

# Biological foundations of music and language: A structural perspective

Teresa Blasco Máñez  
MA Thesis

Master en Ciència Cognitiva i Llenguatge  
Universitat Autònoma de Barcelona  
2010

Director: Sergi Balari Ravera



## Table of contents

Abstract .....	2
1. Introduction.....	3
<i>Music matters</i> .....	3
<i>Objective</i> .....	5
2. The faculty of music and the faculty of language.....	6
2.1. Acquisition and Development.....	7
<i>Reorganization of the perceptual space</i> .....	7
2.2. Syntactic processing .....	9
<i>Interlude: the syntax of music</i> .....	9
<i>Overlap in syntactic processing</i> .....	12
<i>"Selective" impairment of musical pitch processing</i> .....	13
2.3 Formal resources.....	15
2.4. The case of musical rhythm.....	18
<i>Entrainment to a musical beat</i> .....	19
<i>Neural substrates for BPS —a brief sketch</i> .....	22
<i>Beyond (or below) the Vocal Learning Hypothesis</i> .....	24
<i>Beat-based processing and other cognitive deficits</i> .....	25
3. Precursors in non-human animals .....	27
3.1. Perceptual constraints.....	29
<i>Auditory processing</i> .....	29
3.2. Computation.....	31
<i>(So-called) human-specificity: coming to grips with sequencing capacities</i> .....	32
4. Concluding remarks.....	36
References .....	37

# **Biological foundations of music and language: a structural perspective.**

Teresa Blasco Máñez

---

Abstract: The objective of this work is to undertake a comparative approach to the evolutionary biology of music and language as cognitive capacities following a structural, internalist perspective. The first part of the work aims at retrieving insight on the neurocomputational substrate shared by both capacities through the characterization of mechanisms relevant to the acquisition and implementation of knowledge in both domains. In the second part of the work comparative data is reviewed in order to establish possible structural homologies with other species. It is argued that the integration of different kinds of comparative data (developmental, anatomical, genetic...) according to this structural criterion allows us to gain insight into the evolutionary origin of the organic structures that support these capacities and thus, into the nature of human musical and linguistic capacities.

## 1. Introduction

### *Music matters*

Both language and music are universals in human culture that reach deep into our species' past (Nettl, 2000). The fact that these traits are exclusive to our species and that they seem to share a series of formal characteristics at different levels makes the biological relation between both capacities a fascinating topic of research, for the differences setting language and music apart are also many. Indeed, when it comes to approaching the study of music from an evolutionary perspective, the recurrent point of departure involves the lack of an apparent specific utility that –unlike language– this capacity poses for our species' survival. This concern is already present in Darwin's *The Descent of Man*, where he devoted a chapter to 'Musical Powers', providing an often-quoted reflection on the presence of this ability in humans:

“As neither the enjoyment nor the capacity of producing musical notes are faculties of the least direct use to man in reference to his ordinary habits of life, they must be ranked among the most mysterious with which he is endowed. They are present, though in a very rude and as it appears almost latent condition, in men of all races; even the most savage; but so different is the taste of the different races, that our music gives not the least pleasure to savages, and their music is to us hideous and unmeaning.” (Darwin, 1871).

The emergence of the integrating approach required by cognitive science has favoured an increasing interest in the study of music from a biological point of view, so that this capacity has come to be regarded as a product of human cognition which can provide much valuable scientific insight where, formerly, a humanistic and historical perspective to this topic of study had prevailed (Zatorre, 2005). Actually, at least in the non-trivial aspect that both phenomena involve the projection of a hierarchical structure on a linear acoustic stimulus, music provides a privileged standpoint for cross-domain comparison with the Language Faculty as another paradigmatic instance of an inherent ability to make sense out of sound.

Nowadays a substantial corpus of research on the mechanisms that support music –even if still minimum when compared to language– is available, allowing for hypotheses on the evolutionary status of this capacity (indeed, two volumes of essays have been devoted to the evolution of music in the last decade; Wallin et al., 2000; Vitouch & Ladinig, 2009) and on its relation to the human faculty of language (e.g. Patel, 2008). In these respects, Charles Darwin’s quotation above provides, almost a century and a half later, an excellent introduction to the concerns that still today dominate this field of study.

The logic behind Neo-Darwinism and the way in which it has been applied to the study of language and music as cognitive capacities have generally placed emphasis on the search for innate universals underlying these behaviours, with the eventual objective of pinpointing particular features that might have been selected upon for conferring an adaptive value. However, as more comes to be known on the complexity of biological systems and of the nature of the processes that interact upon their evolution, the view emerges –or rather, is brought back into attention– that a satisfactory account for these complex cognitive faculties cannot be achieved in the absence of a realistic biology-grounded framework, which takes into account architectural and developmental factors and “assumes less” in terms of the weight granted to natural selection as a creative force.

In this regard, although significant contributions have aimed at setting the ground for addressing the evolutionary study of the human faculties of language and music from a comparative structural perspective (e.g., Hauser, Chomsky & Fitch, 2002; McDermott & Hauser, 2003), it can be argued (as Balari & Lorenzo, 2009a, do) that much of the discussion has become obscured due to the centrality granted to assumptions on functional continuity. This has led to a predominantly functional application of the comparative method (communication-oriented in the case of language, albeit more imprecise for music) when it comes to determining the evolutionary status of the mechanisms underlying these cognitive capacities. Hence, the emphasis on the selection of particular, domain-specific functions or components as the main driving force in evolutionary processes continues to characterise current debates on the biological foundations of both phenomena and on their relationship as cognitive capacities.

The latter conception of evolutionary processes is put into question within the approach fostered by modern evolutionary developmental biology (*Evo-Devo*; Hall, 1999), where attention is devoted to the mechanisms implied in developmental processes (such as epigenetic mechanisms or organism-environment interactions) and to their active role in the evolution of species. This approach brings in the notion that, from a biological standpoint, as argued by Balari and Lorenzo (2009a), the concept of function can be regarded as an evolutionary epiphenomenon that does not define the origin of the organic structure that supports its activity.

### *Objective*

Bearing this premise in mind, the general object of this paper is to undertake the evolutionary study of music and language by approaching their comparison from a structural (non-functionalist), internalist perspective. Hence, it will be argued that this structural criterion allows for integrating different kinds of comparative data –developmental, anatomical, genetic, etc.— in such a way that deep parallels at apparently unrelated levels can be established, allowing for potential insight as to the biological nature of both capacities.

As a framework for this proposal, I will observe the model put forward by Balari and Lorenzo (2009b) concerning the natural system of computation underlying language, from which two main assumptions derive to this work:

- i) musical and linguistic capacities share a common neurocomputational substrate, localisable, at minimum, in the basal ganglia, with the functions of an ‘universal sequencing’ engine (Lieberman, 2006), but which probably extends to other centres in the cortex.
- ii) given that the organic structures that support these faculties are, to some extent, shared by vertebrates (Striedter, 2005), it is possible to find homologues for these morphological substrates in other species, by comparing their neuroanatomy and through a formal characterization of the cognitive processes they subserve.

The subsequent pages have been organised as follows. The first part of the work (section 2) provides a characterization of mechanisms relevant to the acquisition and implementation of knowledge in both domains, aimed at retrieving insight on the overlap between these cognitive capacities. The next part of the work (section 3) consists of a review on comparative data related to the aforementioned mechanisms, followed by concluding remarks.

## **2. The faculty of music and the faculty of language**

The question of whether the human capacity for music constitutes an evolutionary adaptation seems to arise, as in the case of our linguistic capacity, quite strong feelings among theorists involved in these fields of study. The non-adaptationist position that music is biologically useless —“an auditory cheesecake” (Pinker, 1997), is based on evidence which suggests that it builds from pre-existing brain functions, such as linguistic mechanisms. Proponents of adaptationist views, on the other hand, posit that the human capacity for music is a product of natural selection that reflects the survival value that this capacity would have posed for the human species, for example, as a mechanism favouring social cohesion (e.g., Brown, 2000) or sexual competition (e.g. Miller, 2000, retakes the Darwinian sexual selection hypothesis, drawing a functional analogy between birdsong and music). The notion that language and music are tightly intertwined is reflected by theories that postulate a common origin for both capacities (e.g. Mithen, 2005).

Generally, peculiarities or commonalities of these capacities are emphasised or underscored by proponents of each theory. However, it might be the case that, as pointed out by Fitch (2006), the dichotomy between ‘adaptation’ or ‘frill’ provides an imperfect match when addressing the evolutionary status of certain mechanisms or capacities, and that more parsimonious accounts are possible.

In this section, cross-comparison between music and language is undertaken focusing on the general processes and mechanisms that seem to underlie the development of these faculties and on their similarities at a formal level, suggesting that music and language display common neurocomputational substrates to an important extent.



## 2.1. Acquisition and Development

### *Reorganization of the perceptual space*

A widespread notion that emerged from research on early language acquisition is that infants are born with a pre-developed sensitivity to the prosodic aspects of language. This sensitivity seems to serve as guidance in the process of learning the fine-grained distinctions in the acoustic patterns of language, so that the subsequent acquisition of the lexicon resides in the attribution of meaning to these segments. Thus, for example, a series of studies by Mehler and colleagues (e.g., Mehler et al., 1996; Ramus & Mehler, 1999; Nazzi et al., 1998) showed how during an early phase of development, children are able to discriminate among different languages on the basis of a rhythmic typology. Not very surprisingly, infant-directed speech appears to take advantage of infants' responsiveness to these aspects of speech and thus, this register is characterised cross-linguistically by exaggerated prosodic contours (Fernald, 1992) which seem to facilitate learning (Thiessen, Hill, & Saffran, 2004).

The possibility of adventuring a connection between these early linguistic sensitivities and a capacity for music has not gone unnoticed, even though the traditional interpretation of these results was once that of emphasising the role of innate abilities or mechanisms specific for language acquisition. This view, however, has been countered by more recent results on other forms of auditory processing —as music— and also by comparative evidence from other species. In this regard, it is important not to forget that our perceptual system shares a number of properties and capacities with that of other mammals, as will be noted below.

From a very early stage of development humans are able to extract regularities from an acoustic sequence such as speech. Given that our auditory efficacy will depend upon an efficient perceptual categorization, this can be regarded as one of the first challenges we encounter in the task of acquisition. This process can be understood as the transformation in our perceptual system of an “absolute pitch” initial state in which we are able to establish very subtle distinctions to a progressive formation of discrete categories conditioned by exposure,

characteristic of the mature state<sup>1</sup>. See Locke (1993) for a detailed study of these early stages in the development of auditory development.

