Sharing and Nonsharing of Brain Resources for Language and Music

Aniruddh D. Patel

Abstract

Several theoretical and practical issues in cognitive neuroscience motivate research into the relations of language and music in the brain. Such research faces a puzzle. Currently, evidence for striking dissociations between language and music coexists with evidence for similar processing mechanisms (e.g., Peretz 2006; Patel 2008). The intent of this chapter is to initiate a dialog about how such conflicting results can be reconciled. Clearly, a coherent picture of language–music relations in the brain requires a framework that can explain both kinds of evidence. Such a framework should also generate hypothesis to guide future research. As a step toward such a framework, three distinct ways are put forth in which language and music can be dissociated by neurological abnormalities, yet have closely related cortical processing mechanisms. It is proposed that this relationship can occur when the two domains use a related functional computation and this computation relies on (a) the same brain network, but one domain is much more robust to impairments in this network than the other, (b) the interaction of shared brain networks with distinct, domain-specific brain networks, or (c) separate but anatomically homologous brain networks in opposite cerebral hemispheres. These proposals are used to explore relations between language and music in the processing of relative pitch, syntactic structure, and word articulation in speech versus song, respectively.

Background: Why Study Language–Music Relations in the Brain?

From the standpoint of cognitive neuroscience, there are at least five distinct reasons to study language–music relations. Such studies are relevant to (a) comparative research on the neurobiological foundations of language, (b) debates over the modularity of cognitive processing, (c) evolutionary questions surrounding the origins of language and music, (d) the clinical use of
music-based treatments for language disorders, and (e) educational issues concerning the impact of musical training on linguistic abilities, such as reading and second language learning.

In terms of comparative research on language, a common strategy for exploring the neurobiological foundations of human behavior is to examine animal models. For example, the brain mechanisms of decision making are increasingly being studied using animal models (Glimcher 2003). This comparative approach, however, has certain limitations when it comes to human language. Animals communicate in very diverse ways, but human language differs in important respects from all known nonhuman systems. To be sure, certain abilities relevant to language, such as vocal learning and the perception of pitch, are shared with other species and have good animal models (e.g., Bendor and Wang 2005; Jarvis 2007b; Fitch and Jarvis, this volume). Furthermore, some of the basic auditory cortical mechanisms for decoding and sequencing complex sounds may be similar in humans and other primates or birds, making these species important model systems for studying the evolutionary foundations of speech perception (Doupe and Kuhl 1999; Rauschecker and Scott 2009; Tsunada et al. 2011). Nevertheless, some of the core features of language, such as complex syntax and rich semantics, have no known counterparts in the communication systems of other species. It is notable that even with enculturation in human settings and intensive training over many years, other animals attain very modest syntactic and semantic abilities compared to humans (Fitch 2010). This sets certain limits on what we can learn about the neurobiology of high-level language processing via research on other brains.

Yet within our own brain lies a mental faculty which offers the chance for rich comparative research on high-level language processing, namely music. Like language, music involves the production and interpretation of novel, complex sequences that unfold rapidly in time. In terms of syntactic structure, these sequences have three key properties in common with language (cf. Lerdahl and Jackendoff 1983):

1. **Generativity:** novel sequences are built from discrete elements combined in principled ways.
2. **Hierarchy:** sequences have a rich internal structure, with multiple levels of organization.
3. **Abstract structural relations:** elements fill distinct structural roles depending on the context in which they occur. (For example, in English, the word “ball” can be the subject or object of a verb depending on context, and in tonal music the note B-flat can be the tonic or leading tone in a musical sequence, depending on context).

In terms of semantics, musical sequences, like linguistic sequences, can convey complex and nuanced meanings to listeners (Patel 2008:300–351). While the nature of musical meaning and its relationship to linguistic meaning is a topic of active discussion and debate (e.g., Antović 2009; Koelsch 2011b and
responses in the same journal), the critical point is that the meanings of musical sequences are rich and varied. This stands in contrast to animal songs (e.g., birdsong, whale songs), which convey simple meanings such as “this is my territory” or “I’m seeking a mate.” Thus, both in terms of syntax and semantics, music offers a rich domain for the comparative neurobiological study of language.

Turning to issues of modularity, language–music research is relevant to a persistent question in cognitive science (Fodor 1983; Elman et al. 1996; Peretz and Coltheart 2003; Patel 2003): To what extent are the functional computations underlying particular mental faculties (e.g., language or music) unique to those domains? For example, are certain aspects of linguistic syntactic processing supported by brain mechanisms shared by other types of complex sequence processing? Instrumental (nonverbal) music provides an excellent tool for studying this question, as it is a nonlinguistic system based on hierarchically organized, rule-governed sequences. Hence, if instrumental music processing and linguistic syntactic processing share neural mechanisms, this would inform modularity debates and (more importantly) provide new, comparative methods for studying the neurobiology of syntactic processing.

In terms of evolution, there is a long-standing debate over the role that music or music-like vocalizations played in the evolution of language. Darwin (1871) proposed that human ancestors sang before they spoke; that is, they had a nonsemantic “musical protolanguage” which laid the foundation for the evolution of articulate speech. His contemporary, Herbert Spencer, disagreed and argued that music was a cultural elaboration of the sounds of emotional speech (Spencer 1857). Spencer foreshadowed thinkers such as James (1884/1968) and Pinker (1997), who argue that our musicality is a byproduct of other cognitive and motor abilities, rather than an evolved trait which was selected for in evolution. This debate is very much alive today. Current proponents of the musical protolanguage theory (e.g., Mithen 2005; Fitch 2010) have updated Darwin’s ideas in the light of modern research in anthropology, archeology, linguistics, and cognitive science. As noted by Fitch (2010:506), “The core hypothesis of musical protolanguage models is that (propositionally meaningless) song was once the main communication system of prelinguistic hominids.” Fitch proposes that the neural mechanisms underlying song were the precursors of phonological mechanisms in spoken language, a view that predicts “considerable overlap between phonological and musical abilities (within individuals) and mechanisms (across individuals).” Fitch argues that such a prediction is not made by lexical and gestural protolanguage hypotheses. In other words, Fitch regards music–language relations in the brain as important evidence for resolving debates over the evolution of language (see Fitch and Jarvis, this volume; for a critique see Arbib and Iriki, this volume).

In terms of clinical issues, there is growing interest in the use of music as a tool for language remediation in certain developmental and acquired language disorders (e.g., dyslexia and nonfluent aphasia, Goswami 2011; Schlaug et al.)
To determine which types of language problems might respond to musical training, and to explore how such benefits might take place, it is important to have a basic understanding of how music and language processing are related at the neurobiological level. This same kind of basic understanding is important for educational questions surrounding the benefits of music training to linguistic abilities in normal, healthy individuals. A growing body of evidence suggests that musical training enhances certain linguistic skills in normal individuals, including reading abilities, second language learning, and hearing speech in noise (Moreno et al. 2009; Moreno et al. 2011; Slevc and Miyake 2006; Parbery-Clark et al. 2009). To optimize such effects and, ultimately, to help influence educational policy, we need a neurobiological understanding of how and why such effects occur.

