of yet, very little work that directly investigates if and how music perception relies on cognitive control. Some indirect evidence comes from findings that musical training is associated with advantages in cognitive control ability (Bialystok & DePape, 2009; Pallesen et al., 2010; Moreno et al., 2011; Travis, Harung, & Lagrosen, 2011; but see Schellenberg, 2011), among other types of cognitive advantages (Schellenberg & Weiss, 2012). Transfer from musical training to cognitive control is predicted only if the demands of musical processing tax (and thus potentially strengthen) cognitive control processes (cf. Hussey & Novick, 2012). If so, this “musician advantage” in cognitive control may occur because extensive training and experience with the aspects of music discussed above place additional demands on cognitive control mechanisms, thus serving as a sort of naturalistic cognitive control training (cf. discussions of enhanced cognitive control associated with bilingualism; e.g., Bialystok, Craik, Green, & Gollan, 2009).⁵ Consistent with this link, musicians have greater grey matter density than nonmusician controls in LIFG (AbdulKareem, Stancak, Parkes, & Sluming, 2011; Gaser & Schlaug, 2003; Sluming et al., 2002), an area associated with cognitive control (Badre & Wagner, 2007; Botvinick et al., 2001; Miller & Cohen, 2001). Functional neuroimaging studies that manipulate musical structure—typically in terms of tonal (Koelsch et al., 2002; Koelsch et al., 2005a; Oechslin et al., 2013; Tillmann et al., 2003; 2006; Seger et al., 2013) or rhythmic ambiguity (Vuust et al., 2006; Vuust et al., 2011) find activation in left and right lateral prefrontal areas also associated with cognitive control. This apparent overlap is illustrated in Fig. 6, which shows peak activations from these studies along with regions that are consistently reported in studies of a prototypical cognitive control task (the Stroop task, based on an automated meta-analysis from the Neurosynth database;

Yarkoni, Poldrack, Nichols, Van Essen, & Wager, 2011). Although overlap should be interpreted with caution because these data come from different studies, it does appear that frontal peak activations cluster within or near areas associated with cognitive control in both hemispheres. Given evidence for a posterior-anterior gradient of abstractness in the prefrontal cortex (see above), it is somewhat surprising that the frontal activation peaks from these few studies of musical ambiguity do not appear to be clustered in anterior regions but are spread relatively evenly across inferior frontal regions bilaterally. (In contrast, note that language processing does seem to show a posterior-anterior gradient of abstractness: phonological processing engages more posterior regions of the LIFG, namely BA 44/45, whereas semantic and syntactic processing engage more anterior regions, namely BA 45/47; e.g., Hagoort, 2005; 2013; Poldrack, Wagner, Prull, Desmond, Glover, & Gabrieli, 1999). This apparent overlap for frontal regions involved in active (task-relevant) processing of musical structure and in resolving Stroop interference is suggestive of a neural relationship between musical structure and cognitive control; however, it remains only suggestive without studies investigating these processes in the same participants (cf. January et al., 2009; Ye & Zhou, 2009). In fact, some recent work has not found significant overlap between musical and linguistic manipulations within participants (Fedorenko et al., 2011; 2012; Rogalsky et al., 2011), perhaps because these studies used passive listening instead of tasks and manipulations that would be expected to recruit cognitive control. Thus, an 5 It is important to note that these cognitive control advantages (and the neuroanatomical differences discussed below) have largely been reported in correlational studies; thus, it is possible that they reflect—at least in part—preexisting differences between people who do and do not decide to pursue musical training (e.g., Corrigan et al., 2013; but see Norton et al., 2005).

Fig. 3 The two-note chord in the first measure is harmonically ambiguous, because it contains only the notes C and E. (a) A context typical of the key of C Major, where the ambiguous chord is thus perceived as C Major. (b) A context typical of the key of A minor, where the ambiguous chord is thus perceived as A Minor

644 *Psychon Bull Rev* (2015) 22:637–652

important future direction will be to investigate potential colocalization using tasks requiring active processing and manipulations likely to lead to conflict resolution in music and language (e.g., comparing garden path sentences with musical garden paths). More direct evidence for a role of cognitive control in musical processing comes from recent findings of interference between harmonic manipulations and a classic cognitive control task (Masataka & Perlovsky, 2013; Slevc, Reitman, & Okada, 2013). These experiments relied on the Stroop effect (MacLeod, 1991; Stroop, 1935), where participants are slower to name the ink (or font) color of printed stimuli when the word and color are incongruent (e.g., the word “BLUE” printed in green font) than for neutral conditions (e.g., the string “####” printed in green font). This Stroop interference is a prototypical measure of cognitive control, as participants must override a welllearned and automatized response (reading a printed word) to produce a task-relevant (but nonautomatic) response (naming the color of the printed word). Masataka and Perlovsky (2013) found greater Stroop interference when participants heard music containing harmonically unexpected intervals compared to when they heard consonant, harmonically expected, music. Slevc et al. (2013) similarly found that participants showed significantly greater Stroop interference following short musical chorales that ended in an unexpected key compared with chorales that ended on the tonic chord. However, the Stroop effect was not larger when paired with a final chord of surprising timbre, indicating that this interaction did not reflect shared reliance on attention. Instead, these data suggest that unexpected harmonic information taxed cognitive control resources, thereby reducing the resources available to mitigate Stroop interference. These are (to our knowledge) the only direct findings linking cognitive control and musical processing, and clearly more work is needed. Nevertheless, this, combined with suggestive evidence for LIFG involvement in

musical structural processing and advantages in cognitive control associated with musical training, suggests that cognitive control may indeed play an important role in structural processing in music as well as in language. An important future direction will be to investigate the processing of musical structure in populations with limited cognitive control abilities, such as children, who show protracted development of prefrontal cortex (Huttenlocher & Dabholkar, 1997) and correspondingly protracted development of cognitive control (e.g., Bunge et al., 2002), or patients with cognitive control deficits due to constrained LIFG damage (e.g., Hamilton & Martin, 2005). These approaches have already helped elucidate the role of cognitive control in language processing (e.g., Khanna & Boland, 2010; Novick et al., 2009; Thompson-Schill et al., 2002) and are likely to provide an important window onto the cognitive control/music relationship as well.

Fig. 5 (a) A melodic line that can be perceived with different contrapuntal analyses (i.e., as coming from different numbers of voices). Because of the large differences in pitch, the blue and red notes are likely perceived as two separate streams. (b) An example of a fugue with a “contrapuntal garden path.” The first three measures (the subject of the fugue) would likely be initially perceived as two voices (notated in dark blue and light blue); however, in measure 4, when the answer (notated in red) begins (and the countersubject, notated in blue, continues) the subject and countersubject may be reinterpreted as representing a single voice.

Fig. 4 A chorale beginning in the key of C Major, which then modulates into G Major. The transition occurs via the circled A minor pivot chord, which is common to both keys: it is likely initially perceived as a vi^6 chord (i.e., is a minor six chord in C Major), but may be reinterpreted as a ii^6 chord (a minor two chord) in G Major, thus acting as a tonal garden path.

Psychon Bull Rev (2015) 22:637–652

645 Conclusions

We take the basic tenet from the SSIRH that structural processing in music and language relies on shared processing resources, but suggest that those shared resources are not limited to syntactic integration, but are rather more basic mechanisms of cognitive control that subserve both domains (cf. Novick et al., 2005; 2010). This proposal is not new, but follows earlier suggestions that music and language interactions reflect shared reliance on domain-general mechanisms (Hoch et al., 2011; Koelsch, 2012; Poulin-Charronnat et al., 2005; Tillmann, 2012; among others). However, this proposal differs from previous work: cognitive control is a different shared mechanism than attentional resources (e.g., Chun et al., 2011; Seeley et al., 2007), and conflict resolution and reinterpretation is a more mechanistic explanation than shared mechanisms of structural and temporal integration. An underlying reliance on cognitive control thus has somewhat more explanatory power: it predicts both when interactions between music and language arise (specifically, when harmonic and linguistic reinterpretation co-occur) and when harmonic and linguistic manipulations produce independent effects (e.g., with manipulations that are surprising but produce relatively little need for conflict resolution and reinterpretation, such as manipulations of musical timbre or amplitude or semantically improbable words). Note, however, that not all evidence clearly fits this prediction. Although most work has not found interactions between the processing cost of semantically unexpected words (i.e., words with low cloze probability) and structural manipulations in music (Besson et al., 1998; Bonnel et al., 2001; Hoch et al., 2011; Koelsch et al., 2005b; Perruchet & Poulin-Charronnat, 2013; Slevc et al., 2009), there are two studies that have found such interactive effects. Poulin-Charronnat et al. (2005) found harmonic priming effects (i.e., faster responses to an expected tonic chord than a less expected subdominant chord) only when an accompanying sentence ended on an expected (high cloze) word; harmonic priming was absent when the sentence ended in a semantically unexpected way (but see Hoch et al., 2011). Steinbeis and Koelsch (2008) reported a similar pattern: an ERP effect associated with harmonic unexpectancy (the N500) was reduced when paired with a semantically unexpected sentence ending; however, an ERP signature of semantic unexpectancy (the N400) was not affected by a harmonically

unexpected chord. These findings suggest an asymmetrical relationship between musical structure and semantic comprehension such that semantically surprising words can draw cognitive or attentional resources away from chord processing, but unexpected chords do not appear to distract from processing of linguistic meaning (at least in the nonmusician participants tested in these paradigms; cf. Loui & Wessel, 2007). This suggests that effects of semantic unexpectedness on harmonic processing may reflect asymmetric attentional demands (cf. Poulin-Charronnat et al., 2005), whereas the effects of harmonic processing on linguistic reinterpretation reflect additional demands on cognitive control as argued above. Fig. 6 Regions consistently reported in fMRI studies of the Stroop task—a prototypical measure of cognitive control—and locations of peak activations from fMRI studies of harmonic and rhythmic ambiguity. The activation map of the Stroop task comes from an automated meta-analysis of 101 studies from the Neurosynth database (forward inference map with a threshold of $p < 0.05$ and FDR corrections for multiple comparisons downloaded 6/17/2014 from <http://neurosynth.org>; Yarkoni et al., 2011). Blue circles indicate peak activations from six fMRI studies of harmonic structure (Koelsch et al., 2002; Koelsch et al., 2005a; Oechslin et al., 2013; Tillmann et al., 2003; 2006; Seger et al., 2013) and green circles indicate peak activations from two fMRI studies of rhythmic ambiguity (Vuust et al., 2006; Vuust et al., 2011).

646 *Psychon Bull Rev* (2015) 22:637–652 A second (nonexclusive) possibility is that there is an important distinction between the types of musical manipulations used in studies where harmonic/semantic interactions have and have not been found. Experiments reporting harmonic/semantic interactions manipulated the expectancy of a chord at the end of a cadence (i.e., tonic vs. non-tonic; Poulin-Charronnat et al., 2005; Steinbeis & Koelsch, 2008), whereas most cases where harmonic/semantic interactions have not been found manipulated the occurrence of an incongruous chord embedded within an otherwise harmonically consistent context. This may indicate an important distinction between the expectation of a cadential figure (i.e., the facilitative effect of a tonic chord after a dominant at the end of the sequence) and the broader expectancy induced by an activated tonal hierarchy (i.e., the processing cost imposed by a midsequence chord from an unexpected key). Featherstone, Morrison, Waterman, and MacGregor (2013) make a similar distinction in an attempt to reconcile conflicting electrophysiological patterns associated with harmonic manipulations: they differentiate resolved harmonic incongruities, where there is a return to the original key following an incongruous element (as in within-sequence manipulations), from unresolved incongruities, where there is no such return (as in final-chord manipulations). Resolved harmonic incongruities are associated with a late positive ERP component characteristic of reanalysis, perhaps reflecting an attempt to integrate the unexpected element into its local context via engagement of cognitive control. Unresolved incongruities, however, are not associated with late positive waves, but instead typically associated with a negative component (i.e., the N500). This suggests that cognitive control mechanisms may be engaged primarily for within-sequence manipulations, and sequencefinal manipulations might instead reflect engagement of more general aspects of sensory attention (cf. end-of-sentence wrapup effects; Just & Carpenter, 1980). Of course, it is also possible that processing a semantically anomalous word does, in fact, draw somewhat on cognitive control to resolve conflict between a predicted and actual word but that this resolution is relatively undemanding and so leads to relatively little cost. If so, semantic unexpectedness might interact only weakly with harmonic manipulations; in support of this final possibility, Hoch et al. (2011) point out that many of the reported null interactions between semantic and harmonic expectancy are, numerically, suggestive of such effects. There is clearly need for more work to test exactly when and how specific aspects of musical and linguistic processing interact (cf. Koelsch, 2012). Additional research also is needed to determine if (and if so, when) temporary musical ambiguities are indeed reinterpreted (i.e., if listeners do in fact experience “musical garden

paths”; Vazan & Schober, 2004). Nevertheless, it is striking that interactive effects have been demonstrated in precisely those situations where conflict resolution and revision likely play an important role. This does not, of course, imply that the only resource shared between music and language is cognitive control; both language and music involve processing complex structural relationships that likely place demands on a variety of cognitive abilities. For example, it is clear that perceptual attention plays a role in both domains (e.g., Escoffier & Tillmann, 2008; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). Two other systems that are particularly likely to play a role in both domains are implicit learning and working memory. Implicit learning plays an important role in the acquisition of complex structural knowledge, both in language (e.g., Saffran, Aslin, & Newport, 1996; see Kuhl, 2004, for a review) and in music (e.g., Ettlinger, Margulis, & Wong, 2011; Loui, Wu, Wessel, & Knight, 2009; Loui, Wessel, & Husdon Kam, 2010; Loui, 2012; Rohrmeier & Rebuschat, 2012). Support for shared reliance on implicit learning mechanisms comes from the finding that musical training (which presumably places additional demands on implicit learning mechanisms) leads to better implicit learning of both musical and linguistic structure (Francois & Schön, 2011). Working memory also has been linked to processing of syntax in language (e.g., Just & Carpenter, 1992; Lewis et al., 2006) and in music (Koelsch et al., 2009; Schulze, Zysset, Mueller, Friederici, & Koelsch, 2011; Williamson et al., 2010) and is associated with the inferior frontal regions that are implicated in both domains (Koelsch et al., 2009; Schulze et al., 2011; but see Fedorenko et al., 2011). The role inferior frontal regions (and especially Broca’s area) play in structural processing is controversial (see Rogalsky & Hickok, 2011, for discussion from the language perspective); LIFG, in particular, has been associated with cognitive control and working memory but also has been claimed to support syntax-specific processes (at least in language; e.g., Grodzinsky & Santi, 2008) and/or more general types of complex hierarchical relationships, such as action sequences (e.g., Farag et al., 2010; Fitch & Martins, 2014; Koechlin & Jubault, 2006) and mathematical structure (Friedrich & Friederici, 2009; Maruyama, Pallier, Jobert, Sigman, & Dehaene, 2012). Thus, these frontal regions that may be associated with shared musical/linguistic processing likely reflect a variety of underlying cognitive processes; a greater understanding of the ways in which linguistic and musical manipulations involve LIFG (and its righthemisphere homologue) will likely add important data to this debate. A related prediction is that the processing of both music and language should interact with—and show neural overlap with—other domains that rely on cognitive control mechanisms. There is already evidence for some relationships of this type; for example, structural processing in music interacts with arithmetic processing (Hoch & Tillmann, 2012) and with Stroop interference (Masataka & Perlovsky, 2013; Slevc et al., 2013). In addition, “action syntax,” or meaningful structured sequences of actions, may be related to structural processing in both music and language (Harding et al., 2011; Fazio et al., 2009; Fadiga, Craighero, & D’ausilio, 2009; Fitch & Martins, 2014; Jackendoff, 2009; Sammler, Novembre, Koelsch, & Keller, 2013). Of course, cognitive control processes are not restricted to LIFG; it is clear that both left and right frontal mechanisms are involved in cognitive control (e.g., Aron, 2008; Gläscher et al., 2012), including in the sorts of revision-demanding situations discussed here (e.g., there is bilateral IFG involvement in the processing of lexical ambiguity; Klepousniotou, Gracco, & Pike, 2013). Because musical manipulations often involve bilateral frontal activation (see Fig. 6, and Koelsch, 2011, for review), musical processing may be particularly well suited to investigate the role of right frontal regions in complex cognition. The claim that interactive effects of musical and linguistic structure reflect conflict resolution and revision via a shared reliance on cognitive control mechanisms can be taken in at least two ways. One conclusion might be that linguistic and musical syntax are largely distinct, domain-specific “competence” systems that

can place similar “performance” demands on domain-general cognitive processes (cf. Chomsky, 1965). This fits with the idea that language and music are domainspecific modular systems that only interact with general cognitive abilities in limited ways (e.g., Fedorenko et al., 2011; Peretz & Coltheart, 2003). Alternatively, one could conclude that linguistic syntax, musical structure, action sequences, and the like are all assemblies of more general cognitive processes. To borrow a phrase from Liz Bates, both language and music might be viewed as “new machine[s] constructed entirely out of old parts” (Bates & MacWhinney, 1989, p. 10). By this second theory, there may be few (or even no) processes specific to linguistic or musical parsing per se; instead both may recruit an assembly of more basic underlying cognitive mechanisms to deal with similar cognitive demands. This debate has a long and sometimes acrimonious history; however, both theoretical approaches will benefit from more specific theories of the cognitive demands imposed by musical and linguistic structure, data from more sophisticated experimental techniques (e.g., Grahn, 2012), and insights from developmental perspectives (cf. Brandt et al., 2012; Hannon & Trainor, 2007; McMullen & Saffran, 2004). Music and language are complex, multifaceted systems, and research on their relationship is beginning to go beyond questions of shared versus distinct processing to question which specific aspects of structural processing in music and language recruit shared cognitive and neural systems, and what those systems might be. We believe this change in focus is important and that a deeper understanding of the cognitive and neural basis of these domains is impossible without moving away from monolithic conceptions of “music” and “language.” Instead, we advocate a reductionist approach to investigate the specific cognitive demands imposed by different aspects of music and language and/or imposed by any other type of complex cognitive system. We take a step in this direction by proposing that the ability to flexibly control our behavior and cognition (i.e., cognitive control) plays a critical role in resolving conflict and allowing for reinterpretation in both music and language.

