LINGUISTIC AND GRAMMATICAL MUSIC

From a musical protolanguage to rhythm and tonality

Alexandre Celma Miralles

Dir. Joana Rosselló Ximenes

MA Thesis in Ciència Cognitiva i Llenguatge

Presented to Universitat de Barcelona

August, 2014

Universitat de Barcelona

Aquesta tesina no hauria estat possible

sense el suport incessant i pacient de la meva tutora,

sense les atentes reflexions i llargues discussions dels membres

del grup de recerca Grammar, Mind and Reference,

ni sense la constant pressió, la incondicional escolta

i el reconfortant somriure d'amistats i familiars.

Castellvell del Camp, 25th August 2014

ABSTRACT

Music and language are two faculties that have only evolved in humans, and by mutual interaction. As Darwin (1871) suggested, before speaking, our ancestors were able to sing in a way structurally and functionally similar to what birds do. At that stage, a musical protolanguage with beat yielded a common basis for music and language. Hierarchical recursion along with grammar and lexical meaning joined this musical protolanguage and gave rise to language. Linguistic recursion, in turn, made meter possible. Rhythm therefore would have preceded tonality. Subsequently, in parallel to the emergence of grammar, harmony and tonality were added to the meter. That beat is more primitive than meter is suggested by the fact that some animals perceive but do not externalize it. Crucially, they are all *vocal learners*. Externalization, either in musical rhythm or language, requires a complex social behaviour, which for rhythm is already present in the *drumming* behaviour of certain primates. The role of vocalizations, in turn, goes even further: their harmonic spectrum underpinned the tones of our musical scales. Thus, driven to a large extent by language, music has turned out to be as we know it nowadays.

RESUM

La música i el llenguatge són dues facultats exclusivament humanes que han evolucionat alimentant-se mútuament. Com Darwin (1871) ja va suggerir, abans de parlar, els nostres ancestres tenien cants similars funcionalment i estructuralment al cant dels ocells. En aquest estadi, un protollenguatge musical amb pulsació es consolidà com a base comuna de la música i el llenguatge. La recursió jeràrquica, juntament amb la gramàtica i el significat lèxic, es van afegir a aquest protollenguatge musical i van donar lloc al llenguatge. Aquesta recursió lingüística féu possible el metre. El ritme, doncs, va precedir la tonalitat. Ulteriorment, en paral·lel al sorgiment de la gramàtica, l'harmonia i la tonalitat s'afegeixen al metre (compàs). Que la pulsació és més primitiva ho indica el fet que certs animals la perceben però no l'externalitzen espontàniament. Crucialment, tots són *vocal learners*. L'externalització, tant del ritme com del llenguatge, requereix una conducta social complexa, que ja s'observa en el conducta percutiva (*drumming*) de certs primats. El paper de les vocalitzacions, per la seva banda, va encara més enllà: l'espectre harmònic que presenten és el fonament de les notes a les escales musical. Així doncs, a remolc del llenguatge, és com s'arriba a la música tal i com l'entenem avui en dia.

RESUMEN

La música y el lenguaje son dos capacidades exclusivamente humanas que han evolucionado alimentándose mutuamente. Como Darwin (1871) ya sugirió, antes de hablar, nuestros ancestros, tenían cantos similares funcionalmente y estructuralmente al canto de los pájaros. En este estadio, un protolenguaje musical con pulsación se consolidó como la base común de la música y el lenguaje. La recursión jerárquica, junto con la gramática y el significado léxico, se añadieron a este protolenguaje musical y dieron lugar al lenguaje. Esta recursión lingüística hace posible el metro. El ritmo, pues, precedió la tonalidad. Ulteriormente, en paralelo al surgimiento de la gramática, la armonía y la tonalidad se añaden al metro (compás). Que la pulsación es más primitiva lo indica el hecho de que ciertos animales la perciben pero no la externalizan espontáneamente. Crucialmente, todos son *vocal learners*. La externalización, tanto del ritmo como del lenguaje, requiere una conducta social compleja, que ya se observa en la conducta percutiva (*drumming*) de ciertos primates. El papel de las vocalizaciones, por su parte, va aún más allá: el espectro armónico que presentan es la base de las notas en las escaleras musicales. Así, a remolque del lenguaje, es como se llega a la música tal y como la entendemos hoy en día.

TABLE OF CONTENTS

INTRODUCTION

How has music evolved? What is its relation to language? Was there a musical protolanguage giving rise to both language and music? What would a musical protolanguage provide and supply to speech? Has language, in turn, influenced music? How is music implemented in the brain? Can music shed light on mental diseases and, in turn, be therapeutic?

In order to give an explanative answer to these questions, we appeal to the existence of a musical protolanguage that worked as a rudimentary communication system of the first *Homo sapiens* (and perhaps other extinct ancestors) and that evolved into protomusic, before language emergence. This early system, along with others (such as body-gestural communication), might have enabled the expression of emotions, needs and motivations between conspecifics, as well as sexual-affective behaviour.

Assuming that our current music faculty derives from this musical protolanguage yet was altered and evolved by language (our human mode of thinking), we have to look for fundamental, musical underpinnings in properties of this musical protolanguage. To do that, we must take comparative evidence from the animal kingdom, where "music" appears to be involved in different functions through different formalizations, as well as from current neuroscientific research, studying music neural correlates in comprehension and production. Thus, pulling apart the linguistic elements found in music, we will be able to distinguish the bare components and elements of this faculty from its properties.

Afterwards, this exercise of teasing apart the musical faculty traits will allow us to understand (1) the boundary between language and music, (2) the different functions and meanings they convey, (3) the brain correlates of each, (4) the similarities to other animal beings and, perhaps, (5) the possible (contrasting) differences and correlations between them in brain injuries and mental disorders.