**Nonsharing: Neuropsychological and Formal Differences between Language and Music**

As cognitive and neural systems, music and language have a number of important differences. Any systematic exploration of music–language relations in the brain should be informed by an awareness of these differences. This section discusses three types of differences between music and language: differences in acoustic structure (which are reflected in differences in hemispheric asymmetries), neuropsychological dissociations, and formal differences.

Beginning with acoustic structure, speech and music tend to exploit different aspects of sound. Speech relies heavily on rapidly changing spectral patterns in creating the cognitive categories that underlie word recognition (e.g., phonetic features, phonemes). Much music, on the other hand, relies on more slowly changing spectral patterns to create the cognitive categories that underlie music recognition (e.g., pitch classes, pitch intervals). Animal studies have revealed that different areas of auditory cortex show preferences for rapidly versus slowly changing spectral patterns (Tian and Rauschecker 1998; Rauschecker and Tian 2004), and fMRI research has pointed to distinct regions within the anterior superior temporal cortex for speech and musical instrument timbres (Leaver and Rauschecker 2010). Furthermore, a large body of evidence from patient and neuroimaging research suggests that musical pitch perception has a right-hemisphere bias in auditory cortex (e.g., Zatorre and Gandour 2007; Klein and Zatorre 2011). For example, Figure 14.1 shows the anatomical locations of lesions producing different kinds of pitch-related musical deficits, based on a review of the neuropsychological literature (Stewart et al. 2006).

Zatorre et al. (2002) have suggested that the right-hemisphere bias in musical pitch processing reflects a trade-off in specialization between the right and left auditory cortex (rooted in neuroanatomy), with right-hemisphere circuits having enhanced spectral resolution and left-hemisphere circuits having...
enhanced temporal resolution.\(^1\) Another proposal which could account for the right-hemisphere bias in musical pitch processing is Poeppel’s (2003) “asymmetric sampling in time” hypothesis, which proposes that the right auditory cortex has a longer temporal integration window for acoustic analysis than the left hemisphere, which would predispose it to fine-grained pitch analysis. (These proposals are largely equivalent from a systems-theoretical standpoint, since frequency resolution and time resolution are inversely related.) The right-hemisphere bias for musical pitch processing is not just present in auditory cortex, but is also observed in frontal and temporal regions to which the auditory cortex is connected via long reciprocal fiber tracts. For example, in Tillmann et al.’s (2003) fMRI study of harmonic processing, Brodmann areas 44 and 45 in both hemispheres showed an increased response to harmonically unexpected (vs. expected) chords, with a stronger response in the right hemisphere. The right-hemisphere bias for musical pitch processing stands in sharp contrast to the well-known left-hemisphere bias for many aspects of language.

---

1 Note, however, that when pitch is used to make linguistic distinctions (e.g., between the words of a tone language), pitch perception is often associated with significant left-hemisphere cortical activations. Zatorre and Gandour (2007) propose a framework combining bottom-up and top-down processing to account for the differences in cortical activation patterns observed when pitch plays a musical versus a linguistic role in communication.
processing, such as language production and syntactic processing (Hickok and Poeppel 2007; Hagoort and Poeppel, this volume).

Apart from acoustic structure and hemispheric asymmetry, another striking difference between music and language concerns the neuropsychological dissociations between these domains. For example, individuals with “congenital amusia” have severe, lifelong problems with basic aspects of music perception, such as discriminating between simple melodies or recognizing culturally common tunes. Such individuals may, however, function normally (or even excel) in other cognitive domains, and their everyday language abilities can seem perfectly intact. Structural neuroimaging has revealed subtle abnormalities in the brains of such individuals (Hyde et al. 2007), which likely have a genetic origin (Peretz et al. 2007). These abnormalities occur in multiple regions, including right-frontal and temporal cortices and their connections via the right arcuate fasciculus (Loui et al. 2009). Thus it appears that a neurogenetic condition can selectively affect musical but not linguistic development.

Neuropsychological dissociations between music and language can also occur after brain damage (e.g., stroke), as in cases of amusia without aphasia (deficits in music perception without any obvious language problems) and aphasia without amusia (impaired linguistic abilities but spared musical abilities). Such “double dissociations” are often considered strong evidence for the nonsharing of brain resources for music and language. Since dissociations between music and language due to brain damage have been extensively discussed by Peretz and others (e.g., Peretz and Coltheart 2003; Peretz 2012), they will not be discussed further here. For the current purposes, the key point is that any purported connections between music and language processing must be able to account for such dissociations.

The final set of differences between music and language that I wish to discuss concerns the formal properties of linguistic and musical systems. Let us begin with important aspects of language not reflected in music. One recent statement of such aspects is reproduced below. This list comes from a leading researcher in the field of child language acquisition (via personal communication):

1. The defining trait of human languages is the fact that there are two (largely) parallel structures. There are (a) elementary expressions (“words,” “morphemes”) and rules, according to which these can be combined to form more complex expressions, and (b) these elementary expressions have a conventionally fixed meaning, and compound expressions build up their meaning according to the way in which they are constructed. (There are many violations of this “principle of compositionality” but it is always the basis). Nothing like this exists in music.
2. There are typical functions which are expressed and systematically marked in all languages (e.g., statements, questions, commands). There is nothing comparable in music, except in a very metaphorical sense.

One may perhaps interpret a musical phrase as a “question” but it is not systematically marked, as in all languages. All natural languages have negation, but music does not have anything similar. Music does have many functions—and not just aesthetic—but they are quite different to those of language.

3. Linguistic expressions are systematically built on an asymmetry of various types of constructions (head-dependent element, operator–operand). Again, there is nothing comparable in music.

4. Linguistic expressions serve to express a particular cognitive content, which the interlocutor is supposed to understand. This is not true for (instrumental) music, again with minor exceptions. If you do not believe this, try to transform your last published paper into a piece of music so that someone (who has not read the paper) is able to understand its content.

While some theorists may dispute the claim that music lacks all of the above properties (e.g., in terms of Pt. 3, see Rohrmeier 2011), by and large the claims are likely to be acceptable to many experts on linguistic and musical structure (cf. Jackendoff 2009).

Shifting the focus to music, it is easy to identify properties of music that are not reflected in ordinary language. One useful compendium of common features in music is Brown and Jordania (2011). As noted by these authors, due to the vast and ever-growing diversity of world music, musical universals are “statistical universals” (i.e., widespread patterns reflecting the way music “tends to be”) rather than features present in every musical utterance. Some widespread features of music not found in human language are:

- Use of discrete pitches and intervals, drawn from underlying scales.
- Rooting of songs/melodies in the tonic (ground-pitch) of whatever scale type is being used.
- Predominance of isometric (beat-based) rhythms.
- Use of repetitive rhythmic and melodic patterns.
- Frequent occurrence of coordinated group production (i.e., group singing or playing).

The above formal differences between music and language, together with the neuropsychological differences discussed earlier, clearly demonstrate that music and language are not trivial variants of each other as cognitive systems. This is precisely what makes the discovery of related processing mechanisms in the two domains interesting. Indeed, given the differences outlined above, any related processing mechanisms are likely to be fundamental to human communication.

