Parallels and Nonparallels between Language and Music

Ray Jackendoff
Tufts University

THE PARALLELS BETWEEN LANGUAGE AND MUSIC CAN BE explored only in the context of (a) the differences between them, and (b) those parallels that are also shared with other cognitive capacities. The two differ in many aspects of structure and function, and, with the exception of the metrical grid, all aspects they share appear to be instances of more general capacities.

Received May 6, 2008, accepted November 4, 2008.

Key words: rhythm, pitch, syntax, prolongational structure, recursion

A FUNDAMENTAL QUESTION THAT HAS ANIMATED A great deal of thought and research is:

- What does music share with language that makes them distinct from other human activities?

Focusing on this question leads to an emphasis on the similarities between language and music (e.g., Patel, 2008), sometimes to the point of people coming to believe that they are (almost) the same thing. For instance, the prospectus for the 2008 Dijon conference "Musique Langage Cerveau" states, "Les similitudes entre ces deux activités ne sont donc pas superficielles: la musique et le langage pourraient être deux expressions d'une même compétence pour la communication humaine" ("The similarities between these two activities are therefore not superficial: music and language could be two expressions of the same competence for human communication"). However, the divergences between music and language are also quite striking. So we need also to ask:

- How are language and music different?
- In the respects that language and music are the same, are they genuinely distinct from other human activities?

The emphasis of the present article will be on these latter two questions.

These questions have been sharpened by the "Chomskyan turn" in linguistics, in which the focus is no longer on "the language" as a disembodied social object. Rather, it is on how language is instantiated in speakers' minds, such that they can produce and understand utterances in unlimited profusion—and on how speakers acquire this ability. GTTM (Lerdahl & Jackendoff, 1983) advocates a similar approach to music: the central issue is what constitutes musical understanding, such that individuals can understand an unlimited number of pieces of music in a style with which they are experienced—and how through experience individuals acquire fluency in a musical style.

Through this lens, music and language can be compared in the following terms:

- Every normal individual has knowledge of language and music.
- Everyone learns the local variant(s) of both language and music. Normal adults are at ceiling for language, but they are more variable in musical ability, depending on exposure and talent.

Then the important question becomes:

- What cognitive capacities are involved in acquiring and using a language, and what capacities are involved in acquiring and using knowledge of a musical idiom?

The problem is the same for both capacities, but this does not mean that the *answer* is the same.

The issue of particular interest here is:

- What cognitive capacities are shared by language and music, but not by other cognitive domains?

Similar issues arise with other human capacities, for example, the capacity for social and cultural interaction (Jackendoff, 2007). Like languages and musical idioms, cultures differ widely, and an individual's ability to function in a culture requires a great deal of learning and the use of multiple cognitive capacities. Moreover, the use of language and music is certainly embedded in

Music Perception VOLUME 26, ISSUE 3, PP. 195–204, ISSN 0730-7829, ELECTRONIC ISSN 1533-8312 DOI:10.1525/MP.2009.26.3.195

social and cultural interaction, but that does not entail that the capacity for either language or music is simply a subset of the social/cultural capacity.

A major dispute in the theory of language, of course, is how much of the language acquisition capacity is special-purpose. Many people (e.g., Christiansen & Chater, 2008; Tomasello, 2003) think that language is acquired through general-purpose learning plus abilities for social interaction. This view is explicitly in opposition to the claims of generative grammar up to the early 1990s to the effect that there must be a rich innate language-specific Universal Grammar (Chomsky, 1965, 1981). In between these two extremes are all manner of intermediate views (Hauser, Chomsky, & Fitch, 2002; Jackendoff, 2002; Pinker & Jackendoff, 2005).

There seems to be less vehement dispute about the parallel issue in music cognition and acquisition, perhaps only because claims for an innate music capacity have been less highly politicized—but perhaps also because claims that music is an adaptation favored by natural selection are considerably weaker than those for language. Thus at one extreme we find Pinker's (1997) hypothesis that music is "auditory cheesecake," constructed adventitiously from parts of other capacities; at the other might be the fairly rich claims of GTTM. In between is for example Patel's (2008) view that music is a social construction, but that the capacity for pitch discrimination and formation of tonally oriented scales is nevertheless *sui generis* to music.

General Capacities Shared by Language and Music

Some similarities between language and music are easily enumerated.

- Although many animals have communication systems, no nonhumans have either language or music in the human sense, and in particular there are no obvious evolutionary precursors for either in nonhuman primates.
- Language and music both involve sound production (although care is necessary here, for language also exists in the signed modality and music does not).
- Every culture has a local variant of language, and every culture has a local variant of music. The differences among local variants are moreover quite striking; this contrasts with other species, whose communication systems show very limited variation at best.
- In every culture (I believe), language and music can be combined in song.