Acknowledgments The authors thank David Bashwiner, Nina Hsu, Eugene Montague, Mattson Ogg, Aniruddh Patel, Elizabeth Redcay, and Jason Reitman for helpful comments about earlier versions of this manuscript.

References Abdul-Kareem, I. A., Stancak, A., Parkes, L. M., & Sluming, V. (2011). Increased gray matter volume of left Pars Opercularis in male orchestral musicians correlate positively with years of musical performance. *Journal of Magnetic Resonance Imaging*, 33(1), 24– 32. doi:10.1002/jmri.22391

Abrams, D. A., Bhatara, A., Ryali, S., Balaban, E., Levitin, D. J., & Menon, V. (2011). Decoding temporal structure in music and speech relies on shared brain resources but elicits different fine-scale spatial patterns. *Cerebral Cortex*, 21(7), 1507–1518. doi:10.1093/cercor/ bhq198

Aron, A. R. (2008). Progress in executive-function research. *Current Directions in Psychological Science*, 17(2), 124–129.

Badre, D., & D’Esposito, M. D. (2009). Is the rostro-caudal axis of the frontal lobe hierarchical? *Nature Reviews Neuroscience*, 10, 659– 669.

Badre, D., & Wagner, A. D. (2007). Left ventrolateral prefrontal cortex and the control of memory. *Neuropsychologia*, 45(13), 2883–2901.

Basso, A., & Capitani, E. (1985). Spared musical abilities in a conductor with global aphasia and ideomotor apraxia. *Journal of Neurology, Neurosurgery and Psychiatry*, 48(5), 407–412. doi:10.1136/jnnp.48. 5.407

Bates, E., & MacWhinney, B. (1989). Functionalism and the competition model. In B. MacWhinney & E. Bates (Eds.), *The Crosslinguistic Study of Sentence Processing* (pp. 3–73). New York: Cambridge University Press.

Beauchamp, M. S., Haxby, J. V., Jennings, J. E., & DeYoe, E. A. (1999). An fMRI version of the Farnsworth-Munsell 100-hue test reveals multiple color-selective areas in human ventral occipitotemporal cortex. *Cerebral Cortex*, 9, 257–263.

Bedny, M., Hulbert, J. C., & Thompson-Schill, S. L. (2007). Understanding words in context: The role of Broca’s area in word comprehension. *Brain Research, Special Issue: Mysteries of Meaning*, 1146, 101–114.

Bernstein, L. (1976). *The Unanswered Question: Six Talks at Harvard*. Cambridge: Harvard University Press.

Besson, M., Faïta, F., Peretz, I., Bonnel, A.-

M., & Requin, J. (1998). Singing in the brain: Independence of lyrics and tunes. *Psychological Science*, 9(6), 494–498.

Bialystok, E., Craik, F. I. M., Green, D. W., & Gollan, T. H. (2009). Bilingual minds. *Psychological Science in the Public Interest*, 10(3), 89–129. doi:10.1177/1529100610387084

Bialystok, E., & DePape, A. (2009). Musical expertise, bilingualism, and executive functioning. *Journal of Experimental Psychology: Human Perception and Performance*, 35(2), 565–574.

Blake, R. (1982). The Monk piano style. In R. van der Blik (Ed.), *The Thelonious Monk Reader* (pp. 248–260). New York: Oxford University Press.

Bonnell, A.-M., Faita, F., Peretz, I., & Besson, M. (2001). Divided attention between lyrics and tunes of operatic songs: Evidence for independent processing. *Perception & Psychophysics*, 63(7), 1201–1213.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, 108(3), 624–652.

Brandt, A. K., Gebrian, M., & Slevc, L. R. (2012). Music and early language acquisition. *Frontiers in Psychology*, 3, 1–17. doi:10.3389/fpsyg.2012.00327

Bregman, A. S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound*. Cambridge: MIT Press.

Bunge, S. A., Dudukovic, N. M., Thomason, M. E., Vaidya, C. J., & Gabrieli, J. D. E. (2002). Immature frontal lobe contributions to cognitive control in children: Evidence from fMRI. *Neuron*, 33, 301–311.

Butler, M. J. (2001). Turning the beat around: Reinterpretation, metrical dissonance, and asymmetry in electronic dance music. *Music Theory Online*, 7(6).

Butler, M. J. (2006). *Unlocking the groove: Rhythm, meter, and musical design in electronic dance music*. Bloomington: Indiana University Press.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge: MIT Press.

Chun, M. M., Golomb, J. D., & Turk-Browne, N. B. (2011). A taxonomy of external and internal attention. *Annual Review of Psychology*, 62, 73–101.

Collins, T., Tillmann, B., Barrett, F. S., Delbé, C., & Janata, P. (2014). A combined model of sensory and cognitive representations underlying tonal expectations in music: From audio signals to behavior. *Psychological Review*, 121(1), 33–65.

Cook, N. (1987). The perception of large-scale tonal closure. *Music Perception*, 5(2), 197–205.

Corrigall, K. A., Schellenberg, E. G., & Misura, N. M. (2013). Music training, cognition, and personality. *Frontiers in Psychology*, 4, 1–10. doi:10.3389/fpsyg.2013.00222

Corrigall, K. A., & Trainor, L. J. (2010). Musical enculturation in preschool children: Acquisition of key and harmonic knowledge. *Music Perception*, 28(2), 195–200.

Deutsch, D. (1987). Illusions for stereo headphones. *Audio Magazine*, 36–48.

Deutsch, D. (1999). Grouping mechanisms in music. In D. Deutsch (Ed.), *The psychology of music* (2nd ed., pp. 299–348). San Diego, CA: Academic Press.

Dowling, W. J., Lung, K. M., & Herrbold, S. (1987). Aiming attention in pitch and time in the perception of interleaved melodies. *Perception and Psychophysics*, 41(6), 642–656.

Elton, A., & Gao, W. (2014). Divergent task-dependent functional connectivity of executive control and salience networks. *Cortex*, 51, 56–66. doi:10.1016/j.cortex.2013.10.012

Escoffier, N., & Tillmann, B. (2008). The tonal function of a task-irrelevant chord modulates speed of visual processing. *Cognition*, 107(3), 1070–1083.

Ettlinger, M., Margulis, E. H., & Wong, P. C. M. (2011). Implicit memory in music and language. *Frontiers in Psychology: Auditory Cognitive Neuroscience*, 2, 1–10. doi:10.3389/fpsyg.2011.00211

Fadiga, L., Craighero, L., & D’Ausilio, A. (2009). Broca’s area in language, action, and music. *Annals of the New York Academy of Sciences*, 1169, 448–458.

Farag, C., Troiani, V., Bonner, M., Powers, C., Avants, B., Gee, J., & Grossman, M. (2010). Hierarchical organization of scripts: Converging evidence from fMRI and frontotemporal degeneration. *Cerebral Cortex*, 20(10), 2453–2463.

Fazio, P., Cantagallo, A., Craighero, L., D’Ausilio, A., Roy, A. C., Pozzo, T., ... Fadiga, L. (2009). Encoding of human action in Broca’s area. *Brain*, 132(7), 1980–1988.

Featherstone, C. R., Morrison, C. M., Waterman, M. G., & MacGregor, L. J. (2013). Semantics, syntax, or neither? A case for resolution in the interpretation of N500 and P600 responses to harmonic incongruities. *PLoS ONE*, 8(11), 1–

13. doi:10.1371/journal.pone.0076600 Fedorenko, E., Behr, M. K., & Kanwisher, N. (2011). Functional specificity for high-level linguistic processing in the human brain. *Proceedings of the National Academy of Sciences*, 108, 16428–16433. doi:10.1073/pnas.1112937108

Fedorenko, E., & Kanwisher, N. (2009). Neuroimaging of language: Why hasn't a clearer picture emerged? *Language and Linguistics Compass*, 3, 839–865. doi:10.1111/j.1749-818x.2009.00143.x

Fedorenko, E., McDermott, J. H., Norman-Haignere, S., & Kanwisher, N. (2012). Sensitivity to musical structure in the human brain. *Journal of Neurophysiology*, 108(12), 3289–3300. doi:10.1152/jn.00209.2012

Fedorenko, E., Patel, A. D., Casasanto, D., Winawer, J., & Gibson, E. (2009). Structural integration in language and music: Evidence for a shared system. *Memory and Cognition*, 37(1), 1–9. doi:10.3758/MC.37.1.1

Fernandez-Duque, D. (2009). Cognitive and neural underpinnings of syntactic complexity. In T. Givon & M. Shibatani (Eds.), *The genesis of syntactic complexity* (pp. 433–460). Philadelphia, PA: John Benjamins Publishing.

Ferreira, F., & Clifton, C., Jr. (1986). The independence of syntactic processing. *Journal of Memory and Language*, 25(3), 348–368.

Fitch, W. T., & Martins, M. D. (2014). Hierarchical processing in music, language, and action: Lashley revisited. *Annals of the New York Academy of Sciences*, 1–18. doi:10.1111/nyas.12406

Fiveash, A., & Pammer, K. (2014). Music and language: Do they draw on similar syntactic working memory resources? *Psychology of Music*, 42(2), 190–209.

Francois, C., & Schön, D. (2011). Musical expertise boosts implicit learning of both musical and linguistic structures. *Cerebral Cortex*, 21(10), 2357–2365.

Frazier, L. (1987). Sentence processing: A tutorial review. In M. Colheart (Ed.), *Attention and performance XII* (pp. 559–586). Hillsdale, NJ: Lawrence Erlbaum Associates.

Friederici, A. D. (2011). The brain basis of language processing: From structure to function. *Physiological Reviews*, 91(4), 1357–1392.

Friedrich, R., & Friederici, A. D. (2009). Mathematical logic in the human brain: syntax. *PLoS ONE*, 4(5), 1–7.

Gaser, C., & Schlaug, G. (2003). Brain structures differ between musicians and non-musicians. *The Journal of Neuroscience*, 23(27), 9240–9245.

Gläscher J., Adolphs, R., Damasio, H., Bechara, A., Rudrauf, D., Calamia, M., ... Tranel, D. (2012). Lesion mapping of cognitive control and value-based decision making in the prefrontal cortex. *Proceedings of the National Academy of Sciences*, 109(36), 14681–14686.

Gomez, R. L., & Gerken, L. (1999). Artificial grammar learning by 1- year-olds leads to specific and abstract knowledge. *Cognition*, 70, 109–135.

Grahn, J. A. (2012). Advances in neuroimaging techniques: Implications for the shared syntactic integration resource hypothesis. In P. Rebuschat, M. Rohrmeier, J. Hawkins, & I. Cross (Eds.), *Language and music as cognitive systems* (pp. 235–241). Oxford: Oxford University Press.

Grodzinsky, Y., & Santi, A. (2008). The battle for Broca's region. *Trends in Cognitive Science*, 12(12), 474–480.

Hagoort, P. (2005). On Broca, brain, and binding: A new framework. *Trends in Cognitive Sciences*, 9(9), 416–423.

Hagoort, P. (2013). MUC (memory, unification, control) and beyond. *Frontiers in Psychology*, 4, 416. doi:10.3389/fpsyg.2013.00416

Hamanaka, M., Hirata, K., & Tojo, S. (2006). Implementing “a generative theory of tonal music”. *Journal of New Music Research*, 35(4), 249–277.

Hamilton, A. C., & Martin, R. C. (2005). Dissociations among tasks involving inhibition: A single case study. *Cognitive, Affective and Behavioral Neuroscience*, 5(1), 1–13.

Psychon Bull Rev (2015) 22:637–652

Hannon, E. E., & Trainor, L. J. (2007). Music acquisition: Effects of enculturation and formal training on development. *Trends in Cognitive Sciences*, 11(11), 466–472.

Harding, E., Sammler, D., D'Ausilio, A., Friederici, A., Fadiga, L., & Koelsch, S. (2011). Explicit action perception shares resources with music syntax: A controlled behavioral study. Poster presented at The Neurosciences and Music IV: Learning and Memory, Edinburgh, UK.

Hesselink, N. D. (2013). Radiohead's “pyramid song”: Ambiguity, rhythm, and participation. *Music Theory Online*, 19, 1–25.

Hoch, L., Poulin-Charronnat, B., & Tillmann, B. (2011). The influence of task-irrelevant music on language processing: Syntactic and semantic structures. *Frontiers in*

Psychology, 2, 1–10. doi:10.3389/fpsyg.2011.00112 Hoch, L., & Tillmann, B. (2012). Shared structural and temporal integration resources for music and arithmetic processing. *Acta Psychologica*, 140(3), 230–235. Hussey, E. K., & Novick, J. M. (2012). The benefits of executive control training and the implications for language processing. *Frontiers in Psychology*, 3, 1–14. doi:10.3389/fpsyg.2012.00158 Huttenlocher, P. R., & Dabholkar, A. S. (1997). Regional differences in synaptogenesis in human cerebral cortex. *Journal of Comparative Neurology*, 387(2), 167–178. Iverson, P. (1995). Auditory stream segregation by musical timbre: Effects of static and dynamic acoustic attributes. *Journal of Experimental Psychology*, 21(4), 751–763. Jackendoff, R. S. (1991). Musical parsing and musical affect. *Music Perception*, 9(2), 199–230. Jackendoff, R. S. (2009). Parallels and nonparallels between language and music. *Music Perception*, 26(3), 195–204. doi:10.1525/MP.2009.26.3.195 Janata, P., Tillmann, B., & Bharucha, J. J. (2002). Listening to polyphonic music recruits domain-general attention and working memory circuits. *Cognitive, Affective and Behavioral Neuroscience*, 2(2), 121–140. January, D., Trueswell, J. C., & Thompson-Schill, S. L. (2009). Colocalization of stroop and syntactic ambiguity resolution in Broca's area: Implications for the neural basis of sentence processing. *Journal of Cognitive Neuroscience*, 21(12), 2434–2444. Jentschke, S., Koelsch, S., Sallat, S., & Friederici, A. D. (2008). Children with specific language impairment also show impairment of music syntactic processing. *Journal of Cognitive Neuroscience*, 20(11), 1940–1951. Jonides, J., Smith, E. E., Marshuetz, C., Koeppe, R. A., & Reuter-Lorenz, P. A. (1998). Inhibition in verbal working memory revealed by brain activation. *Proceedings of the National Academy of Sciences of the United States of America*, 95(14), 8410–8413. Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, 87(4), 329–354. Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, 99(1), 122–149. Katz, J., & Pesetsky, D. (2011). The Identity Thesis for Language and Music. Retrieved from <http://ling.auf.net/lingbuzz/000959> Kerns, J. G., Cohen, J. D., MacDonald, A. W., III, Cho, R. Y., Stenger, V. A., & Carter, C. S. (2004). Anterior cingulate conflict monitoring and adjustments in control. *Science*, 303, 1023–1026. Khanna, M. M., & Boland, J. E. (2010). Children's use of language context in lexical ambiguity resolution. *The Quarterly Journal of Experimental Psychology*, 63(1), 160–193. Klepousniotou, E., Gracco, V. L., & Pike, G. B. (2013). Pathways to lexical ambiguity: fMRI evidence for bilateral fronto-parietal involvement in language processing. *Brain and Language*, 131, 56–64. Koechlin, E., & Jubault, T. (2006). Broca's area and the hierarchical organization of human behavior. *Neuron*, 50(6), 963–974. Koechlin, E., & Summerfield, C. (2011). An information theoretical approach to prefrontal executive function. *Trends in Cognitive Sciences*, 11(6), 229–236. Koelsch, S. (2011). Toward a neural basis of music perception—A review and update model. *Frontiers in Psychology*, 2, 1–20. Koelsch, S. (2012). Response to target article “Language, music, and the brain: A resource-sharing framework”. In P. Rebuschat, M. Rohrmeier, J. Hawkins, & I. Cross (Eds.), *Language and music as cognitive systems* (pp. 224–234). Oxford: Oxford University Press. Koelsch, S., Fritz, T., Schulze, K., Alsop, D., & Schlaug, G. (2005a). Adults and children processing music: An fMRI study. *NeuroImage*, 25(4), 1068–1076. Koelsch, S., Gunter, T. C., von Cramon, D. Y., Zysset, S., Lohmann, G., & Friederici, A. D. (2002). Bach speaks: A cortical “language network” serves the processing of music. *NeuroImage*, 17(2), 956–966. Koelsch, S., Gunter, T. C., Wittfoth, M., & Sammler, D. (2005b). Interaction between syntax processing in language and in music: An ERP Study. *Journal of Cognitive Neuroscience*, 17(10), 1565–1577. Koelsch, S., Schulze, K., Sammler, D., Fritz, T., Muller, K., & Gruber, O. (2009). Functional architecture of verbal and tonal working memory: An FMRI study. *Human Brain Mapping*, 30(3), 859–873. Kouneiher, F., Charron, S., & Koechlin, E. (2009). Motivation and cognitive control in the human prefrontal cortex. *Nature Neuroscience*, 12(7),