---

2 Negation does not mean contrast! It means to deny the truth of something.
Sharing: Three Types of Hidden Connections between Linguistic and Musical Processing

The neuropsychological and formal differences reviewed above suggest that relations between musical and linguistic cortical processing are not likely to be obvious; that is, they will be “hidden connections.” In the remainder of this chapter, I illustrate three types of hidden connections that can exist between linguistic and musical processing in the brain. In each case, musical and linguistic cortical processing rely on a similar functional computation, yet musical and linguistic abilities can be dissociated by brain damage. This relationship is possible when a functional computation in language and music relies on:

1. The same brain network, but developmentally one domain (language, music) is much more robust to impairments in this network.

The following sections illustrate these three situations, focusing on prosody, syntax, and the motor control of speech versus song, respectively.

Before proceeding, however, it is worth noting that there is one aspect of brain processing where music and language can, a priori, be expected to overlap. At subcortical auditory processing stages (e.g., in structures such as the cochlear nucleus and inferior colliculus; see Figure 14.2), the processing of music and spoken language can be expected to overlap to a large degree, since both domains rely on temporally and spectrally complex signals, often with salient harmonic structure and a discernable pitch. Support for this overlap comes from recent research, which shows that subcortical encoding of linguistic sounds is superior in individuals who have been trained on a musical instrument and that the quality of encoding is related to number of years of musical training (Kraus and Chandrasekaran 2010). In other words, learning a musical instrument appears to enhance the early auditory processing of linguistic sounds. This presumably occurs via mechanisms of neural plasticity, perhaps driven by top-down “corticofugal” pathways from the cortex to subcortical structures (cf. dashed lines in Figure 14.1). The relationship between musical training and brainstem encoding of speech is now an active area of research. Such research has important practical consequences since the quality of brainstem speech encoding has been associated with real-world language skills such as reading ability and hearing speech in noise. While there is growing empirical research in this area, there is a need for theoretical frameworks specifying why musical training would benefit the neural encoding of speech (for one such framework, see Patel 2011). Since the focus of this chapter is on cortical processing, language–music overlap at the subcortical level is not discussed further here.
Shared Networks: Relative Pitch Processing

As an example of a functional computation in language and music which may rely on the same brain network, we examine relative pitch processing. Humans effortlessly recognize the same speech intonation pattern, such as a “question” contour with a final pitch rise, when heard at different absolute frequency levels, for example, as spoken by an adult male or a small child (Ladd 2008). Humans also easily recognize the same musical melody (e.g., the “Happy Birthday” tune) when heard at different absolute frequency levels (e.g., played on a piccolo or a tuba). These skills rely on relative pitch: the ability to encode and recognize a pattern of ups and downs of pitch independent of absolute frequency level. Humans take this ability for granted: it requires no special training and is present in infancy (Trehub and Hannon 2006). Yet comparative research indicates that this ability is not universal among animals (McDermott and Hauser 2005). Starlings, for example, are animals that use acoustically
complex, learned songs for communication, yet have great difficulty recognizing a tone sequence when it is transposed up or down in pitch (Bregman et al. 2012). Unlike most humans, starlings appear to favor absolute frequency cues in tone sequence recognition (e.g., Page et al. 1989). This suggests that relative pitch processing requires special brain mechanisms, which abstract the pattern of ups and downs from pitch patterns, and use these as an important cue for sound sequence recognition.

Do speech and music share brain mechanisms for relative pitch processing? Evidence from the study of congenital amusia (henceforth, amusia) is relevant to this question. Amusics have severe deficits in musical melody perception, yet their speech abilities often seem normal in everyday life. Early research on their ability to discriminate pitch contours in speech on the basis of falling versus rising intonation contours (e.g., discriminating statements from questions) suggested that they had no deficits (Ayotte et al. 2002). However, such research used sentences with very large pitch movements (e.g., statements and questions with pitch falls or rises of 5–12 semitones). More recent research with smaller, but still natural-sounding pitch movements (4–5 semitones) has demonstrated that most amusics have deficits in discriminating between statements and questions (Liu et al. 2010). Furthermore, their performance on this task correlates with their psychophysically determined thresholds for discriminating upward from downward pitch glides. These findings are consistent with the idea that speech and music share a brain network for relative pitch processing and that, in amusics, this network is impaired, leading to a higher threshold for discriminating the direction of pitch movements than in normal listeners. According to the “melodic contour deafness hypothesis” (Patel 2008:233–238), this higher threshold disrupts the development of music perception, which relies heavily on the ability to discriminate the direction of small pitch movements (since most musical intervals are 1 or 2 semitones in size). However, it does not disrupt the development of speech perception, which is quite robust to modest problems in pitch direction discrimination, as speech tends to rely on larger pitch movements and often has redundant information which can compensate for insensitivity to pitch direction. Thus the melodic contour deafness hypothesis suggests that music and language use the same brain network for relative pitch processing, but that musical and linguistic development place very different demands on this network, resulting in different developmental outcomes.

What brain regions are involved in computing pitch direction? Lesion studies in humans point to the importance of right-hemisphere circuits on the boundary of primary and secondary auditory cortex (Johnsrude et al. 2000). It is likely, however, that these regions are part of a larger network that extracts relative pitch patterns and (when necessary) stores them in short-term memory for the purpose of recognition or comparison. Structural neuroimaging studies of amusic individuals have revealed abnormalities in a number of frontal and temporal brain regions in both hemispheres (Hyde et al. 2007; Mandell et al. 2007), as well as in connections between right frontal and temporal regions.
Sharing and Nonsharing of Brain Resources (Loui et al. 2009). Furthermore, functional neuroimaging shows that the right inferior frontal gyrus is underactivated in amusics during tone sequence perception (Hyde et al. 2010). Hence, right frontal and temporal regions and their connections are good candidates for being part of a network that processes relative pitch patterns. Furthermore, the intraparietal sulcus may be important in comparing pitch sequences on the basis of patterns of relative pitch (Foster and Zatorre 2010).

Identifying the different anatomical components of this network, and the role played by each, is an enterprise that could benefit from an integrated approach combining neuroimaging with computational models of relative pitch perception. Husain et al. (2004) have proposed, for example, a computational model of auditory processing that incorporates up and down frequency modulation selective units, as well as contour units, which are responsive to changes in sweep direction (Figure 14.3). The model includes auditory, superior temporal, and frontal brain regions, and builds on neurophysiological data on pitch direction processing, including several studies by Rauschecker (e.g., Rauschecker 1998). It has also been related to fMRI studies of auditory processing. This model could be used to simulate the pitch direction deficits of amusics (e.g., by disrupting the function of selected areas or by manipulating connectivity between areas) and may help guide the search for brain regions supporting relative pitch processing.