Looking at the two domains more cognitively, the acquisition and processing of language and the acquisition and processing of music call for a number of capacities that are also shared with other cognitive domains. I'll mention seven.

First, both language and music require substantial memory capacity for storing representations—words in language (tens of thousands) and recognizable melodies in music (number unknown, but my informal estimate easily runs into the thousands[1]). But this is not specific to music and language. Similar massive storage is necessary for encoding the appearance of familiar objects, the detailed geography of one's environment, the actions appropriate to thousands of kinds of artifacts (Jackendoff, 2007), and one's interactions with thousands of people—not just what they look like but also their personalities and their roles in one's social milieu (Jackendoff, 2007).

Second, in order to account for the perception and comprehension of novel stimuli, both language and music require the ability to integrate stored representations combinatorially in working memory by means of a system of rules or structural schemata. Again, this characteristic is not specific to language and music. Understanding a complex visual environment requires a capacity to integrate multiple objects into a structured scene; and creating a plan for complex action requires hierarchical integration of more elementary action schemata, in many cases bringing in complex social information as well. We return to this point later.[2]

Third (a parallel stressed by Patel, 2008), the processing of both language and music involves creating expectations of what is to come. But visual perception involves expectation too: if we see a car heading for a tree, we expect a crash.

Fourth, producing both language and music requires an ability to achieve fine-scale voluntary control of vocal production. No other faculties place similar demands on vocal production per se. However, voluntary control of vocal production is plausibly a cognitive extension of the enhanced voluntary control of the

[1]Consider nursery tunes, camp songs, folk songs, popular tunes, liturgical music, plus—for some of us—classical pieces of 30 minutes or more in length. Also consider: How many tunes do people put on their iPods?

[2]This point is in strong opposition to the hypothesis of Hauser, Chomsky, and Fitch (2002) that recursion is the cognitive characteristic that makes language unique. All cognitive capacities of any complexity have recursion. Language is unique primarily in that combinatoriality in the communicative signal maps into combinatoriality in the message being communicated. For discussion, see Jackendoff and Pinker (2005).

hands in our species, crucial to tool-making and tool use (Calvin, 1990; Wilkins, 2005). Hand control, of course, is also necessary for signed language, drumming, and playing musical instruments.

Fifth, learning to produce both language and music relies on an ability—and desire—to imitate others' vocal production. In the case of music, one may imitate other sound-producing actions as well (e.g., drumming, birdsong). It is this ability to incorporate others' inventions that enables both language and music to build a culturally shared repertoire of words and songs. But this is not specific to language and music: the richness of human culture is a consequence of the ability to imitate and integrate others' actions (not just others' words) into one's own repertoire.

Sixth, at least some individuals must have the ability to invent new items—words or tunes—that others can imitate. This too extends to cultural practices, be they tools, food, types of clothing, or praxis (customs, trade, games, rituals, etc.).

Seventh, individuals must have the ability to engage in jointly intended actions—actions that are understood not just as me doing *this* and you doing *that*, but the two of us doing something *together*, each with a particular role in the composite (Bratman, 1999; Gilbert, 1989; Searle, 1995). This ability lies behind the human ability for widespread cooperation (Boyd & Richerson, 2005; Tomasello, Carpenter, Call, Behne, & Moll, 2005), and it is necessary in language use in order to hold a conversation (Clark, 1996) and in music for any sort of group singing, playing, dancing, and/or performing for an audience.

The only capacity on this list that is not shared with other domains is fine-scale voluntary vocal production, which of course is not necessary either for signed languages or instrumental music. The first three—large-scale memory, combinatoriality, and expectation—are arguably present to a considerable extent in other primates (though not in their communication systems). The last three—imitation, innovation, and joint action—are not shared with other primates, but are generally necessary for all sorts of cultural cognition and culturally guided action. The point is that these general abilities alone do not specifically determine the form of either language or music.

Differences in Ecological Function

One fundamental difference between language and music concerns their ecological functions in human life. In brief, language conveys propositional thought, and music enhances affect (I prefer the broader term *affect* to the more usual *emotion*; see Jackendoff & Lerdahl, 2006). Although this point is hardly new, it is worth expanding on in order to point up the extent of the difference.

Language is essentially a mapping between sound and "propositional" or "conceptual" thought. The messages it conveys can be about people, objects, places, actions, or any manner of abstraction. Language can convey information about past, future, visible things, invisible things, and what is *not* the case. Linguistic utterances can be used to offer information, make requests for action, ask questions, give instructions and orders, negotiate, undertake obligations (including promises), assert authority, and construct arguments about the differences between language and music. Linguistic messages distinguish information taken to be new to the hearer ("focus") from information taken to be shared with the hearer ("common ground"), and they can incorporate social distinctions between speaker and hearer (as in French *tu* versus *vous* or Japanese honorifics). The gist of a linguistic utterance can be translated from any language into any other, given appropriate vocabulary.