939–945. Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience*, 5(11), 831–843. Large, E. W., & Palmer, C. (2002). Perceiving temporal regularity in music. *Cognitive Science*, 26, 1–37. Lavie, N., Hirst, A., de Fockert, J. W., & Viding, E. (2004). Load theory of selective attention and cognitive control. *Journal of Experimental Psychology: General*, 133(3), 339–354. Lerdahl, F. A. (2001). *Tonal pitch space*. New York: Oxford University Press. Lerdahl, F. A., & Jackendoff, R. S. (1983). *A generative theory of tonal music*. Cambridge: MIT Press. Levitin, D. J., & Menon, V. (2003). Musical structure is processed in “language” areas of the brain: A possible role for Brodmann Area 47 in temporal coherence. *NeuroImage*, 20, 2142–2152. Lewin, D. (1986). Music theory, phenomenology, and modes of perception. *Music Perception*, 3(4), 327–392. Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences*, 10(10), 44–54. London, J. (2012a). *Hearing in time: Psychological aspects of musical meter*. New York: Oxford University Press. London, J. (2012b). Schemas, not syntax: A reply to Patel. In P. Rebuschat, M. Rohrmeier, J. Hawkins, & I. Cross (Eds.), *Language and music as cognitive systems* (pp. 242–247). Oxford: Oxford University Press. Longuet-Higgins, H. C. (1976). The perception of melodies. *Nature*, 263, 646–653. Loui, P. (2012). Learning and liking of melody and harmony: Further studies in artificial grammar. *Topics in Cognitive Science*, 4, 1–14. Loui, P., & Wessel, D. L. (2007). Harmonic expectation and affect in Western music: Effects of attention and training. *Perception & Psychophysics*, 69(7), 1084–1092. Loui, P., Wessel, D. L., & Hudson Kam, C. L. (2010). Humans rapidly learn grammatical structure in a new musical scale. *Music Perception*, 27(5), 377–388. Loui, P., Wu, E. H., Wessel, D. L., & Knight, R. T. (2009). A generalized mechanism for perception of pitch patterns. *Journal of Neuroscience*, 29(2), 454–459. MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). Lexical nature of syntactic ambiguity resolution. *Psychological Review*, 101(4), 676–703. MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, 109(2), 163–203. *Psychon Bull Rev* (2015) 22:637–652. Maruyama, M., Pallier, C., Jobert, A., Sigman, M., & Dehaene, S. (2012). The cortical representation of simple mathematical expressions. *NeuroImage*, 61(4), 1444–1460. Masataka, N., & Perlovsky, L. (2013). Cognitive interference can be mitigated by consonant music and facilitated by dissonant music. *Scientific Reports*, 3, 1–6. McClelland, J. L., St. John, M., & Taraban, R. (1989). Sentence comprehension: A parallel distributed processing approach. *Language and Cognitive Processes*, 4, 287–335. McDonald, C. (2000). Exploring modal subversions in alternative music. *Popular Music*, 19(3), 355–363. McMullen, E., & Saffran, J. R. (2004). Music and language: A developmental comparison. *Music Perception*, 21(3), 289–311. Micheyl, C., Hanson, C., Demany, L., Shamma, S., & Oxenham, A. J. (2013). Auditory stream segregation for alternating and synchronous tones. *Journal of Experimental Psychology: Human Perception and Performance*, 39(6), 1568–1580. doi:10.1037/a0032241 Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, 24, 167–202. Minati, L., Rosazza, C., D’incerti, L., Pietrocini, E., Valentini, L., Scaioli, V., ... Bruzzone, M. G. (2008). fMRI/ERP of musical syntax: comparison of melodies and unstructured note sequences. *NeuroReport*, 19(14), 1381–1385. Moore, B. C., & Gockel, H. E. (2012). Properties of auditory stream formation. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 367, 919–931. Moreno, S., Bialystok, E., Barac, R., Schellenberg, E. G., Cepeda, N. J., & Chau, T. (2011). Short-term music training enhances verbal intelligence and executive function. *Psychological Science*, 22, 1425–1433. doi:10.1177/0956797611416999 Norton, A., Winner, E., Cronin, K., Overy, K., Lee, D. J., & Schlaug, G. (2005). Are there pre-existing neural, cognitive, or motoric markers for musical ability? *Brain and Cognition*, 59(2), 124–134. Novick, J. M., Kan, I. P., Trueswell, J. C., &

Thompson-Schill, S. L. (2009). A case for conflict across multiple domains: Memory and language impairments following damage to ventrolateral prefrontal cortex. *Cognitive Neuropsychology*, 26(6), 527–567.

Novick, J. M., Trueswell, J. C., & Thompson-Schill, S. L. (2005). Cognitive control and parsing: Reexamining the role of Broca's area in sentence comprehension. *Cognitive, Affective and Behavioral Neuroscience*, 5(3), 263–281.

Novick, J. M., Trueswell, J. C., & Thompson-Schill, S. L. (2010). Broca's area and language processing: Evidence for the cognitive control connection. *Language and Linguistics Compass*, 4(10), 906–924.

Oechslin, M. S., Van De Ville, D., Lazeyras, F., Hauert, C.-A., & James, C. E. (2013). Degree of musical expertise modulates higher order brain functioning. *Cerebral Cortex*, 23, 2213–2224.

Pallesen, K. J., Brattico, E., Bailey, C. J., Korvenoja, A., Koivisto, J., Gjedde, A., & Carlson, S. (2010). Cognitive control in auditory working memory is enhanced in musicians. *PLoS ONE*, 5, 1–12.

Patel, A. D. (2003). Language, music, syntax and the brain. *Nature Neuroscience*, 6, 674–681.

Patel, A. D. (2008). *Music, language, and the brain*. New York: Oxford University Press.

Patel, A. D. (2012). Language, music, and the brain: A resource-sharing framework. In P. Rebuschat, M. Rohrmeier, J. Hawkins, & I. Cross (Eds.), *Language and music as cognitive systems* (pp. 204–223). New York: Oxford University Press.

Patel, A. D., Gibson, E., Ratner, J., Besson, M., & Holcomb, P. J. (1998). Processing syntactic relations in language and music: An event-related potential study. *Journal of Cognitive Neuroscience*, 10(6), 717–733.

Patel, A. D., Iversen, J. R., Wassenaar, M., & Hagoort, P. (2008). Musical syntactic processing in agrammatic Broca's aphasia. *Aphasiology*, 22, 776–789. doi:10.1080/02687030701803804

Peretz, I. (1993). Auditory atonalia for melodies. *Cognitive Neuropsychology*, 10(1), 21–56.

Peretz, I. (2006). The nature of music from a biological perspective. *Cognition*, 100, 1–32.

Peretz, I., & Coltheart, M. (2003). Modularity of music processing. *Nature Neuroscience*, 6(7), 688–691.

Perruchet, P., & Poulin-Charronnat, B. (2013). Challenging prior evidence for a shared syntactic processor for language and music. *Psychonomic Bulletin and Review*, 20(2), 310–317.

Pickering, M. J., & van Gompel, R. P. G. (2006). Syntactic parsing. In M. J. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (2nd ed., pp. 455–503). Amsterdam: Academic Press.

Poldrack, R. A., Wagner, A. D., Prull, M. W., Desmond, J. E., Glover, G. H., & Gabrieli, J. D. E. (1999). Functional specialization for semantic and phonological processing in the left inferior prefrontal cortex. *NeuroImage*, 10, 15–35.

Poulin-Charronnat, B., Bigand, E., Madurell, F., & Peereman, R. (2005). Musical structure modulates semantic priming in vocal music. *Cognition*, 94(3), B67–B78.

Repp, B. H. (2007). Hearing a melody in different ways: Multistability of metrical interpretation, reflected in rate limits of sensorimotor synchronization. *Cognition*, 102, 434–454.

Ripani, R. J. (2006). *The New Blue Music: Changes in Rhythm & Blues, 1950-1999*. Jackson: University Press of Mississippi.

Rodd, J. M., Johnsrude, I. S., & Davis, M. H. (2010). The role of domain-general frontal systems in language comprehension: Evidence from dual-task interference and semantic ambiguity. *Brain and Language*, 115(3), 182–188.

Rogalsky, C., & Hickok, G. (2011). The role of Broca's area in sentence comprehension. *Journal of Cognitive Neuroscience*, 23(7), 1664–1680.

Rogalsky, C., Rong, F., Saberi, K., & Hickok, G. (2011). Functional anatomy of language and music perception: Temporal and structural factors investigated using functional Magnetic Resonance Imaging. *The Journal of Neuroscience*, 31(10), 3843–3852. doi:10.1523/JNEUROSCI.4515-10.2011

Rohrmeier, M. (2011). Towards a generative syntax of tonal harmony. *Journal of Mathematics and Music*, 5(1), 35–53.

Rohrmeier, M., & Koelsch, S. (2012). Predictive information processing in music cognition. A critical review. *International Journal of Psychophysiology*, 83, 164–175.

Rohrmeier, M., & Rebuschat, P. (2012). Implicit learning and acquisition of music. *Topics in Cognitive Science*, 4, 525–553.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274, 1926–1928.

Saffran, J. R., Senghas,

A., & Trueswell, J. C. (2001). The acquisition of language by children. *Proceedings of the National Academy of Sciences*, 98(23), 12874–12875. Sammler, D., Koelsch, S., & Friederici, A. D. (2011). Are left frontotemporal brain areas a prerequisite for normal music-syntactic processing? *Cortex*, 47(6), 659–673. Sammler, D., Novembre, G., Koelsch, S., & Keller, P. E. (2013). Syntax in a pianist's hand: ERP signatures of “embodied” syntax processing in music. *Cortex*, 49(5), 1325–1339. Schellenberg, E. G. (2011). Examining the association between music lessons and intelligence. *British Journal of Psychology*, 102, 283–302. Schellenberg, E. G., & Weiss, M. W. (2012). Music and cognitive abilities. In D. Deutsch (Ed.), *The psychology of music* (3rd ed., pp. 499–550). Amsterdam: Elsevier. Schenker, H. (1935/1979). *Free composition*. New York: Longman. Scherzinger, M. (2010). Temporal geometries of an African music: A preliminary sketch. *Music Theory Online*, 16(4). Schnur, T. T., Schwartz, M. F., Kimberg, D. Y., Hirshorn, E., Coslett, H. B., & Thompson-Schill, S. L. (2009). Localizing interference during naming: Convergent neuroimaging and neuropsychological Psychon Bull Rev (2015) 22:637–652 651 evidence for the function of Broca's area. *Proceedings of the National Academy of Sciences*, 106(1), 322–327. Schulze, K., Zysset, S., Mueller, K., Friederici, A. D., & Koelsch, S. (2011). Neuroarchitecture of verbal and tonal working memory in nonmusicians and musicians. *Human Brain Mapping*, 32(5), 771–783. Seeley, W. W., Menon, V., Schatzberg, A. F., Keller, J., Glover, G. H., Kenna, H., ... Greicius, M. D. (2007). Dissociable intrinsic connectivity networks for salience processing and executive control. *The Journal of Neuroscience*, 27(9), 2349–2356. doi:10.1523/JNEUROSCI.5587-06.2007 Seger, C. A., Spiering, B. J., Sares, A. G., Quraini, S. I., Alpeter, C., David, J., & Thaut, M. H. (2013). Corticostriatal contributions to musical expectancy perception. *Journal of Cognitive Neuroscience*, 25(7), 1062–1077. Shenhav, A., Botvinick, M. M., & Cohen, J. D. (2013). The expected value of control: An integrative theory of anterior cingulate cortex function. *Neuron*, 79, 217–240. doi:10.1016/j.neuron.2013.07.007 Slevc, L. R. (2012). Language and music: Sound, structure, and meaning. *Wiley Interdisciplinary Reviews: Cognitive Science*, 3(4), 483–492. Slevc, L. R., Reitman, J. G., & Okada, B. M. (2013). Syntax in music and language: The role of cognitive control. *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 3414–3419). Berlin, Germany: Cognitive Science Society. Slevc, L. R., Rosenberg, J. C., & Patel, A. D. (2009). Making psycholinguistics musical: Self-paced reading time evidence for shared processing of linguistic and musical syntax. *Psychonomic Bulletin and Review*, 16(2), 374–381. doi:10.3758/16.2.374 Sluming, V., Barrick, T., Howard, M., Cezayirli, E., Mayes, A., & Roberts, N. (2002). Voxel-based morphometry reveals increased gray matter density in Broca's area in male symphony orchestra musicians. *NeuroImage*, 17(3), 1613–1622. Smith, P. H. (2006). You reap what you sow: Some instances of rhythmic and harmonic ambiguity in Brahms. *Music Theory Spectrum*, 28(1), 57–97. Steinbeis, N., & Koelsch, S. (2008). Shared neural resources between music and language indicate semantic processing of musical tension-resolution patterns. *Cerebral Cortex*, 18(5), 1169–1178. Sternberg, S. (1969). Memory-scanning: Mental processes revealed by reaction-time experiments. *American Scientist*, 57(4), 421–457. Stevens, C. J. (2012). Music perception and cognition: A review of recent cross-cultural research. *Topics in Cognitive Science*, 4, 653–667. Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18, 643–662. Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268, 1632–1634. Temperley, D. (2001). *The cognition of basic musical structures*. Cambridge: MIT Press. Thompson, W. (1983). Functional ambiguity in musical structures. *Music Perception*, 1(1), 3–27. Thompson-Schill, S. L., Jonides, J., Marshuetz, C., Smith, E. E., D'Esposito, M., Kan, I. P., ... Swick, D. (2002). Effects of frontal lobe damage on

interference effects in working memory. *Cognitive, Affective, and Behavioral Neuroscience*, 2(2), 109–120. Thothathiri, M., Kim, A., Trueswell, J. C., & Thompson-Schill, S. L. (2012). Parametric effects of syntactic-semantic conflict in Broca's area during sentence processing. *Brain and Language*, 120(3), 259–264. Tillmann, B. (2012). Music and language perception: Expectations, structural integration, and cognitive sequencing. *Topics in Cognitive Science*, 4, 568–584. doi:10.1111/j.1756-8765.2012.01209.x Tillmann, B., Bigand, E., & Madurell, F. (1998). Local versus global processing of harmonic cadences in the solution of musical puzzles. *Psychological Research*, 61, 157–174. Tillmann, B., Janata, P., & Bharucha, J. J. (2003). Activation of the inferior frontal cortex in musical priming. *Cognitive Brain Research*, 16, 145–161. Tillmann, B., Koelsch, S., Escoffier, N., Bigand, E., Lalitte, P., Friederici, A. D., & von Cramon, D. Y. (2006). Cognitive priming in sung and instrumental music: Activation of inferior frontal cortex. *NeuroImage*, 31(4), 1771–1782. Travis, F., Harung, H. S., & Lagrosen, Y. (2011). Moral development, executive functioning, peak experiences and brain patterns in professional and amateur classical musicians: Interpreted in light of a unified theory of performance. *Consciousness and Cognition*, 20(4), 1256–1264. Vazan, P., & Schober, M. F. (2004). Detecting and resolving metrical ambiguity in a rock song upon multiple rehearsals. *Proceedings of the 8th International Conference on Music Perception and Cognition* (pp. 426–432). Adelaide, Australia: Causal Productions. Vuong, L. C., & Martin, R. C. (2011). LIFG-based attentional control and the resolution of lexical ambiguities in sentence context. *Brain and Language*, 116(1), 22–32. Vuust, P., Roepstorff, A., Wallentin, M., Mouridsen, K., & Østergaard, L. (2006). It don't mean a thing... Keeping the rhythm during polyrhythmic tension, activates language areas (BA47). *NeuroImage*, 31(2), 832–841. Vuust, P., Wallentin, M., Mouridsen, K., Østergaard, L., & Roepstorff, A. (2011). Tapping polyrhythms in music activates language areas. *Neuroscience Letters*, 494(3), 211–216. Williamson, V. J., Baddeley, A., & Hitch, G. J. (2010). Musicians' and nonmusicians' short-term memory for verbal and musical sequences: Comparing phonological similarity and pitch proximity. *Memory and Cognition*, 38(2), 163–175. Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C., & Wager, T. D. (2011). Large-scale automated synthesis of human functional neuroimaging data. *Nature Methods*, 8(8), 665–670. doi:10.1038/nmeth.1635 Ye, Z., & Zhou, X. (2008). Involvement of cognitive control in sentence comprehension: Evidence from ERPs. *Brain Research*, 1203, 103–115. Ye, Z., & Zhou, X. (2009). Executive control in language processing. *Neuroscience and Biobehavioral Reviews*, 33(8), 1168–1177. Yeung, N., Botvinick, M. M., & Cohen, J. D. (2004). The neural basis of error detection: Conflict monitoring and the error-related negativity. *Psychological Review*, 111(4), 931–959. 652 *Psychon Bull Rev* (2015) 22:637–652