![Network model](image)

**Figure 14.3** Network diagram of the computational model of auditory pattern recognition proposed by Husain et al. (2004). Regions include medial geniculate nucleus (MGN), two regions in primary auditory cortex (Ai, Aii), superior temporal gyrus/sulcus (ST), and prefrontal cortex. In the prefrontal cortex region, C contains cue-sensitive units, D1 and D2 contain delay units and R contains response units. Reprinted from Husain et al. (2004) with permission from Elsevier.
Shared Networks Interacting with Domain-Specific Networks: Syntactic Processing

As an example of a functional computation in language and music that may rely on shared networks interacting with domain-specific networks, we examine structural integration in linguistic syntax and tonal harmony. In the past decade or so, a surprising amount of neuroimaging evidence has accumulated suggesting that the cortical processing of tonal harmonic relations in instrumental music shares neural resources with the cortical processing of linguistic syntax. This evidence is surprising for three reasons. First, a number of behavioral investigations of patients with musical deficits (following brain damage or due to congenital amusia) have demonstrated clear dissociations between tonal harmonic processing and linguistic syntactic processing (Peretz and Coltheart 2003). Second, tonal harmonic music (henceforth, “tonal music” or “tonality”) does not have the grammatical categories that are fundamental to language, such as nouns, verbs, subjects, or objects. Third, tonality and linguistic syntax serve very different ends. The former helps articulate the nuanced ebb and flow of tension and resolution in nonpropositional sound sequences, in the service of emotional and aesthetic communication (for an example using the music of the Beatles, see Jackendoff and Lerdahl 2006). The latter, in contrast, helps articulate argument structure within referential propositions, by specifying who did what to whom, as well as where, when, and why. (For a brief introduction to the structure of tonal harmonic music, oriented toward comparison with language, see Patel et al. 2008, Section 5.2).

I begin by briefly reviewing some of the key neuroimaging evidence for overlap in the processing of tonality and linguistic syntax (for related material, see Koelsch, this volume). This evidence raises the question of why these two seemingly very different kinds of processing should overlap in this way. I offer my own perspective on this subject and outline the “shared syntactic integration resource hypothesis” or SSIRH (Patel 2003). The SSIRH makes specific, testable predictions, including predictions about interference between tonality and linguistic syntax processing. Thus far, the predictions have been supported by empirical work, but this is a young line of research and more work is needed to explore precisely how and why tonality and linguistic syntax processing are related in the human brain.

Evidence from Neuroimaging Studies

The first neurobiological study to compare tonality processing and linguistic syntactic processing directly used event-related potentials or ERPs (Patel et al. 1998). ERPs have played an important role in neurolinguistic research. Two extensively studied ERP components are the N400, which is thought to reflect the cost of integrating a word’s meaning into the meaning representation of a sentence (or the ease of accessing information in semantic memory), and the
Sharing and Nonsharing of Brain Resources

P600, which is associated with syntactic processing (e.g., syntactic violations or ambiguities, or structural complexity, Gouvea et al. 2010).³

Patel et al. (1998) used music-theoretic principles of harmony and key-relatedness to construct chord sequences in which a target chord in the sequence was either in key, from a nearby key, or from a distant key compared to the rest of the phrase. All target chords were well-tuned major chords which sounded consonant in isolation; in the context of a chord sequence, however, the out-of-key chords sounded contextually dissonant due to their departure from the prevailing key. Participants in the study also heard spoken sentences in which a target word was easy, difficult, or impossible to integrate into the preceding syntactic structure. The critical finding was that the out-of-key chords elicited a bilateral P600 that was statistically indistinguishable from the P600 elicited by syntactically incongruous words (i.e., in terms of latency, amplitude, and scalp distribution in the 600 ms range). This was interpreted as evidence that similar processes of structural integration were involved in tonality and linguistic syntax processing.

ERPs do not provide firm information of the physical location of underlying neural generators, and thus direct tests for shared brain regions involved in linguistic syntactic and tonal harmonic processing await studies that use localizationist techniques (such as fMRI). In recent years, a few fMRI studies have compared sentence processing to instrumental musical sequence processing (e.g., Abrams et al. 2011; Fedorenko et al. 2011; Rogalsky et al. 2011b) and have reported salient differences in activation patterns associated with the processing of structure in the two domains. However, the manipulations of musical structure in these studies did not focus specifically on tonality, and the manipulations of language structure did not specifically target syntactic complexity in meaningful sentences. Hence, the door is open for future fMRI studies which focus on comparing linguistic syntactic processing to musical tonality processing, based on established principles of structural complexity in each domain.

At the moment, the strongest neuroimaging evidence for overlap in tonality and linguistic syntax processing comes from the research done by Koelsch and colleagues on brain responses to out-of-key chords in chord sequences. Patel et al. (1998) found that such chords, in addition to eliciting a P600, also elicited a right-hemisphere antero-temporal negativity (RATN) peaking around 350 ms, reminiscent of the left anterior negativity (LAN) associated with linguistic syntactic processing. While Patel et al. (1998) studied musically trained individuals, Koelsch et al. (2000) examined ERP responses to out-of-key chords in nonmusicians (i.e., individuals with no musical training, outside of normal

³ Recently, P600s and N400s have also been observed in cases where there are conflicts between semantic and syntactic information. For recent data and a theory which explains these findings while maintaining the distinction between the N400 as an index of semantic processing and the P600 as an index of syntactic processin, see Kos et al. (2010).
exposure to music in school). They discovered an early right anterior negativity (ERAN), peaking around 200 ms, in response to such chords, again reminiscent of a LAN associated with linguistic syntax processing, i.e., the early left anterior negativity or ELAN, which is associated with word category violations. (For evidence that the RATN and ERAN are two variants of a similar brain response, with different latency due to the rhythmic context in which a chord occurs, see Koelsch and Mulder 2002). In subsequent studies, Koelsch and colleagues have demonstrated that the ERAN occurs in both musicians and nonmusicians. In a series of pioneering studies on the neurobiology of tonality processing in musicians and nonmusicians (largely based on the perception of chord sequences), they have shown that the ERAN has several attributes reminiscent of syntactic processing:

1. The ERAN is elicited by structurally unexpected events, not by psychoacoustically unexpected events. Using a psychoacoustic model of sensory processing, they have shown that the ERAN is not simply a response to a sensory mismatch between the pitches of an incoming chord and a pitch distribution in short-term memory created by previous chords (Koelsch et al. 2007).
2. The ERAN is a response to departures from learned structural norms, not to deviations from local sound patterns in sensory memory. The ERAN is distinct from the mismatch negativity, in terms of both the nature of its eliciting events and in terms of its underlying neural generators (Koelsch 2009).
3. The amplitude of the ERAN is modulated by the degree of structural incongruity between target chord and local context (Koelsch et al. 2000).
4. Based on magnetoencephalography (MEG) studies, the generators of the ERAN include the inferior part of left BA 44 in Broca’s area (Maess et al. 2001), as well as its right-hemisphere homolog, which is an even stronger generator. This is of interest because left BA 44 (part of Broca’s region) appears to be involved in the processing of linguistic syntax (Hagoort 2005; Meltzer et al. 2010). (Note that the right-hemispheric bias in the scalp topography of the ERAN is rather weak and is even absent in some studies; for a review, see Koelsch 2009).
5. The ERAN is abnormal in individuals with lesions in left Broca’s area (Sammler et al. 2011).
6. The ERAN is absent in children with specific language impairment, a disorder which includes problems with linguistic syntax processing (Jentschke et al. 2008).
7. An ERP marker of linguistic syntactic processing (the ELAN) is enhanced in musically trained children, who also show an enhanced

MEG measures the magnetic fields generated by bioelectric currents in cortical neurons, and unlike EEG, can be used to reconstruct the location of underlying sources of brain activity.

ERAN compared to musically untrained children (Jentschke and Koeslch 2009).