None of these functions can be satisfied by music. In particular, although one can translate from, say, Japanese into Swahili or Quechua, it makes no sense to think of translating from Japanese into raga, rock and roll, or Japanese *gagaku*. (I take it that various drumming languages and whistling languages, although they use media normally deployed for musical purposes, are codes for language and are not forms of music.)

Consider now what music is used for. Probably furthest from the evolutionary roots of music are the uses in which people sit and passively listen to a performance (just as paintings in a museum are probably the uses of art furthest from the roots of human pictorial abilities). Many different uses of music in traditional cultures have been proposed as the original adaptive function of music (e.g., the articles in Wallin, Merker, & Brown, 2000), but actual evidence is scanty.

It is worth compiling a list of these different uses of music to see their variety. What they seem to me to all have in common is the enhancement of affect associated with an activity. If one wishes to call this "musical meaning" (e.g., Raffman, 1993), I have no objection, as long as it is clear that it bears little or no relation to linguistic meaning.

One sort of music involves one person directing music at another. Lullabies convey a sense of soothing intimacy; love songs convey affection and passion; ballads convey the emotional impact of a story.

Another sort of music is meant to be sung or played together. Work songs convey the rhythm of work, often coordinated rhythm, and they convey as well the affect of coordinated action. Marches convey the coordinated action of walking, often militaristically or ceremonially. Religious music conveys transcendence and spirituality, with associated affect anywhere from meditative to frenzied. Dance tunes stimulate affective or expressive body movement. There is a genre of songs that are sung in collective situations such as around a campfire or in a bar; as far as I can tell, the function of such songs is to instill a sense of fellowship. Another genre is children's songs, including nursery tunes; it is not clear to me what their function is.

Yet another sort of music, background music, is addressed to no one in particular, and is perhaps perceived subliminally. Its function is evidently to enhance mood. Muzak is perhaps the most offensive example, but café music serves a similar function. Film music, whose effects can be quite powerful, also fits into this genre. There is no comparable subliminal use of language.

The distinction between propositional and affective content is actually not quite this cut and dried. Music cannot be put to propositional use, but language obviously *can* be put to affective use. One route goes through the propositional system. For instance, utterances such as "You are an idiot" and "I love you" obviously convey affective content—though not of a sort music is good at. But there is also a wide range of rhetorical devices in language that borrow from music. Poetry (especially but not exclusively poetry of a "folk" character) makes use of isochrony or strict rhythm, which brings linguistic utterances closer to the metrical character of music. The use of rhyme in poetry likewise parallels the rhythmic patterns of harmonic/melodic expectation in music (Lerdahl, 2003). The appeal of poetry—even to children—presumably comes from the affect evoked by such rhythmic patterning. Similarly, the use of call and response patterns (as in certain styles of preaching) evokes strong affect, paralleled by the experience of choral singing. And of course combinations of music and language are ubiquitous—in song, where language follows musical rhythms, in chant (e.g., recitative), where melody follows speech rhythms, and in rap, where words without melody follow musical rhythms.

Lerdahl (2003) suggests (and I concur) that these are all hybrids. The hypothesis is that poetry is the result of superimposing musical principles to some degree on linguistic utterances. Thus to the extent that poetic form conveys affect, it is precisely because it invokes principles of *musical* perception that are not normally associated with language.[3]

In addition, language and music both convey affect through tone of voice. However, this concurrence is not evidence that the two capacities are the same. Rather, they have both incorporated in different ways some of the character of mammalian call systems. Mammalian calls *do* convey both affect (like music) and some very limited sort of conceptual information (like language). But this does not lead inevitably to the conclusion that language and music evolved as a single capacity that later split (the "musilanguage" hypothesis of Brown, 2000). It is equally possible that they are independent evolutionary specializations of primate communication. The differences enumerated in the next section to my mind favor the latter hypothesis.

From this brief survey, my conclusion is that once we abstract away from the use of tone of voice shared by language and music, and once we recognize poetry and other rhetorical devices as mixtures of language and music, the specialization of language to conceptual information and music to affect is actually quite extreme.

Similarities and Differences in Formal Structure

Another take on similarities and differences between language and music comes from looking at the formal devices out of which language and music are constructed. These emerge from examination of the fine-scale organization of the two domains, and they are the areas where formal linguistics and approaches to music such as GTTM are most pertinent.

Unlike any other cognitive capacities, both language and music involve a sequence of digitized sounds: speech sounds in language, tones or pitch events in music. This is certainly one reason for the strong intuition that they are alike. However, the resemblance ends there. In phonological structure, the repertoire of speech sounds forms a structured space of timbres—vowels and consonants. To a lesser degree speech sounds are distinguished by length (short to long a factor of 2 or so). In music, by contrast, the notes are distinguished by the way they form a structured space of *pitches* and by a broad range of relative lengths

[3]Poetry in signed languages (Klima & Bellugi, 1979) makes use of alliteration (such as deliberate choice of parallel handshapes), rhythmic patterning, and—unlike spoken poetry—counterpoint (overlapping of signs). Again, arguably musical types of structure are superimposed on language.