The aim of this thesis is to separate the structural components of music, rhythm and pitch, in order to analyse their origins, which may be done by contrasting them to language phylogenetically and neurally. The first part of this thesis will report music and language interrelations, their evolution from a musical protolanguage and

their neural correlates. The second part will focus on rhythm and tonality origins, attending to animal comparisons and brain studies.

First, we will compare these human-unique faculties, music and language, and their implementation in the brain. After reviewing the biological (rather than cultural) origins of music, we will argue that music is an exaptation. That is to say that the ancestor of music was selected as a communicational system but later, recently in evolutionary time, its function has been overtaken by language. Looking for music in nature implies a search for *musicality*: structurally and functionally separated elements of our current music, such as rhythm and pitch —which permit melody and harmony—, and which may have been selected for other purposes.

Since tonal structure has been widely studied in music, we propose to take a look at recent studies on rhythmic cognition, shifting the structural component of music from "tonality" to "rhythm". This latter component involves distinct (1) constitutive elements: beat, meter, grouping and tempo, and (2) separated cognitive processes: meter induction, beat perception and synchronization (BPS, in Patel, 2010) [also called pulse extraction and entrainment (PEE, in Fitch, 2012)]. While BPS appears in some animals, more concretely in some vocal learners such as songbirds and mammals, meter induction —*a hierarchical way of categorizing the beat in music production and perception*— seems to be unique to humans. Although we defend that music meter depends on recursion, whose hierarchy may organize the beat, an alternative view such as that proposed by the so called *Demanding Attentional Theories*.[1]

Notice that, neurally, what is involved in BPS requires the connection between auditory-motor regions in the brain. This sensory-motor integration could be then as essential for music as it is for language acquisition and speech production. Apart from vocal learning, other theories claim that primate rhythmic behaviour (drumming) and primate rhythmic perception (grouping capacities) must be the precedents of our music. We argue that both, in fact, constitute a rhythmic protomusic, an intermediate stage between musical protolanguage and music.

---

[1] These theories explain the neural correlates of musical meter as beta-band oscillations synchronized to stimulus by generating cyclical, attentional expectancies.

Apart from rhythm, we also briefly look at pitch and tonality origins, so as to conclude that the harmonic spectra of human vocalizations have underpinned the discrete tones of worldwide musical scales. Scale notes stem from a selected auditory specialization for our conspecific vocalizations; which, again, points to a communicative musical protolanguage. In contrast, tonality will be argued to be a by-product of our referential system. Thus, the relation between pitches and chords according to their harmonic spectra and their hierarchical position within the scale would yield a musical grammar.

In short, meter and tonality, because of their hierarchical organization, are argued to be deeply related to the emergence of merge and grammar, respectively. As such, rhythm is the structural component of music while pitch and tonal-harmony are their grammatical counterpart. Both structural components of music are tightly related to language evolution, that is, the emergence of hierarchical merge and referential grammar. At the same time, some properties of language are strongly tied to ancestral musical properties, such as prosody or syllabic rhythm.

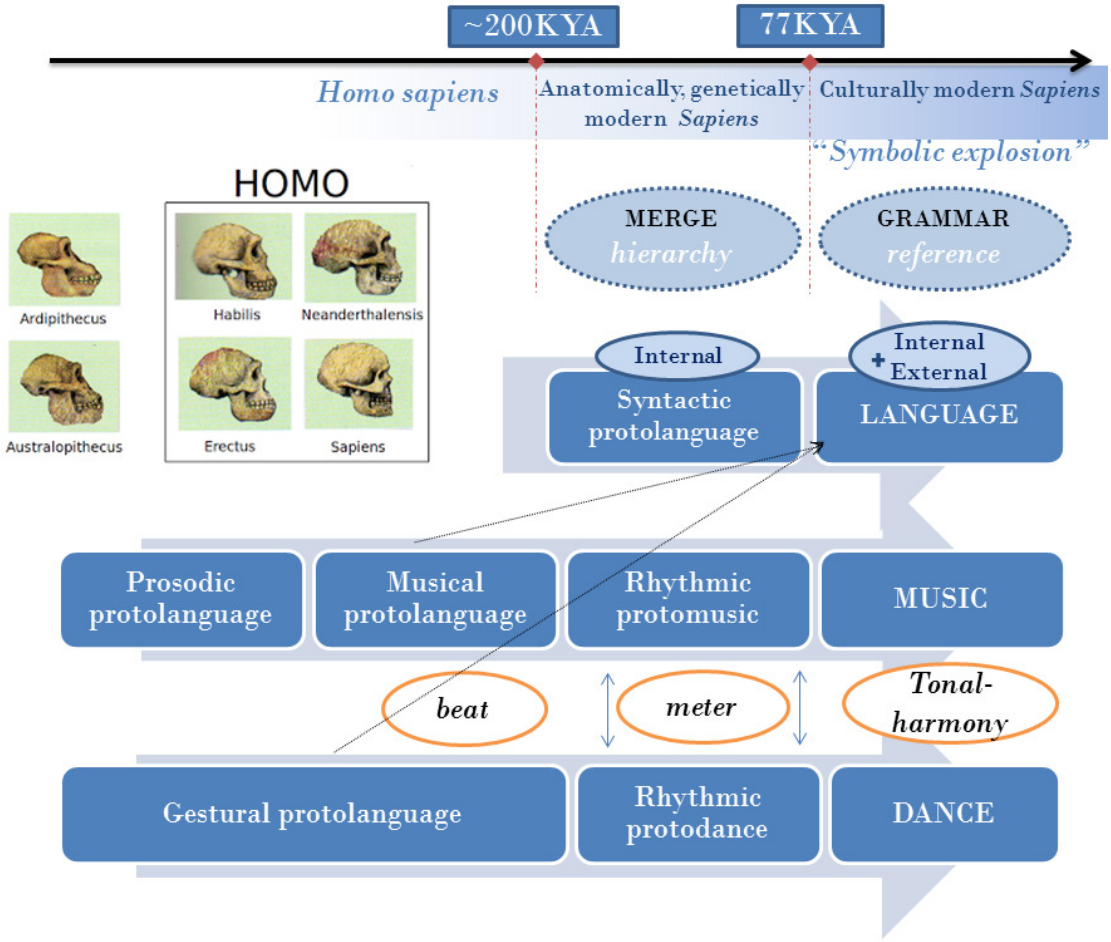


Figure 1

PART I

PART I: MUSIC AND LANGUAGE

This first part briefly compares music and language: their common properties, their specific structural components and their social nature which gives rise to different meanings and the ability to evoke emotions, in the case of music, or express concepts, in the case of language.