71 *Ann. N.Y. Acad. Sci.* 1060: 71–81 (2005). © 2005 New York Academy of Sciences. doi: 10.1196/annals.1360.006 Musical and Linguistic Processing in Song Perception DANIELE SCHÖN,^a REYNA LEIGH GORDON,^{a,b} AND MIREILLE BESSON^a ^aInstitute of Cognitive Neurosciences of the Mediterranean, CNRS, Marseille, France ^bCenter for Complex Systems and Brain Sciences, Florida Atlantic University, Boca Raton, Florida 33431, USA **ABSTRACT:** One approach to comparing the neural bases of language and music is through the use of song, which is a unique and ecological combination of these two cognitive domains. In song, language and music are merged into one acoustic signal with two salient dimensions. By manipulating either the linguistic or musical dimensions (or both) of song and studying their relationships, it is possible to gain important information about the neural networks underlying language and music cognition. We will present a brief review followed by recent behavioral, electrophysiological, and neuroimaging studies concerned with the functional and structural relationships of music and language. These results, together

with the previous studies in the field, help understanding whether the different levels of music and language processing are independent or interactive. **KEYWORDS:** singing; linguistics; music; cognition; neuroanatomy; brain; neuroimaging

INTRODUCTION Singing, like music, language, and other human activities is a universal human behavior. Nonetheless, it has been too scarcely exploited as a subject of neurocognition research until now.¹ This is somewhat surprising since singing is particularly well suited to the study of the relationship between language and music. However, most of the studies investigating the relationship between music and language have used language and music separately.^{2–5} This is somewhat problematic insofar as it is difficult to compare results issued from different tasks, different subjects, different types of analyses, and different statistical thresholds to define what is considered significant. Other experiments have been carried out using musical and linguistic tasks with the same participants,⁶ or a similar task with different materials.^{7–10} The comparison is easier since participants are the same and the tasks are as similar as possible. However, this approach also has some drawbacks. First, the models that are used to describe the two tasks as “similar” are often too simple and reductive. For Address for correspondence: Daniele Schön, INCM-CNRS, 31, Chemin J. Aiguier, 13402 Marseille, France. schon@incm.cnrs-mrs.fr 72

ANNALS NEW YORK ACADEMY OF SCIENCES instance, the comparison of musical and linguistic syntax rarely goes beyond a general and loosely defined concept of syntax (see Patel¹¹ for an interesting discussion). Second, the materials sometimes sacrifice some of their ecological validity because of the need to make tasks/materials as comparable as possible. The advantage of singing is that both linguistic and musical information are merged into one acoustic signal with two salient dimensions, allowing for a direct comparison within the same experimental material. In other words, it is possible to ask subjects to perform a linguistic task while manipulating the musical dimension, or vice-versa, to study potential interferences from one dimension on the other. The flip side of the coin is that in song, the phonological and metrical structures of language are strongly influenced by the type of melody that is used to sing. For instance, vowels sung in a very high register have the tendency to be deformed and less recognizable.¹² Moreover, the duration and accentuation of syllables will depend upon the musical structure.¹³ These complications are further increased when songs are in tone languages, in which pitch concurrently contains lexical and musical information.^{14,15} In the following section, we will review several studies that have investigated song cognition and discuss the findings relevant to the comparison of language and music processing. We seek to show how recent research has dealt with the interaction or independence of language and music processing, modulation of attention, and functional mapping of brain areas involved in singing. Then, we will briefly present three experiments that we recently carried out in our laboratory on linguistic and musical processing in song perception.

RELATIONSHIP BETWEEN THE LINGUISTIC AND MUSICAL DIMENSIONS OF SONG Results showing independent processing of language and music in song would provide support for the idea that separate neural substrates are responsible for processing these two different types of information. By contrast, results showing interactive processing would argue in favor of common neural resources to process language and music. The early neuropsychological and imaging studies of song perception examined the linguistic and musical aspects of song separately, by comparing normal speech and vocalises (nonlinguistic, melodic singing).^{16–18} More recently, researchers have compared sung and spoken language.^{19–22} Behavioral and electrophysiological methods have granted the possibility of using a more finegrained approach, by studying the musical and linguistic dimensions of song as they occur in sung language.^{23–28} Just as attentional processes generally exert a powerful influence over the functional organization of virtually all cognitive behaviors, attention has played a key role in the latter group of studies. A series of experiments has used song to study how harmony in music interacts with phonology and

semantics in language.^{25,28} In both experiments, musician and nonmusicians listened to sung chord progressions that ended on the expected (tonic) chord, or the less expected but still congruous (subdominant) chord. However, participants were asked to ignore the harmonic dimension and focus their attention on the linguistic one (phonology or semantics). In the first study, the behavioral data SCHÖN et al.: SONG PERCEPTION 73 showed that the less-expected harmonic stimulus did interfere with performance on the phoneme-monitoring task, in which subjects focused on nonword syllables sung by four voices.²⁵ The follow-up study's results also showed that the less-expected musical cadence interfered with semantic priming, measured by performance in a lexical decision task.²⁸ Measured implicitly, these interactions between language and music in song not only provide evidence for shared cognitive resources but also demonstrate how musical and linguistic priming occur in song. Interestingly, another series of experiments on harmony and semantics in song perception yielded findings that point to independence of language and music processing.^{24,26} Listeners showed distinct event-related brain potential (ERP) components for semantic (N400) and harmonic (P300) violations when simultaneously attending to the linguistic and musical dimensions of very ecological opera excerpts.²⁴ Indeed, by successfully fitting the data to an additive model of the two components, the authors concluded that the perception of lyrics and melodies was independent. Further support for the independence of semantic and harmonic processing comes from another experiment using the same stimuli, in which participants showed similar levels of performance when they simultaneously attended the lyrics and melodies, compared to when attending only one dimension.²⁶ Although no direct comparison can be made, it seems likely that crucial differences in stimuli and experimental designs would account for these differing findings. While a subtle musical change, which was acceptable but not highly expected in the harmonic context, was introduced in the former set of studies, the latter used harmonically and semantically incongruous endings that might have captured the listener's attention more dramatically to the point of causing independent effects. The above-mentioned research seems to substantiate the idea that the dynamic relationship between the multiple levels of structural and functional organization of language and music in song hinges on the allocation of attentional resources to different dimensions of song. In addition, findings obtained with the dichotic listening method show task- and attention-related effects when the musician's attention was modulated between phonemes and melodies in sung syllables/digits.^{23,29} Finally, results revealed typical lateralization patterns of left-ear (right hemisphere) advantage for the melodic task and right-ear (left hemisphere) advantage for the phonological task. Another approach to investigating music and language specificity through song is to consider the role of human voice specificity. Following up on some recent research using brain imaging to reveal a voice-selective area in the supratemporal sulcus (STS; see Belin³⁰ for a review), Levy, Granot, and Bentin^{31,32} used sung tones to distinguish ERPs occurring in response to musical pitches produced by voices and various instruments. In an oddball paradigm, listeners were first instructed to respond when they heard a piano tone presented among a series of sustained pitches played by instruments or sung by different voices on a neutral vowel. Results showed that a positive component, peaking at 320 ms (P320), called a voice-specific response (VSR), was elicited only by sung tones, even when subjects were attending a piano tone.³¹ However, in follow-up experiments, the VSR occurred only when participants were attending to timbre but not when they were attending to duration.³² Again, the use of singing builds a nice bridge between language and music, in addition to highlighting how task and attention to different acoustic parameters serve to allocate neural resources to different types of auditory stimuli. ⁷⁴ ANNALS NEW YORK ACADEMY OF SCIENCES CEREBRAL STRUCTURES INVOLVED IN SONG PROCESSING One of the main goals of studying the nature of the relationship between the linguistic and musical

dimensions of song is to determine whether music and language are processed by separate or integrated cerebral structures. Recent functional magnetic resonance imaging (fMRI) and positron emission tomography (PET) studies of singing production have sought to establish the neural bases of language and music in song. In particular, the spatial locations correlating with the linguistic and musical dimensions of song do seem to follow the classic patterns of left dominance for language and right dominance for music, heretofore described in the literature on language and music processing.^{33,34} Recent methodological advances with transcranial magnetic stimulation (TMS), which allows experimenters to collect pseudoneuropsychological data by provoking temporary, reversible lesions of localized cortical areas in healthy subjects, have brought forth evidence for right hemisphere dominance in singing.^{20,35,36} Speech is inhibited by applying TMS to certain areas of the left hemisphere; conversely, when TMS is performed on the right cortex, some participants are unable to sing melodies. Speech and nonlinguistic melodic singing were also compared by Riecker and colleagues,¹⁸ using fMRI. They showed opposite hemispheric activations with divergent networks for speaking (left lateralization) and singing (right lateralization). Thus, it seems that compared to speech, the production of sung vocalises uses more right hemispheric resources. However, their design did not include a more ecological condition of sung language, and it is therefore difficult to assess whether the right hemispheric dominance was due to the vocal motor aspects or to the musical aspects of singing. Sung language was also studied by Jeffries and colleagues²¹ using PET. Nonmusicians performed two different tasks: singing an “overlearned” song and speaking the text of the same song. As predicted, results also showed hemispheric lateralization (right dominance for singing, left for speaking). Thus, the authors were able to speculate that a right-lateralized network of brain area, typically involved in music perception and cognition, is also crucial for producing language in song. However, since no “vocalises” were included in the protocol, it is impossible to determine whether the right-hemispheric activations reflect music processing or linguistic production specific to song, or a combination of the two. Again, a direct comparison would be necessary to clarify the status of sung language. Brown and colleagues³⁷ took an alternative approach to studying nonlinguistic singing, by focusing on the different aspects of the structural organization of music. Indeed, they were able to use PET to differentiate brain activity (in the planum polare, BA38), associated with complex musical tasks such as harmonization and melody repetition, from activity associated with a simpler musical task (e.g., monotonic vocalization). While this elegant study did show the brain’s sensitivity to various musical elements of singing, the phonological and semantic conditions of singing also have to be investigated to determine if that same musical vocalization network would also be mobilized to sing songs with words. In summary, song perception and production do indeed provide a powerful model for comparing the neurocognition of language and music in the brain. While previous studies of song demonstrate some lateralization evidence for left-dominant language processing and right-dominant musical processing, on the whole it remains to SCHÖN et al.: SONG PERCEPTION 75

be determined whether these two dimensions of song are interactive or independent. It is essential to keep in mind that task demands and attention modulations exert an influence on this relationship. Furthermore, results also differ as a function of which levels of processing are being examined. Finally, song perception has not been studied outside of the context of a singing production task with either the fMRI or PET methods, so the spatial localization of the perception of singing also remains to be explored. Most importantly, no one study has yet directly compared spoken language, sung language, and vocalises within the same design and with the same participants.

EXPERIMENTS ON THE NEUROCOGNITION OF SINGING

In this last part, we will present recent data of three experiments that we recently carried out to investigate the relationship of the linguistic and

musical dimensions in song perception. In the first study, we directly compared spoken language, sung language, and vocalises. The main question was whether we would confirm a previous finding of different and lateralized networks for language and music. Nonmusician participants listened to pairs of spoken words (trisyllabic French words, spoken), sung words (trisyllabic words sung on three-note melodies), and vocalises (the syllable “vi” sung on three-note melodies). They had to judge whether the two stimuli of a pair were the same or different. Note that, in order to make the task more demanding, the first syllable or the first note of the two stimuli in a given pair was always the same. We also used pairs of noises as a control condition. fMRI results for each experimental condition contrasted with the control condition (noise) showed a very similar bilateral pattern of activation, mostly in temporal regions (see FIG. 1). Thus, results did not reveal a clearly lateralized pattern of activation. Interestingly, careful examination of the activity of a single voxel in the right middle temporal gyrus shows that its level of activity is modulated by the experimental condition (FIG. 2). Indeed, this is highest when participants are listening to sung words (top curve), and decreases for vocalises (middle curve) and spoken words (low curve). Note that this does not mean that this voxel is “specific” to song perception, but rather that it is involved in all three conditions, with maximal amplitude of the blood oxygen level–dependent (BOLD) response for sung words. When we directly compared sung and spoken words, for instance, many voxels are significantly more involved in sung than spoken word processing. While only performing such a type of comparison may have led us to believe that these voxels are specific to song perception, the use of a control condition clearly allowed us to show that this is not the case. Nonetheless, if we do not overinterpret differences, it is interesting to make direct comparisons of the experimental conditions. The comparison of sung words to vocalises gives a pattern of activation lateralized to the left, mostly in regions well known to be involved in lexical semantic processing. The comparison of sung words to spoken words gives a pattern of activation more lateralized to the right, mostly in regions well known to be involved in pitch processing, pitch memory, and voice processing. Interestingly, the largest differences between sung and spoken words are found in the anterior portion of the STS, a region that seems to be dedicated to voice processing and related areas.³⁸ What may be singing, both from a perceptual and emotional point of view, is a sort of supervoice! 76

ANNALS NEW YORK ACADEMY OF SCIENCES FIGURE 1. Each experimental condition is contrasted with the control condition. Activations are projected onto a transparent brain in the three different planes. SCHÖN et al.: SONG PERCEPTION 77

In summary, we found a strong overlap of the regions involved in language, music, and song processing. Moreover, the lateralization that was predicted on the basis of previous results was rather small. It only appears as a difference in the degree of activation when sung words are compared to vocalises (especially in linguistic areas) and to spoken words (especially in musical areas). While these first results are interesting, they did not shed light on the relationship between musical and linguistic dimensions in song perception. To determine whether these two dimensions of songs are treated independently or interactively, we conducted an ERP study. Nonmusician participants listened to pairs of sung words and, in different blocks of trials, had to focus their attention on the words to decide whether they were same or different, or on the melody to decide whether the notes were same or different. Therefore, depending upon the instructions given to the subjects, the linguistic or the musical dimension was relevant to the task at hand. For instance, if subjects have to pay attention to language, language is the relevant dimension and music the irrelevant one. We used four experimental conditions (both dimensions same; music same/language different; music different/language same; both dimensions different). Behavioral data showed small and similar error rates in both attentional conditions. By contrast, reaction times were longer in the musical task. Moreover, reaction times were longer when the irrelevant dimension was

incongruent with the relevant dimension, and this effect was stronger when participants paid attention to music. In other words, we found some interference effects of the irrelevant dimension on the relevant one, and language produced more interference on music than vice-versa. ERP results also point to an effect of the irrelevant dimension on the relevant one, in both attentional situations. As expected from the literature and colleagues,^{39,40} when participants focused attention on linguistic information, the N400 to the second word of the pair was larger for other than the same words (FIG. 3, dotted line). Most interestingly, a change in the irrelevant dimension (different melody/same word) elicits FIGURE 2. Adjusted amplitude of the BOLD response of a single voxel in the right middle temporal gyrus. 78

ANNALS NEW YORK ACADEMY OF SCIENCES an early negative effect, followed by a positive-going component (FIG. 3, dashed line). Thus, nonmusician participants seem to process the music dimension even when irrelevant for the task at hand. When participants paid attention to the melody, a change in the relevant dimension (different melody) is, as expected from the literature and colleagues⁴¹ associated with a late positive-going component (FIG. 3, dashed line). Interestingly again, a change in the irrelevant dimension (different word/same melody) also elicits a more negative-going component than in the control condition (all same). Thus, linguistic information seems to be processed (as reflected by what we interpret as an N400-like component) even when irrelevant for the task at hand (FIG. 3, dotted line). To summarize, results show that, in both the linguistic and musical tasks, processing of the relevant dimension seems influenced by the irrelevant dimension. In other words phonological/lexical processing and pitch processing cannot be processed independently. In order to determine which cerebral structures are involved in these interactions, we conducted the same study using fMRI. Specifically, the aim was to test the following three anatomical hypotheses. First, would lateralization differ depending upon the orientation of attention on the linguistic or melodic dimensions? Second, when language is the irrelevant dimension, would an effect be found in typical linguistic areas? Conversely, when music is the irrelevant dimension, would an effect be found in musical areas? Finally, and perhaps most interestingly, would language, when irrelevant, influence the degree of activation in musical areas, and would music, when irrelevant, influence activation in linguistic areas? FIGURE 3. Top: ERPs when participants are focusing their attention on linguistic information. Bottom: ERPs when participants are focusing their attention on musical information. Solid line = both dimensions same; dotted line = music same/language different; dashed line = music different/language same. SCHÖN et al.: SONG PERCEPTION

79 Regarding the first hypothesis, results show that when paying attention to language, linguistic processing is rather bilateral, except for a left frontal opercular activation. By contrast, when paying attention to music, melodic processing is rather right lateralized to middle and superior temporal regions. Thus, linguistic processing in song seems less lateralized than typically reported for spoken words. Regarding the second hypothesis, language as an irrelevant dimension induces strong and bilateral activation, not restricted to the typical linguistic areas. Music as an irrelevant dimension induces more activation in the right hemisphere. Finally, regarding the third hypothesis, when language is irrelevant and music relevant, overlaps of activation are found over bilateral temporal and prefrontal opercular regions, and they are larger in the right hemisphere. When music is irrelevant and language relevant, there is an overlap of activations over bilateral temporal regions. Therefore, it seems that both language and music, when irrelevant, influence the degree of activation in areas that are typically activated by the relevant dimensions. CONCLUSION Using different and complementary brain imaging methods provided interesting insights into the neurocognition of singing. Overall, the results obtained with fMRI and ERPs were compatible, in that they showed that the linguistic and musical dimensions of songs are processed by similar, overlapping brain areas. Moreover, both the time course of linguistic

and musical processing, and the degree of activation in areas typically involved in the processing of the relevant dimension, are influenced by the irrelevant dimension. Thus, taken together, these findings support the hypothesis that the linguistic and melodic components of songs are processed in interaction. While they support and extend the results, recently reported by Bigand and colleagues²⁵ and Poulin-Charronnat and colleagues,²⁸ they stand in contrast to those previously reported by Besson et al.²⁴ and Bonnel et al.²⁶ However, as mentioned above, both the task at hand and the stimuli play a crucial role in the ongoing perceptual and cognitive processes involved in human information processing. The materials used by Besson et al.²⁴ were ecological and complex (excerpts from opera sung a cappella), and participants were only asked to listen to these excerpts. Because of the methodological constraints linked with fMRI data acquisition, and since it was very important to use the same task for the fMRI and ERPs experiments, here we used very simple materials (pairs of stimuli), which may only reflect part of the high-level processes involved in song perception. Using words as stimuli may, for instance, have limited linguistic processing for the phonological and lexical aspects. Similarly, by using three notes, we limited musical processing to very simple melodic processing. Therefore, methodological advances are strongly needed to be able to use more ecological materials with fMRI. Finally, musical expertise has been shown to exert a strong influence on brain structural and functional organization.⁴² Differences between musicians and nonmusicians are even more likely to emerge when participants are asked to focus attention on one dimension of songs. In future experiments, it will be important to test musicians to further track the intriguing nature of the relationship between two of the most human cognitive abilities, music and language, when they are naturally combined in songs.