8. Intracranial EEG measurements (from epileptic patients awaiting surgery) reveal partial overlap of the sources of the ERAN and ELAN in bilateral superior temporal gyrus and, to a lesser extent, in the left inferior frontal gyrus (although electrode coverage was not extensive in frontal cortex; Sammler et al. 2009).

Resolving a Paradox

From the evidence just reviewed, it is clear that any attempt to understand the relationship between tonality processing and linguistic syntactic processing must address a paradox: Evidence from behavioral studies of patients with musical deficits (either due to brain damage or lifelong congenital amusia) points to the independence of tonality and linguistic syntax processing (Peretz 2006), while evidence from neuroimaging points to an overlap.

To resolve this paradox and to guide future empirical research on the relationship between tonality and syntax, SSIRH (Patel 2003) posits a distinction between domain-specific representations in long-term memory (e.g., stored knowledge of words and their syntactic features, and of chords and their harmonic features) and shared neural resources which act upon these representations as part of structural processing. This “dual-system” model considers syntactic processing to involve the interaction (via long-distance neural connections) of “resource networks” and “representation networks” (Figure 14.4).

Figure 14.4  Schematic diagram of the functional relationship between shared resource networks and domain-specific representation networks, according to the SSIRH. L = Language, M = Music. The diagram represents the hypothesis that linguistic and musical long-term knowledge are stored in anatomically distinct “representation networks,” which can be selectively damaged, whereas there is overlap in the “resource networks” that help activate structural information in representation networks during sequence processing. Arrows indicate functional connections between networks. Note that the circles do not necessarily imply highly focal brain areas. For example, linguistic and musical representation networks could extend across a number of brain regions, or exist as functionally segregated networks within the same brain regions. Reprinted from Patel (2008) with permission from Oxford University Press.

Such a view contrasts with an alternative view, common in artificial neural network models, that syntactic representation and processing involve a single network (e.g., Elman et al. 1996).

Why posit a dual-system model? The primary reason is that such a model can explain dissociations between tonality and linguistic syntax processing (via damage to domain-specific representation networks) as well as overlap (via activations in shared resource networks). Furthermore, the idea that complex cognitive processing involves temporally coordinated activity between spatially segregated brain regions is part of current neurobiological theory. For example, a core feature of the theory of neural Darwinism (Edelman 1993) is the concept of reentry, “a process of temporally ongoing parallel signaling between separate maps along ordered anatomical connections” (Edelman 1989:65). Neurobiologically realistic computational models of reentrant interactions between functionally segregated brain regions have been explored by Edelman and colleagues (e.g., Seth et al. 2004). In such models, different cortical regions have distinct functions, but higher-level processing abilities (such as object recognition) are an emergent property of a network of reentrantly connected regions. Within neurolinguistics, reentrant models involving interactions between distant cortical regions have also been proposed; for a recent example, see Baggio and Hagoort’s (2011) discussion of the brain mechanisms behind the N400.

Based on evidence from neuropsychological dissociations and neuroimaging, I suggested that the domain-specific representation networks involved in language and music processing were located in temporal regions of the brain, while the shared resource networks were located in frontal regions (Patel 2003). Further, I posited that resource networks are recruited when structural integration of incoming elements in a sequence is costly; that is, when it involves the rapid and selective activation of low-activation items in representation networks. I used specific cognitive theories of syntactic processing in language (dependency locality theory; Gibson 2000) and of tonal harmonic processing in music (tonal pitch space theory; Lerdahl 2001b) to specify the notion of processing cost. In both models, incoming elements incur large processing (activation) costs when they need to be mentally connected to existing elements from which they are “distant” in a cognitive sense (e.g., in music, distant in tonal pitch space rather than in terms of physical distance in Hz). According to the SSIRH, in such circumstances, activity in frontal brain regions increases in order to rapidly activate specific low-activation representations in temporal regions via reentrant connections. Put another way, music and language share limited neural resources in frontal brain regions for the activation of stored structural information in temporal brain regions.

The SSIRH has a resemblance to another independently proposed dual-system framework for linguistic syntactic processing: the memory, unification, and control (MUC) model (Hagoort 2005; Hagoort and Poeppel, this volume). As noted by Baggio and Hagoort (2011:1357), “the MUC model assigns the
storage of lexical items to temporal cortex and the unification of retrieved structures to frontal cortex.” Hence, no permanent memory patterns are stored in frontal cortex. Instead, “unification is essentially the result of the coactivation of different tokens in inferior frontal gyrus, dynamically linked to their lexical types in middle temporal gyrus/superior temporal gyrus via persistent neuronal firing and feedback connections” (Baggio and Hagoort 2011). Figure 14.5a shows the components of the MUC model projected onto the left cerebral hemisphere, and Figure 14.5b shows some of the anatomical connections between frontal and temporal brain regions which could support interactions between these regions.

The MUC model views different parts of inferior frontal gyrus as involved in different aspects of linguistic unification (i.e., phonological, semantic, and syntactic), with semantic unification relying more heavily on BA 45/47, and syntactic on BA 44/45 (Hagoort 2005; for recent empirical data, see Snijders et al. 2009). It is thus of interest that tonal harmonic processing appears to activate the BA 44/45 region, in line with the idea that tonality processing has more in common with linguistic syntactic processing than with semantic processing.

![Figure 14.5](image-url)

**Figure 14.5** (a) Schematic of brain regions involved in Hagoort’s *memory unification and control* model of language processing, projected onto the left hemisphere: Memory (yellow) in left temporal cortex, unification (blue) in left inferior frontal gyrus, and control (gray) in dorsolateral prefrontal cortex (control is involved in verbal action planning and attentional control, but is not discussed in this article). For the memory region, associated Brodmann areas include BA 21, 22, and 42. For the unification region, associated Brodmann areas include BA 44, 45, and 47. Reprinted from Hagoort (2005) with permission from Elsevier. (b) Simplified illustration of the anatomy and connectivity of the left-hemisphere language network. Cortical areas are represented as red circles: pars orbitalis (or), pars triangularis (tr), and pars opercularis (op) of the left inferior frontal gyrus; angular gyrus (ag), superior and middle temporal gyri (tg), fusiform gyrus (fg) and temporal pole (tp). White matter fibers are shown in gray, arrows emphasize bidirectional connectivity: arcuate fasciculus (AF), extreme capsule (EC), inferior longitudinal fasciculus (ILF) and uncinate fasciculus (UC). Interfaces with sensorimotor systems are shown in green: visual cortex (vc), auditory cortex (ac) and motor cortex (mc). Reprinted from Baggio and Hagoort (2011) with permission from Taylor and Francis Ltd.

One important issue for future work relating the SSIRH and MUC frameworks concerns hemispheric asymmetries. The SSIRH diagram in Figure 14.4 and the MUC diagram in Figure 14.5 represent areas and connections within one hemisphere. However, the processing of tonality likely involves frontotemporal networks in both hemispheres, though likely with a right-hemisphere bias, reflecting the fact that tonality involves structured pitch patterns. How strong is this bias, and how does it compare to the strength of left-hemisphere bias in frontotemporal interactions that support linguistic syntax? Within each domain, how is the processing of structural relations coordinated in the left and right hemispheres? Are the “resource” regions shared by language and music in the left hemisphere only (e.g., due to a strong leftward bias in linguistic syntactic processing vs. a more equal hemispheric balance for musical tonality processing)? Or do resource regions overlap in both hemispheres? Does the degree of hemispheric bias for tonality processing depend on the amount of musical training? These interesting questions await future research.