A first acquisitional parallel can be traced between language and music regarding the development and use of category-based auditory perception, where acoustically variable stimuli are “abstracted” into a framework of stable mental categories. Hence, research on the development of the sensitivity to key structure in music shows that the perception of categories within the octave and the time in which it is manifest (6-12 months) parallels the acquisition of phonemic categories in language (Justus & Hustler, 2005). It is at this stage that the characteristic ability for perceiving musical stimuli in terms of relative pitch encoding –instead of in absolute frequency terms– begins to emerge (McMullen & Saffran, 2004), a capacity that is also displayed in the processing of speech sounds. The formation of these categories is conditioned by exposure and reflects infants’ remarkable abilities for keeping track of statistical and distributional cues in the stimulus (cf. Maye, Werker & Gerken, 2002).

This capacity for statistical learning and its role in language acquisition has been repeatedly addressed in Jenny Saffran’s studies (Saffran et al. 2008; Pelucchi, Hay, & Saffran 2009), focusing on the extent to which similar learning and memory mechanisms might mediate the acquisition of knowledge in this and other domains such as music. Children, for instance, are able to extract transitional probabilities from sequences made up of syllables, but also of discrete pitch tones (Saffran, Aslin & Newport, 1996; Saffran et al. 1999) or visual sequences (Kirkham et al. 2002). This suggests that the acquisition of these systems relies on some mechanisms which rather than being part of a language-specific learning kit, can be viewed as corresponding to more general perceptual and processing capacities, which may be shared with other species.

Therefore, evidence from studies on auditory development (McMullen & Saffran, 2004; Patel, 2008 ch. 2 for overviews) seems to provide support for the idea that although in adults musical and linguistic knowledge might be instantiated as separated stocks in the brain, both domains might share basic developmental mechanisms. As noted by Patel (2008) among others, this points

---

<sup>1</sup> See Locke (1993) for a detailed study of these early stages in auditory development.

at the need for drawing a distinction between the end products of development, which might be domain-specific, and the processes operating during development, which might be domain-general, a view that is consistent with earlier modularity-skeptical views on language development, such as Karmiloff-Smith's, (1992). As we will see, this notion is also consistent with data from dissociations and can extend to a number of domains, such as syntactic processing.

## 2.2. Syntactic processing

The evidence above suggests the need for posing a distinction between the specificity attributed to the output of a given process —e.g., linguistic and musical categories— and that of the cognitive mechanisms involved in this process —auditory processing and memory. Neuropsychological studies on syntactic processing allow us to extend the proposal on the acquisition mechanisms for linguistic and musical sound categories to a wider resource-sharing framework (Patel 2003; 2010), the characteristics of which can be summarised along two assumptions: i) language and music have specific representations for each domain, and ii) when certain cognitive operations work on these representations, the brain makes use of similar resources.

### *Interlude: the syntax of music*

The fact that as listeners we are able to recognise a tune when it is transposed across different tonalities, or even when we are faced with an entirely different version of it (as it might be the case in jazz, where often little more than chord progression structure is preserved), already provides an indication that we are able to abstract away from the mere acoustic characteristics of the stimulus in important ways. For Western tradition in general, these ways have to do with a number of dimensions of pitch organization implied in the tonal system —e.g., discrete pitches are organised into unequally stepped scales of seven degrees,

which determine the formation of chords and harmonic relations in terms of their perceived proximity and stability with a tonal centre<sup>2</sup>.

Implicit knowledge of this system on the part of listeners becomes evident in our ability to detect ‘sour’ notes in a melody —i.e., its ‘well-formedness’— or to anticipate certain kinds of events as the music unfolds. In other words, we are able to make restricted predictions concerning temporal and harmonic aspects of music (e.g. the final chord of a song or musical piece; Krumhansl, 1990 also Huron, 2006<sup>3</sup>), in the same way that as language speakers or listeners we have expectations as to the kind of word that will come after a determiner, for example. Our implicit knowledge of music resembles grammatical knowledge in this sense, for relations are established on the basis of abstract structural properties of its ‘building blocks’ —for instance, it is the harmonic context that determines whether a pitch is encoded as the structural category of the tonic (the most stable pitch) or the leading tone (a highly unstable pitch).

It is important to bear in mind that although certain features of the Western tonal idiom —e.g., the use of scales with seven degrees— are idiosyncratic to this system, it can be taken to reflect different organisational biases presented by other musical idioms, which suggests that perceptual and psychoacoustic factors, just as limitations on processing, act as constraints on variance for musical systems cross-culturally<sup>4</sup>:

Helmholtz (1863), for instance, argued for grounding harmony and consonance in physiology, a topic that has driven the attention of contemporary researchers such as Krumhansl (e.g. 2000). Support for a physiological basis for certain features of music is provided by infant and primate studies with regard to, for example, the distinction between dissonance and consonance (Trainor, Tsang, & Cheung, 2002; Izumi, 2000). Also, the perceived equivalence between pitches separated by a doubling in frequency —a 2:1 frequency ratio, corresponding to the octave interval in the diatonic scale— is reflected by most musical systems,

---

<sup>2</sup> For a brief introduction to these and other features of the Western tonal system, see Harkleroad (2006).

<sup>3</sup> Taking on a different issue, Huron (2006) suggests that the way in which music ‘plays’ with the different expectations that become engaged during listening is precisely one of the main sources for emotional responses to music.

<sup>4</sup> In this sense, ‘third-factor’ explanations (Chomsky, 2005) might provide a ground for addressing convergences between language and music.

and in studies on young infants<sup>5</sup> and macaques (Trainor, 1997; Wright et al., 2000). Other of the universals in pitch organization that have been put forward in comparative studies (Justus & Hustler, 2005; McDermott & Hauser, 2005) concern the asymmetry of interval patterns in scales (i.e. of the use of unequal steps between the scale degrees), which have been proposed to enhance ‘orientation’ with respect to the tonal centre (Balzano, 1980). As noted by Patel (2008, ch. 2), although symmetric scales do exist —e.g., in Javanese Gamelan music—, the predominance of asymmetric scales suggests that musical systems tend to favour an organization that promotes a sense of tonal orientation (Patel, 2008). In this sense, the clearly defined melodic and rhythmic cycles that are characteristic of Gamelan music could be taken to reflect a trade-off between the different dimensions. Along the same lines, another feature related to processing and learning constraints concerns the use of a limited number of categories per octave —typically scales are built using between 5 and 7 tones, regardless the differences in the number of discrete steps separating each octave (Miller, 1956). Other factors, such as short-term memory limitations on the length of melodic groupings that can be directly perceived as a unit, might translate into the use of phrases, which, in Western music, typically have a length of 4 or 8 bars (Snyder, 2000). Thus, it is possible to argue that many of the virtually universal features of music organization can likely be accounted for by general constraints imposed by our perceptual and cognitive endowment, without the need to resort to evolutionary processes involving the selection of domain-specific mechanisms.

Leaving aside for the time being the cognitive and interface constraints that intervene in shaping the features presented by musical idioms, it is now important to note the different levels at which music organises pitch and timing yielding a rule-governed system with a generative potential, which bears substantial resemblance with linguistic syntax. Not only in the neural resources implied in imposing a hierarchical structure into these auditory sequences, as we shall see, but also at a ‘deeper’ formal level.

---

<sup>5</sup> Peter et al. (2008) suggest that the perceived similarity between pitches one octave apart is not restricted to musical stimuli. This equivalence seems also to be used by children when imitating spoken stimuli in their study.

### *Overlap in syntactic processing*

As explained in Patel (2003; 2008), neuropsychological research points at the existence of distinct, domain-specific representations for linguistic and musical syntax. That is to say, linguistic knowledge of words and their syntactic properties appears to recruit a series of representations which are different from those regarding chords and their harmonic relations, as shown by the fact that deficits in musical skills like tonality processing might occur which do not disrupt language processing abilities and vice versa (Peretz, 1993). However, this evidence seems at odds with neuroimaging studies of musical and linguistic processing showing overlap in the neural resources engaged during the activation and integration of these representations during syntactic processing, especially in Broca's area –traditionally considered the cerebral locus of syntax– and the frontal inferior circumvolution (cf. Maess et al. 2001; Koelsch & Siebel, 2005).

Scientific support for the notion that harmonic processing could involve brain operations of the sort that subserve linguistic syntax was first provided by an ERP study by Patel and colleagues (Patel, Gibson et al. 1998). In this study it was observed that placing out-of-key chords in a tonal sequence elicited a P600 component, a language-relevant ERP associated to grammatical and syntactic integration (which is elicited, for instance, by garden-path sentences; Osterhout & Holcomb, 1992). The fact that this peak can be elicited in “non-linguistic (but rule-governed) sequences”, as noted by Patel and colleagues, shows that it does not correspond to language-specific brain processes and –together with the imaging studies showing overlap– can be taken to imply that similar cognitive operations, even if drawing on domain-specific representations, are subserved by a common pool of limited neural resources in both domains. Patel (2003) labels this idea the “shared syntactic integration resource hypothesis” (SSIRH).

The SSIRH thus formulated (see Patel, 2008: 276-297, for a detailed account) yields specific hypotheses that can be tested empirically. Two relevant lines of evidence that have provided support for this view concern the interference in simultaneous linguistic and musical syntactic processing (Koelsch et al., 2005; Fedorenko et al. 2009), suggesting that structural integration actually demands neural resources from a shared pool; and behavioural studies of individuals

with agrammatic Broca's aphasia<sup>6</sup> which display difficulties in processing harmonic syntax (Patel, Iversen et al. 2008).

Although further research along these lines is required — which, ideally, should also take into account the implication of subcortical regions in these cognitive sequencing operations and in Broca's aphasia, this preliminary evidence from linguistic syntactic processing and musical harmonic processing suggests that, in both modalities, the process of bringing long-term, domain-specific knowledge into working memory is carried out by the same neural resources. This view also has implications for traditional considerations of agrammatism, since the available evidence would speak against the picture of a language-specific syntactic function of Broca's area (as the one fostered by Grodzinsky, e.g. 2000), suggesting that processing in this area is not limited to linguistic syntax.

It is also important to note that, as we shall see in more detail, the part played by this region in language cognition is starting to be redefined within the context of a distributed network in which cortical-striatal-cortical circuits are implicated in cognitive — an hence, linguistic, among other — sequencing operations (Lieberman, 2006). Attending to the proposal put forward by Balari and Lorenzo (2009) this network can be characterised as a natural system of computation, based in the distinction between (a) a reiterative 'sequencing engine', constituted by the basal ganglia and, (b) a working memory space, provided by the cortical component. Broca's area would thus be implied in these linguistic and musical sequencing operations, as part of the cortical circuit providing the memory space for the system.