(shortest to longest a factor of 16 or more).[4] I will return to pitches in a moment.

Rhythm

One of the important discoveries of GTTM was the extent to which phonology and music are structured rhythmically by very similar metrical systems, both based on a hierarchical metrical grid. This is an important formal parallel between the two domains, perhaps shared by only music and language. Even so, the domains do not use the grid quite the same way. In phonology, the minimal metrical unit is the syllable, a structured sequence of speech sounds. Each syllable corresponds to a beat in the metrical grid. The performance of the metrical grid need not be isochronous and usually is not (Patel, 2008, and references therein). In music, by contrast, a single note can subtend multiple beats, and a beat can be subdivided by multiple notes. And within certain degrees of tolerance (depending on the style), the metrical grid is isochronous, which is what permits the possibility of syncopation.

The other component of rhythm in music is grouping, a recursive segmentation of the musical stream into motives, phrases, and sections. GTTM stresses the similarity of musical grouping to visual segmentation, which configures multiple objects in space and governs the segmentation of objects into parts and parts of parts. Grouping structure, though recursive, is not a headed hierarchy; that is, there is no distinguished element in a group that functions as head. Musical groups simply contain a collection of elements, which may be either individual notes or smaller groups.

The closest analogue in language to musical grouping is intonational phrasing (or breath groups). Unlike musical grouping, intonational phrasing forms a relatively flat structure, bounded by the extent of a single sentence; there is none of the deep recursion found in music from the smallest motivic units all the way up to an entire piece. Moreover, intonational phrases are made up of smaller prosodic constituents such as phonological words and phonological phrases, each with its own specific properties (Selkirk, 1984). Such a differentiation of grouping units into distinct types is not found in music.

My conclusion is that in the rhythmic domain, the metrical grid may well be a genuine capacity unique to language and music; musical grouping is shared more with vision than with language, and linguistic intonation contours are partly specific to language.

Pitch

As mentioned above, the organization of sounds in language is governed by the structure of phonological space: consonants and vowels are distinguished by how they are articulated in the vocal tract. In contrast, in all traditional musical cultures that use pitch at all, the organization of sound in music is built around a tonal pitch space: there is a fixed collection of pitches that differ in stability with respect to a tonic pitch. It is well established that the structure of tonal systems is explained only in part by psychoacoustics; the rest is culture-specific (see Jackendoff & Lerdahl, 2006, and references therein).

Are the characteristics of tonal pitch spaces shared at all with language? Consider a number of possible parallels. First, prosodic contours in language tend to go down at the end, and so do melodies. But both probably inherit this from the form of human and mammalian calls (and possibly from the physiology that gives rise to calls, e.g., the drop in air pressure as the lungs are emptied). Thus this common characteristic could be the product of independent inheritance from a common ancestor.

An important difference between prosodic contours and melodies is that only melodies have discrete pitches, while prosodic contours usually involve a continuous rise and fall. Of course there are mixtures. On the one hand, many vocal and instrumental traditions incorporate bending of pitches and sliding between them. On the other hand, intonation (in many languages) is now commonly analyzed in terms of high and low pitches anchored on prominent accents and the ends of breath-groups (Pierrehumbert, 1980). Thus, one might be able to think of intonation in language as fundamentally a two-pitch tonal system, modulated by continuous transitions between the anchoring pitches. However, the high and low pitches are not fixed in frequency throughout an utterance in the way the dominant and tonic are fixed in musical pitch space. So the analogy between intonational systems and tonal pitch spaces is strained at best.

[4]Patel (2008) offers evidence that the local variation in note lengths in a culture's music is sometimes correlated with the local variation in syllable lengths in its language (which in turn is linked to the variety of possible syllable types). In order to rule out a direct infusion of linguistic rhythms into music, Patel is careful to look at music most distant from potential vocal influences, e.g., classical instrumental music of England versus that of France. Even so, it is hard to discount the influence of vocal music, including folk and popular music, on the imaginations of composers.

Another use of pitch in language is for tone in tone languages such as the Chinese languages and many West African languages. In such languages, the tones form a fixed set that might be seen as analogous to tonal pitch space. There are two arguments against this analogy. First, since the choice of tones is determined by the words in the sentence, none of the tones can play the role of a tonic, which constitutes a point of maximum stability at which melodies typically come to rest. Second, tones are not fixed through the sentence; rather, they are superimposed on an overall intonation contour. As a breath group continues, all tones drift down, and the intervals between them get smaller (Ladd, 1996). This is an entirely different use of pitch than in the pitch spaces of music.[5]

Finally, it must be added that evidence from tone deafness and amusia (Peretz & Coltheart, 2003) suggests that linguistic intonation and musical pitch are controlled by distinct brain areas.