1. FORMAL AND FUNCTIONAL COMMONALITIES

Music and language are universal human faculties, neurobiologically constrained and culturally transmitted during sensitive periods of acquisition, which permit a human-specific way of thinking and communicating. Music is an organized arrangement of sounds and silences that evoke emotions and involves people in a social interactive performance, made of gestures, sounds and shared intentions and moods. Fundamentally, music is "governed by structural principles that specify the relationships among *notes* that make up melodies and chords[,] and *beats* that make up rhythms" (Fedorenko, McDermott, Norman-Haignere, and Kanwisher, 2012).

As it occurs with language, the music faculty is located in the mind and is internally and externally constrained by genetic, developmental and structurally physical factors. This faculty must be distinguished from musical idioms,[2] which are found in every culture and in every era (classified by genre, style or ethnographical locations), which consist of culturally-driven, learned systems of a musical grammar.

Music is acquired through a Music Acquiring Device (Liu, Jiang & Li, 2014) —a homologue of Language Acquisition Device—, which is found to interact with language acquisition by facilitating pronunciation skills, accelerating the mastery of language rhythm and promoting syntax acquisition. Following a general capacity to make sense of the world through grammar, the human brain (even in babies) is able to sort out distinct musical sounds so as to yield systems of rules.

Paralleling the multicomponent Faculty of Language (Hauser, Chomsky and Fitch, 2002, Fitch, Hauser and Chomsky, 2005), the music faculty is also made of different components, each with their own evolutionary history. Both faculties could be seen as mosaics of traits, some of them shared. In this line, Hockett (1960)

[2] Musical idioms are with respect to music what languages are with respect to language.

proposed a set of language design features —from spoken language—, and already pointed out which of them were shared with music, either vocal or instrumental.

Language	Music		Innate human calls
Design feature	Instrumental	Vocal	
1. Vocal auditory channel	No	Yes	Yes
2. Broadcast transmission	Yes	Yes	Yes
3. Rapid fading	Yes	Yes	Yes
4. Interchangeability	No	Yes?	Yes
5. Total feedback	Yes	Yes	Yes
6. Specialization	Yes	Yes	Yes
7. Semanticity	No	No	No?
8. Arbitrariness	No	No	No?
9. Displacement	No	No	No
10. Duality of patterning	No	No	No
11. Productivity	Yes	Yes	No
12. Discreteness	Yes	Yes	No
13. Cultural transmission	Yes	Yes	No

Table 1

The preceding table shows Fitch's application of Hockett's language design features on vocal and instrumental music, as well as innate human calls. Those non-shared traits include: semanticity, arbitrariness, displacement and duality of pattern, which directly derive from referentiality —or, in Hockett's terms, semanticity. The following table consists of Fitch's (2005) list of music design features that are applied to spoken language and innate calls.

Design features of music

Design feature	Language?	Innate calls?
1. Complexity	Yes	No
2. Generativity	Yes	No
3. Culturally transmitted	Yes	No
4. Discrete Pitches	No	No
5. Isochronic	No	No
6. Transposability	Yes	?
7. Performative context	No	No
8. Repeatable (repertoire)	No	No
9. A-referentially expressive	No	Yes

Table 2

From the table, it seems that music-specific features —except for the trait 9— include pitch discreteness[3] (a discrete set of pitches yielding a scale) and isochrony (a regular periodic pulse or beat), as well as performative context (cultural rituals depending on distinct societal behaviours), repeatability (multiple performance of

[3] Comparing music to language, Anirudh Patel (2008) denies that any language (even tonal languages) organize pitches in terms of musical scales, which are cultural frameworks for musical performance and perception. Then, as every language has its own set of phonemes, distinguished by timbre, all music has its own set of notes, distinguished by pitch, separated by different intervals.

identifiable pieces from a repertoire) and a-referentially expression (a gestural form including flexible mappings to movement and mood).

Looking across cultures, some general properties of music (broadly called musical universals) are found: discrete pitch levels, octave equivalence, a moderate number of pitches (5 to 7) repeated in every octave, a tonal hierarchy of pitches functioning as either stable or unstable referential points, the notion of a deep-structural idea, reference pulses, the induction of rhythmic patterns by asymmetrical subdivision of time pulses, relational pitch and time features (i.e. contour), small integer frequency ratios (relative proportions as 1:3, 2:1, 3:2), unequal scale steps of pitches, and the musical genre for infants called lullabies (Isabelle Peretz, 2006).

Since music interactive performance is broadly found cross-culturally, music may be considered as "a communicative medium complementary to language that is deeply embedded in [...] the species-specific human capacity to manage complex social relationships" (Cross, 2009). As a mode of human interaction, music is optimal to "manage situations of uncertainty by virtue of its semantic indeterminacy" or "floating intentionality" (Cross, 2009). Although music is not 'about' events in the world, the emotional physiological reactions that it provokes are otherwise similar to those elicited by them, perhaps coming from an earlier emotional mechanism which is still preserved in pitch and timbre across species.

Following Cross (2009), musical meaning is driven by three simultaneous dimensions evolved in different moments: the culturally-enactive meaning[4] (based on cultural-contextual links), the socio-intentional (based on cross-cultural interpersonal, communicational cues, or prosody) and the motivational-structural (based on the acoustical signal and it involuntary affective variation). This three dimensional meaning of music operates in musical performances enabling collective musical behaviour and, consequently, promoting group affiliation.