**Why Link Tonality to Syntax?**

The SSIRH provides a conceptual framework for understanding how tonality processing and linguistic syntactic processing might be related in the brain, yet a deeper theoretical question is: Why should tonality processing be related to linguistic syntactic processing? The SSIRH is framed in terms of integration cost, but in parallel models of language architecture (e.g., Jackendoff 2002; Hagoort 2005), linguistic integration (or unification) occurs at multiple levels in parallel (e.g., phonological, semantic, and syntactic). Why is tonality more cognitively akin to linguistic syntax than to phonology or semantics?

Considering phonology, it is notable that speech, like tonal music, has rule-like processes for combining sounds (Chomsky and Halle 1968). Phonological rules, however, typically involve changes to the forms of sounds as a result of the context in which they occur. For example, French has a phonological process known as voicing assimilation, in which a phoneme changes its feature “+/– voiced” depending on the local context in which it occurs. Ramus and Szenkovits (2008) provide the following illustration. In French, the voicing feature may spread backward from obstruents or fricatives to the preceding consonant: `cape grise` [kapgriz] → [kabgriz] (gray cloak). This assimilation process is both context specific (it does not occur before nasals: `cape noire` is always [kapnwar]; black cloak) and language specific (it does not occur in English, which instead shows assimilation of place of articulation: `brown bag` [brownbag] → [browmbag]). In tonal music, in contrast, principles of combination can result in an individual note changing its structural category without necessarily changing its physical characteristics in any way (e.g., the note B4 as played on a piano can have the identical physical frequency spectrum when played as the tonic of the key of B or the leading tone in the key of C, yet be perceived as sounding very different in these two contexts in terms of
tension or resolution). Empirical research indicates that listeners are sensitive to changes in the structural categories of musical sounds as a function of context, even when the sounds themselves are physically identical in the different contexts (e.g., Bigand 1993).

Turning to semantics, an obvious difference between tones and words is in their referential qualities. Words are linked to concepts in a very intricate way: words have many phonological and semantic features which give them a specific position in a vast network of meaningful concepts in long-term memory. This ensures that when we hear a word, we activate the relevant, specific concept; when we wish to communicate about some concept, we use the appropriate word form. This highly complex form of cognition involves multiple brain areas in the perisylvian region (with a left-hemisphere bias) and is likely supported by a network that has been specialized over evolutionary time for this semantic processing. Consistent with this idea, recent neuroimaging research using MEG combined with structural MRI has shown that a similar frontotemporal brain network, with a left-hemisphere bias, is used for word understanding by adults and 12- to 18-month-old infants, meaning that this network is in place early in life (Travis et al. 2011).

The pitches of instrumental tonal music do not have rich semantic properties. To be sure, tonal sequences can sometimes convey general semantic concepts to an enculturated audience: a passage of instrumental music can sound “heroic” and may activate associated semantic concepts in the brain of musically untrained listeners (Koelsch 2011b; Koelsch et al. 2004). Even in such cases, however, one cannot pin down the semantic meaning of music in a precise fashion (Slevc and Patel 2011). Swain (1997:140) has argued that “the difference between musical and linguistic reference lies not in quality but in range”; one requires, for example, little effort to decide if Beethoven’s Appassionata connotes “explosive fury” or “peaceful contemplation.” However, deciding if it has to do with “explosive fury” or “passionate determination” is more difficult. In other words, if the distinction is binary, and concepts underlying a musical motive are diametrically opposed, musical meaning is grasped easily. If, however, there is a finer nuance, agreement among listeners is lost. Therefore, compared to linguistic meaning, the range of musical semantics is rather limited, (Swain 1997:49), as musical and linguistic structures have “varying degrees of semantic potential” (Antović 2009).

What of syntax? Here I would like to argue for significant connections between tonality and language. Notably, in both linguistic and tonal harmonic sequences, the brain interprets incoming events in terms of a small number of abstract structural categories and relations. For example, a word in a sentence is not just a semantic reference to some entity in the world, it also belongs to a structural category (e.g., noun, verb) and can enter into certain structural relations with respect to other words (e.g., subject, object). Importantly, while many languages mark structural relations by phonetically “tagging” the word

in question (e.g., via distinct case markers for subjects and objects), this is not a necessary feature of language, as illustrated by English, where the same lexical form (e.g., “chapter”) can be the subject or object of a sentence, depending on context.

In tonal music, pitches also become part of abstract structural categories and relations. In Western tonal harmonic music, abstract structural categories are formed at the level of chords (i.e., simultaneous, or near simultaneous, soundings of pitches). For example, particular collections of pitches define a “major chord” versus a “minor chord,” but the precise physical instantiation of a chord (e.g., its component frequencies) depends on the particular key of the passage, which pitch serves as the root of the chord, the current musical tessitura (pitch range), and the “inversion” of the chord. Thus (implicitly) recognizing a certain chord type (e.g., a major chord) requires a certain kind of abstraction over different events, which can vary widely in their physical structure.

Turning from abstract structural categories to abstract structural relations, the same musical sound (e.g., a C major chord, C-E-G) can be an in-key or out-of-key chord and when it is in-key can serve different “harmonic functions.” For example, it can be a “tonic chord” (i.e., the structurally most central and stable chord in a key) or when it occurs in a different key, it can be a “dominant chord” (i.e., a less stable chord built on the fifth note of scale). Tonic and dominant chords play a central role in Western tonal music, since movement from tonic to dominant back to tonic is an organizing structural progression for chord sequences. Musically untrained listeners are sensitive to the structural categories of chords (Bigand et al. 2003). This sensitivity is thought to be acquired via implicit learning (Tillmann et al. 2000) and plays an important role in feelings of tension and resolution created by chords (Lerdahl and Krumhansl 2007) and in the emotional responses to chords in musical context (Steinbeis et al. 2006). In other words, the harmonic functions of chords are abstract structural relations that influence the “meaning” of tone sequences in music. It is important to note that abstract structural relations in tonal music apply to individual pitches, not just chords. For example, a well-tuned note (e.g., A4, a pitch of 440 Hz) can be in-key or out-of-key, and when in-key can vary in the structural role it plays (e.g., a stable “tonic note” or an unstable “leading tone”), depending on the context in which it occurs (Krumhansl 1990). Just as with chords, sensitivity to abstract relations among individual pitches emerges via implicit learning in enculturated listeners and requires no formal musical training (Krumhansl and Cuddy 2010; Trainor and Corrigal 2010). Thus such sensitivity is likely to be part of many of the world’s musical systems, since many such systems (e.g., the classical music of North India) have rich melodic but not harmonic structures (cf. Brown and Jordanian 2011).

Thus the reason that tonality may have something in common with linguistic syntax is that both involve the interpretation of rapidly unfolding sequences in terms of abstract structural categories and relations. Such categories and
relations are *structural* because they are not bound to specific semantic meanings, nor are they necessarily signaled by the physical structure of an event, yet their existence and organization strongly influences the overall meaning of the sequence (e.g., who did what to whom in language, or patterns of tension and resolution in music).