I conclude that there is no convincing analogue in language to the musical use of pitch space, despite their making use of the same motor capacities in the vocal tract.

Words

Beyond elements of the sound system, language and music diverge more radically. Linguistic utterances are built up from words and syntax; pieces of music are built up from individual tones, some formulaic patterns, and prolongational structure. Let us go through the possibilities for parallelism here.

Words are conventionalized sound patterns associated in long-term memory with pieces of meaning (or concepts). Sentences are made up entirely of words plus "grammatical glue" such as inflectional morphology (e.g., agreement, grammatical gender, case). Is there a musical analogue? Musical idioms do incorporate conventionalized sound patterns such as stylistic clichés and standard forms of cadences, as well as more abstract patterns such as 12-bar blues or sonata form. But unlike words, these patterns are not associated with concepts. Moreover, except in some limited musical styles,[6] melodies are not made up exclusively of conventionalized patterns in the way that sentences are made up of words.

In fact, the function of conventionalized patterns in music more closely resembles the function of linguistic clichés, idioms, and figures of speech—so-called "prefabs." Like conventionalized musical patterns, prefabs occur with considerable frequency in language, but utterances are not exclusively made of them: there is plenty of opportunity for free choice of words. However, if musical formulas are parallel to prefabs, then there is no musical parallel to words per se. And of course musical formulas do not carry conceptual meaning in any event.

Syntax

What enables language to serve as such a flexible and expressive mode of communication is syntactic structure. Syntactic structure is a hierarchical structure in which each node belongs to a syntactic category such as Noun or Adjective Phrase. There is no musical counterpart to these categories. Each syntactic category has its own characteristic configurations of arguments and modification; for instance, an English verb phrase contains a head verb followed by up to two noun phrases (direct and indirect objects), followed by prepositional phrases, adverbs, and finally subordinate clauses. Syntactic structure is a headed hierarchy, in that one element of most constituents is designated as its head. Moreover, the category of a phrase is determined by the category of its head, so that a noun phrase is headed by a noun, a prepositional phrase by a preposition, and so on; this is the fundamental principle behind the "X-bar" theory of phrase structure (Chomsky, 1970; Jackendoff, 1977).

Syntax also contains multiple devices for encoding the dependencies among its constituents; for example, agreement, case, reflexivity (and other anaphora), ellipsis, and long-distance dependencies (such as wh-displacement). Within words, there is often further differentiation into *morphosyntactic* structure: affixal structures that affect meaning and syntactic category. Sometimes this structure is recursive (as in

[5]Patel (2008) discusses "discrete tone languages" such as the West African language Jukun, described by Welmers (1973), in which tones do not drift downward throughout the utterance. Patel argues that such languages provide a closer parallel to musical pitch. However, they still do not display the anchored property of tonal pitch spaces, in that the low tone does not function as a tonic. Moreover, in a passage quoted by Patel (pp. 43-44), Welmers says that "these tonemic ranges . . . may all tilt downward at the very end of the phrase in a brief final contour." This could never happen in a melody.

[6]One such is the Jewish liturgical tradition for torah chant, which is constructed from 18 musical motives, several of which have two to five variants (Binder, 1959).

antidisestablishmentarianism) and sometimes templatic (for instance French object clitics).

None of this structure has a counterpart in music. All of it is in the service of the function of syntax: to code meaning relations among words in a form amenable to phonological expression, namely linear order and affixation. And of course this function is absent in music as well.

Prolongational Structure

The closest musical counterpart to syntax is GTTM's prolongational structure, which was originally inspired by the recursive reductional hierarchy of Schenkerian theory (Schenker, 1935/1979). Prolongational structure, like syntax, is a recursive headed hierarchy, in which each constituent has a head, and other dependents are modifiers or elaborations of the head. But in other respects the two structures diverge. Prolongational structure has no parts of speech: the tonic/dominant distinction, for instance, is not formally analogous to either noun/verb or subject/predicate/object.[7] There is a sense in which the category of a constituent is determined by its head, but it does not parallel X-bar structure in language. For instance, a phrase headed by the note G or by a G major chord is not a "G-phrase," but simply an elaborated G.[8] The difference between the two structures is illustrated in Figure 1.