#### *"Selective" impairment of musical pitch processing*

Turning now for a moment to cases of selective impairment, these would seem to favour — in the absence of other kind of data — the view that language and music are largely independent cognitive functions. In this regard, it is

---

<sup>6</sup> Even though as Lieberman (2006) remarks, damage to Broca's region alone is not sufficient for inducing permanent agrammatism, a condition that does not occur in the absence of subcortical damage.

important to note that an appropriate consideration of developmental processes (cf. Karmiloff & Karmiloff-Smith, 2002, for language) reveals that the innateness of a given component or the domain-specificity of a cognitive mechanism —let alone their genetic specification—cannot be inferred from the existence of acquired dissociations pointing at modularity (orthographic alexia is a good example) and, as far as it is known today, neither from congenital dissociations.

Perhaps the most instructive case in this regard concerns the discovery of a genetic basis to the deficit known as Selective Language Impairment (SLI), which hastened claims on FOXP2 as the ‘language’ or even the ‘grammar’ gene. Today it is known that perturbations in this gene lead to a broad spectrum of effects —relevant though not specific to language— that result from its nonstandard expression at a molecular and eventually morphological level; see Benítez Burraco (2009) for an overview. A similar insight comes from studies on congenital amusia, also known as ‘musical tone deafness’, a deficit that seems to have a genetic basis and that has been put forward as evidence for modularity of musical processing (Peretz & Coltheart, 2003) understood along Fodor’s (1983) terms.

As described by Peretz and Coltheart (2003), individuals with this condition “suffer from lifelong difficulties with music” and are unable to recognise familiar tunes on the basis of music alone or to discriminate out of key vs. in-key changes in a melody (Ayotte et al. 2002), while auditory processing abilities for language are spared. Subsequent research (Hyde & Peretz, 2004; Foxtan et al. 2004; Patel, Foxtan & Griffiths, 2005), however, has shown that the condition of congenital amusia is due to a sensory deficit that involves an elevated threshold for the detection of changes in pitch direction. The apparent music-specificity of the deficit would be explained by the fact that individual differences aside, speech processing skills would remain largely robust to this deficit, given that linguistically relevant intonation changes mostly involve coarser pitch movements (and other acoustic cues, such as intensity) that exceed this threshold. However, the consequences for music —where most of the melodic transitions involve smaller steps (cf. Huron, 2006)— would be dramatic, in the sense that the failure to detect contrasts between successive tones would render the acquisition of the pitch-class distinctions on which musical syntax builds unattainable. In Patel’s (2008: 392) words, “[d]ue to these elevated thresholds,



individuals would receive a degraded version of the ambient musical input, so that normal cognitive representations of normal pitch would not develop”.

The case of the general deficit underlying an apparently selective impairment such as tone deafness thus once again brings forward the need for caution in establishing direct assumptions between an observable behaviour and the nature of its biological underpinnings. Structural and developmental considerations in place, it is not surprising to find that deficits do not map onto particular functions but onto particular physical structures and on the kind of activities carried by these structures (Love, 2007). This highlights the need for sticking to the approach at hand when addressing the evolutionary study of complex cognitive capacities such as language or music, just as its potential in uncovering the relationship between the activity carried by mechanisms and its mapping onto particular traits.

### 2.3 Formal resources

Once the overlap in syntactic linguistic and musical processing has been pinpointed, it is time to address the relationship between these systems in terms of their formal machinery. Syntax has been at the centre of strong claims on the domain-specificity of linguistic and musical components (e.g. Fodor, 1983; Jackendoff & Lerdhal, 2006), and most notably, on the arguably human-specific character of these components and its evolutionary status in terms with its relationship to the Faculty of Language.

Based on the intuition that music as a rule-governed system could be characterised along the lines of the generativist approach to linguistic grammar, this task was undertaken during the eighties by Lerdhal and Jackendoff, yielding the Generative Theory of Tonal Music or GTTM (Lerdhal & Jackendoff, 1983) –the same enterprise had been influentially undertaken before by Leonard Bernstein (1976), though with limited success given his attempt at

establishing analogies at a predication level between music and language<sup>7</sup>. Interestingly, however, one of the main conclusions reached by Lerdhal and Jackendoff in their GTTM proposal was that musical grammar did not look much like generative grammar, in that the hierarchical structure that organises tones (vs. words) seemed to be quite different.

In this regard, one of the features that have been granted more importance as distinguishing the human Faculty of Language from the rest of animal communication systems is its capacity for generating recursive hierarchical structures. Hence, according to the conception of the Narrow Faculty of language (FLN) proposed by Hauser, Chomsky and Fitch (2002), it is the mechanism of recursion and the mapping to the interfaces that yield the human FL unique. Thus, in their view, recursion is to be distinguished from the rest of components belonging to the sensory-motor and conceptual-intentional interfaces, which might be shared by other domains and species, and which are encompassed in the Broad Faculty of Language or FLB.

It is important to remark that the capacity for generating recursive patterns is also shared by music where, for instance, a pattern can be embedded within a broader pattern with identical geometry (Lerdhal & Jackendoff, 1983:207). The presence of recursive structures in music has been presented against Hauser et al.'s proposal for FLN (e.g., Pinker & Jackendoff, 2005; Jackendoff & Lerdhal, 2006), however, as evidence favouring stronger specificist claims —e.g., on the 'narrowness' of the syntactic component of the music faculty (Jackendoff & Lerdhal, 2006: 25).

The latter notion has recently been put into question by Katz & Pesetsky (2009), for example, who argue that music and language can be shown to display an identical formal component —where musical harmonic structure is derived by applying Merge—, once Lerdahl and Jackendoff's GTTM is realigned in the light of modern generative linguistic theory. According to this proposal, all formal differences between language and music owe to differences in their fundamental building blocks, while both systems are identical in what regards

---

<sup>7</sup> In this respect, musical syntax might be better characterised as leading to the perception of tension and resolution patterns, which are devoted a major component in Lerdahl and Jackendoff's approach.

their combinatorial engine —a central syntactic component which combines elements by means of iterated, recursive Merge.

Katz and Pesetsky's "Identity Thesis" can thus be regarded as a formal account that would favour strongly the kind of resource-sharing framework proposed by Patel, suggesting a convergence between both capacities at an even deeper level —even though, in principle, Patel's resource-sharing framework would not require identical syntactic principles operating in both domains.

At the same time, the similarity between both capacities at the level of their formal resources suggests a link in terms of shared neurocomputational substrates that is consistent with the framework observed here.

As Lieberman (2006) notes, the basal ganglia sequencing engine can form a potentially infinite number of different sentences by reordering, recombining, and modifying a finite set of words —or pitch classes— using a finite set of syntactic "rules." Balari and Lorenzo (2009b) have remarked that the emergence of the degree of computational complexity implied by recursion would not require modifications as to the sequencing engine, but would be yielded by the extension of the working-memory space available to the system —which would allow for the access to more complex sequence patterns. Following Balari and Lorenzo's proposal —and contra adaptationist claims, this quantitative and qualitative change would have resulted from general processes of brain growth and organization and as such, not from an evolutionary event directly related to language —or music—, though crucial for the emergence of these and other complex cognitive capacities.

Bearing this in mind, the main differences between language and music may well be just a matter of the nature of the components interfacing to a common or shared sequencing engine, with the Conceptual-Intentional interface being perhaps one of the distinguishing features of human linguistic capacities (Chomsky, 2004).

## 2.4. The case of musical rhythm

The studies reviewed so far reveal a good deal of overlap between music and language, especially when we focus on the nature of the mechanisms that subserve both abilities. From an evolutionary standpoint, this substantial degree of convergence between both capacities suggests a tight link that, if we were to follow a non-adaptationist line of reasoning, could be taken to point at the “parasitic” or “free-rider” character of music in relation to its more advantageous communicative counterpart –or, to put it in more neutral terms, as support for the more parsimonious, null hypothesis that music was not shaped by natural selection and as such, it cannot be considered an evolutionary adaptation.

In this subsection, then, the focus is on a particular aspect of music cognition that has been highlighted as a candidate that could challenge the non-adaptationist hypothesis for the origins of music (Bispham, 2006; Patel, 2006), given its apparent music-specificity<sup>8</sup>. This is beat-based rhythmic processing, which yields the capacity for motor synchronization to a musical beat (i.e., Beat-based Processing and Synchronization, or BPS; Patel, 2006).

As we shall see, there are reasons to argue that musical rhythmic processing involves cognitive mechanisms that are distinct from those that would play a part in linguistic rhythm, which would favour the claim that the former is not an off-shoot of the latter.

However, here I will suggest that a link might exist at the computational level. A claim that should nevertheless not be taken as a statement in favour of the thesis that music is a by-product of mechanisms that evolved for language, but instead as an argument that the whole notion of domain-specificity must be reconsidered in the light of the versatility of different brain structures and functions. While in this subsection I will not deal with comparative evidence from non-human animals, this picture should become clearer by the end of section 3, once this kind of comparative data is incorporated.

---

<sup>8</sup> As previously noted, other components, most notably, Tonality Processing (Peretz & Coltheart, 2003; Bispham, 2009) have also been presented along the same lines.

### *Entrainment to a musical beat*

Synchronization with music seems to be a universal activity, so that some form of music with an underlying periodic pulse that provides a basis for synchronised performance and movement on the part of listeners can be found in every human culture (Nettl, 2000). Indeed, in the face of cultural variability, this seems to be one of the deeper-rooted aspects of musical behaviour, if we take into account that some languages do not have a term that refers to musical practice —understood as the Western conception of sound alone— without encompassing also dance (Mithen, 2005).

Crucial to this kind of sensorimotor entrainment is the ability to sense a beat (a regular isochronous pulse or, more technically, the *tactus*, Lerdahl & Jackendoff, 1983) in an auditory signal. The process of activation of this pulse, which affords temporal coordination in, for example, dance or ensemble performance, takes place spontaneously, as long as the auditory stimulus meets some really minimal conditions<sup>9</sup>, and it is a skill that arises without instruction.

The fact that language displays a rich rhythmic structure and that, as noted previously, sensitivity to rhythmic cues in language is manifest early in infancy —a sensitivity that, as we shall see, is also shared by other mammals— might lead us to believe that musical and linguistic rhythm are analogous phenomena that build upon the same perceptual or cognitive skills. It is thus convenient to remark the fundamental differences that in spite of the rich rhythmic structure displayed by language make the processing of musical rhythm interesting from a cognitive standpoint:

As noted by Patel (2006; 2008: ch.3), the key element that sets apart the rhythmic properties of language and music is the role played by temporal periodicity in the latter. Thus, although both domains converge in the use of grouping structure, exhibiting a tendency to organise elements into larger units in terms of hierarchical prominence, they differ in that ‘stresses’ in speech do not mark

---

<sup>9</sup> Although the perception of an underlying musical pulse is normally associated to complex auditory stimuli in which a number of cues (as intensity or harmony) are implied in conveying the temporal structure, a beat can also be readily perceived in much simpler stimulus —e.g. rhythmic sequences of clicks or tones of equal intensity—, even if an isochronous pattern is not explicitly present, as in strongly syncopated rhythms. The presence of integer ratios seems, however, to be a necessary condition for perceiving periodicity in a rhythmic pattern (cf. Grahn & Brett, 2004).

out a temporally periodic pulse, i.e., a beat. The induced beat, which Bispham (2006) describes as an “internally generated and/or externally guided attentional pulse”, engages a series of multilayered temporal expectancies which play a basic role in organising both musical perception (cf. Huron, 2006) and production<sup>10</sup> (Palmer & Pfordresher, 2003). These levels of temporal organisation are also implied in determining the relative importance of notes in the harmonic and melodic structure —i.e., in the syntactic component. Beat perception is, moreover, robust to tempo fluctuations, which suggests that it is based on flexible timekeeping mechanisms (Patel, 2008).