80 ANNALS NEW YORK ACADEMY OF SCIENCES ACKNOWLEDGMENTS The series of experiments reported in this chapter were conducted thanks to the support of the Human Frontier Science Program (HSFP 2/15/06RGP0053) and by a grant from the Ministère de la Recherche (Programme Cognitique, ACI “Imagerie cérébrale”) to Mireille Besson. We would like to thank Monique Chiambretto and Vanina Luigi for methodological assistance; the two singers, Véronique Kirkyacharian and Serge Charron, for their precious voices; as well as all the participants in the experiments.

REFERENCES 1. PERETZ, I. et al. 2004. Singing in the brain: insights from cognitive neuropsychology. *Music Percept.* 21: 373–390. 2. MAESS, B. et al. 2001. Musical syntax is processed in Broca’s area: an MEG study. *Nat. Neurosci.* 4: 540–545. 3. KOELSCH, S. et al. 2002. Bach speaks: a cortical “language-network” serves the processing of music. *Neuroimage* 17: 956–966. 4. LEVITIN, D.J. & V. MENON. 2003. Musical structure is processed in “language” areas of the brain: a possible role for Brodmann Area 47 in temporal coherence. *Neuroimage* 20: 2142–2152. 5. KNOSCHE, T.R. et al. 2005. Perception of phrase structure in music. *Hum. Brain Mapping* 24: 259–273. 6. BESSON, M. & F. MACAR. 1987. An event-related potential analysis of incongruity in music and other non-linguistic contexts. *Psychophysiology* 24: 14–25. 7. SCHÖN, D., C. MAGNE & M. BESSON. 2004. The music of speech: music training facilitates pitch processing in both music and language. *Psychophysiology* 41: 341–349. 8. MAGNE, C., D. SCHÖN & M. BESSON. Musician children detect pitch violations in both music and language better than non-musician children: behavioral and electrophysiological approaches. *J. Cognit. Neurosci.* In press. 9. PATEL, A.D. et al. 1998. Processing syntactic relations in language and music: an event-related potential study. *J. Cognit. Neurosci.* 10: 717–733. 10. PATEL, A.D. et al. 1998. Processing prosodic and musical patterns: a neuropsychological investigation. *Brain Lang.* 61: 123–144. 11. PATEL, A.D. 2003. Language, music, syntax and the brain. *Nat. Neurosci.* 6: 674–681. 12. SCOTTO DI CARLO, N. 1994. L’Intelligibilité de la voix chantée. *Médecine des Arts* 10: 2–15. 13. PALMER, C. & M. KELLY. 1992. Linguistic prosody and musical meter in song. *J. Mem. Language* 31: 525–542. 14. CHEN-HAFTECK,

L. 1999. Discussing text-melody relationship in children's song-learning and singing: a Cantonese-speaking perspective. *Psychol. Music* 27: 55–70. 15. WONG, P.C. M. & R.L. DIEHL. 2002. How can the lyrics of a song in a tone language be understood? *Psychol. Music* 30: 202–209. 16. GORDON, H.W. & J.E. BOGEN. 1974. Hemispheric lateralization of singing after intracarotid sodium amylobarbitone. *J. Neurol. Neurosurg. Psychiatry* 37: 727–738. 17. RYDING, E., B. BRADVIK & D.H. INGVAR. 1987. Changes of regional cerebral blood flow measured simultaneously in the right and left hemisphere during automatic speech and humming. *Brain*. 110 (Pt 5): 1345–1358. 18. RIECKER, A. et al. 2000. Opposite hemispheric lateralization effects during speaking and singing at motor cortex, insula and cerebellum. *Neuroreport* 11: 1997–2000. 19. CADALBERT, A. et al. 1994. Singing with and without words: hemispheric asymmetries in motor control. *J. Clin. Exp. Neuropsychol.* 16: 664–670. 20. EPSTEIN, C.M. et al. 1999. Localization and characterization of speech arrest during transcranial magnetic stimulation. *Clin. Neurophysiol.* 110: 1073–1079. 21. JEFFRIES, K.J., J.B. FRITZ & A.R. BRAUN. 2003. Words in melody: an H(2)15O PET study of brain activation during singing and speaking. *Neuroreport* 14: 749–754. SCHÖN et al.: SONG PERCEPTION 81 22. STAGER, S.V., K.J. JEFFRIES & A.R. BRAUN. 2003. Common features of fluency-evoking conditions studied in stuttering subjects and controls: an H(2)15O PET study. *J. Fluency Disord.* 28: 319–336. 23. BARTHOLOMEUS, B. 1974. Effects of task requirements on ear superiority for sung speech. *Cortex* 10: 215–223. 24. BESSON, M. et al. 1998. Singing in the brain: independence of lyrics and tunes. *Psychol. Sci.* 9: 494–498. 25. BIGAND, E. et al. 2001. The effect of harmonic context on phoneme monitoring in vocal music. *Cognition* 81: 11–20. 26. BONNEL, A.M. et al. 2001. Divided attention between lyrics and tunes of operatic songs: evidence for independent processing. *Percept. Psychophys.* 63: 1201–1213. 27. KOLINSKY, R. et al. Perceptual interactions between lyrics and tunes: vowels sing but consonants speak. In preparation. 28. POULIN-CHARRONNAT, B. et al. 2005. Musical structure modulates semantic priming in vocal music. *Cognition* 94: 67–78. 29. GOODGLASS, H. & M. CALDERON. 1977. Parallel processing of verbal and musical stimuli in right and left hemispheres. *Neuropsychologia* 15: 397–407. 30. BELIN, P., S. FECTEAU & C. BEDARD. 2004. Thinking the voice: neural correlates of voice perception. *Trends Cognit. Sci.* 8: 129–135. 31. LEVY, D.A., R. GRANOT & S. BENTIN. 2001. Processing specificity for human voice stimuli: electrophysiological evidence. *Neuroreport* 12: 2653–2657. 32. LEVY, D.A., R. GRANOT & S. BENTIN. 2003. Neural sensitivity to human voices: ERP evidence of task and attentional influences. *Psychophysiology* 40: 291–305. 33. ZATORRE, R.J., P. BELIN & V.B. PENHUNE. 2002. Structure and function of auditory cortex: music and speech. *Trends Cogn. Sci.* 6: 37–46. 34. TERVANIEMI, M. et al. 2000. Lateralized automatic auditory processing of phonetic versus musical information: a PET study. *Hum. Brain Mapp.* 10: 74–79. 35. STEWART, L. et al. 2001. Transcranial magnetic stimulation produces speech arrest but not song arrest. *Ann. N. Y. Acad. Sci.* 930: 433–435. 36. LO, Y.L. & S. FOOK-CHONG. 2004. Ipsilateral and contralateral motor inhibitory control in musical and vocalization tasks. *Exp. Brain Res.* 159: 258–262. 37. BROWN, S. et al. 2004. The song system of the human brain. *Brain Res. Cogn. Brain Res.* 20: 363–375. 38. BELIN, P. et al. 2000. Voice-selective areas in human auditory cortex. *Nature* 403: 309–312. 39. BENTIN, S., G. MCCARTHY & C.C. WOOD. 1985. Event-related potentials, lexical decision and semantic priming. *Electroencephalogr. Clin. Neurophysiol.* 60: 343–355. 40. HOLCOMB, P.J. & H.J. NEVILLE. 1990. Auditory and visual semantic priming in lexical decision: a comparison using event-related brain potentials. *Lang. Cogn. Processes* 5: 281–312. 41. BESSON, M. & F. FAITA. 1995. An event-related potential study of musical expectancy: comparison of musicians with non-musicians. *J. Exp. Psychol. Hum. Percept. Perform.* 21: 1278–1296. 42. PANTEV C. et al. 2003. Music learning-induced cortical plasticity. *Ann. N. Y. Acad. Sci.*

Hierarchical control as a shared neurocognitive mechanism for language and music

Author links open overlay panel

Rie Asano ^a, Cedric Boeckx ^{b c d}, Uwe Seifert ^a

[Show more](#)

Add to Mendeley

Share

Cite

<https://doi.org/10.1016/j.cognition.2021.104847>

[Get rights and content](#)

Abstract

Although comparative research has made substantial progress in clarifying the relationship between language and music as neurocognitive systems from both a theoretical and empirical perspective, there is still no consensus about which mechanisms, if any, are shared and how they bring about different neurocognitive systems. In this paper, we tackle these two questions by focusing on hierarchical control as a neurocognitive mechanism underlying syntax in language and music. We put forward the Coordinated Hierarchical Control (CHC) hypothesis: linguistic and musical syntax rely on hierarchical control, but engage this shared mechanism differently depending on the current control demand. While linguistic syntax preferably engages the abstract rule-based control circuit, musical syntax rather employs the coordination of the abstract rule-based and the more concrete motor-based control circuits. We provide evidence for our hypothesis by reviewing neuroimaging as well as neuropsychological studies on linguistic and musical syntax. The CHC hypothesis makes a set of novel testable predictions to guide future work on the relationship between language and music.

Introduction

Hierarchical cognitive control (hereafter “hierarchical control”) is a neurocognitive mechanism that captures a set of control processes guiding thought and action on the basis of intentions, plans, and goals to enable flexible behavior (Badre & D’Esposito, 2009; Koechlin & Summerfield, 2007). The key idea of hierarchical control is that those control processes are hierarchically organized to select, maintain, and inhibit progressively abstract and temporally extended goals (Badre, 2008; Badre & D’Esposito, 2009; Koechlin & Summerfield, 2007). The goal representations and control processes are hierarchically organized in such a way that the more abstract levels determine the more concrete levels. For example, to execute and understand a complex action (e.g., making coffee), goals should be selected, maintained, and inhibited at multiple hierarchical levels ranging from the abstract, temporally extended goal (e.g., making coffee) to the sub-goals (e.g., filling a pot with water) and to concrete motor acts (e.g., grabbing the pot). We understand a mechanism as composed of different constituent parts performing unique operations and working in tandem to bring about a given phenomenon (Bechtel & Shagrir, 2015). Hierarchical control is a neurocognitive mechanism because its constituent parts (selection, maintenance, and inhibition) build an interface between cognitive and neural operations (Badre & Frank, 2012; Frank, Loughry, & O’Reilly, 2001).

As suggested by Lashley (1951), in a similar way to action, language and music cannot be explained by associative chain theories whereby one element (e.g., movement, word, or note) serially triggers the next one. Rather, there should be a generalized pattern (called ‘syntax’ or ‘schema’) to hierarchically process linguistic and musical sequences. Thus, hierarchical control is not only a neurocognitive mechanism underlying action, but also language and music (Fitch & Martins, 2014; Jeon, 2014; Slevc & Okada, 2015; Uddén & Bahlmann, 2012). However, linguistic and musical syntax differ in items to be combined,

categories to be labeled, and contents to be encoded (Asano & Boeckx, 2015; Patel, 2008), and it is not clear how this shared neurocognitive mechanism relate to such differences.

One possible explanation was provided by the “resource sharing framework” (Patel, 2003, Patel, 2013). This framework suggests that linguistic and musical syntax engage a set of shared neurocognitive processes implemented in the frontal regions and those shared processes operate on domain-specific long-term memory representations in the temporal regions via functional connections between those two regions. In the current paper, we offer a complementary approach, and highlight the oft-neglected fact that even shared processes are utilized differently in processing linguistic and musical syntax. We frame this in terms of a gradient between more abstract temporally extended control and more concrete motor control (hereafter “abstraction gradient”). In processing linguistic and musical sequences, their representations are temporarily maintained, selected, and inhibited at hierarchical levels differing in abstractness to flexibly adapt to the current control demand in a way similar to what we illustrated above for goal representations. Thus, our focus is temporarily stored representations and corresponding processes.

The concepts of abstraction gradient and control demand stem from research on cognitive control. In the literature on cognitive control, control demands are a function of the hierarchical depth of contextual representations and the time span for maintaining contextual representations, which directly corresponds to the degree of abstraction (Badre, 2008; Badre & Nee, 2018). For example, selecting a motor response based on a contextual cue (e.g., geometric shape), which in turn depends on another contextual cue (e.g., color), is more demanding than selecting a motor response based on a contextual cue without any further contextual representations. In this case, the color cue is considered to be more abstract than the shape cue as it does not directly govern the motor response, but instead governs another contextual cue. This type of abstraction in which context generalizes over differently abstract stimulus-response rules is known as “policy abstraction”, where a policy is considered to be akin to a rule (Badre & Nee, 2018). Another type of abstraction concerns “temporal abstraction”: contextual representations are maintained over time and span intervening episodes or subtasks as overarching goal representations (Badre & Nee, 2018). Performing complex actions such as coffee making requires both types of abstraction.

Similarly, we suggest that linguistic and musical syntax involve different levels of policy (or rule) and temporal abstraction depending on the hierarchical depth and the span of representations to be maintained, selected, and inhibited. For example, linguistic syntax interfaces with conceptual (i.e., semantic) and sensory-motor (i.e., phonological) systems, and thus differently engage more abstract and more concrete control processes to link meaning and sound in processing sentences. This is related to an abstraction gradient from semantics to syntax to phonology suggested by Uddén and Bahlmann (2012). Musical syntax (Box 1) – tonal-harmonic and rhythmic syntax – links affect and sound at different levels of abstraction. While affect can be directly linked to the sensory-motor system, it can also represent more abstract, schematic patterns of tension and relaxation (Lerdahl, 2013). Moreover, sentences and musical sequences with center-embedding calls for high control demand and require temporal abstraction (Jeon, 2014). Relevant examples include sentences with center-embedded relative clauses such as “[Researchers [who work hard and sleep little] EMBEDDED Clause need a lot of coffee.] MAIN Clause” and chord progressions with embedding of one key inside another (i.e., modulation) (see the famous Bach's “Little Harmonic Labyrinth” (BWV 591) example by Hofstadter, 1979; see also Rohrmeier, 2011). In any case, the greater the hierarchical and temporal distance between contextual and sensory-motor representations is, the more the current control demand becomes.

In Section 2, based on those concepts of abstraction gradient and control demand, we introduce a hypothesis elucidating the relationship between linguistic and musical syntax in terms of hierarchical control, provide supportive evidence from neuroimaging and neuropsychological studies, and present testable predictions of our hypothesis. In Section 3, we then discuss the previous findings from comparative research on linguistic and musical syntax in light of our hypothesis. Although we acknowledge that prosody is hierarchically organized, can be described through formal rules, influences linguistic syntactic processing, and parallels some aspects of musical syntax (for discussions, see Fabb & Halle, 2012; Heffner & Slevc, 2015), our hypothesis does not treat the relationship between prosody and musical syntax because empirical studies investigating its relationship in terms of hierarchical processing is too sparse at the moment. But we briefly come back to this issue in Section 4.

Section snippets

The Coordinated Hierarchical Control (CHC) hypothesis: Same neurocognitive mechanism but different use

Our hypothesis concerning the relationship between linguistic and musical syntax states that both rely on hierarchical control as an orchestration of the prefrontal maintenance function and the basal ganglia dual control functions (i.e., selection and inhibition), yielding a flexible top-down control from more abstract onto more concrete processes and representations on an abstraction gradient. At the same time, our hypothesis also states that linguistic and musical syntax differ in the extent

Reassessing comparative research on linguistic and musical syntax

In line with the “resource sharing framework” (Patel, 2003), a number of behavioral, neuroimaging, and neuropsychological studies provided evidence for shared neural resources for linguistic and musical syntax in the past decades (Table 2). How do the shared neural resources revealed by those previous studies relate to hierarchical control? In Section 3.1, we suggest that the shared neural resources can be explained in terms of the resources required for the maintenance function and the dual

Conclusions and future perspectives

To conclude, we have claimed that hierarchical control is a shared neurocognitive mechanism underlying syntax in language and music. Brain bases of hierarchical control include 1) functional organization of frontal cortex to maintain more abstract representations in the more anterior regions and more concrete representations in the more posterior regions, 2) dual control functions (selection and inhibition) of the basal ganglia, and 3) top-down influence from abstract temporally extended

The neural control of singing

Jean Mary Zarate*

Department of Psychology, New York University, New York, NY, USA

Singing provides a unique opportunity to examine music performance—the musical instrument is contained wholly within the body, thus eliminating the need for creating artificial instruments or tasks in neuroimaging experiments. Here, more than two decades of voice and singing research will be reviewed to give an overview of the sensory-motor control of the singing voice, starting from the vocal tract and leading up to the brain regions involved in singing. Additionally, to demonstrate how sensory feedback is integrated with vocal motor control, recent functional magnetic resonance imaging (fMRI) research on somatosensory and auditory feedback processing during singing will be presented. The relationship between the brain and singing behavior will be explored also by examining: (1) neuroplasticity as a function of various lengths and types of training, (2) vocal amusia due to a compromised singing network, and (3) singing performance in individuals with congenital amusia. Finally, the auditory-motor control network for singing will be considered alongside dual-stream models of auditory processing in music and speech to refine both these theoretical models and the singing network itself.