With this perspective, the SSIRH can be framed in a more specific way. SSIRH addresses the cost of integrating an event’s *structural status* with the existing structure of the sequence. In language processing, some structural integrations are more difficult to integrate than others. For example, a word may be cognitively distant from another word to which it needs to be conceptually related (e.g., long-distance noun–verb dependencies in dependency locality), or a word’s category may be unexpected and result in a change in the existing structural interpretation of a sentence, as in garden path sentences (e.g., Levy 2008). Similarly, in tonal music, some structural integrations are more difficult to integrate than others. For example, a chord may be cognitively distant from the existing tonal region (e.g., an out-of-key chord) or because it is an unexpected category at that particular point in a sequence (e.g., a minor chord at a point in a sequence where a major chord is highly expected). According to the SSIRH, structural integration difficulty in language and music results in an increased activation cost, and this activation cost is “paid” by increased activity in shared frontal brain regions which are reciprocally connected to domain-specific temporal regions in which linguistic or musical representations reside.

In making this connection between linguistic syntactic processing and musical tonal harmonic processing, it is important to keep in mind that structural integration difficulty in the two domains can have quite distinct consequences. In the case of language, integration can become so difficult that it actually becomes impossible (in certain ungrammatical or highly complex sentences), and this effectively defeats a listener’s attempt to assign a meaningful structure to the sentence. In tonal harmonic music, listeners generally try (implicitly) to make sense of any harmonic sequence they encounter, even when such sequences are highly complex or “ungrammatical.” Perhaps this is because structurally unexpected events in music play an important role in eliciting emotional responses from listeners (Huron 2006; Steinbeis et al. 2006).

**Testing the Predictions of the SSIRH**

The SSIRH makes specific, testable predictions about the relationship between tonality and linguistic syntactic processing. One prediction is that since neural resources for structural integration are limited, simultaneous costly integrations in tonality and language should lead to interference. Testing this prediction requires experiments which present music and language simultaneously, and which align points to difficult structural integration in the two domains. This prediction has now been tested in five studies across three different
A. D. Patel

laboratories: two studies using ERPs (Koelsch et al. 2005; Steinbeis and Koelsch 2008b) and three using behavioral methods (Fedorenko et al. 2009; Hoch et al. 2011; Slevc et al. 2009). All have supported the predictions of the SSIRH. For the sake of brevity, only one experiment is described here.

Fedorenko et al. (2009) manipulated linguistic syntactic integration difficulty via the distance between dependent words. These researchers used fully grammatical sentences of the type shown below, which differ in the structure of an embedded relative clause (italicized):

(a) The cop *that met the spy* wrote a book about the case.
(b) The cop *that the spy met* wrote a book about the case.

The sentences were sung to melodies (one note per word) which did or did not contain an out-of-key note on the last word of the relative clause: “spy” in (a), “met” in (b). According to dependency locality theory (Gibson 2000), this word is associated with a distant structural integration in (b) (between “met” and “that”) but not in (a). A control condition was included for an attention-getting but nonharmonically deviant musical event: a 10 dB increase in volume on the last word of the relative clause. After each sentence, participants were asked a comprehension question, and accuracy was assumed to reflect processing difficulty.

The results revealed an interaction between musical and linguistic processing: comprehension accuracy was lower for sentences with distant versus local syntactic integrations (as expected), but crucially, this difference was larger when melodies contained an out-of-key note. The control condition (loud note) did not produce this effect: the difference between the two sentence types was of the same size as that in the conditions which did not contain an out-of-key note. These results suggest that structural integration in language and music relies on shared processing resources.

Another line of work motivated by the SSIRH concerns tonality processing in agrammatic Broca’s aphasia (Patel 2003). The sensitivity of such aphasics (all with unilateral left-hemisphere lesions, though in variable areas) to tonal harmonic relations was tested using both explicit and implicit tasks (Patel et al. 2008b). The aphasics showed reduced sensitivity to such relations and, in the explicit task, the pooled performance of aphasics and controls on the tonality task predicted their performance on the linguistic syntax task (but not on a linguistic semantic task).

These initial studies call for further work to test reliability of these findings as well as to probe the specificity of the link between tonality and syntax (e.g., as opposed to tonality and semantics or tonality and phonology). There is a particular need for within-subjects fMRI research to compare brain areas involved in tonality versus linguistic syntactic processing. However, even overlapping activation regions in fMRI cannot absolutely prove the existence of shared neural circuits for tonality and linguistic syntax, due to issues of spatial resolution.

It is possible that spatially distinct networks exist at the microscopic level but occur in the same macroscopic brain region (e.g., by exhibiting different mosaics of functional organization interdigitated in the same cortical region). Thus fMRI results should be combined with evidence from other methods, including ERPs, behavioral research, patient studies, and techniques which produce transient “virtual lesions” (i.e., transcranial magnetic stimulation). That is, multiple converging methods, driven by specific hypotheses, are needed to discover the cognitive and neural operations shared by tonality and linguistic syntactic processing. These operations are worth uncovering because they are likely to be fundamental to our complex communicative abilities.

**Related Computations in Homologous Areas of Opposite Hemispheres: Word Articulation**

As an example of a related functional computation in language and music that may rely on homologous networks in opposite hemispheres, let us look at word articulation in speech and song. By “word articulation” I mean the brain mechanisms that convert a lexical item retrieved from memory to a sequence of sounds. Word articulation is one of the most complex motor actions produced by the brain. In both speech and song, it requires rich motor planning, the coordination of multiple articulators and the respiratory system, as well as self-monitoring and online correction based on sensorimotor information (Levelt et al. 1999). Despite this similarity between speech and song, there are salient differences between spoken and sung word articulation. For example, words in song are usually produced more slowly than spoken words (in syllables per second), and require more precise pitch control. These differences (i.e., slower rate, emphasis on pitch precision) are likely to be two reasons why song relies heavily on a right-hemisphere auditory motor network that is involved in the precise control of pitch patterns: a frontotemporal network that connects the superior temporal gyrus to the inferior frontal lobe via the arcuate fasciculus (Wan and Schlaug 2010).

What is the relationship between the brain networks involved in word articulation in speech and song? Evidence from neuropsychology is mixed. It has long been claimed that some nonfluent aphasics, who have severe difficulty with spoken word production, can produce words fluently when they sing familiar songs (Yamadori et al. 1977). However, in a study of nonfluent aphasics, Racette, Bard, and Peretz (2006) reported that word articulation abilities were no better during singing as opposed to speaking, suggesting that the same (impaired) word articulation network was involved in both domains. On the other hand, there is a long-standing observation that individuals with developmental stuttering, which is associated with left inferior frontal cortex structural anomalies (Kell et al. 2009), can often sing words with great fluency, and recent evidence indicates that transmagnetic stimulation over left Broca’s...
A. D. Patel

region disrupts spoken but not sung word production (Stewart et al. 2001). Furthermore, neuroimaging research shows that when individuals sing songs, they activate certain brain regions that are not activated when they speak the same words. For example, Saito et al. (2006) compared overt singing to reciting song lyrics and found that the right inferior frontal gyrus, the right premotor cortex and the right anterior insula were active only during singing (cf. Peretz 2012). Hence, song and speech production show different brain activation patterns, with song production engaging certain right-hemisphere regions not activated by ordinary speech.