This key component of music cognition does not seem to play a part in speech where, as noted by Zatorre et al. (2007), “apart from certain highly elaborated speech forms, such as poetry, there is no ‘beat’ to tap to”. Ordinary speech does not, therefore, generate the kind of temporarily-based attentional framework which is characteristic of music. Instead, we can say that the aforementioned sensitivity to linguistic rhythm concerns the rhythmic cues conveyed by overall frequency contours and by the durations of particular phonemic clusters<sup>11</sup> (Hauser & McDermott, 2003).

There is some inconsistency as to how this cognitive skill is termed in the literature so that, depending on the author, it is alternatively referred to as ‘Beat Induction (BI)’ (e.g. Desain & Honing, 1999), ‘musical pulse’ (Bispham, 2006) or, ‘beat-based rhythm processing’, ‘beat-based processing’ and just ‘beat perception’ (e.g. Patel, 2006; Grahn & Brett, 2009). In this work I will stick to ‘beat-based processing’ in order to differentiate it from ‘beat-based rhythmic processing and synchronization’ (BPS), which Patel uses to refer to the ability for sensory and motor entrainment.

---

<sup>10</sup> Purwins et al. (2008) suggest that the beat can be thought of as a temporal grid that provides a context in which the perceived events take place. At a higher level of organisation, the perceptual saliency of beats in relation to each other gives rise to a metrical structure, which can be thought of as a hierarchical grid of beats. Evidence that musical sequences are planned and executed in terms of metrical structure by musicians (Palmer & Pfordresher, 2003) echoes London (2006) paraphrase: “meter is how you count time, and rhythm is what you count—or what you play while you are counting”.

<sup>11</sup> It is in this respect that overlap between language and music can be found, so that perhaps not very surprisingly, music from a particular culture has been shown to reflect or mimic the rhythmic characteristics of its language at the level, for instance, of average durational contrasts, as shown in Patel & Daniele (2003).

This differentiation allows for considering the possibility that beat-based processing might be in place despite impairment or a lack of accuracy in motor control (which may take longer to develop) required in movement synchronization. This means that the sort of synchronization tasks used to test this ability, which traditionally involved reproducing or tapping along with rhythmic sequences, may often prove insufficient to assess perceptual and processing skills. Hence, infant data showing that the ability for motor synchronization manifests relatively late in development (Eerola et al. 2006) are not informative as to the age onset of beat-based processing capacities.

Indeed, recent evidence from neuroimaging (Winkler et al., 2009) suggests that neonates already seem to engage in the temporal expectations generated by beat-based perception. Winkler and colleagues carried an ERP experiment in which sleeping neonates listened to a sound sequence –based on a typical rock drum accompaniment pattern– where infrequent omissions of sounds in different metrical positions were introduced. In this experiment, the mismatch negativity response (MMN, associated to deviations from expectations) was only elicited when the omission corresponded with the ‘downbeat’, this is, the perceptually most salient position where a beat onset was expected. The fact that the rest of deviations from the standard pattern did not elicit the response (i.e., the omission of the downbeat was not perceived as a mere deviation from the standard pattern) suggests that the brain engages in this sort of timing-sensitive expectancies from birth.

It is worth noting that the results of this experiment bring in the question of whether these early beat perception capabilities belong to the kind of general auditory processing mechanisms that we share for instance, with primates –as we will see in the next section– or, on the contrary, they require a more complex network that integrates also temporal processing and coordination, the details of which will be discussed below.

Likewise, and once the distinct character of musical versus linguistic rhythm has been clarified, it is convenient to remark that the kind of sensorimotor entrainment that concerns us here entails a level of processing that seems to be more complex than that involved in the more general ability for calculating individual temporal intervals (Grahn & Brett, 2007; Patel, 2008). The

construction of the kind of temporal representations involved in beat-based processing requires first the ability to extract the relevant temporal information from a complex auditory stimulus. Then, these temporal schemata must be maintained over time, enabling the planning and execution of synchronised movement. Hence, the cognitive demands on this task –the induction and/or self-generation of this mental framework and its recurring implementation— can be taken to differ non-trivially from the generic ability for gauging individual time intervals. To put it differently, we can say that at least intuitively, the capacity for building periodical expectancies seems to require a degree of computational sophistication different from the ability to construct generic temporal expectancies<sup>12</sup>. This intuition would accord well with the fact that the generic ability for gauging an individual interval is widespread in other species while BPS is a rather restricted phenomenon, and also with data regarding the neuroanatomical substrates for the capacity at hand, as we shall see.

#### *Neural substrates for BPS —a brief sketch*

Given the apparently exceptional character of BPS and the claims put forward on its musical-specificity, it would be sensible to expect the mechanisms that support this ability at a neural level to be similarly singular. Patel (2006, 2008) provides a very interesting proposal in this regard, which links the capacity for beat-based processing and synchronization to the neural circuitry implied in vocal learning.

This proposal, which he labels ‘The Vocal Learning Hypothesis’, partly builds on the observation that BPS seems to bear a special relation with the auditory modality. Visual rhythmic sequences do not seem to induce the kind of structured temporal representations that arise when the same sequences are presented auditorily (Patel, Iversen, et al. 2005) and, even when they consist on a train of isochronous visual patterns, difficulties in synchronization arise at

---

<sup>12</sup> Indeed, the computational modeling of this ability reveals itself a complex task that has been an area of substantial research (cf. Longuet-Higgins & Lee, 1982 or for an overview, Desain & Honing, 1999).



sequence rates —or *tempi*— which can be easily dealt with for auditory stimuli<sup>13</sup> (Repp, 2003). This difference in performance might be related to an advantage of the auditory system in temporal perception, which is reflected by the dominance of this modality when conflicting temporal information is received by the auditory and visual systems<sup>14</sup> (cf. Repp & Penel, 2002).

As noted by Patel, motor entrainment to a beat imposes a special relation between the auditory channel and patterned movement, very much resembling that involved in vocal learning. In anatomical terms, this tight coupling between auditory input and motor output suggests a pathway between the basal ganglia, which subserve motor and timing functions in a wide range of species, and the auditory system. It is this kind of evolutionary ‘modifications’ in terms of brain circuitry that, according to the author, might provide the neural foundations for BPS (it is important to remember that the capacity for complex vocal learning is a relatively rare trait from an evolutionary standpoint, which is not shared by other primates; Egnor & Hauser, 2004). In other words, it is possible that as suggested by Patel, the ‘online integration’ of the auditory and motor systems that affords matching vocal production to a desired model allows also for synchronised movement with a musical beat<sup>15</sup>.

This hypothesis, furthermore, yields the prediction that the capacity for synchronization with an external auditory stimulus is not an exclusively human trait, for such a skill might also be implicitly in other vocal learners —and interestingly, as we will see in more detail, his hypothesis does not seem to be misguided.

The point here, however, is to qualify Patel’s proposal by showing that it can be integrated into the framework of the present work once more attention is devoted to the role of the basal ganglia as a sequencer of motor and cognitive

---

<sup>13</sup> According to Repp’s (2003) results, the synchronization threshold is four times higher for visual than for auditory stimuli.

<sup>14</sup> In any case, recent research supporting this dominance (McAuley & Henry, 2010) seems, however, to counter prior claims on the obligatory and automatic auditory encoding (‘hearing’) of visual rhythms (Guttman et al. 2005).

<sup>15</sup> It must be noted that both in vocal imitation (or learning) and in BPS, sensory feedback plays a central role in the real-time adaptation of our performance. Studies on deaf infants, for instance, reveal that auditory feedback is needed to lead to coordination of phonatory and articulatory system and thus, for the development of normal speech production (cf. Koopmans-van Beinum et al. 2001).

patterns, approaching the relation between BPS and vocal learning from a structural point of view.

*Beyond (or below) the Vocal Learning Hypothesis*

Patel's hypothesis bears on neurobiological data showing that the basal ganglia not only play a basic role in rhythm perception and production, but they are also involved in the kind of modifications associated to the nervous systems of vocal learners across-species (cf. Jarvis, 2004).

Regarding the implication of this deep brain structure in beat perception and motor control, Patel refers to a neuroimaging study by Grahm & Brett (2007) in which activity in different areas of the brain was compared when subjects listened to rhythmic sequences structured so that an underlying beat could be easily induced or not. As remarked by the authors of the study, an increase of activity in the basal ganglia was elicited only by sequences that induced an isochronous pulse, which is suggestive of the role of this structure in beat-based processing.

At this point it is also important to note that according to the broader proposal developed here, the basal ganglia play a critical role in cognitive sequencing operations, which, again, might be of many types. Patel (2008) does hint at this relation: "Importantly, the basal ganglia are also involved in motor control and sequencing (cf. Janata & Grafton, 2003), meaning that a brain structure involved in perceptually 'keeping the beat' is also involved in the coordination of patterned movement".

Thus, a more careful consideration of the structures implied in beat-based processing and synchronization allows us to relate them to the neurocomputational substrate underlying complex sequencing operations of the kind that may also provide for the "reiterative" quality of linguistic and musical syntax.

Various studies (Rao et al. 2001; Janata & Grafton, 2003; Grahm & Brett, 2007; Zatorre et al., 2007; Rao et al. 2001) have implied motor regions of the brain

both in the production and perception of rhythm, showing activation also in passive listening tasks. In particular, as noted by Grahn & Brett, (2007) the timing system seems to be mediated by a set of neural structures connecting the basal ganglia and motor areas via a striato-thalamo-cortical loop.

Activity in the basal ganglia, moreover, increases during the processing of rhythms that require to a greater extent an internal generation of the beat (e.g., strongly syncopated rhythms, in contrast to those where the beat is strongly conveyed by acoustic cues), which highlights the part played by this subcortical structure in the generation of an internally guided regular pulse (Grahn, 2009). The implication of the basal ganglia in the generation of these temporal expectancies seems consistent with the notion that this structure functions as a sequencing engine which releases and inhibits pattern generators (Lieberman, 2006).