However, music is normally built on patterns of tension and release, creating a musical ebb-and-flow. These tension-resolution patterns refer to structural music

[4] Stephan Koelsch (2011) claims that musical meaning emerges from embracing extra-musical sign qualities, intra-musical structural relations, musicogenic effects, the establishment of a unified coherent sense out of 'lower-level' units, and a musical discourse; all of them competing at the same time in brain processes.

properties, which arise from the relationship of musical elements and are based on a hierarchy of stability —quiescence points established with either fulfilled or violated expectations. These patterns yield an absolute musical meaning that solely relies on the interplay of formal musical structures, which point to musical-unique consequences. Hence, this musical meaning is driven by implicitly learned and rule-constrained expectations and is similar to the linguistic structural meaning arising from hierarchical relations. This indeed constitutes the internal meaning of music.

2. MUSIC EVOLUTION

After having compared music and language faculties, and their design features, we will now argue that music may have evolved by selecting independent components bearing musicality separately. Within this gradual evolutionary view, both faculties' similarities in certain components suggest an ancestral communicative system common for language and music: a musical protolanguage. Although currently rediscovered, this musical protolanguage was first proposed by Darwin at 1971. This musical protolanguage may have split into music and language respectively in culturally-modern *Homo sapiens*, after the emergence of our symbolic thinking, which was boosted by language and grammar (providing hierarchy and reference).

2.1 *A mosaic of independent traits*

Laurel Trainor (2008) defends that music has deep evolutionary roots in the underpinnings of universal features of human sound processing,[5] which both (i) constrains rhythmic, melodic and harmonic structures, and (ii) permits variation of these features across cultures. While temporal and spectral organization of music derives from our biology, scales and harmonic structures, for instance, depend on learning, and therefore on environmental exposure. Thus, while every culture seems to show music and dance (suggesting music to be a genetically-coded universal),[6] the emotional response to learned scales and chords is culturally-dependent.

---

[5] The structure of our sensory organs, our basic encoding of organization and our visceral response to emotional sound features constitutes an evolutionary inheritance that may not have changed recently

[6] However, clear evidence for a genetic underpinning for musical traits is lacking. Consequently, it does not favour the adaptationist argument that musical behaviours were specially selected.

Even though cross-cultural universals[7] suggest that music was naturally selected, a null hypothesis for its evolution (McDermott, 2008) should rather be hold, defending that "music's perceptual basis could derive from general-purpose auditory mechanisms, its syntactic components could be co-opted from language, and its effect on our emotions could be driven by the acoustic similarity of music to other sounds of greater biological relevance, such as speech or animal vocalizations". In addition, since animals lack music, it is logical that any music-related trait found in them may represent a general-purpose mechanism. This therefore indicates that common traits among humans and animals have not evolved for music in particular.

In this line, the evolutionary psychology view of Honing and Ploeger (2012) moves the premise "music as biology" to "music as cognition". This view emphasizes the cognitive traits that have evolved in the human mind to solve specific ancestral problems. Since heritable cognitive variation does not fossilize, in order to prove how musical cognition has arisen, spread and changed, it should (i) be separated the notion of *musicality* from *music*,[8] and (ii) be collected evidence showing that cognitive traits are adaptations. However, although musicality could have been selected for biological functions in the past, current functions may differ from them.

The cognitive components making up *musicality* might be those involved in the perception, production and appreciation of music, which are then affected by socio-cultural and psychobiological factors. Despite emerging early in life, these cognitive mechanisms do not need to be specific for music —in fact, it should not be. Current literature proposes some candidates to have been evolved and selected from non-species-specific general domains to species-specific modular domains, which are pitch, tonal encoding of pitch, beat induction and metrical encoding of rhythm.

Assuming that evolution has specifically shaped human mind to support musical behaviour, or rather certain traits bearing *musicality*, three main approaches arise:

---

[7] Universals such as similar slow and repetitive lullabies directed towards infants, the inclination to move and dance to music, the musical meter organizing beats and the hierarchical organization of pitch, giving structural prominence to particular notes. Furthermore, musical diseases [see Annex, 12] affecting specific components of processing music also support the existence of these universals.

[8] While traits bearing *musicality* can be present in animal skills, music, understood as "a social and cultural construct based on that musicality", is unique to humans (Honing&Ploeger, 2012).

1. Music as an adaptation (a selected function for better survival)
    a) By sexual selection (courtship, pair bonding through duetting, group choruses)
    b) By kin selection (parental care, *motherese*)
    c) By group selection (social cohesion, coordinated behaviours)
2. Music as an exaptation (a selected by-product coming from other adaptations)
    a. *Music as a spandrel (just a by-product, without any selective pressure)

The first position considers music as a (naturally, sexually, kin...) selected trait during evolution, playing a fundamental role in survival. Some biological and cognitive functions have been proposed for an ancestral musical system, such as a sexual attractor mechanism for mating (Darwin, 1871; Miller, 2000), a monogamist pair-bonding mechanism through vocalizations in duets (Fitch, 2009), a social mechanism for group cohesion (Cross, 2007; Merker, 200, Mithen, 2005), an emotional bond for parents-offspring relations through *motherese* or infant-directed speech (Dissanayake, 2008), and so on. The second view is an intermediate position, in which music emerges from some existing selected traits being put to new uses. Here, Honing (2011) maintains that music, as a beneficial play, challenges our cognitive functions, promoting diversity and thus creating a cognitive advantage. Finally, the last *position (2.a) considers music to be a side effect of other functions, coming from non-selected traits, either as a by-product of a motivational system dealing with a technological system (position maintained by Pinker (1997, 2007) with his metaphor of music as an "auditory cheesecake"), or as a transformative invention impacting our biology and culture (Patel, 2010).