Most of the literature on sensory-motor control in music production and training-induced plasticity focuses on trained instrumental musicians or learning paradigms

with musical instruments (e.g., learning to play short piano melodies, etc.). Singing, however, provides a unique opportunity to examine sensory-motor processes during musical production, since the instrument is already contained within the body; there is no need to create artificial instruments to assess motor control mechanisms with neuroimaging or any other experimental approach. Moreover, the adult vocal apparatus is highly trained to produce nuanced utterances in both song and speech. Across their lifetime, healthy non-musicians have sung (or have attempted to sing) a full repertoire of songs in socially and culturally specific settings, (“Happy Birthday,” their national anthem, etc.). Additionally, healthy individuals can control their vocal pitch and/or output intensity to indicate the intent of a sentence (e.g., declarative statements vs. questions vs. commands), set the emotional context for a conversation (e.g., happiness, anger, sadness), or in tonal languages, distinguish between words and their meanings. Singers, on the other hand, undergo many years of extensive sensory-motor training and practice to exert much finer vocal control during more difficult tasks, such as singing fast vocal runs (e.g., melismata, melodic embellishments, etc.) or maintaining a melodic passage as someone else simultaneously sings a harmonic line. Therefore, using singing tasks to test groups with different levels of singing experience is a rare opportunity to determine how musical experience specifically enhances sensory-motor control of this particular instrument, beyond the remarkable feats it already can perform. However, the mechanisms by which the vocal instrument is precisely controlled for singing are highly complex and thus require multiple networks for vocal motor control and sensory feedback processing.

Sensory-Motor Control of Vocalization

Sensory-Motor Control Observed From the Vocal Tract

When air passes through the glottis (opening of the larynx) and causes the vocal folds surrounding the glottis to vibrate at a particular rate, the resulting vibration rate determines the fundamental frequency (i.e., perceived pitch) of the voice ([Sundberg, 1987](#)). Different intrinsic and extrinsic laryngeal muscles interact to regulate fundamental frequency by altering the length of the vocal folds, thus changing the rate of vocal-fold vibration ([Hirano et al., 1969](#); [Sundberg, 1987](#)). The precise control of laryngeal muscles is maintained in part by laryngeal reflexogenic control systems, in which receptors within the larynx adjust muscular contractions during perturbations. For instance, during vocalization, the uneven airflow passing through the glottis stimulates the myotatic mechanoreceptors in the intrinsic laryngeal muscles; these stretch-sensitive receptors initiate reflexive muscular adjustments to ensure that the vocal folds remain at the intended length and tension and therefore maintain a steady vocal pitch ([Wyke, 1974](#)). Additional reflexogenic systems work in concert with the intrinsic laryngeal reflexogenic system to ensure a stable vocalization ([Wyke, 1974](#)). Vocalization also involves the coordination of many other muscles, including the diaphragm and abdominal/thoracic muscles to provide airflow and regulate vocal output intensity, and articulatory muscles (e.g., lip, jaw, and tongue muscles, [Hardcastle, 1976](#); [Sundberg, 1987](#)). The articulatory muscles contain somatosensory receptors that play a role in generating different vocal-tract configurations, which shape the formant frequencies that contribute toward vowel

formation and vocal timbre (Sundberg, 1987; Jürgens, 2002; Perkell, 2012).

Similar to the somatosensory contribution to reflexogenic vocal control systems, auditory feedback also plays a role in reflex-like adjustments of ongoing vocal motor control. For instance, a slight *decrease* in auditory feedback amplitude elicits a quick *increase* in vocal output amplitude, which is known as the Lombard reflex (Lombard, 1911). During speech production, when the first formant frequency is shifted so that a produced vowel (e.g., /ε/) sounds like a different one (e.g., /æ/), the vocal motor system immediately compensates for the formant shift (Houde and Jordan, 1998, 2002; Purcell and Munhall, 2006a,b). Arguably, the most relevant auditory-vocal motor correction for singers deals with vocal pitch. When the pitch of auditory feedback is shifted up or down as participants vocalize for a few seconds (either at a comfortable pitch or to match a target pitch), investigators have observed pitch-shift responses, during which vocal pitch is adjusted quickly in the opposite direction of the feedback shift (Anstis and Cavanagh, 1979; Burnett et al., 1998; Larson, 1998; Hain et al., 2000; Jones and Munhall, 2000, 2005; Larson et al., 2000; Burnett and Larson, 2002; Liu and Larson, 2007; Jones and Keough, 2008). These pitch-shift responses often have two components: (1) an early pitch-shift response of 25–50 cents (irrespective of the pitch-shift magnitude) that occurs 100–150 ms after the pitch shift; and (2) a late pitch-shift response with a latency of 250–600 ms, whose magnitude and direction can be under voluntary control, if listeners are instructed to make a specific response (e.g., change pitch to either oppose or follow the pitch shift, etc., Burnett et al., 1998; Larson, 1998; Hain et al., 2000). Interestingly, prolonged exposure to feedback that is incrementally pitch-shifted over numerous trials can produce aftereffects in which intended vocal pitch and vocal output are mismatched, such that vocal pitch is automatically adjusted even when auditory feedback is returned to normal (Jones and Munhall, 2000, 2005; Jones and Keough, 2008).

Neural Networks Governing Sensory-Motor Control of Vocalization

Brain regions involved in vocal motor control

Multiple neural networks are required for precise control of the “phonatory” muscles mentioned above. The reticular formation of the pons and medulla has direct connections to the motoneurons for all phonatory muscles (Figure 1, white boxes, Thoms and Jürgens, 1987), and thus may coordinate phonatory muscle groups to generate complete vocal patterns (Jürgens and Hage, 2007). This region receives excitatory input from two distinct neural pathways of vocal control (Figure 1; Jürgens, 2009; Owren et al., 2011). The first vocal control pathway (Figure 1, green boxes) contains the anterior cingulate cortex (ACC) and the midbrain periaqueductal gray (PAG), both of which produce vocalizations when stimulated electrically or pharmacologically (Müller-Preuss and Jürgens, 1976; Müller-Preuss et al., 1980; Suga and Yajima, 1988; Dujardin and Jürgens, 2005). The second neural pathway includes the primary motor cortex (M1, Figure 1, blue box) and two subcortical loops—comprised of putamen, globus pallidus, pontine gray, and cerebellum—that modulate vocal motor commands from M1 and subsequently send modified motor programs via the ventrolateral thalamus back to M1; electrical stimulation of the ventral part of M1 elicits vocalizations, as well as individual movements of the jaw, tongue, and lips (Penfield and Rasmussen, 1950).

figure 1

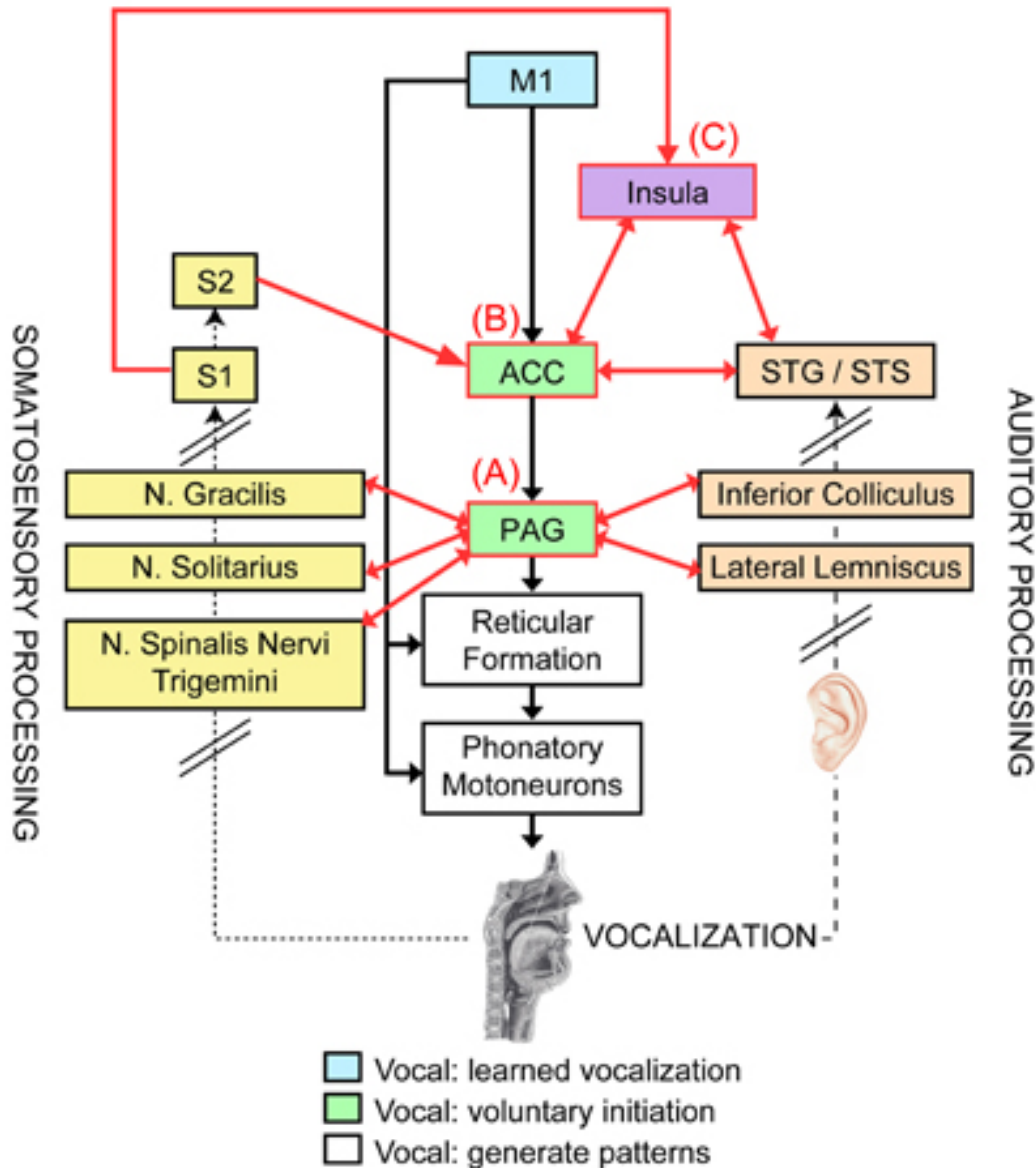


FIGURE 1. NEURAL NETWORKS OF VOCAL MOTOR CONTROL (CENTRAL COLUMN), SOMATOSENSORY (LEFT) AND AUDITORY FEEDBACK PROCESSING (RIGHT), AND HYPOTHESIZED REGIONS OF SENSORY-MOTOR CONTROL OF VOICE [MODIFIED FROM A MODEL PROPOSED BY JÜRGENS (2009)]. The vocal motor control hierarchy starts with the generation of complete vocal patterns from the reticular formation and phonatory motoneurons (white boxes), and then the next highest level of control (green boxes) stems from the anterior cingulate cortex (ACC) and periaqueductal gray (PAG), which can initiate and emotionally motivate vocal responses. The highest level of vocal control comes from the primary motor cortex (M1, blue box; its modulatory brain regions are not depicted), which is responsible for producing learned vocalizations (i.e., speech and song). Somatosensory feedback (dotted arrow) from various receptors distributed throughout the vocal tract is processed in the ascending somatosensory pathway (yellow boxes, left; black slanted lines indicate that only selected regions of this pathway are shown) and transmitted to the primary and secondary somatosensory cortex (S1, S2). Auditory feedback (dashed arrow) from the vocalization is processed by the ascending auditory pathway and auditory cortical regions (orange boxes,

right). Potential neural regions that integrate sensory feedback processing with vocal motor control are indicated with red-outlined boxes, and their shared connections are represented by red arrows: (A) the PAG, (B) ACC, and (C) the insula (in purple, classified as a higher-order associative area).

In humans, these networks form a tripartite hierarchy of vocal motor control (Figure 1, center column, [Simonyan and Horwitz, 2011](#)): (1) the reticular formation constitutes the lowest level at which complete vocal patterns are generated; (2) the next level is comprised of the ACC and the PAG, which are attributed with the voluntary initiation and emotional/motivational control of vocalizations ([Jürgens, 2002, 2009](#)); and (3) the highest level of vocal control occurs in M1 (and its modulatory brain regions), which is associated with the generation of learned vocalizations, such as speech and song ([Jürgens, 2002, 2009](#)). Importantly, this functional distinction of M1 is based on humans' unique possession of direct connections between the phonatory region of M1 (i.e., the ventral portion) and the motoneurons of phonatory muscles (see Figure 1); bilateral lesions to this M1 region destroys the ability to speak and sing ([Jürgens, 2009](#)), while innate vocalizations (e.g., shrieking, crying, etc.) that may be controlled by the ACC and PAG are left intact. In contrast, damage to the modulatory brain regions associated with M1 (e.g., putamen, globus pallidus, pontine gray, and cerebellum) can result in speech disorders such as stuttering and dysarthria ([Ackermann et al., 1992](#); [Jürgens, 2002](#); [Alm, 2004](#)). Lesions in the second level of vocal control may lead to mutism (attributed to PAG damage, [Esposito et al., 1999](#)) or loss of emotional/motivational intonation in speech (following damage to the ACC, [Simonyan and Horwitz, 2011](#)). Importantly, the functional organization of vocal motor control in humans is concurrently hierarchical and parallel, since damage to brain regions within the second or third levels does not abolish all vocalizations.

Neural processing of somatosensory feedback

Various somatosensory receptors transmit feedback about the current state of the vocal motor system (e.g., placement of articulators, respiration, etc.) via the glossopharyngeal and vagus nerves and the ascending somatosensory pathway, which includes the nuclei gracilis, solitarius, and spinalis nervi trigemini and the medial lemniscus in the medulla, and the ventral posteromedial nucleus in the thalamus ([Jürgens and Kirzinger, 1985](#); [Willis, 1986](#)). The thalamus sends somatosensory information to primary and secondary somatosensory cortex (S1 and S2), as well as the insula ([Jones and Powell, 1970](#); [Augustine, 1996](#); [Jürgens, 2002](#); [Ackermann and Riecker, 2004, 2010](#)). More specifically, the ventral portion of the primary somatosensory cortex (S1)—posteriorly adjacent to the M1 phonatory area that governs vocalizations and individual movements of the articulators ([Penfield and Rasmussen, 1950](#))—processes somatosensory information about articulatory movements ([Grabski et al., 2012](#)), while the anterior portion of the insula is recruited particularly during overt vocalizations (compared to covert speech and song, [Riecker et al., 2000](#)) and may contribute to voluntarily controlled respiration during vocalizations in general ([Ackermann and Riecker, 2010](#)).

Neural processing of auditory feedback during singing

As each sung note reaches a singer's ear as auditory feedback, each of the different

frequencies within that particular vocal pitch are transduced by the organ of Corti on the basilar membrane of the cochlea (Hudspeth, 2000). The frequency characteristics that are required to perceive the pitch are transmitted and/or processed along different parts of the ascending auditory pathway—comprised of the cochlear nucleus, lateral lemniscus, inferior colliculus, and the medial geniculate nucleus of the thalamus (Griffiths et al., 2001)—before the extracted frequencies (and many other attributes of sounds) are further processed in primary and secondary auditory cortex within Heschl's gyrus. In particular, pitch information may be processed specifically by a (rightward lateralized) pitch-sensitive area located in lateral Heschl's gyrus, reported to be involved in conscious pitch perception (Griffiths, 2003; Bendor and Wang, 2006). This region may also be involved in organizing pitches in a hierarchical fashion, since patients with lesions in this region displayed much higher discrimination thresholds than controls when asked to indicate the direction of pitch change between two notes (Johnsrude et al., 2000). Processing pitch changes or melodic phrases within a sung passage recruits additional auditory cortical regions outside of Heschl's gyrus, including regions in the right superior temporal gyrus (STG), planum polare, and planum temporale (Zatorre et al., 1994; Patterson et al., 2002; Hyde et al., 2008). When pitch comparisons are performed within a sequence of tones or short melodies, increased activity is observed within right auditory and frontal cortical regions presumably during tonal working memory processes, compared to passive melody perception (Zatorre et al., 1994). Melodic phrase comparisons in the same key, which may be done to ensure correct melodic reproduction, engages extensive activity within several auditory cortical regions along bilateral STG, whereas melodic phrase comparisons across a pitch transposition (i.e., a key change) engages additional activity from the intraparietal sulcus (IPS, Foster and Zatorre, 2010).

Aside from providing details about vocal pitch, auditory feedback can also provide information about vocal timbre, which is argued to be processed specifically along the superior temporal sulcus (STS, Belin et al., 2000). Kriegstein and Giraud (2004) discovered three functionally distinct regions along the STS. The anterior STS is associated with familiar voice recognition, while the mid/anterior STS preferentially responds to the spectral characteristics of voices. The posterior STS (pSTS), which is recruited during recognition of unfamiliar voices, may be involved in analyzing spectral details (or the changes therein) of voices over time (Kriegstein and Giraud, 2004; Warren et al., 2006). Given that the pSTS is also recruited in response to presentation of frequency-modulated sweeps of pure tones (Poeppel et al., 2004) and phonological processing (Hickok and Poeppel, 2007), this region may be involved generally in processing spectrotemporal fluctuations in sound, including notable changes in auditory feedback.