This subcortical structure supports circuits that project to cortical areas, linking the basal ganglia to the working-memory space and the interfaces. As previously noted, and bearing in mind that the basal ganglia are a highly conservative structure in evolutionary terms, it is at the latter level that we can expect to find inter-specific differences, related to connectivity to the interfaces and to the amount of working-memory space available to the system (Balari & Lorenzo, 2009b). This view captures the intuition that beat-based processing requires a greater working-memory capacity than the more general ability to calculate a time interval —e.g., as in catching a ball—, which must allow for the reiteration of timing-based patterns. At the same time, it also brings forward the importance of the cortical-striatal-cortical network associated to auditorymotor interactions that, as noted by Patel, is characteristic of vocal learners in other species.

#### *Beat-based processing and other cognitive deficits*

Data from studies on different deficits seem to provide support for the view that rather than a domain-specific adaptation for music or an off-shoot of vocal learning, beat-based processing can be identified along with other basic cognitive skills involving complex cognitive sequencing operations in terms of the neural structures involved.

For instance, studies on Parkinson's Disease show that individuals affected with this deficit present, along with other cognitive sequencing problems, syntactic comprehension deficits (Lieberman, 2006: 182-185). Parkinson Disease's patients also display poor performance in rhythm discrimination tasks (involving no motor production), according to a study by Grahn and Brett (2009). Importantly, a significant difference in performance was found only for beat-based rhythms —i.e., their performance did not differ significantly from controls in the rhythmic sequences that did not induce a periodic pulse<sup>16</sup>, which suggests impairment at encoding the rhythmic sequences in terms of beat structure.

The 'linguistic' impairment in members of the KE family correlates also with a deficit in tasks involving discrimination and reproduction of rhythmic patterns<sup>17</sup> (Alcock et al., 2000), consistent with the problems in complex temporal sequencing reflected in to oral movements. However, although the poor performance in the rhythmic tasks reveals impairment at encoding the relationship between time intervals, which is generally facilitated by beat-based processing, this skill (i.e., the extraction of an underlying pulse) was not explicitly addressed in Alcock et al.'s experiment.

The relationship between the genetic deficit in the KE family and rhythmic performance brings forward the relevance of the neural structures related to FOXP2 expression in fine-grained sequencing and timing.

Neurobiological evidence from birds, as shall be noted in the next section, shows that the pattern of expression of this gene in avian species differs for those that learn vs. do not learn their song, linking its expression to modifications in the basal ganglia that play a key role in mediating the connection between auditory perception and motor production during learning (Gale & Perkel, 2005).

---

<sup>16</sup> In rhythmic sequencing tasks, behavioural measures improve when a regular beat can be perceived. As noted by Grahn and Brett (2009), detection of a timing structure allows for encoding the temporal intervals according to the beat, instead of as a sequence of unrelated time durations.

<sup>17</sup> Performance in pitch/intonation discrimination and reproduction tasks, however, does not seem to be impaired (Alcock 2000).

This ultimately raises a broader question as to the part played by the circuits associated to sensory-motor functions and motor-skill learning in providing for the reiterative quality that is characteristic of human cognitive skills.

### **3. Precursors in non-human animals**

The section above shows that from a structural perspective, language and music appear to share very important aspects that go from the mechanisms playing a role in the development of both capacities to the kind of computational sequencing operations that characterise the production of both musical and linguistic representations. In this section we will continue to focus on the mechanisms that subserve these capacities by inquiring in the extent to which they might be shared by other species.

Given the centrality that is granted to the human linguistic capacity in the study of cognition, language and its components have received more attention than music in research concerning the precursors for these faculties. The comparative approach to the evolutionary origins of language advocated by Hauser, Chomsky and Fitch (2002), which places an emphasis on the possibly shared nature of a number of components within what they term the Faculty of Language in the broad sense (FLB) has favoured an increase in the amount of work aimed at providing insight into the “biology of music” and ultimately, at clarifying the evolutionary status of the cognitive underpinnings for human faculty of music; see, for example, Hauser & McDermott, 2003 or Justus & Hustler, 2005 in very much the same spirit of HCF.

However, even though it is generally acknowledged that the intimate link between music and language is suggestive of some sort of evolutionary bonding, it is rare to find comparative studies that do not make a clear-cut distinction in the ‘original’ role of certain mechanisms, based on assumptions on the functional continuity of both capacities. Similarly, functional considerations tend to prevail when pondering the fitness of comparison between certain animal traits and the human cognitive mechanisms under study.

A good example concerns the debate on the traditional analogy between animal song displays and human music. Hauser and McDermott (2003: 667; 2005) reject both homology –since none of the other great apes sing– and analogy for animal song and music, on the basis that animal song is predominantly male and produced in “extremely limited” behavioural contexts, having a solely communicative function, whereas music is “characteristically produced for pure enjoyment”. The arguments against analogy are countered by Fitch (2006: 184-185), who regards studies on animal song as a source for potential insights into general and perceptual constraints on the evolution of complex signalling systems. It must be noted, however, that in subsequent work McDermott and Hauser (2005: 39) acknowledge the parallels between these “communication signals” with human music on a structural level –non-trivially, the generation of songs by rule-based systems and innate constraints on sequencing. However, they continue to regard as unlikely the possibility that any of the resemblances between both behaviours are due to a homology, for the reasons mentioned above.

At this point we must remember that these and other questions of adaptive function remain largely orthogonal to the structural approach fostered here, and that we are looking for the precursors of these cognitive faculties in minds that are by definition non-musical and non-linguistic in the human sense of the terms. One of the premises of the framework at hand is that all components of the ‘broad’ faculties under study (including ‘narrow’, if any) might well subserve behaviours having little to do the function granted to human linguistic and musical behaviours as we construe them. In this sense, as argued in Balari and Lorenzo (2009a) it is possible to suppose that homologies might exist among organic structures that carry very different functions, but which nevertheless display the same ‘functioning’.

### 3.1. Perceptual constraints

#### *Auditory processing*

A well-known instance of the insights provided by comparative research concerns the previously mentioned series of experiments on the categorical perception of speech by infants, which were rapidly interpreted as evidence for a language-specific learning mechanism (Liberman et al., 1967). The fact that a perceptual ability that seemed so appropriately tailored to the particularities of language was later proven to be present in primates, chinchillas and birds (Kuhl & Miller, 1975; Kuhl & Padden, 1982; Kojima & Kiritani 1989; Kluender et al. 1987) shows that the mechanism underlying these perceptual discontinuities responds instead to features of the vertebrate auditory system.

This case can be taken as exemplifying a tendency as we progress in the evolutionary study of complex cognitive capacities, where former assumptions on the, often taken for granted, specificity of particular mechanisms turn out to be contested by more recent comparative data<sup>18</sup>. Indeed, comparative evidence suggests that both language and music seem to be constrained by sensitivities of our perceptual system to a great extent.

For instance, as noted in the previous section, the privileged status of the octave in musical idioms could be explained along a perceptual basis, as suggested by the fact that rhesus monkeys have been shown to generalise along transpositions by this (vs. other) particular interval (Wright et al. 2002). The notion that the prevalence of the octave interval might have a biological basis, however, is not incompatible with other evidence illustrating the ubiquity of absolute (vs. relative) pitch encoding in most species<sup>19</sup>, which is mirrored by human infants. Rather, it brings forward that relative pitch perception depends to some extent on the formation of a representational framework that facilitates the encoding of pitch in relational terms. Curiously enough, as McDermott and Hauser (2003) point out, macaques in the experiment by Wright and colleagues

---

<sup>18</sup> Another example is that of the Perceptual Magnet effect, which P. Kuhl hypothesized to be uniquely human, but more recent studies have shown it to be present in macaque monkeys (Kuhl 1991) and some avian species (Kluender et al 1998); see Fitch et al. 2005 for discussion.

<sup>19</sup> Although results from a study with starlings suggest that these birds can switch from relying on absolute to relative pitch strategies in adapting to the demands of certain tasks (MacDougall-Shackleton & Hulse, 1996).

showed octave generalization only for tonal melodies but not for the atonal ones, something that raises the question of whether the primates could extract some key structure from exposure, or else tonal melodies are “naturally” easier to encode. Both options ultimately relate to the constraints shaping musical systems and thus would deserve further investigation.

Continuing with the biological foundations of certain harmonic features, neural research by Tramo et al. (2001) shows that the different acoustic properties of consonant and dissonant intervals correlate with distinct patterns of activation in auditory nerve fibres in humans<sup>20</sup>, consistent with research on primates and humans (Fishman et al. 2001) showing the same effect. Although behavioural research (McDermott & Hauser, 2004) has highlighted that monkeys do not display a preference for consonant intervals<sup>21</sup>, as 2-month-old human infants seem to do (Trainor et al., 2002), there should be no reason to assume that this preference in humans reflects some sort of adaptation for music rather than, for example, reflecting an acquired association to the affective cues in infant-directed speech (cf. Thiessen, Hill & Saffran, 2004).

Replication of other experiments in primates provides a hint that more than just a perceptual basis might be shared with our ancestors. As we saw before, by 8 months of age infants are capable of computing transitional probabilities from an auditory stream such as speech, an ability that is not restricted to speech sounds but that also applies to pitch tones and visual sequences. Using the synthetic speech stimuli from the human infant study (Saffran, Aslin & Newport, 1996), Hauser, Newport and Aslin (2001) reproduced the same experiment with adult cotton-top tamarins in order to check whether this ability is shared by these primates<sup>22</sup>. Tamarin monkeys, as noted by the authors, use sequential calls as a means of interspecific communication. The results of the experiment parallel those obtained in the original experiment, which shows that like human infants, these primates can spontaneously keep track of statistical regularities in a relatively fast and complex stream of sounds<sup>23</sup>. This finding

---

<sup>20</sup> Tramo et al. (2001) report a correlation between tonal dissonance of musical intervals and the total number of auditory nerve fibres that show beating patterns.

<sup>21</sup> Though see Sugimoto et al. (2010) for recent counter-evidence in an infant chimpanzee.

<sup>22</sup> Tamarin monkeys, as noted by Hauser, Newport and Aslin (2001) use conspecific calls which display some sequential structure, which derives from the combination of two basic elements.

<sup>23</sup> The synthetic languages used in the original experiment by Saffran et al. (1996) consisted in 12 distinct syllables and 20 different transitional probabilities (or 20 distinct syllable pairs). Three-



entails that certain computational ability —distinct from the overlap in perceptual sensitivities— that allows for processing and retaining these aspects of serial order information is common to both species. Interestingly, available evidence from rats shows that these rodents are also able of segmenting this kind of speech streams (Toro & Trobalón, 2005) and thus of learning statistical relations between adjacent elements. However, these rats’ ability for tracking distributional regularities seems to be more constrained than that of tamarins in that the former did not succeed in discriminating sequences involving dependencies between non-adjacent elements as tamarins appear to be able to do (Newport et al. 2004).