Prolongational relations do not concern the regimentation of conceptual relations; rather, they encode the relative stability of pitch-events in local and global contexts. Prolongational structure creates patterns of tensing and relaxing as the music moves away from stability and back towards a new point of stability. GTTM—and in much more detail, Lerdahl (2001)—argue that these patterns of tensing and relaxation have a great deal to do

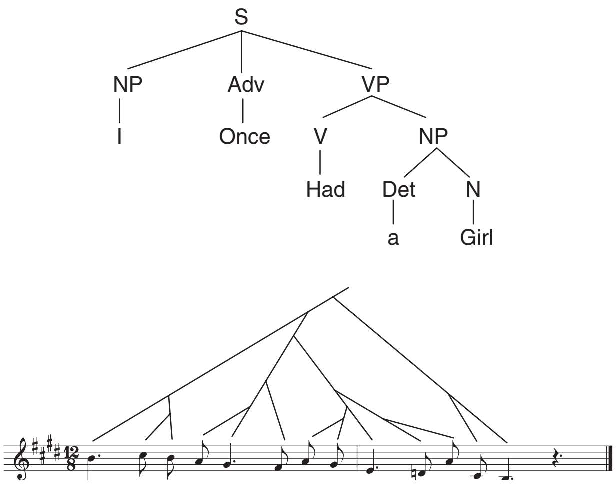

FIGURE 1. Contrast between syntactic and prolongational structure.

with affect in music. There is no counterpart to this function in language: it codes something entirely different than syntax does. Thus, both on formal and functional grounds, syntax and prolongational structure have little in common beyond both being headed hierarchies.

Following the general intuition that the components of music ought not to be *sui generis*, it would be delightful to find a stronger analogue of prolongational structure in some other cognitive capacity. Evaluating the strength of any potential parallels with another capacity requires a detailed analysis and comparison of faculties, of the sort I have done here. At the moment, this is impossible: we simply don't know about the mental representations for any other cognitive capacity in the way we do for music and language.

Another Recursive Headed Hierarchy: Complex Action

However, a candidate comparison has recently emerged. Jackendoff (2007), drawing in part on work in robotics, suggests that the formulation and execution of complex actions—actions as ordinary as shaking hands or making coffee—invokes elaborate hierarchical structure that integrates and modulates many subactions stored in long-term memory. Figure 2 shows some of the structure involved in making coffee in an electric coffeemaker.

Like syntax and prolongational structure, this is a recursive headed hierarchy. The basic element in the tree is an event consisting of a Head (the main action), with an optional Preparation (things that have to be done before the Head can be begun) and an optional

[7]Except in the relatively superficial fashion suggested by Patel (2008), whereby the same chord, say G major, can function as tonic in G major and dominant in C major. Patel likens this to a noun phrase functioning either as subject or object, depending on its position. But this context-dependence of function makes music no more like language than like, say, social interaction. For example, shaking hands has the function of greeting in some contexts and of congratulating in others, and it can be "ungrammatical" in still others, such as in the middle of a joke.

[8]Chomsky's (1995) theory of "bare phrase structure" proposes syntactic trees that are more like prolongational reductions, in that, for example, a phrase headed by *dog* is not an NP but simply an elaborated *dog*. However, to my knowledge it has never been shown how this approach captures the standard insights of phrase structure grammar, and the proposal has had virtually no impact on subsequent empirical research.