In sum, the neuroscientific evidence on word articulation in speech versus song is paradoxical: some evidence point to overlap (e.g., Racette et al. 2006), some to differentiation (e.g., Saito et al. 2006). A full resolution of this paradox remains to be proposed. Here I would like to suggest that some of the right-hemisphere brain circuits involved in word articulation in song are anatomically distinct from homologous left-hemisphere circuits involved in spoken word articulation, but that they carry out similar functional computations in terms of motor control. The idea of similar functional computations in homologous motor regions of opposite hemispheres is well accepted in neuroscience: after all, the left and right hands are largely controlled by homologous motor regions in opposite hemispheres. In the case of hand control, this opposite lateralization is largely driven by the decussating pattern of neuroanatomical connections between the hands and the brain. In contrast, the opposite lateralization of certain parts of the word articulation network for song versus speech may be driven by the differences in rate and pitch precision with which words are articulated in the two domains (Zatorre et al. 2002; Poeppel 2003).

To make my suggestion more concrete, I would like to place it in the context of a specific model of speech production; namely, the gradient-order DIVA (GODIVA) model (Bohland et al. 2010). GODIVA is an update of Guenther’s original directions into velocities of articulators (DIVA) model—a neural network model of speech motor control and acquisition that offers unified explanations for a number of speech phenomena including motor equivalence, contextual variability, speaking rate effects, and coarticulation. Among computational models of speech production, GODIVA is notable for making detailed reference to known neuroanatomy and neurophysiology. A schematic of the model is provided in Figure 14.6.

A salient feature of the model, motivated by neurobiological research on speech production, is the left-hemisphere lateralization of several components. For example, the model posits that a speech sound map (SSM) exists in the left ventral premotor cortex and/or posterior inferior frontal gyrus pars opercularis (“frontal operculum” in Figure 14.6). SSM is the interface between the phonological encoding system and the phonetic/articulatory system and contains cell groups that code for well-learned speech sounds. SSM representations are functionally similar to a mental syllabary (Levelt and Wheeldon 1994; Crompton 1982), suggested by Levelt et al. (1999:5) to consist of a “repository

Figure 14.6  Schematic diagram of the GODIVA model of speech production, including hypothesized cortical and subcortical substrates. Boxes with dotted borders are given explicit computational treatment, whereas other boxes are treated conceptually. Lines with arrows represent excitatory pathways, and lines with filled circles represent inhibitory pathways. Lines with both arrowheads and filled circles indicate that connectivity between these modules features top-down excitatory connections as well as bottom-up inhibitory connections. For further details, see Bohland et al. (2010). Reprinted from Bohland et al. (2010) with permission from MIT Press.
of gestural scores for the frequently used syllables of the language.” Using alternative terminology, SSM representations can be thought of as sensorimotor chunks or programs, learned higher-order representations of frequently specified spatiotemporal motor patterns. Recent fMRI research motivated by this model has supported the idea of syllable-level motor programs for speech which rely heavily on left ventral premotor cortex (Peeva et al. 2010).

My suggestion is that the right hemisphere has a “song sound map” (SGSM) that involves the right premotor cortex, complementary to the left-hemisphere SSM. The left-hemisphere SSM and the right-hemisphere SGSM carry out similar functional computations for word articulation in speech and song, though the SGSM part of the network normally operates at slower rates and with strong functional coupling to right-hemisphere regions involved in precise pitch control.

Such a view is relevant to recent research which has used a singing-based therapy called melodic intonation therapy or MIT (Albert et al. 1973) to help nonfluent aphasics recover some of their spoken language abilities. MIT embeds short phrases (e.g., “I love you”) in “melodic” speech intonation patterns that rely on up-and-down movements between two discrete pitches. Patients practice such utterances intensively and regularly with a therapist, who gradually lengthens the phrases to span more syllables (Norton et al. 2009). The goal of the therapy is to improve fluency for both the trained phrases and for spontaneous, untrained utterances spoken in a normal fashion. Two features of MIT that distinguish it from nonmusical speech therapy are the use of melodic speech intonation and rhythmic tapping (i.e., while speaking the utterance, the patient also taps its syllabic rhythm using the hand unaffected by the stroke, typically the left hand).

Schlaug and colleagues have recently begun a set of studies aimed at measuring the efficacy of MIT versus a matched speech repetition therapy (SRT) without melodic intonation and tapping. In addition to quantifying the effects of MIT versus SRT on posttherapy measures of verbal fluency, the researchers are also measuring changes in brain physiology associated with the two therapies, by conducting fMRI and structural neuroimaging studies before and after the therapy. Of particular interest in this regard is the extent to which MIT patients shift toward using right-hemisphere “song” circuits for speech after therapy. That is, can such patients retrain right-hemisphere song networks to take over the functions of left-hemisphere speech networks? From a theoretical standpoint, this might be possible if the song network is already doing functional computations similar to the damaged speech network.

Preliminary data reported by Schlaug, Marchina, and Norton (2008) support the idea of a right-hemisphere shift in speech control with MIT; a patient who had undergone forty sessions of MIT showed substantially increased verbal fluency as well as increased speech-related activation in a right-hemisphere network involving the premotor, inferior frontal, and temporal lobes. Schlaug
et al. (2009) have also found structural enhancement of the right arcuate fasciculus in patients who underwent MIT. The arcuate fasciculus is a large fiber tract that connects the frontal and superior temporal lobes and which is thought to be important for auditory sensorimotor integration.

At the moment, the relative contributions of the vocal/melodic versus rhythmic/hand movement components of MIT to these neural changes is unknown. An important issue for future research is to study how vocal melody and manual tapping might act synergistically to recruit neural plasticity in right-hemisphere word articulation circuits; that is, how vocal and manual motor circuits might be interacting in the brain (cf. Arbib 2006a). For current purposes, however, the findings of Schlaug and colleagues support the idea that right-hemisphere regions normally involved in word articulation in song can take over for damaged left-hemisphere regions normally involved in word articulation in speech. I would argue that this sort of neural plasticity is possible because the two sets of networks must have initially been carrying out similar functional computations. If this is indeed the case, it illustrates how music and language cortical processing can be related via similar functional computations in anatomically distinct circuits.

**Conclusion**

In this chapter, I have offered three conceptual solutions to address the paradoxical evidence on language–music relations in the brain. All three solutions illustrate how language and music may rely on related functional computations, despite neuropsychological dissociations between linguistic and musical abilities. Related functional computations used by language and music are likely to be fundamental to human cognition. The presented solutions are, however, not exhaustive. Additional solutions may emerge when other relations (e.g., in rhythmic processing) are considered. My intent in offering these solutions is to open up a dialog about the paradox of language–music relations in the brain. A full resolution of this paradox will yield significant new insights into the mechanisms underlying our species’ uniquely powerful communicative abilities.

**Acknowledgments**

The author’s work was supported by the Neurosciences Research Foundation as part of its research program on music and the brain at The Neurosciences Institute, where he was the Esther J. Burnham Senior Fellow. Helpful comments were provided by David Perlmutter, Bob Slevc, and two anonymous reviewers.

Chapter 14 References


Chapter 14 References


Chapter 14 References


Chapter 14 References