### 3.2. Computation

Thus, as noted by the authors of the previous experiment, the results derived from comparative studies on these spontaneous processing capacities suggest that some basic statistical learning mechanism generalised over nonprimate species (Toro & Trobalón, 2005). We must note, however, that in spite of its usefulness in the task of segmenting an auditory stream as speech, the capacity for extracting this kind of serial information is still far from rendering a system like human language learnable. Therefore, the perceptual similarities that we share with other primates do not extend to the computational domain.

The combinatorial nature of human language displays, as discussed above, a generative power located within a higher level of complexity (Context-Sensitive Grammars), that requires a greater capacity in terms of working memory in order to carry out operations that involve the extraction of regularities at levels ‘higher’ or beyond element adjacency —beyond, to put it in terms of Chomsky’s hierarchy, Finite State Grammars. It is at this point that we find the divergence between the computational abilities of human and non-human primates.

In this regard, Fitch and Hauser (2004) showed that tamarin monkeys display no difficulties in discriminating sequences made up of syllables within the

---

syllable words were to be distinguished from part- or non-words on the basis of adjacent co-occurrence, i.e. transitional probabilities. Learning was tested with an orientation response; for details, see Saffran, Aslin & Newport, 1996.

range of a regular grammar, but they are unable of discriminating more complex sequences within more complex context-free grammars. The fact that this computational capacity is absent in primates, as our closer ancestors, has been taken to imply that it is this computational ability that cognitively singles us out as species, singling language out at the same time as the apparently obvious target of natural evolution. Indeed, in the proposal put forward by Hauser, Chomsky and Fitch (2002), the capacity for recursion is isolated as the computational core of the human faculty of language, thus drawing a distinction between the Faculty of Language in a Narrow sense (FLN)—essentially, the recursive engine— and the Faculty of Language in a Broad sense (FLB), which encompasses FLN together with the rest of components belonging to the conceptual-intentional and sensory-motor interfaces.

*(So-called) human-specificity: coming to grips with sequencing capacities*

In a parallel line of reasoning to that of the human-specificity of the recursion component, lack of evidence for BPS in our closer relatives has been taken to favour the claim that it is a uniquely human capacity (e.g. Bispham 2006), which could have constituted an evolutionary adaptation for music.

If as suggested above (section 2.4.), the cognitive ability for extracting an isochronous temporal pattern and synchronising to it (BPS) shares the same neurocomputational substrate that supports the reiteration of complex sequencing patterns, we would expect primates to show no indices of this capacity, and this seems to be the case. Indeed, although for example, Zarco et al. (2009) record that the temporal performance of rhesus macaques was equivalent to that of human subjects in a task involving the production of single intervals, macaques succeeded in the task of producing multiple intervals only after months of intensive training, displaying more variability and less overall accuracy<sup>24</sup>. Moreover, while humans synchronised their performance to the metronome displaying a tendency to tap slightly 'ahead' of the beat —what indicates self-pacing, the time asynchronies for rhesus were positive, corresponding to taps after the stimulus onset. The study by Zarco et al. also

---

<sup>24</sup> It is important to remind here that, as emphasized by Patel et al. (2009), synchronization to pulse trains does not involve the extraction of a regular beat from a complex auditory signal.

reveals that unlike humans, rhesus macaques did not show an advantage in the auditory vs. visual condition of the experiment, what once again brings forward the computational advantage implied by the neural structures supporting auditory-motor interactions in this kind of structured temporal processing tasks.

These results also seem to provide support for the neurocomputational model observed here in terms of the part that is granted to the basal ganglia and cortical structures. The ‘universal’ and ‘amodal’ nature of the network (including basal ganglia, thalamus and areas of the cortex) that subserves the capacity for gauging more general temporal expectancies —as single intervals— is probably shared among vertebrates<sup>25</sup> (Matell & Meck, 2000). Differences between humans and primates thus can be taken to correspond to the circuitry connecting the basal ganglia to the interfaces (conferring an advantage to the auditory modality) and to the working-memory available for processing, which make possible the ‘rehearsal’ of this timing-based information.

At this stage it is also convenient to distinguish the capacity for beat-based processing and synchronization from superficially similar ‘spontaneous’ behaviours in other species, such as those involving the production of periodic signals in frogs or crickets. At the level of cognitive implications, again, it is important to note that the fact that episodes of synchrony emerge in these choruses does not entail a form of beat-based processing, but rather synchrony results from phase adjustment mechanisms (Bispham, 2006; Patel, 2008).

Therefore, returning to the point at hand, it might be the case that, as suggested by Patel, we have to look in different groups in order to find evidence for this sensorimotor coupling, and for other forms of evidence regarding the use of complex sequencing skills that the corresponding neurocomputational substrate may afford.

As explained above, McDermott and Hauser (2005) reject the notion of homology or homoplasy for human music and birdsong due the discontinuity of this trait, as a singing behaviour comparable to that of birds is not present in the primate lineage and thus would not have been ‘passed on’ to our species.

---

<sup>25</sup> Patel (2008) notes that rabbits can also be trained to gauge the duration of short time intervals.

This claim may be rebutted if we stick to a classical notion of homology, that is a strictly structural resemblance relation between unequals. In this context it may then be possible to postulate a homology at the morphological level between the structures that subserve the use of motor and “melodic” complex sequencing patterns in these different species. An idea also shared by Jarvis (2004), who argues that vocal learning might have evolved “independently among birds and humans, [...] under strong genetic constraints of a pre-existing basic neural network of the vertebrate brain.”

In the same vein, Patel’s Vocal Learning Hypothesis might therefore turn out to be correct, but for deeper reasons: the difference would be basically at the level of interpretation, as this circuitry wouldn’t necessarily have been selected for language and, thus, it wouldn’t follow that human entrainment evolved as a by-product of vocal learning —as claimed in Schachner et al. 2009— or of anything else, for that matter. Hence, a capacity that is generally regarded as a functional phenotype serving an adaptive role —i.e, vocal learning— might provide insight as to its relation with other cognitive skills once it is considered in terms of its underlying morphological or computational phenotype.

In this case, the kind of modifications associated to the expression of *FoxP2* in vocal learners seem to provide the basis for a neural circuit that allows for the access to relatively complex motor-melodic and cognitive sequencing patterns. This consideration affords for a unified approach to the study of seemingly unrelated aspects of animal behaviour which, nevertheless, provide insight as to the evolutionary status of the mechanisms that underlie these human linguistic and musical capacities.

Actually, deep parallels seem to arise at developmental, mechanistic and formal levels between birdsong, speech and music (Fitch, 2006). The famous ‘sensitive’ or ‘critical’ period for linguistic acquisition runs also for songbirds, where exposure to conspecific song is required early in life in order to develop normal singing behaviour (Marler, 1987). Similarly, vocal production in these species is characterised by an immature state —equivalent to babbling— known as subsong, where sensorimotor feedback plays a key role in matching production to a template, and which seems to be essential for the development of normal singing performance in some species (cf. Marler & Slabbekoorn, 2004).

Indeed, Patel's hint at the relationship between the capacity for beat-based processing and vocal learning is favoured by evidence showing that avian learners can move in synchrony with musical stimuli. Patel et al. (2009) and Schachner et al. (2009) report instances of entrainment to music in a number of vocal mimicking species, whereas no evidence for this behaviour was found for non-mimicking species. As a related form of evidence, pigeons (which do not learn their song), display little ability to perceive grouping structure and seem unable to learn discriminations between rhythmic and arrhythmic patterns of sounds –although, not without difficulty, these animals learned to discriminate among two instances of musical metres (Hagman & Cook, 2010).

Concerning a different species of avian learners, the human-specificity of the computational capacity for recursion has been put into question in a study by Gentner et al. (2006) showing that starlings can be trained to recognise center-embedding structures –which correspond to a range of computational complexity close to the one attributed to human language. However, whether these sequences were parsed by starlings using a recursive procedure is still under debate (cf. Perruchet & Rey, 2005; ten Cate et al., 2010).

At a formal level, the combinatorial characteristics of birdsong might be more straightforwardly compared to human phonology, as noted in Samuels et al. (2010). The generation of a song (a process to which Marler, 2000, refers as 'phonocoding') involves the recombination of learned segments into more complex sequences, which differ in terms of the notes selected and their arrangement. Birdsong, and also whalesong (Suzuki et al., 2006), display a multi-level organization described as "linear hierarchy" which appears to be rule-governed in some species (Marler, 1984).

This would lead us again to the role of *FOXP2* and the different patterns of expression in avian learners vs. non-learners, providing insight as to the kind of morphological and genetic constraints that may participate in the development of the putatively shared structures in human and non-human species.

Recall that, in Lieberman's model, the basal ganglia constitute the sequencing engine of a complex cortico-striato-cortical circuit and that, as we have extensively discussed above, they appear to participate in other processing

tasks different from language all requiring the sequencing of cognitive patterns, notably those which form the building blocks of our musical abilities. At this point, and given recent findings concerning the fact that avian *FoxP2* is also expressed in the basal ganglia, both during development and during song production (Benítez Burraco, 2009; Rochefort et al. 2007), an exciting area for further research opens up, pointing at the existence of far deeper homologies between language, music, birdsong, and perhaps other non-human abilities.

#### **4. Concluding remarks**

The evolutionary study of language and music has traditionally been addressed within the context of a selectionist framework, in which the emphasis placed on the functional uniqueness of mechanisms often obscures parallelisms at the level of organic structure, on which evolutionary accounts should be grounded.

This research proposal has thus aimed at showing that substantial insight into the biological foundations of language and music can be gained by tackling the comparison between both capacities from a structural standpoint. As has been argued, this criterion yields a promising approach that actually allows for tracing the evolutionary history of the organic structures that support these human capacities, paving the ground for an account of their emergence in terms of common mechanisms at cognitive, neural and ultimately, genetic levels.

At same time, this structural position makes it possible to isolate the points of contact and divergence between both cognitive capacities, helping us to understand how they differ. In this respect, it must be noted that this paper has focused mainly on the common substrates for musical and linguistic capacities, and that important differences between these phenomena have not been addressed —basically involving the conceptual-symbolic interface—, which should be devoted further attention.

Finally, as this work has tried to put forward, the adoption of a structural perspective where data from anatomy, genetics and developmental studies can be integrated makes it possible to establish deep homology relationships among different species, which can shed light on the origins of human and musical

capacities, just as on the different processes that constrain natural evolution. As remarked by Fitch (2006: 206), “while studying the biological basis of music and language simultaneously may seem daunting, comparisons should ultimately result in more parsimonious models of human nature.”