Potential substrates for integrating sensory feedback with vocal motor control

The constituents of the vocal motor network associated with voluntary initiation and emotional/motivational control of vocalizations—the PAG and ACC—receive both somatosensory and auditory input, and thus form two potential substrates for sensory-motor control of vocalization (Figure 1, red-outlined boxes and arrows). The PAG (Figure 1A) receives somatosensory input via afferent projections from the nucleus gracilis (implicated in respiratory control, Hannig and Jürgens, 2006) and nuclei solitarius and spinalis nervi trigemini (kinesthetic and proprioceptive information, Jürgens and Kirzinger, 1985; Yoshida et al., 2000), as well as auditory

information from the inferior colliculus and lateral lemniscus (Dujardin and Jürgens, 2005), all of which may facilitate initiating vocalizations in response to external stimuli or adjusting vocalizations based on sensory feedback. For example, when connections to the cerebrum are severed, the Lombard reflex is preserved during PAG-induced vocalizations coupled with auditory masking, suggesting that the PAG may govern auditory-motor control during involuntary auditory-vocal reflexes (e.g., Lombard reflex, formant- and pitch-shift responses) without additional control from cortical regions (Nonaka et al., 1997). The ACC (Figure 1B) directly receives somatosensory input from S2 and auditory input from auditory cortical regions along the STG and STS (Jürgens, 1983; Barbas et al., 1999). This region also receives these types of sensory input indirectly from S1 and auditory association areas via the insula (Mesulam and Mufson, 1982; Augustine, 1996). Since the insula is a gateway of both somatosensory and auditory information for the ACC, this region itself may provide another substrate for sensory-motor control of vocalization (Figure 1C, purple box). In particular, the anterior insula, whose cytoarchitecture and projections classify it as an association area that integrates different modalities (e.g., auditory, visual, somatosensory, motor, etc., Rivier and Clarke, 1997; Lewis et al., 2000; Bamiou et al., 2003; Ackermann and Riecker, 2004), is engaged specifically during voiced speech and song, relative to covert or internal versions (Riecker et al., 2000; but see Hillis et al., 2004; Ackermann and Riecker, 2010 for conflicting clinical evidence of the insula's role in speech production).

Neuroimaging evidence: a general functional network for human vocalization

Neuroimaging studies from the past two decades have confirmed that many regions within vocal motor and sensory networks are recruited during various *overt* speech and song tasks, including: word or letter generation (Paus et al., 1993); syllable repetition (Riecker et al., 2005); singing a note repeatedly (Perry et al., 1999), in a sustained fashion (Zarate and Zatorre, 2008), or while changing vowels in particular rhythms (Jungblut et al., 2012); repeating syllables, spoken words, and sung or hummed melodies (Özdemir et al., 2006); humming, speaking, or singing lyrics of a well-known song (Formby et al., 1989; Jeffries et al., 2003); reciting the months of the year or singing a familiar melody (Riecker et al., 2000); telling a story (Schulz et al., 2005); improvising word phrases, melodies, or harmonies (Brown et al., 2004, 2006); spontaneous and synchronized speaking and singing (Saito et al., 2006); and singing an Italian aria (Kleber et al., 2007). Summarized from the neuroimaging evidence above, a general functional network for human vocalization (including speech and song) is comprised of the brain regions reviewed in the preceding sections: M1, ACC, basal ganglia, thalamus, and cerebellum for vocal motor control; S1 and S2 for somatosensory feedback processing; bilateral auditory cortical regions (primary auditory cortex and a pitch-sensitive region within Heschl's gyrus, various portions of STG and STS) for auditory feedback processing; and the insula presumably during multimodal processing of sensory feedback. In addition, premotor and parietal areas are recruited during human vocalization, and their functional roles will be further discussed below.

Until this point, both speech and song studies have been included to outline the brain regions associated with general vocal control in humans, since speaking and singing employ common mechanisms involved in vocal production. Moving forward, we will focus more on singing studies to examine how musical training modulates the

general functional network for human vocalization as it is used for singing.

Training Effects on the Sensory-Motor Control of Singing

Vocal Training Effects on the Neural Correlates of Sensory-Motor Control of Singing

In general, due to their extensive auditory-motor training and experience, musicians excel in various auditory and motor tasks. For instance, previous studies report that musicians perform better at pitch, timbre, and voice discrimination tasks than non-musicians (Kishon-Rabin et al., 2001; Tervaniemi et al., 2005; Chartrand and Belin, 2006; Micheyl et al., 2006). In addition to possessing better auditory discrimination skills than non-musicians, musicians also display more precise control over the vocal apparatus in the absence of proper auditory feedback. For example, trained singers sang more accurately with masked auditory feedback than non-musicians (Schultz-Coulton, 1978), yet one study reported the reverse (Watts et al., 2003). However, Watts' group of singers may have had less vocal training than the singers in Schultz-Coulton's study; Watts suggested that during the earlier stages of vocal training, more emphasis is placed on monitoring auditory feedback for vocal accuracy (Watts et al., 2003), which may account for their recruited singers' greater vocal inaccuracy with masked feedback compared to non-musicians. In fact, in a longitudinal study with trained singers performing various slow and fast singing tasks, vocal accuracy was not differentially affected by masked auditory feedback neither before nor after 3 years of vocal training (Mürbe et al., 2004), which suggests that auditory feedback may not play a crucial role in vocal accuracy after extensive vocal training. Nevertheless, vocal accuracy did improve during slow singing tasks with masked feedback after vocal training, which Mürbe et al. (2004) attributed to training-enhanced “neuromuscular memory of pitch” (p. 240). This implies that trained singers may rely more on somatosensory feedback to make sure that notes are produced properly, since they can still sing accurately for some time after losing their hearing (Wyke, 1974). Indeed, a functional magnetic resonance imaging (fMRI) singing study demonstrated that both vocal students (enrolled in a performance program) and professional opera singers recruited more activity within S1 and somatosensory association cortex than amateur singers, and moreover, the amount of singing practice positively correlated with the activity in these regions (Kleber et al., 2010). In a more recent fMRI study, Kleber et al. (2013) effectively reduced the amount of somatosensory feedback available by applying a topical anesthetic to the vocal folds just prior to singing in the MR scanner. The investigators determined that under vocal-fold anesthesia, singers displayed reduced activity in the right anterior insula than non-musicians, who had enhanced insular activity with anesthesia. Additionally, this region exhibited decreased functional connectivity to M1, S1, and auditory cortex in singers under topical anesthesia, while functional connectivity increased between these regions in non-musicians with anesthetized vocal folds. Notably, singers still sang more accurately under anesthesia than non-musicians, despite the observed reduction of insular activity and functional connectivity. Both of Kleber's experiments provide evidence that: (1) singers may rely more heavily on somatosensory feedback as a function of vocal training and practice, and (2) singers, perhaps by virtue of their training, can regulate activity within the right anterior insula to “disengage” or ignore somatosensory feedback when it is perturbed or

deemed unreliable and thus may significantly alter their singing performance.

Similar to the somatosensory feedback perturbation induced in Kleber's recent study, [Zarate and colleagues \(2008, 2010b\)](#) utilized pitch-shifted auditory feedback with fMRI techniques to target explicitly the brain regions involved in auditory-vocal motor control in singing. As discussed earlier, pitch-altered feedback elicits pitch-shift responses that often contain early and late components. Larson and colleagues suggested that the early pitch-shift response, which may be governed by the midbrain PAG, is a more automatic reaction used to stabilize vocal output by correcting small, unexpected fluctuations in vocal pitch; the late pitch-shift response, on the other hand, may be under more voluntary control—perhaps controlled by the auditory cortex, ACC, etc.,—and thus may contribute to vocal pitch control during speaking and singing ([Burnett et al., 1998](#); [Larson, 1998](#); [Hain et al., 2000](#); [Liu and Larson, 2007](#)). Indeed, although trained singers exhibit early pitch-shift responses to briefly pitch-shifted feedback, they were still able to maintain their intended goal for vocalization (either sustaining a steady pitch or glissandos, [Burnett and Larson, 2002](#); [Hafke, 2008](#)), perhaps due to enhanced top-down control of the late pitch-shift response that resulted from years of vocal training. In contrast, non-musicians may not exhibit such precise vocal control over the late pitch-shift response. To assess the effects of extensive vocal training on pitch control in singing, [Zarate and colleagues \(2008, 2010b\)](#) tested singers and non-musicians with two singing tasks that required different types of top-down voluntary control: (1) an “ignore” task where subjects were required to hold their pitch steady, despite hearing pitch-shifted auditory feedback; and (2) a “compensate” task in which subjects had to voluntarily adjust their vocal pitch precisely to correct for the pitch shift. The authors hypothesized that ignoring a small pitch shift would not only elicit an early pitch-shift response, but also target the PAG relative to the compensate task, which was specifically designed to engage their proposed cortical substrates for auditory-motor control of vocal pitch—auditory cortex, insula, and ACC ([Zarate and Zatorre, 2008](#); [Zarate et al., 2010b](#)).

Due to the temporal limitations of fMRI methodology, [Zarate et al. \(2010b\)](#) were not able to determine whether the PAG is involved particularly with eliciting early pitch-shift responses, since these responses have a latency that is shorter than the best temporal resolution for fMRI. Nevertheless, two interesting cortical findings from their singing tasks were observed. First, both groups recruited the IPS and dorsal premotor cortex (dPMC) in each pitch-shifted singing task, compared to singing with normal feedback ([Zarate and Zatorre, 2008](#)). The authors suggested that since the IPS is associated with transformations of sensory input for motor preparation ([Astafiev et al., 2003](#); [Grefkes et al., 2004](#); [Tanabe et al., 2005](#)), it was recruited specifically during transformations of auditory input (see [Foster and Zatorre, 2010](#); [Zatorre et al., 2010](#); [Foster et al., 2013](#)) into spatial information within the frequency domain (i.e., up or down). This “frequency spatial information” can then be used by the dPMC—an area that receives indirect connections from auditory and parietal areas via the insula ([Mufson and Mesulam, 1982](#)), and is attributed to conditional sensory-motor associations ([Petrides, 1986](#); [Chouinard and Paus, 2006](#))—to prepare a vocal response (e.g., maintain steady vocal output or correct for the pitch shift). Second, despite the observed lack of performance differences in the compensate task—i.e., both groups voluntarily adjusted for the pitch-shifted feedback to a similar extent—different neural substrates for auditory-motor control

were recruited in each group. Compared to singers, the non-musicians exhibited more activity within the dPMC while voluntarily correcting for the pitch shift (Figure 2A; Zarate and Zatorre, 2008); the authors proposed that the dPMC was recruited selectively in non-musicians as they learned to associate a pitch-shift “cue” in auditory feedback with a corrective adjustment in vocal pitch. Therefore, this region may constitute a basic substrate for voluntary auditory-motor control of vocal pitch (Zarate and Zatorre, 2008) and perhaps music production in general—after more training and practice, the dPMC is recruited less in non-musicians during the same musical production task that was learned (and assessed with fMRI) at earlier stages of an experiment (Chen et al., 2012). Indeed, rather than recruiting the dPMC, singers engaged auditory cortex within the pSTS, anterior insula, and ACC for this task (Figure 2B; Zarate and Zatorre, 2008; Zarate et al., 2010b). Moreover, voluntary vocal-control singing tasks (i.e., compensating for and ignoring large pitch shifts in feedback) specifically enhanced the functional connectivity between the pSTS and IPS (Figure 2C; Zarate et al., 2010b). Given the IPS' role in sensory-motor transformations, Zarate and colleagues suggested that within singers, the auditory cortex and IPS jointly process and extract pitch-shift information that can be used to control vocal pitch (e.g., magnitude and direction of the pitch shift). Since the auditory cortex is functionally connected to the insula and ACC (Zarate and Zatorre, 2008; Zarate et al., 2010b), the pitch-shift information may be sent via the anterior insula to the ACC for initiation of the task-appropriate vocal motor program (i.e., maintain the originally produced note or correct for the shift). The authors proposed that these four cortical regions constitute an experience-dependent network for auditory-motor control of the singing voice, which may be recruited increasingly as a function of more vocal training and practice.

figure 2

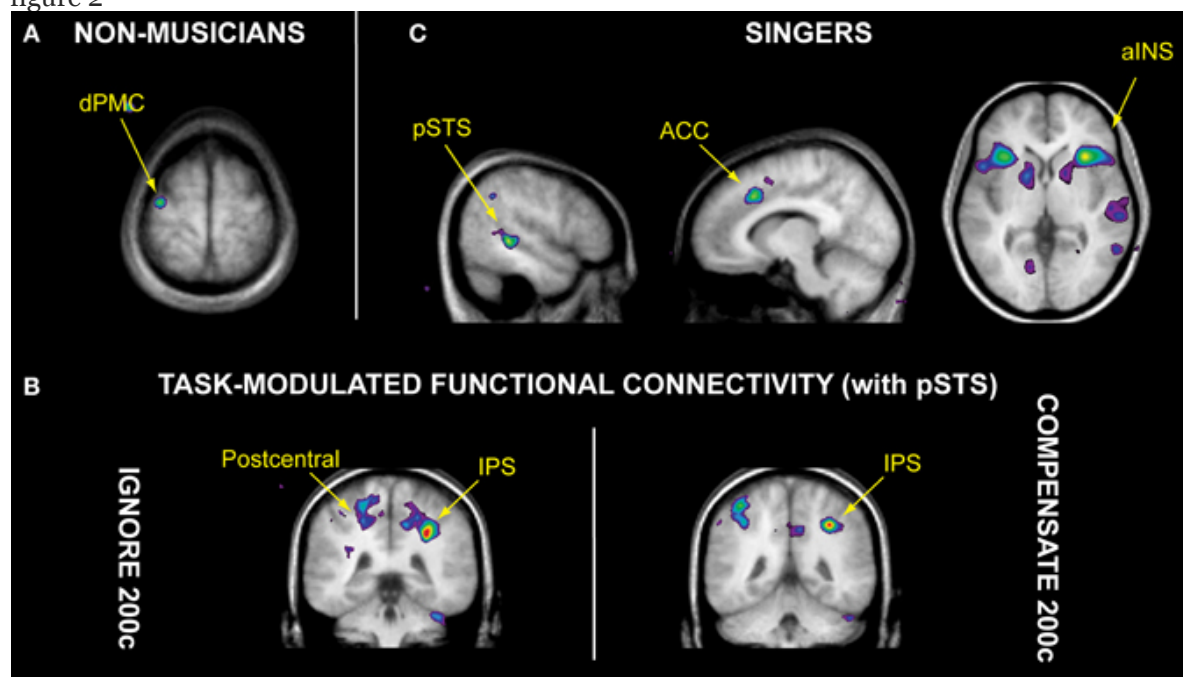


FIGURE 2. BRAIN REGIONS INVOLVED IN AUDITORY-MOTOR CONTROL OF SINGING, AS OBSERVED IN NON-MUSICIANS AND SINGERS. (A) When voluntarily correcting for a 200-cent pitch shift in auditory feedback (“compensate 200c” task), non-musicians recruited more activity within the dorsal premotor cortex (dPMC) than singers. (B) Singers engaged the posterior superior temporal sulcus

(pSTS), anterior cingulate cortex (ACC), and anterior insula (aINS) when performing the “compensate 200c” task. (C) Analyses of task-modulated functional connectivity revealed that relative to singing with normal auditory feedback, the 200-cent pitch shift specifically enhanced functional connectivity between right pSTS and intraparietal sulcus (IPS) during both the “ignore 200c” and “compensate 200c” tasks, as well as the postcentral gyrus (containing somatosensory cortex) during the “ignore 200c” task. Data from [Zarate and colleagues \(2008, 2010b\)](#).

Short-Term Training Effects on Auditory and Vocal Skills and their Neural Correlates

Based on the studies above, trained singers may have more precise vocal control compared to non-musicians, due to extensive vocal training that recruits an experience-dependent cortical network and/or selectively gates access to sensory feedback within this network. However, [Amir et al. \(2003\)](#) determined that instrumental musicians (without formal vocal training) also sang more accurately than non-musicians in a simple pitch-matching task, in which subjects were required to sing a note that was just presented. Additionally, two studies report a significant correlation between pitch discrimination and vocal accuracy in both instrumental musicians and non-musicians—individuals who sang more accurately also had better discrimination skills ([Amir et al., 2003](#); [Watts et al., 2005](#)). If this observed correlational relationship is a causal one, as these studies suggest, then refining pitch-discrimination skills may lead to better vocal accuracy. For instance, many studies have reported that auditory training improves pitch discrimination both at the training frequency and at other non-trained frequencies ([Demany, 1985](#); [Delhommeau et al., 2002, 2005](#); [Ari-Even Roth et al., 2003](#)). Furthermore, the effects of auditory training with pure tones also generalize to more complex tones ([Grimault et al., 2003](#)). In light of these observations and the proposed causal relationship between pitch discrimination and vocal accuracy, the newly enhanced ability to discriminate between pitches (following training) may increase the likelihood of detecting slight errors in vocal output, which may result in increased vocal accuracy. In turn, these training-induced behavioral changes are often accompanied by neural plasticity. For example, after non-musicians had received pitch-discrimination training, improved pitch discrimination was accompanied by enhanced auditory cortical responses ([Bosnyak et al., 2004](#)). Additionally, when non-musicians were trained to associate specific piano keys with their corresponding pitches and play short piano melodies, significant training-induced increases in cortical activity were observed within auditory, sensorimotor, frontal, and parietal regions ([Bangert and Altenmüller, 2003](#); [Lahav et al., 2007](#)).