The debate on music origins between adaptationists and non-adaptationists is still opened [see Annex, 14], but we will adopt an intermediate position, in which traits bearing musicality were independently selected for purposes other than music. In this line, we will defend that current music is therefore an exaptation —with a wide range of current functions that we will not discuss here— that puts traits also evolved in animals together, which were further modified by language and, in particular, grammar.

2.2 *Language evolution*

The evolution of humans is inseparable from the emergence of language and grammar. Human linguistic and grammatical way of thinking is our fundamental distinctive trait as a species. The transition to symbolic reasoning,[9] from a non-symbolic and non-linguistic ancestor, occurred very late in the hominids coming into place in a genetically and anatomically modern *Homo sapiens* (Tattersall, 2013).

Language emergence requires the development of (i) a "conceptual system with abstract and symbolic meanings referring to general concepts away from the immediate sensory experience", using discrete units to label them (phonology), and (ii) the cognitive computation *merge*, "with the outcome to generate hierarchical structures for the purpose to cluster complex computations" (Hillert, 2014). In this line, Boeckx (2012) calls for the emergence of (i) merge and (ii) a lexical-envelope mechanism that permits to homogenize concepts to make them cognitively mixable.

Assuming Tattersall (2013), our symbolic reasoning should be logically paired with archaeological findings of figurative production. The earliest *Homo sapiens*, who appeared in Ethiopia between 200 and 160 KYA,[10] behaved alike their hominid contemporaries and have not left any trace of modern cognitive behaviour in the archaeological record. It was later, over 100KYA, that unprecedented behavioural proclivities and symbolic production appeared in Africa,[11] due to a developmental reorganization which led to a brain that was capable of complex symbolic manipulation, and therefore Universal Grammar. With figurative minds, these humans left Africa 60KYA and took over the world displacing other hominids.

Since the reorganized neural structure was in place 200KYA, what allowed the onset of symbolic thinking among the second wave of migrating *Homo sapiens* might have been a cultural stimulus: the "invention of language in an African isolate *Homo sapiens*" at (approx.) 100KYA (Tattersall, 2013). In turn, language may have acted as a cognitive trigger which suddenly produced a new cognitive phenotype.

---

[9] Symbolic reasoning is the ability to process and rearrange symbolic information following certain rules so as to envision multiple realities, through forming and manipulating symbols in the mind.

[10] KYA= thousand years ago.

[11] These modern cognitive behaviours consists of pierced marine shell beads, ochre deposits for paint and engraved geometric designs (Blombos Cave, 77KYA, southern African coast).

Regarding music, since this capacity constitutes figurative art that implies cultural behaviours and is structured by a culturally learned, grammatically-ruled system, it is highly plausible that it has appeared hand in hand with language and grammar. As a consequence, music may have also appeared recently[12] in an evolutionary time-scale, thus exapting anatomical traits and organs already present in humans for other uses, in the same way that language and symbolic thought[13] proceeded.

2.3 *A musical protolanguage*

The term *protolanguage*, according to Hewes (1973) and Bickerton (1990, 1995), refers to a communicative system of our lineage that has preceded our current language capacity,[14] proportioning basic formal and structural properties, as well as physiological traits and neural-computational mechanisms. Many protolanguages have been proposed, such as (i) musical protolanguage giving rise to the linguistic phonology and prosody, (ii) gestural protolanguage giving rise to the intentionality and signs, (iii) lexical protolanguage giving rise to the lexical referential words (previous to syntax), or (iv) syntactical protolanguage[15] —which may consist of recursive merge as an internal computational capacity without externalization. Excluding (iii), these protolanguages could have interacted with each other.

In *The descent of man and selection in relation to sex* (1871), Darwin observed that music, despite being a human universal carrying a physiological cost and playing an important role in society, does not show any obvious function. For that reason, music would be better seen as a fossil remaining from a former adaptation, that is, a communicational system used by earlier hominids whose core original function was later overtaken by language. This original common stage was termed *musical protolanguage*, which subsequent investigators have reviewed (Jespersen, 1922;

---

[12] Instrumental music is at least 40.000 years old (Fitch, 2005), taking as a reference a flute which has been found in a Slovenian Neanderthal cave.

[13] For instance, the wide range of formant frequencies exapted for our contemporary speech requires a descended larynx into the throat, the right position of the hyoid bone, the barrel-like structure of human rib cage and an innocuous breathing control, which were elected for a musical protolanguage.

[14] If we understand language as a complex multicomponent system, every gradually evolved property added to the system may configure a slightly different protolanguage, until arriving to our language.

[15] This syntactic protolanguage has been recently proposed by Boeckx et al. (2013), and is not yet in well known in the field. It consists of a computational mechanism emergence, an unrestricted Merge, contributing to the language-ready brain.

Livingstone, 1973; Richman, 1993; Brown, 2000; Mithen, 2005; Fitch, 2006). Although it may have been present in other hominids, for *Homo sapiens,* a musical protolanguage[16] may imply (at least) a vocal producing system, pitch contours dealing with emotional content, and a social structure protecting the members from predators attracted by the sounds (i.e. social bonding).

Darwin's theory of language evolution could be divided into three stages: (i) an "increase in intelligence and complex mental abilities", (ii) a "sexually selected attainment of the specific capacity for vocal control: singing", and (iii) an "addition of meaning to the songs", driven by a further intelligence increase (Fitch, 2013a) [see Annex, 15]. While the first step refers to the progress of cognitive power from an ape-like ancestor to modern humans (fuelled by social and technological factors), the second requires the evolution of vocal imitation abilities, which were used in courtship and territoriality, and in expressing emotions. Conversely, the third step, the transition of non-propositional songs to propositional speech, was dubiously resolved by Darwin appealing to signs and gestures combinations,[17] onomatopoeias and controlled imitation of modified instinctive cries and emotional vocalizations.