## References

- Alcock, K. J., Passingham, R. E., Watkins, K., & Vargha-Khardem, F. 2000. Pitch and timing abilities in inherited speech and language impairment. *Brain and Language*, 75, 34-46.
- Allen, G. 1878. Note-deafness. *Mind*, 3, 157-167.
- Ayotte, J., Peretz, I., & Hyde. 2002. Congenital Amusia: A group study of adults afflicted with a music-specific disorder. *Brain*, 125, 238-251.
- Balari, S., & Lorenzo, G. 2009a. Comunicación. Donde la lingüística evolutiva se equivocó, *Report de recerca, Centre de Lingüística Teòrica/UAB*, CGT-09-10.
- Balari, S., & Lorenzo, G. 2009b. Computational phenotypes: Where the theory of computation meets Evo-Devo. *Biolinguistics*, 3(1), 2-61.
- Balzano, G. J. 1980. The group-theoretic description of 12-fold and microtonal pitch systems. *Computer Music Journal*, 4, 66-84.
- Benítez Burraco, A. 2009. *Genes y lenguaje. Aspectos ontogenéticos, filogenéticos y cognitivos*. Barcelona: Reverté.
- Bernstein, L. 1976. *The unanswered question*. Cambridge, MA: Harvard University Press.
- Bispham, J. C. 2006. Rhythm in music: What is it? Who has it? And why? *Music Perception*, 24, 125-134.
- Bispham, J. C. 2009. Music's "design features": musical motivation, musical pulse, musical pitch. *Musicae Scientiae*, special issue 2009-2010, 29-44.
- Brown, S. 2000. The "musilanguage" model of music evolution. In N. L. Wallin, B. Merker, and S. Brown (Eds.), *The Origins of Music* (271-300). Cambridge, MA: MIT Press.
- Chomsky, N. 2004. *Three Factors in Language Design*. Manuscript: MIT.

- Chomsky, N. 2005. Some simple evo-devo theses how true might they be for language? In *Symposium of the evolution of language*. State University of New York.
- Darwin, C. 1871. *The Descent of Man, and Selection in Relation to Sex*. London: John Murray.
- Desain, P., & Honing, H. 1999. Computational models of Beat Induction: the Rule-Based Approach. *Journal of New Music Research*, 28(1), 29-42.
- Eerola, T., Luck, G., & Toiviainen, P. 2006. An investigation of pre-schoolers' corporeal synchronization with music. In: M. Baroni, A. R. Addessi, R. Caterina, M. Costa (Eds.), *Proceedings of the 9th International Conference on Music Perception and Cognition* (472-476). Bologna, Italy.
- Egnor, S. E. R., & Hauser, M. D. 2004. A paradox in the evolution of primate vocal learning. *Trends in Neurosciences*, 27, 649-654.
- Fedorenko, E., Patel, A. D., Casasanto, D., Winawer, J., & Gibson, E. 2009. Structural integration in language and music: Evidence for a shared system. *Memory & Cognition*, 37, 1-9.
- Fernald, A. 1992. Meaningful melodies in mothers' speech to infants. In H. Papousek, U. Jurgens, & M. Papousek (Eds.), *Nonverbal Vocal Communication: Comparative and Developmental Aspects* (262-282). Cambridge, UK: Cambridge University Press.
- Fisher, S. E., & Marcus, G. F. 2006. The eloquent ape: genes, brains and the evolution of language. *Nature Reviews Genetics*, 7, 9-20.
- Fishman, Y., Volkov, I., Noh, M., Garell, P., & Bakken, H. 2001. Consonance and dissonance of musical chords: neural correlates in auditory cortex of monkeys and humans. *Journal of Neurophysiology*, 86, 2761-2788.
- Fitch, W. T. 2006. The biology and evolution of music: A comparative perspective. *Cognition*, 100, 173-215.
- Fitch, W. T., & Hauser, M. D. 2004. Computational constraints on syntactic processing in a nonhuman primate. *Science*, 303, 377-380.
- Fitch, W. T., Hauser, M. D., & Chomsky, N. 2005. The Evolution of the Language Faculty: Clarifications and Implications. *Cognition*, 97(2), 179-210.
- Fodor, J. A. 1983. *The Modularity of Mind*. Cambridge, MA: MIT Press.
- Foxton, J. M., Dean, J. L., Gee, R., Peretz, I., & Griffiths, T. D. 2004. Characterisation of deficits in pitch perception underlying "tone deafness". *Brain*, 127, 801-810.



- Gentner, T., Fenn, K. M., Margoliash, D., & Nusbaum, H. C. 2006. Recursive Syntactic pattern learning by songbirds. *Nature*, 440, 1204-1207.
- Grahn, J. A., & Brett, M. 2004. Beat-based rhythm processing in the brain. *Proceedings of the 8<sup>th</sup> International Conference on Music Perception & Cognition* (207-208). Evanston, IL.
- Grahn, J. A., & Brett, M. 2009. Impairment of beat-based rhythm discrimination in Parkinson's disease. *Cortex*, 45(1), 54-61.
- Grahn, J. A. 2009. The role of the basal ganglia in beat perception: neuroimaging and neuropsychological investigations. *Annals of the New York Academy of Sciences*, 1169, 35-45.
- Grodzinsky, Y. 2000. The neurology of syntax: language use without Broca's area. *Behavioral and Brain Sciences*, 23, 1-71.
- Guttman, S. E., Gilroy, L. A., & Blake, R. 2005. Hearing what the eyes see: auditory encoding of visual temporal sequences. *Psychological Science*, 16, 228-235.
- Harkleroad, L. 2006. *The Math Behind the Music*. Cambridge, UK: Cambridge University Press.
- Hagmann, C. E., & Cook, R. G. 2010. Testing meter, rhythm, and tempo discriminations in pigeons. *Behavioural processes*, 85(2), 99-110.
- Hall, B. K. 1999. *Evolutionary Developmental Biology*. Second Edition. Dordrecht: Kluwer Academic.
- Hauser, M. D., Newport, E. L., & Aslin, R. N. 2001. Segmentation of the speech stream in a nonhuman primate: Statistical learning in cotton top tamarins. *Cognition*, 72, B53-B64.
- 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.
- Hauser, M. D., & McDermott, J. 2003. The evolution of the music faculty: A comparative perspective. *Nature Neuroscience*, 6, 663-668.
- Helmholtz, H. von. 1954. *On the sensations of Tone as a Physiological Basis for the Theory of Music* (2<sup>nd</sup> ed., A. J. Ellis, Trans. Original work published 1885) New York: Dover.
- Huron, D. 2006. *Sweet anticipation: music and the psychology of expectation*. Cambridge, MA: MIT Press.
- Hyde, K. L., & Peretz, I. 2004. Brains that are out of tune but in time. *Psychological Science*, 15, 356-360.

- Izumi, A. 2000. Japanese monkeys perceive sensory consonance of chords. *Journal of the Acoustical Society of America*, 108, 3073-3078.
- Jackendoff, R., & Lerdahl, F. 2006. The capacity for music: What is it, and what's special about it? *Cognition*, 100, 33-72.
- Janata, P., & Grafton, S. T. 2003. Swining in the brain: Shared neural substrates for behaviours related to sequencing and music. *Natural Neuroscience*, 6, 682-687.
- Jarvis, E. D. 2004. Learned birdsong and the neurobiology of human language. *Annals of the New York Academy of Sciences*, 1016, 749-777.
- Jentschke, S., Koelsch, S., Sallat, S., & Friederici, A. 2008. Children with Specific Language Impairment Also Show Impairment of Music-s yntactic Processing. *Journal of Cognitive Neuroscience*, 20, 1940-1951.
- Justus, T., & Hutsler, J. J. 2005. Fundamental issues in the evolutionary psychology of music: Assessing innateness and domain-specificity. *Music Perception*, 23, 1-27.
- Karmiloff-Smith, A. 1992. *Beyond modularity: A Developmental Perspective on Cognitive Science*. Cambridge, MA: MIT Press.
- Karmiloff, K., & Karmiloff-Smith, A. 2002. *Pathways to language: from fetus to adolescent*. Cambridge, MA: MIT Press.
- Katz, J., & Pesetsky, D. 2009. The identity thesis for language and music. Draft published online, :lingBuzz/000959.
- Kirkham, N.Z., Slemmer, J.A., & Johnson, S. P. 2002. Visual statistical learning in infancy: evidence of a domain general learning mechanism. *Cognition*, 83, B35-B42.
- Kluender, K. R., Diehl, R., & Killeen, P. R. 1987. Japanese quail can learn phonetic categories. *Science*, 237, 1195-1197
- Kluender, K. R., Lotto, A. J., Holt, L. L., & Bloedel, S. L. 1998. Role of experience for language-specific functional mappings of vowel sounds. *Journal of the Acoustical Society of America*, 104, 3568-3582.
- Koelsch, S., Gunter, T., von Cramon, D. Y., Zysset, S. Lohmann, G., & Friederici, A. D. 2002. Back speacks: A cortical "language-network" serves the processing of music. *NeuroImage*, 17, 956-966.
- Koelsch, S., Gunter, T., Wittforth, M., & Sammler, D. 2005. Interaction between syntax processing in language and music: An ERP study. *Journal of cognitive Neuroscience*, 17, 1565-1577.

- Koelsch, S., & Siebel, W. A. 2005. Towards a neural basis of music perception. *Trends in cognitive sciences*, 12, 578-584.
- Kojima, S. & Kiritani, S. 1989. Vocal-auditory functions of the chimpanzee: vowel perception. *International Journal of Primatology*, 10, 199-213.
- Kojima, S., Tatsumi, I. F., Kiritani, S. & Hirose, H. 1989. Vocal-auditory functions of the chimpanzee: consonant perception. *Human Evolution*, 4, 403-416.
- Koopmans-van Beinum, F. J., Clement, C. J., & Van Den Dikkenberg-Pot, I. 2001. Babbling and the lack of auditory speech perception: a matter of coordination? *Developmental Science*, 4(1), 61-70.
- Krumhansl, C. L. 1990. *Cognitive Foundations of Musical Pitch*. New York: Oxford University Press.
- Krumhansl, C. L. 2000. Tonality induction: A statistical approach applied cross-culturally. *Music Perception*, 17, 461-479.
- Kuhl, P. K., & Miller, J. D. 1975 . Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonants. *Science*, 190, 69-72.
- Kuhl, P. K., & Padden, D. M. 1982. Enhanced discriminability at the phonetic boundaries for the voicing feature in macaques. *Perception & Psychophysics*, 32, 542-550.
- Kuhl, P.K. 1991. Humans adults and human infants show a “perceptual magnet effect” for the prototypes of speech categories, monkeys do not. *Perception and Psychophysics*, 50, 93-107.
- Lerdahl, F., & Jackendoff, R. A., 1983. *A Generative Theory of Tonal Music*. Cambridge, MA: MIT Press.
- Lieberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. 1967. *Psychological Review*, 74, 431.
- Lieberman, P. 2006. *Toward an Evolutionary Biology of Language*. Cambridge, MA: Harvard University Press.
- Locke, J. L. 1993. *The Child's Path to Spoken Language*. Cambridge, MA: Harvard University Press.
- Longuet-Higgins, H. C., & Lee, C. S. 1982. Perception of musical rhythms. *Perception*, 11, 115-128.
- Love, A. C. 2007. Functional homology and homology of function: biological concepts and philosophical consequences. *Biology and Philosophy*, 22, 691-708.