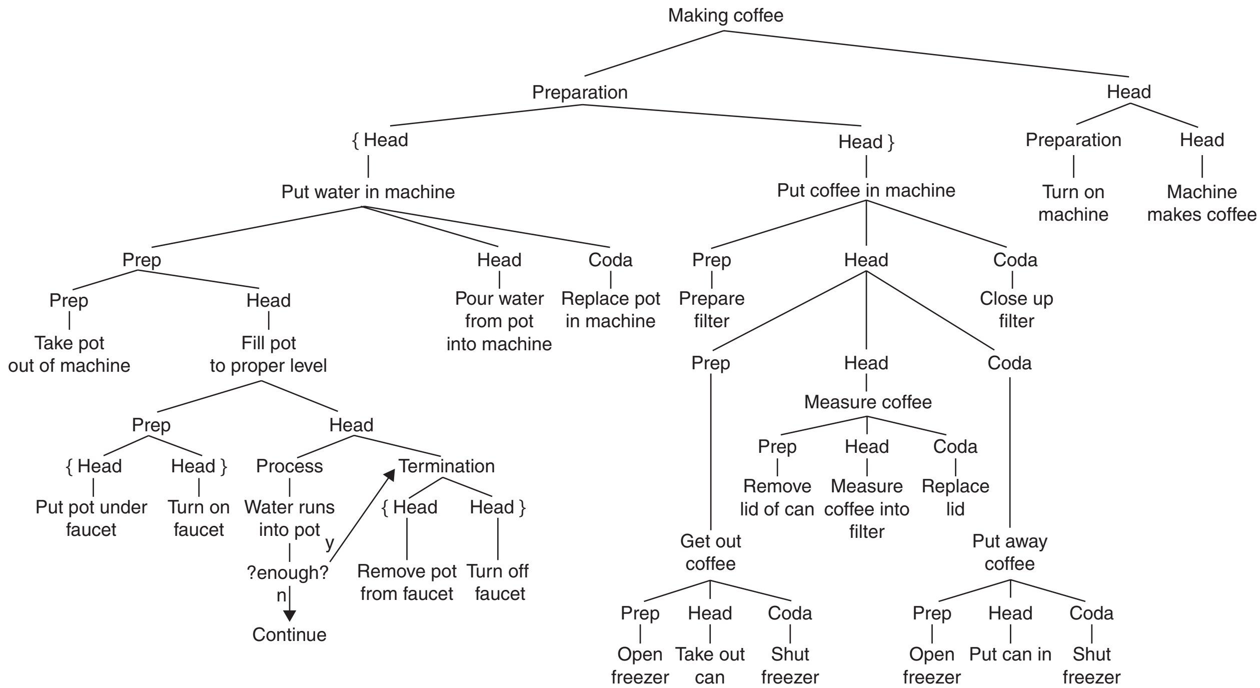

FIGURE 2. Structure of the complex action of making coffee.

Coda (things that are done to restore the status quo ante). For instance, consider the constituent "put water in machine" in Figure 2. The Head consists of actually pouring water into the machine from the pot. But in order to do this, one must first measure water into the pot—the Preparation—which in turn is organized into Preparation plus Head, and each of these has further organization. And once one has poured the water into the machine, one must replace the empty pot in the machine—the Coda.

There are some variations on this basic structure. For instance, it doesn't matter what order one puts the water in the machine and puts the coffee in the machine, but they are both necessary Preparations for turning on the machine so it will make coffee. These thus constitute co-Heads of the Preparation, and they are notated in Figure 2 with { } brackets around them. Another variant is the Head of "fill pot": one is filling the pot from the faucet, and one has to monitor it to see when it has reached the right level, at which point one terminates the process. This is bears some resemblance to the "TOTE units" of Miller, Galanter, and Pribram (1960).

In short, Figure 2 shares features with both syntactic and prolongational trees. This gives us a little better perspective on the extent to which we should consider musical and linguistic structure related, and whether the relations are specific to these two domains or more general.

Consider also how Figure 2 is processed in the course of actually making coffee. How much of it is stored in memory as part of the "coffee-making" routine? My guess would be that the stored routine does not include, for instance, how one turns on the faucet in order to fill the pot with water. Rather, this is another routine that has to be integrated with the stored part of the coffee-making routine in the course of executing the constituent "fill pot." An even more basic piece, not shown in Figure 2, is walking from the coffeemaker to the sink in order to get water. This surely is a separate routine that gets integrated on the fly. Jackendoff (2007) suggests that the online integration of these structures has many parallels to speech production.

Patel (2003, 2008) presents experimental evidence that the hierarchical structures of language and music, although formally distinct, are integrated by the same part of the brain, roughly Broca's area. If so, I would hazard a guess that complex action structures are too. That this area is usually considered premotor would add some plausibility to this speculation. In fact, the integration and execution of complex action might be a strong candidate for a more general, evolutionarily older function that could be appropriated by both language and music, quite possibly independently.

Conclusion

Language and music share a considerable number of general characteristics and one detailed formal one, namely metrical structure. They also share some brain areas. I have argued, however, that most of what they share does not indicate a particularly close relation that makes them distinct from other cognitive domains. Many shared characteristics are domain-general, for instance recursion, the use of memory, and the need for learning and for a social context. Moreover, the fact that language and music are both conveyed through the auditory-vocal modality, though it places constraints on both of them, does not have much to do with their formal structure. This is pointed up especially by the alternative signed modality for language, which preserves most of the standard formal properties of language. Finally, language and music differ substantially in their rhythmic structure, in their use of pitch, in their "meaning" (propositional versus affective), and in the form and function of their hierarchical structures.

The conclusion, then, is to urge caution in drawing strong connections between language and music, both in the contemporary human brain and in their evolutionary roots. This is not to say we should not attempt to draw such connections. For example, Patel (2008), surveying much the same evidence as I have here, concludes the glass is half full rather than half (or three-quarters) empty. But if one wishes to draw connections, it is important to do so on the basis of more than speculation. In particular, at the moment we do not have a properly laid out account of even one other capacity against which to compare language and music. It is an interesting question when and how cognitive science will approach such analyses, so that we may eventually have a fair basis for comparison.

Author Note

I am immensely grateful to Emmanuel Bigand for conceiving and organizing the conference "Musique Langage Cerveau" in Dijon in honor of the 25th anniversary of GTTM, as well as for organizing the present special issue of *Music Perception* in the same cause. Such recognition of one's work comes rarely in one's lifetime, and I am deeply honored.

Much of the material in the present article is drawn from Lerdahl and Jackendoff (1983), Jackendoff (1987), and Jackendoff and Lerdahl (2006). Many similar points are made in Patel (2008), often with considerably more thorough documentation. I wish to thank Fred Lerdahl and three reviewers for *Music Perception* for many useful comments on an earlier version of this article.