Language clearly does not come from an evolved vocal communicative system, but from an intelligence increase in humans. Considering language as "an instinctive tendency to acquire an art" (Darwin, 1871), biological and environmental (cultural) factors are unified. Since articulate speech does not suffice for explaining language, we should better look at the role of language in developing mental faculties, which permitted to connect sounds to ideas. Once meaning was in place, "words" impacted our mind enabling to carry long chains of complex thoughts. Then, the rise of language seems to be due to a cognitive development of social intelligence.

Ahead of his time, Darwin recognized the importance of learned vocalizations. Although complex vocal learning is unusually found in mammals and virtually absent in primates, which runs against a continuity view between non-human

[16] A musical protolanguage may have been triggered by (i) sexual selection, based on courtship or pair-bonding preferences, by (ii) parent care through infant directed speech or *motherese* to comfort the offspring, or by (iii) group bonding which promoted social cohesion.

[17] As other authors point out (e.g. Tomasello, Call, Arbib...), besides being primary triggered by communicative vocalizations, gestural signals and strategic planning in tool use probably contributed to language as well. Thus, a complex thinking may have been reinforced by a gestural protolanguage.

primate calls and language, it is shared with many birds. Then, beyond our closest phylogenetic relatives, Darwin took learned birdsongs as analogues to these putative vocalizations because they show parallelisms such as an innate *babbling* or *subsong stage* during critical periods of cultural transmission, as well as the final production of dialects and idiolects. These vocalizations may have expressed fitness, high-status position, territory maintenance, male-female pair-bonding, child care, and so on.

While musical protolanguage seems to require a vocal learning capacity,[18] a mirror system hypothesis for language origin is otherwise based on gestural behaviour and social life.[19] The interaction of both protolanguages may have contributed to language as well. In line with Arbib and Iriki (2013), we consider that music might have evolved inseparably from dance. Moreover, we claim that both are possible externalizations of the same faculty, which evokes emotions within social contexts. Their non-propositional, free-floating meaningfulness allows music and dance to be attached to several group activities and cohesive events.

Darwin (1871) expressed that "the progenitors of man [...], before acquiring the power of expressing their mutual love in articulate language, endeavoured to charm each other with musical notes and rhythms", thus favouring the idea of a musical protolanguage preceding the emergence of language. Although Darwin involved musical rhythms and notes into his pre-semantic model of musical protolanguage, both music and language currently show different properties due to the fact that they have changed in the course of evolution. What both speech and song share is prosodic and phonological aspects (Fitch, 2013a): "the use of a set of primitives (syllables) to produce larger, hierarchically structured units (phrases) that are discretely distinctive; but not the musical key aspects of discrete-pitched notes and temporal isochrony".[20] For that reason, Fitch (2013a) suggests to rename the

---

[18] Notice that it may have lacked in human and chimpanzee last common ancestor (five to seven million years ago) and it has not given propositional meaning to any other vocal learning species.

[19] It "builds upon skills for imitation, tool use and the development of novel communicative manual gestures by chimpanzees" (Arbib & Iriki, 2013), as well as upon the rich social structures, which are found in monkeys, apes and humans, but not in songbirds. From a mirror neuron hypothesis, a complex imitation system may have been developed and later converted to pantomime, protosign and protospeech (Arbib, Liebal and Pika (Arbib, 2013)). We do not agree with this view.

[20] A protolanguage made of isochronous rhythms and discrete pitches (and a tone-based meaning), is proposed by Brown (2000). His *musilanguage* hypothesis gathers together all these aspects, but only

Darwinian musical protolanguage as prosodic protolanguage, consisting of sung syllables not arranged in a scale nor produced with steady rhythm (Fitch, 2006). Assuming that, notes and rhythms should be considered as a more recent development in music, likely appearing within a protomusic[21] stage [see Part II].

Human learned vocalizations, given their syllabic structure and their melodic and rhythmic nature expressing emotionally prosodic features, seem to be a perfect initial substrate for phonology (specially, its phonetics and prosody). Probably, human protosyllabic vocalizations were similar to geladas vocalizations produced during grooming, which acoustically can be analysed as sequences of consonant and vowel-like elements. Furthermore, as it is defended in this thesis, vocalizations not only offer a *protophonology*, but also the underpinnings for rhythmic and harmonic structures. These random syllabic and rhythmic protophrases —associated to emotional states through prosody and intonation— were present in human communication before the creation of symbolic concepts, (Hillert, 2014), and could have promoted a functional hemispheric asymmetry.[22] In fact, this asymmetry is found in chimpanzees and Old World Monkeys processing species-specific vocalizations (Tagliatela et al., 2009), which reinforces the biological right-hemisphere origin of prosody (and music).

Human perception of voices, faces, gestures, smells and pheromones are allegedly lateralized to the right hemisphere, which is usually considered the place of social perception.[23] For example, primate vocalizations (similarly to auditory faces) carry paralinguistic information in its structure which permits to identify conspecific individuals. The neural mechanisms involved in these social interactions are lateralized to the right superior temporal sulcus, which indeed combines information from vocalizations and face displays (Belin, 2014). In humans, brain responses to affectively-laden animal vocalizations and speech reveal similar unconscious

relies on group-selection to explain their evolution. From him, music may have evolved by increasing the expression of emotion, while language may have enhanced the expression of lexical meaning.

[21] The term *protomusic* was coined to indicate a precedent stage of the music faculty.

[22] Human planum temporale, which is larger in the left hemisphere and involves the Wernicke's area located in the temporo-parietal junction, has a homolog in chimpanzees and macaques called areas Tpt. They are also asymmetrically left-sided and process multisensory information.

[23] The functional lateralization of the social brain involves the orienting of attention to emotional cues and the establishment of the first person perspective versus others (Brancucci et al., 2014).

orbitofrontal activations, related to the limbic system. These similarities support continuity in affective responses to vocal productions across mammals.