Therefore, to examine whether: (1) singing accuracy improves subsequent to auditory training, and (2) auditory-training enhanced singing specifically engaged the experience-dependent network for auditory-motor control in singing (i.e., auditory cortex, IPS, anterior insula, and ACC), [Zarate et al. \(2010a\)](#) tested two groups of non-musicians—an experimental group that received training to improve their auditory discrimination skills, and a control group that received no training—with auditory discrimination and singing tasks. In this study, the investigators employed more naturalistic melodic singing tasks to target the experience-dependent network, since accurate production of novel melodies requires auditory-motor control in a similar

fashion as voluntarily correcting for pitch-shifted feedback; the auditory feedback of the currently produced note may be monitored in order to produce the correct pitch interval to the next note. Although the experimental group displayed enhanced auditory discrimination skills and training-induced changes in auditory task-associated neural activity (Zatorre et al., 2012), they did not show significant improvements in singing performance or recruit the experience-dependent network for auditory-motor control in singing (Zarate et al., 2010a). Consequently, Zarate et al. (2010a) concluded that auditory training alone (at least in an experimental setting) is not sufficient to improve vocal performance or recruit the experience-dependent network for auditory-motor control of singing (auditory cortex, IPS, anterior insula, and ACC); perhaps only simultaneous enhancements in both auditory and vocal motor skills via extensive training (e.g., voice lessons) would bring forth improvements in vocal performance and engage this particular network.

Sensory-Motor Control of Singing in Other Populations

Acquired Vocal Amusia

Clinical evidence that complements the proposed roles of the auditory cortex, IPS, S1, insula, and premotor regions during singing comes from case reports of brain lesions that result in vocal amusia or oral-expressive amusia (for a review, see Berkowska and Dalla Bella, 2009; Stewart et al., 2009). For instance, a woman with cortical atrophy in the right temporal lobe and insula, as well as diminished blood flow to right frontal and temporal regions, exhibited signs of progressive amusia and aprosodia—she gradually was incapable of perceiving and producing well-known melodies and affective intonation or prosody in speech (Confavreux et al., 1992). Additionally, a female tango singer who suffered a right-lateralized cerebral infarction presented with damage to right Heschl's gyrus and STG, inferior parietal regions including supramarginal gyrus and S1, and posterior insula; her music perception was greatly diminished post-stroke (relative to speech discrimination), and her singing was considered less stable within single notes, less accurate in pitch, and monotonous in affect (Terao et al., 2006).

While the two previous cases with damage to auditory cortex, insula, and other regions within the singing network presented with deficits in both music perception and production, two additional cases present perhaps the strongest evidence for these regions' involvement specifically for singing in the absence of impaired auditory perception. In a female patient who suffered a stroke in the right hemisphere affecting the lateral frontal lobe and M1, STG, insula, S1, and inferior parietal lobe, investigators observed impaired affective intonation in speech and the inability to sing pitch intervals accurately, while familiar-song perception and singing rhythms or melodic contour were relatively preserved (Murayama et al., 2004). Finally, a male amateur singer with right-lateralized damage to his posterior temporal lobe, inferior parietal lobe, insula, and inferior frontal gyrus presented with relatively spared speech comprehension and production, prosodic perception and production, music perception, and rhythm production; however, he exhibited specifically impaired pitch-interval production (Schön et al., 2004). This rather pure case of vocal amusia—in the absence of aphasia, aprosodia, and “perceptual” amusia—demonstrates that the damaged brain regions, which overlap with the areas

outlined by [Zarate and colleagues \(2008, 2010b\)](#), contribute to the finely-grained sensory-motor control of singing.

Congenital Amusia

Recall that the same neural network is recruited for singing in healthy individuals, irrespective of the amount of vocal training or experience (see section Neuroimaging Evidence: A General Functional Network For Human Vocalization). However, when pitch processing is compromised as observed in congenital amusia ([Ayotte et al., 2002](#); [Peretz and Hyde, 2003](#); [Foxton et al., 2004](#))—due to cortical malformations in the STG and inferior frontal gyrus ([Hyde et al., 2007](#)) and disrupted structural and functional connectivity ([Loui et al., 2009](#); [Hyde et al., 2011](#))—it may be assumed that pitch production in singing would similarly be affected as well. Yet, as observed in [Murayama's et al. \(2004\)](#) and [Schön's et al. \(2004\)](#) case reports, a dissociation between pitch perception and production skills can exist—following a stroke, spared pitch perception does not necessarily preclude inaccurate pitch production. Conversely, some individuals with congenital amusia still can sing pitch changes in the correct direction (e.g., up vs. down), match target notes, and sing familiar song excerpts somewhat accurately, despite observed problems with pitch perception ([Ayotte et al., 2002](#); [Loui et al., 2008](#); [Dalla Bella et al., 2009](#); [Hutchins et al., 2010](#)).

Based on this behavioral evidence, as well as observations of singing in the general population, Berkowska and Dalla Bella proffered a “vocal sensorimotor loop” model to outline two functional pathways within the song system that may explain observations of accurate-pitch and poor-pitch singing ([Berkowska and Dalla Bella, 2009](#); [Dalla Bella et al., 2011](#)). In this model, the authors list potential brain regions—based on previous neuroimaging studies, many of which are included in the section Neuroimaging Evidence: A General Functional Network For Human Vocalization—that contribute to mechanisms underlying singing, such as: regions within the STG for processing auditory input, which includes the auditory target to be reproduced and auditory feedback; dorsal prefrontal cortex, inferior sensorimotor cortex, area “Spt” within the planum temporale, and insula for auditory-motor mapping and memory access; supplementary motor area, ACC, and insula for motor preparation; and ventral M1 for vocal motor execution. Berkowska and colleagues also make distinctions between two pathways—a covert pathway involved in pitch discrimination (that can be compromised in congenital amusia), and an overt pathway involved in pitch production—but they do not clarify which of the aforementioned brain regions belong to each pathway. Congenital amusia may be due to a structural and functional “disconnection” between right auditory and inferior frontal cortical regions that contribute to pitch processing—although the right auditory cortex exhibits differential responses to pitch changes, the right inferior frontal cortex does not show a correlated increase in activity, as it does in normal listeners ([Hyde et al., 2011](#)). Even though this particular covert pathway is affected, auditory input (e.g., presented auditory targets, auditory feedback, etc.) can still be processed by auditory cortex ([Moreau et al., 2009](#); [Peretz et al., 2009](#); [Moreau et al., 2013](#)). Hypothetically speaking, auditory input may then be processed further by IPS (depending on the amount of vocal training), anterior insula, and premotor regions (dPMC or ACC) for auditory-motor control of singing based on Zarate's findings ([Zarate and Zatorre, 2008](#); [Zarate et al., 2010b](#)),

rendering vocal production relatively spared in some instances of congenital amusia.

Comparisons with Models of Auditory Processing

Berkowska and Dalla Bella's (2009), Dalla Bella et al.'s (2011) vocal sensorimotor loop model for singing, when enriched with neuroimaging evidence from Zarate and Zatorre (2008), Hyde et al. (2011), and Loui et al. (2009), potentially consists of auditory and inferior frontal cortex in the covert perception pathway (Figure 3, blue arrow), and auditory cortex, IPS, anterior insula, and premotor areas in the overt production pathway (Figure 3, red arrows). These updated pathways resemble the more recognized (and widely debated) dual-stream model for auditory processing, which was first proposed by Rauschecker and Tian (2000). The dorsal stream was originally suggested to be specialized for processing auditory spatial information (the “where” pathway), while the ventral stream was attributed with processing auditory object/sound identity information (the “what” pathway). The scientific debate focuses mostly on competing accounts and hypotheses of the dorsal stream's contributions, which include: (1) processing spectral changes over time (the “where in frequency” or “how” pathway, Belin and Zatorre, 2000); (2) extracting relevant sound features and matching them with stored templates of motor responses (the “do” pathway, Warren et al., 2005); (3) transforming auditory representations of speech into motor programs for speech gestures (Hickok and Poeppel, 2000, 2004, 2007); and (4) comparing between feedforward and feedback mechanisms (Rauschecker and Scott, 2009).

figure 3

UPDATED VOCAL SENSORIMOTOR LOOP MODEL

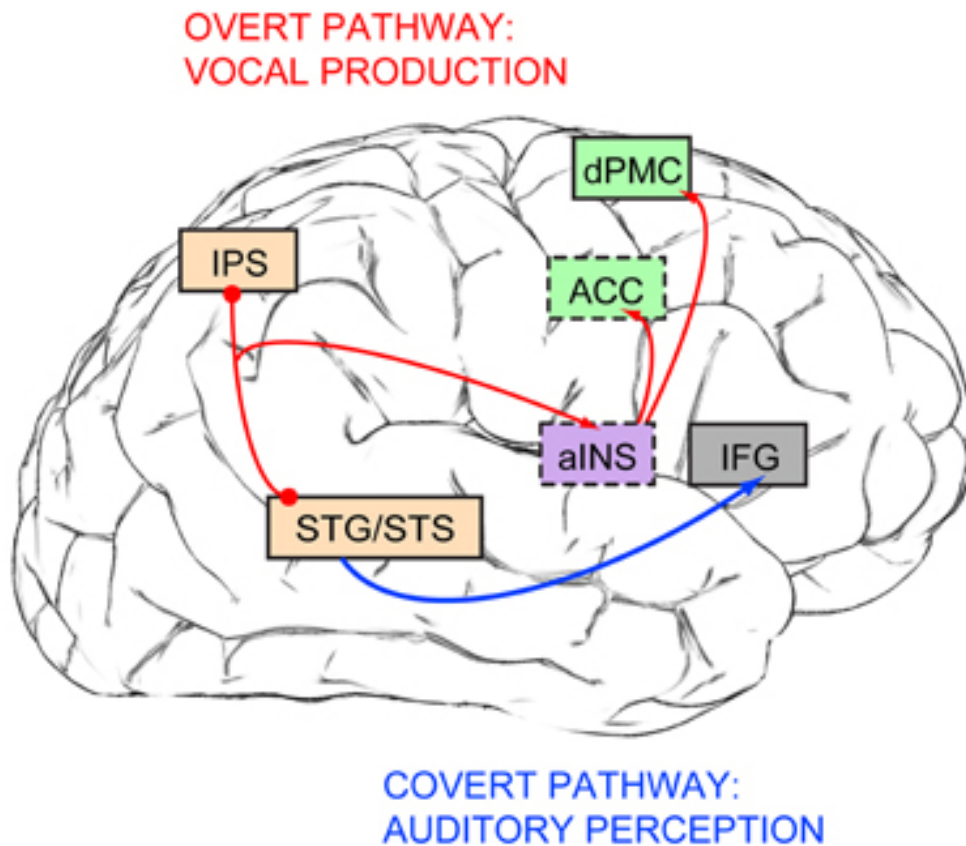


FIGURE 3. A REVISED VERSION OF [BERKOWSKA AND DALLA BELLA'S, DALLA BELLA, AND COLLEAGUES' \(2009, 2011\)](#) VOCAL SENSORIMOTOR LOOP MODEL FOR SINGING, UPDATED WITH FINDINGS FROM [ZARATE AND COLLEAGUES \(2008, 2010B\)](#) FMRI STUDIES. The covert pathway for pitch production (blue arrow) includes auditory cortex and inferior frontal gyrus (IFG), while the overt pathway for vocal pitch production (red arrows) is comprised of auditory cortex (STG/STS), intraparietal sulcus (IPS), anterior insula (aINS), anterior cingulate cortex (ACC), and dorsal premotor cortex (dPMC). Brain regions that are not visible normally from this lateral brain view are indicated in boxes outlined with dashes. Box colors are retained from Figure 1: light orange for auditory processing, green for vocal motor control, purple for multimodal processing.

For our purposes here, the most relevant dorsal-stream models are the spectrotemporal processing account from [Belin and Zatorre \(2000\)](#) and auditory-motor transformation hypotheses for auditory spatial processing and speech from [Warren et al. \(2005\)](#) and [Hickok and Poeppel \(2000, 2004, 2007\)](#). It should be noted, however that the auditory-motor control network for singing conflicts with the latter two models, in which area Spt in the planum temporale is the sole neural substrate for auditory-motor transformations ([Hickok and Poeppel, 2000, 2004; Warren et al., 2005; Hickok and Poeppel, 2007](#)). Zarate's singing research ([2008, 2010b](#)) provides empirical evidence both supporting, and perhaps, updating these dorsal-stream models—auditory cortex and IPS process and extract pitch changes from feedback, and the pitch information is sent from these regions via

the insula to premotor areas for vocal motor adjustments. Therefore, according to these neuroimaging findings, transformations of task-relevant auditory features into subsequent motor responses may not take place in only one brain region, as purported by the Warren et al. and Hickok/Poeppel models, but rather may be parceled among a network of different areas within the dorsal auditory stream. Thus, it could be argued that many brain regions along the dorsal auditory stream are involved in processing “how” auditory features change over time before executing or “doing” a specific motor act in response to these auditory events, regardless of the particular modality—be it information related to auditory space, speech, or music.

Conclusion

In this review, findings from over 20 years of research have been reviewed to outline a general neural network for song and speech production (section Neuroimaging Evidence: A General Functional Network For Human Vocalization). Within this functional network, cortical substrates that are specific for the sensory-motor control of singing pitch and are sensitive to the amount of vocal training have been identified (Figure 4): the pSTS and IPS for auditory processing and transformation for motor output (light orange boxes), S1 for somatosensory processing (yellow box), anterior insula (in purple, both for auditory-motor integration and somatosensory feedback gating), and premotor regions for vocal motor preparation and response initiation (dPMC and ACC, in green). When the auditory-related findings are placed within a larger framework—a dual-pathway (i.e., perception vs. production), sensory-motor model for singing ([Berkowska and Dalla Bella, 2009](#))—these music-specific findings can then be linked to broader research interests in auditory cognition, such as auditory spatial localization and speech perception/production, due to the auditory-motor control network's similarity to prevalent dual-stream models of auditory processing as a whole.

figure 4

TRAINING-SENSITIVE SENSORY-MOTOR AREAS FOR SINGING

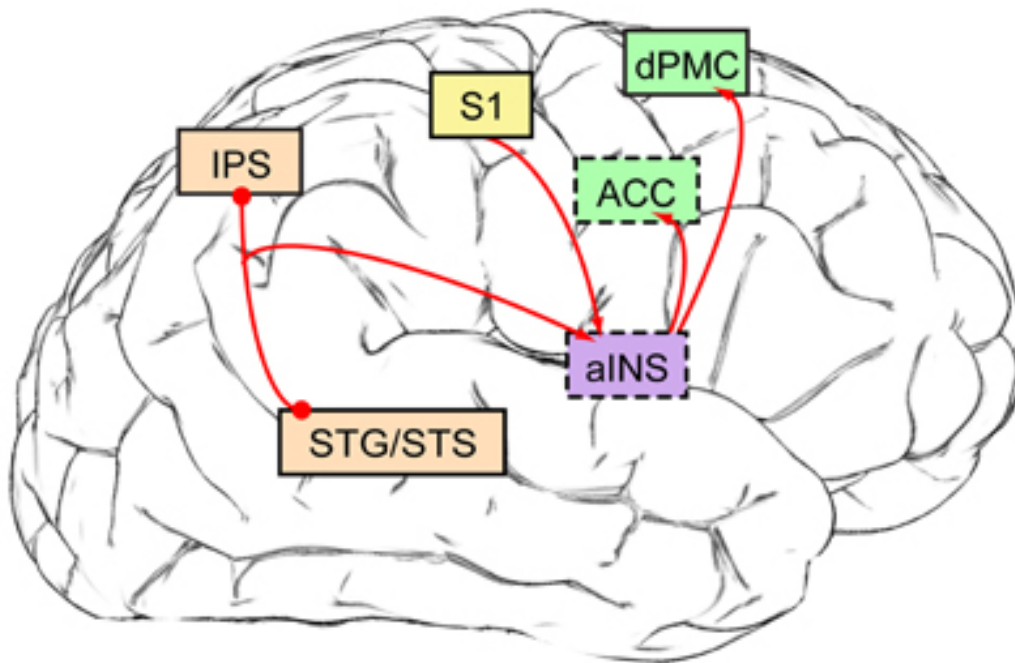


FIGURE 4. NEURAL SUBSTRATES FOR SENSORY-MOTOR CONTROL OF SINGING THAT ARE SENSITIVE TO THE AMOUNT OF VOCAL TRAINING [BASED ON FINDINGS FROM [KLEBER ET AL. \(2010, 2013\)](#), [ZARATE AND ZATORRE \(2008\)](#), [ZARATE ET AL. \(2010B\)](#)]. Brain regions that are not visible normally from this lateral brain view are indicated in boxes outlined with dashes, and box colors are retained from Figures 1 and 3. Activity within primary somatosensory cortex (S1) increases as a function of the amount of weekly vocal practice, suggesting a greater reliance on somatosensory feedback with more training and experience. After extensive vocal training and practice, the anterior insula (aINS) can serve a gating function for somatosensory feedback. Features within auditory feedback are processed and extracted by auditory cortex (STG/STS) and the intraparietal sulcus (IPS), and task-relevant auditory information is sent via the aINS to the dorsal premotor cortex (dPMC)—in people with little to no formal vocal training—or to the anterior cingulate cortex (ACC) in experienced singers to voluntarily adjust vocal output according to the singing task demands.