Correspondence concerning this article should be addressed to Ray Jackendoff, Center for Cognitive Studies, Tufts University, Medford, MA 02155. E-MAIL: ray.jackendoff@tufts.edu

References

Binder, A. W. (1959). *Biblical chant*. New York: Philosophical Library.

Boyd, R., & Richerson, P. J. (2005). *The origin and evolution of cultures*. Oxford: Oxford University Press.

Bratman, M. E. (1999). *Faces of intention*. Cambridge: Cambridge University Press.

Brown, S. (2000). The "musilanguage" model of music evolution. In N. L. Wallin, B. Merker, & S. Brown, (Eds.). *The origins of music* (pp. 271-300). Cambridge, MA: MIT Press.

Calvin, W. (1990). *The cerebral symphony*. New York: Bantam Books.

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

Chomsky, N. (1970). Remarks on nominalizations. In R. Jacobs & P. Rosenbaum (Eds.), *Readings in English transformational grammar* (pp. 184-221). Waltham, MA: Ginn.

Chomsky, N. (1981). *Lectures on government and binding.* Dordrecht: Foris.

Chomsky, N. (1995). Bare phrase structure. In H. Campos & P. Kempchinsky (Eds.), *Evolution and revolution in linguistic theory* (pp. 51-109). Washington: Georgetown University Press.

Christiansen, M., & Chater, N. (2008). Language as shaped by the brain. *Behavioral and Brain Sciences, 31*, 537-558.

Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.

Gilbert, M. (1989). *On social facts*. Princeton: Princeton University Press.

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science, 298*, 1569-1579.

Jackendoff, R. (1977). *X-bar syntax*. Cambridge, MA: MIT Press.

Jackendoff, R. (1987). *Consciousness and the computational mind*. Cambridge, MA: MIT Press.

Jackendoff, R. (2002). *Foundations of language*. Oxford: Oxford University Press.

JACKENDOFF, R. (2007). *Language, consciousness, culture*. Cambridge, MA: MIT Press.

JACKENDOFF, R., & LERDAHL, F. (2006). The capacity for music: What's special about it? *Cognition, 100*, 33-72.

JACKENDOFF, R., & PINKER, S. (2005). The nature of the language faculty and its implications for the evolution of language (Reply to Fitch, Hauser, and Chomsky). *Cognition, 97*, 211-225.

KLIMA, E., & BELLUGI, U. (1979). *The signs of language*. Cambridge, MA: Harvard University Press.

LADD, R. (1996). *Intonational phonology*. Cambridge: Cambridge University Press.

LERDAHL, F. (2001). *Tonal pitch space*. New York: Oxford University Press.

LERDAHL, F. (2003). The sounds of poetry viewed as music. In J. Sundberg, L. Nord, & R. Carlson (Eds.), *Music, language, speech and brain* (pp. 34-47). London: Macmillan.

LERDAHL, F., & JACKENDOFF, R. (1983). *A generative theory of tonal music.* Cambridge, MA: MIT Press.

MILLER, G. A., GALANTER, E., & PRIBRAM, K. (1960). *Plans and the structure of behavior*. New York: Holt, Rinehart & Winston.

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

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

PERETZ, I., & COLTHEART, M. (2003). Modularity of music processing. *Nature Neuroscience, 6*, 688-691.

PIERREHUMBERT, J. (1980). *The phonetics and phonology of English intonation*. Doctoral dissertation, Massachusetts Institute of Technology. [Published by Indiana University Linguistics Club, 1987].

PINKER, S. (1997). *How the mind works*. New York: W. W. Norton.

PINKER, S., & JACKENDOFF, R. (2005). The faculty of language: What's special about it? *Cognition, 95*, 201-236.

RAFFMAN, D. (1993). *Language, music, and mind*. Cambridge, MA: MIT Press.

SCHENKER, H. (1979). *Free composition* (E. Oster, Trans.). New York: Longman. (Original work published 1935)

SEARLE, J. (1995). *The construction of social reality*. New York: Free Press.

SELKIRK, E. O. (1984). *Phonology and syntax: The relation between sound and structure*. Cambridge, MA: MIT Press.

TOMASELLO, M. (2003). *Constructing a language*. Cambridge, MA: Harvard University Press.

TOMASELLO, M., CARPENTER, M., CALL, J., BEHNE, T., & MOLL, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences, 28*, 675-691.

WALLIN, N. L., MERKER, B., & BROWN, S. (Eds.). (2000). *The origins of music*. Cambridge, MA: MIT Press.

WELMERS, W. E. (1973). *African language structures*. Berkeley: University of California Press.

WILKINS, W. K. (2005). Anatomy matters. *The Linguistic Review, 22*, 271-288.