The natural melody of speech (i.e. prosody), which encompasses "overall pitch level and pitch range, pitch contour, loudness variation, rhythm and tempo" (Deutsch, 2010), reflects the speaker's emotional state and intention, similar to what occurs to musical pitch and timing features. Given that, Christensen (2004) proposes a close connection between music, rhetoric speeches and prosody,[24] defined as the emotional, non-semantic, and slowly varying pitch contours and rhythms of speech. Moreover, Christensen (2004) states that "music is not the language of emotions, but prosody is, and as far as music emulates prosody, it can also encode emotions". Given that, it is highly plausible that prosody has been a selected trait of a musical protolanguage. In fact, a deficiency in detecting and understanding the emotional qualities of speech is found in alexithymia,[25] which makes this feature discriminable. Hence, processing prosody could have been somehow selected.[26]

What allows us to detect prosody and acoustical patterns in speech is our fine processing of the spectral structure.[27] In this connection, it is well established that the human brain has developed two parallel and complementary systems: one, in the right hemisphere, processes slowly varying contours fitted in with spectral structure and prosody and the other one, in the left hemisphere, processes rapidly-paced inputs (see Zatorre et al., 2002).[28]

---

[24] Prosody is seen as an effective way of manipulating listener emotions within a group.

[25] Alexithymia (literally, 'no words for feelings') is a personality trait characterized by impairments in the experience of emotion and its cognitive processing: difficulties in emotionalizing and fantasizing (its emotional dimension), as well as in identifying and verbalizing feelings (its affective dimension). For more information about Alexithymia, see Annex, 13.

[26] Human frequency discrimination in hearing may have been selected for (i) responding quickly to loud, sudden sounds (meaning danger), for (ii) locating sources and for (iii) detecting conspecifics (Christensen, 2004).

[27] A cortical specialization for spectral and temporal resolution in auditory cortices seems to be similarly found in other mammals, which suggests that speech and music might have co-opted these ancestral structures.

[28] According to Christensen (2004), these two components support "a prosodic-semantic distinction" even in physiology since prosodic and semantic meanings "are processed by different brain centres".

A MUSICAL BRAIN

The human brain processes music both as an input and as an output, perceiving[29] and producing it through different neural processes. Music engages a variety of non-domain-specific skills, such as memorization or motor mechanisms, as well as general mental processes, such as executive function or abstract reasoning.

3.1 *Shared vs. specific areas*

On the one hand, the music faculty depends on specialized cerebral processes, which are neurobiologically determined, making it "an autonomous function, innately constrained and made up of multiple modules that overlap minimally with other functions" (Peretz, 2006). On the other hand, music and language also share certain components. For example, Brodmann Areas 44 and 45 (i.e. Broca's area), within the inferior frontal gyrus (IFG), are involved in processing linguistic hierarchy[30] as well as fine-grained musical pitch structures and rhythmic synchronization. Broca's and Wernicke's areas process harmony, rhythm and instrumental performance.

Although both hemispheres are involved in music production, melody and timbre discrimination activate right-hemispheric temporal and frontal regions in passive listening, and pitch and rhythm processing activate left-hemispheric linguistic areas. Melody and pauses[31] are processed in right-hemisphere temporal areas. Musical memory involves the right area of the hippocampus, bilateral temporal regions, IFG and the left precuneus. Finally, the neural representation of tones resides in the lateral margin of the primary auditory cortex and non-primary auditory cortex.

Studies coming from brain-damaged patients reveal the implication of temporal lobes in music processing of pitch and rhythm that is distinct from lower-level perceptual abilities. In contrast, neuroimaging research can isolate structure processing in music from generic auditory processing and highlights the implication of frontal lobes in musical structure violations. In other words, while patient

[29] According to Montinaro (2010), music perception follows three stages: (i) elementary auditory musical perception, (ii) musical structural analysis, at an elementary level (consisting of pitch, intensity, rhythm, duration, timbre) and an advanced level (consisting of phrasing, timing, themes), and (iii) played piece identification.

[30] Hierarchical-structure resources (involving Broca's area and its right homotope) are used to process musical syntax (Sammler et al. 2011).

[31] The perception of pauses could also occur in the left-hemisphere, together with rhythm and pitch.

literature points to temporal cortices for musical processing, neurological studies implicate the inferior frontal gyrus (IFG), Broca's area, and anterior, orbital parts of IFG in BA 47. This discrepancy should be studied in more depth in order to correctly discriminate different musical traits and locate their brain regions, so as to yield a clear theory of musical processing.

Regarding music-specific areas, Fedorenko et al. (2012) analyses report that bilaterally temporal activations are sensitive to pitch and rhythm though they are insensitive to high-level linguistic structure.[32] Seven cortical parcels are found to be sensitive to musical structure:[33] the bilateral anterior superior temporal gyrus (STG), the bilateral posterior STG (spanning the right middle temporal gyrus), the bilateral premotor cortex and the supplementary motor area (SMA). Regions anterior and posterior to Heschl's gyrus in the superior temporal plane, together with superior and middle temporal gyri, respond more to intact than scrambled musical stimuli, suggesting that they may play a role in musical structure analyses and representation, such as key, meter, harmony, melodic contour…