- Maess, B., Koelsch, S., Gunter, T., & Friederici, A.D. 2001. Musical Syntax is processed in Broca's area: An MEG-study. *Nature Neuroscience*, 4, 540–545.
- Maye, J., Werker, J., & Gerken, L. 2002. Infant sensitivity to distributional information can effect phonetic discrimination. *Cognition*, 82, B101-B111.
- Marler, P. 1984. Song learning: Innate species differences in the learning process. In P. Marler & H. S. Terrace (Eds.), *The biology of learning*, (289–309). Berlin, Germany: Springer-Verlag.
- Marler, P. 1987. Sensitive periods and the roles of specific and general sensory stimulation in birdsong learning. In J. Rauschecker & P. Marler (Eds.), *Imprinting and cortical plasticity* (99–135). New York, NY: Springer-Verlag.
- Marler, P. 2000. Origins of music and speech: Insights from animals. In N. L. Wallin, B. Merker, & S. Brown (Eds.), *The origins of music* (31–48). Cambridge, MA: MIT Press.
- Mcauley, J. D., & Henry, M. J. 2010. Modality effects in rhythm processing: auditory encoding of visual rhythms is neither obligatory nor automatic. *Attention, Perception, & Psychophysics*, 72, 1377-1389.
- McDermott, J., & Hauser, M., 2004. Are consonants intervals music to their ears? Spontaneous acoustic preferences in a nonhuman primate. *Cognition*, 94, B11-B21.
- McDermott, J., & Hauser, M., 2005. The origins of music: Innateness, uniqueness, and evolution. *Music Perception*, 23(1), 29-59.
- McDougall-Shackleton, S., & Hulse, S. 1996. Concurrent absolute and relative pitch processing by European starlings (*Sturnus vulgaris*). *Journal of Comparative Psychology*, 110, 139–146.
- McMullen, E., & Saffran, J. R. 2004. Music and language: A developmental comparison. *Music Perception*, 21, 289-311.
- Mehler, J., Dupoux, E., Nazzi, T., & Dehaene-Lambertz, D. 1996. Coping with linguistic diversity: The infant's viewpoint. In J. L. Morgan & D. Demuth (Eds.), *Signal to Syntax* (101-116). Mahwah, NJ: Lawrence Erlbaum.
- Miller, G. A. 1956. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63, 81-97.

- Mithen, S. 2005. *The Singing Neanderthals: The Origins of Music, Language, Mind and Body*. London: Weidenfeld & Nicolson.
- Nazzi, T., Bertoni, J., & Mehler, J. 1998. Language discrimination in newborns: Toward an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 756-777.
- Nettl, B. 2000. An ethnomusicologist contemplates universals in musical sound and musical culture. In N. L. Wallin, B. Merker, and S. Brown (Eds.), *The Origins of Music* (463–472). Cambridge, MA: MIT Press.
- Newport, E. L., & Aslin, R. N. 2004. Learning at a distance: I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology*, 48, 127–162.
- Newport, E. L., Hauser, M. D., Spaepen, G., & Aslin, R. N. 2004. Learning at a distance: II. Statistical learning of non-adjacent dependencies in a non-human primate. *Cognitive Psychology*, 49, 85–117.
- Osterhout, L., & Holcomb, P. 1992., Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, 31, 785-806.
- Palmer, C., & Pfordresher, P. Q. 2003. Incremental planning in sequence production. *Psychological Review*, 110, 683–712.
- Patel, A. D. 2010. Language, music, and the brain: A resource-sharing framework. In P. Rebuschat, M. Rohrmeier, J. Hawkins, & I. Cross (Eds.), *Language and Music as Cognitive Systems*. Oxford: Oxford University Press.
- Patel, A. D. 2003. Language, music, syntax, and the brain. *Nature Neuroscience*, 6, 674–681.
- Patel, A. D. 2006. Musical rhythm, linguistic rhythm, and human evolution. *Music Perception*, 24, 99-104.
- Patel, A. D. 2008. *Music, Language, and the Brain*. New York: Oxford University Press.
- Patel, A. D., Gibson, E., Ratner, J., Besson, M., & Holcomb, P. 1998. Processing syntactic relations in language and music: An event-related potential study. *Journal of Cognitive Neuroscience*, 10, 717–733.
- Patel, A. D., & Daniele, J. R. 2003. An empirical comparison of rhythm in language and music. *Cognition*, 87, B35-B45.
- Patel, A. D., Iversen, J. R., Bregman, M. R., & Schulz, I. 2009. Studying synchronization to a musical beat in nonhuman animals. *Annals of the New York Academy of Sciences*, 1169, 459-469.

- Patel, A. D., Iversen, J. R., Chen, Y., & Repp, B. H. 2005. The influence of metricality and modality on synchronization with a beat. *Experimental Brain Research*, 163, 226-238.
- Patel, A. D., Iversen, J. R., Wassenaar, M., & Hagoort, P. 2008. Musical syntactic processing in agrammatic Broca's aphasia. *Aphasiology*, 22, 776-789.
- Patel, A. D., Foxton, J. M., & Griffiths, T. D. 2005. Musically tone-deaf individuals have difficulty discriminating intonation contours extracted from speech. *Brain and Cognition*, 59, 310-313.
- Pelucchi, B., Hay, J. F., & Saffran, J. R. 2009, Statistical Learning in a Natural Language by 8-Month-Old Infants. *Child Development*, 80, 674-685.
- Peretz, I. 1993. Auditory atonalia for melodies. *Cognitive Neuropsychology*, 10, 21-56.
- Peretz, I., & Coltheart, M. 2003. Modularity of music processing. *Nature Neuroscience*, 6, 688-691.
- Perruchet, P., & Rey, A. 2005. Does the mastery of center-embedded linguistic structures distinguish humans from nonhuman primates? *Psychonomic Bulletin & Review*, 12(2), 307-313.
- Peter, B., Stoel-Gammon, C., & Kim, D. 2008. Octave equivalence as an aspect of stimulus-response similarity during nonword and sentence imitations in young children. *Speech Prosody*, 2008, 731-734.
- Pinker, S. 1997. *How the Mind Works*. London: Allen Lane.
- Pinker, S., & Jackendoff, R. 2005. The faculty of language: What's special about it? *Cognition*, 95, 201-236.
- Ramus, F., & Mehler, J. 1999. Correlates of linguistic rhythm in the speech signal. *Cognition*, 73, 265-292.
- Repp, B. H. 2003. Rate limits in sensorimotor synchronization with auditory and visual sequences: the synchronization threshold and the benefits and costs of interaval subdivision. *Journal of Motor Behaviour*, 35, 355-370.
- Repp, B. H. & Penel, A. 2002. Auditory dominance in temporal processing: new evidence from synchronization with simultaneous visual and auditory sequences. *Journal of Experimental Psychology: Human Perception and Performance*, 28(5), 1085-1099.

- Rocheffort, C., He, X., Scotto-Lomassese, S. & Scharff, C. 2007. Recruitment of FoxP2-expressing neurons to Area X varies during song development. *Developmental Neurobiology*, 67, 805-817
- Saffran, J. R., Hauser, M., Seibel, R., Kapfhamer, J., Tsao, F., & Cushman, F. 2008. Grammatical pattern learning by human infants and cotton-top tamarin monkeys. *Cognition*, 107, 479-500.
- Samuels, B., Hauser, M., & Boeckx, C. 2010. Do animals have Universal Grammar? A case study in phonology. In I. Roberts (Ed.), *The Oxford Handbook of Universal Grammar*. Oxford: Oxford University Press.
- Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. 1999. Statistical learning of tone sequences by human infants and adults. *Cognition*, 70, 27-52.
- Schachner, A., Brady, T.F., Pepperberg, I., & Hauser, M. 2009. Spontaneous motor entrainment to music in multiple vocal mimicking species. *Current Biology*, 19, 831–836.
- Snyder, B. 2000. *Music and Memory: An Introduction*. Cambridge, MA: MIT Press.
- Suzuki, R., Buck, J. R., & Tyack, P.L. 2006. Information entropy of humpback whale. *Journal of Acoustic Society of America*, 119, 1849–1866.
- Ten Cate, C., van Heijningen, C., & Zuidema, W. 2010. Reply to Gentner et al.: As simple as possible, but not simpler. *PNAS*, 107, E66-E67.
- Thiessen, E.D., Hill, E. A., & Saffran, J.R.. 2005. Infant directed speech facilitates word segmentation. *Infancy*, 7, 49-67.
- Toro, J. M., & Trobalón, J. B. 2005. Statistical computations over a speech stream in a rodent. *Perception & Psychophysics*, 67, 867-875.
- Trainor, L. J. 1997. Effect of frequency ratio on infants' and adults' discrimination of simultaneous intervals. *Journal of Experimental Psychology: Human Perception and Performance*, 23, 1427-1438.
- Trainor, L. J., Tsang, D. D., & Cheung, V. H. W. 2002. Preference for consonance in 2 month-old infants. *Music Perception*, 20, 185-192.
- Tramo, M. J, Cariani, P. A, Delgutte, B., & Braidà, L. D. 2001. Neurobiological foundations for the theory of harmony in western tonal music. *Annals of the New York Academy of Sciences*, 930, 92–116.
- Vitouch, O. & Ladining, O. (Eds.) 2009. *Music and Evolution. Musicae Scientae*, (Special Issue), 2009-2010.
- Wallin, N. L., Merker, B., & Brown, S. (Eds.) 2000. *The Origins of Music*. Cambridge, MA: MIT Press.

- Winkler, I., Haden, G. P., Ladinig, O., Szillere, I., & Honing, H. 2009. Newborn infants detect the beat in music. *PNAS*, 7, 2468–2471.
- Wright, A. A., Rivera, J. J., Hulse, S. H., Shyan, M., & Neiworth, J. J. 2000. Music perception and octave generalization in rhesus monkeys. *Journal of Experimental Psychology: General*, 129, 291-307.
- Zarco, W., Merchant, H., Prado, L., & Mendez, J. C. 2009. Subsecond timing in primates: Comparison of interval production between human subjects and rhesus monkeys. *Journal of Neurophysiology*, 102, 3191–3202,
- Zatorre, R. J. 2005. Music, the food of neuroscience? *Nature*, 434, 312-315.
- Zatorre, R. J., Chen, J. L., & Penhune, V. 2007. When the brain plays music: auditory–motor interactions in music perception and production. *Neuroscience*, 8, 547-558.