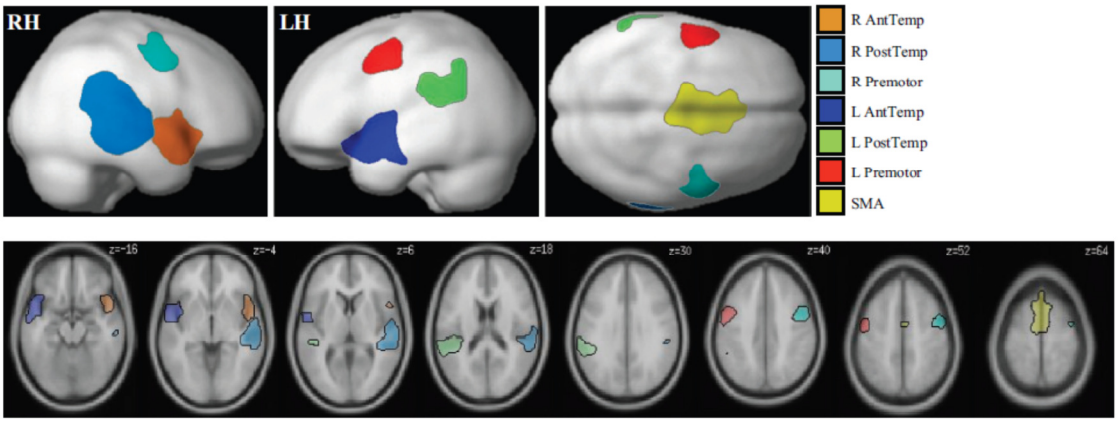


Figure 2

As Fedorenko et al. (2012) point out, their function is not well established yet, but they could store musical knowledge (i.e., information of prototypical musical patterns of melodies, rhythms and sequences), or responses to generic structures (such as consonance-dissonance discrimination), rather than pitch processing (McDermott et al. 2010). However, in experiments of pitch and rhythm scrambling, the activation of temporal lobe regions and bilateral premotor and supplementary

[32] Taking sentences as syntactically-complex, rather than simple lexical lists.

[33] These areas reveal a neural specialization for music-associated mental processes that is distinct from lower-level acoustic representations and high-level linguistic representations.

motor areas[34] is affected, which indicates that pitch and rhythm are inextricably linked (Jones and Boltz, 1989), thus constituting interdependent structures in music.

Nevertheless, these unique regions which are sensitive to music do not preclude overlapping regions to be engaged in linguistic and musical processing (Koelsch et al., 2002; Patel, 2003). Given that, these overlapping areas may be widely recruited in other cognitive tasks as well, either in general executive functions dealing with working memory and attention (Duncan, 2001, 2010), or in lower-level acoustic processes shared by speech and music, such as pitch processing and its encoding mechanisms in the auditory brainstem (Krizman et al., 2012).

3.2 *Emotions and the limbic system*

Music exploits brain mechanisms which have evolved to perceive and respond to vocal affects[35] (Patel, 2008), although none of them seem to be unique to music. In fact the mechanism leading to the pleasure sensation of music is an evolutionarily ancient neural circuit involved in survival and in mediating rewarding stimuli as food or sex,[36] involving the basal forebrain, the brainstem nuclei, the orbitofrontal and insular regions. Besides, medial temporal areas (integrating ventral and dorsal striatum) and the anterior cingulated are also activated during musical emotional processing (Montinaro, 2010). Thus, music elicits a response in the limbic system, a brain region which is evolutionary ancient and shared with most animals.

Menon and Levintin (2005) relate the activation of a dopaminergic mesocorticolimbic system by music to positive arousal in mood and cognitive tasks' performance.[37] Healthy individuals listening to music after a stressful event reveal a cortisol level reduction (Khalfa et al. 2003). This reduction facilitates hippocampal function, which is involved in verbal memory (Zimmerman et al., 2008). Since

---

[34] Brain regions which are also involved in beat perception and synchronization.

[35] In this line, Koelsch (2010), Salimpoor et al. (2012), Peretz (2010) and Perani et al. (2010) also maintain that music has recycled emotional circuits which have evolved for processing biologically relevant stimuli, provided that "musical emotions engage core brain structures devoted to emotional processing, such as the amygdale and ventral striatum, even in new-borns" (Aubé et al., 2013).

[36] Perhaps there is a link here between the earlier function of human protomusic for courtship and pair-bonding and its implicit sexual and food stability reward.

[37] An increase in verbal memory and focused attention, as well as a decrease in depression, among patients who engaged in music while recovering (Särkämö et al. 2008).

cortisol production is regulated by signals from the hypothalamus, and this is, in turn, influenced by projections from the limbic system regulating emotions (Koelsch, 2010; Peretz, 2010), music voice-like acoustic cues may, in some way, affect the limbic system, unfolding the following chain of reactions.

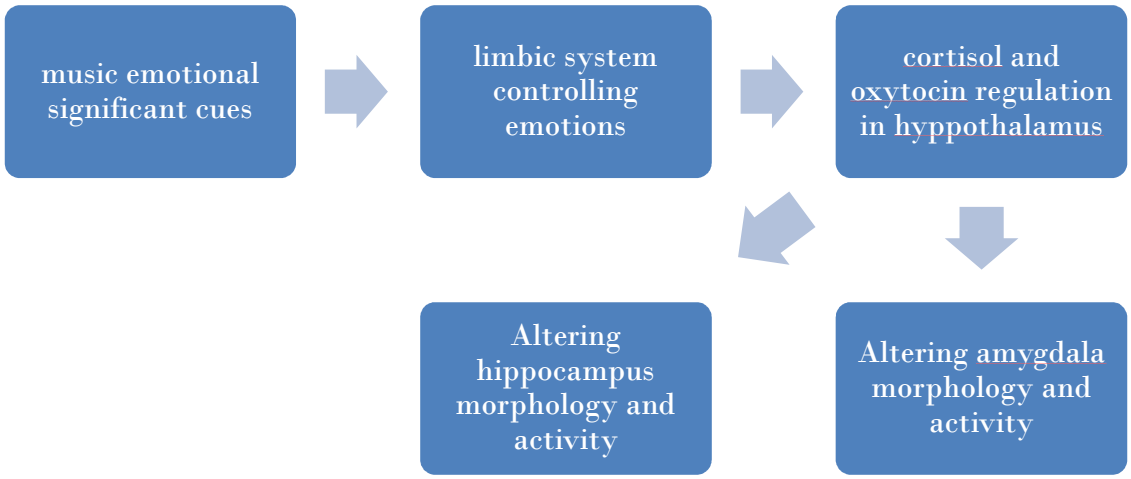


Figure 3

This scheme shows that the neuroendocrine system, via hormonal regulation, has an important role in neural morphology and activity. The hypothalamus regulation of cortisol and oxytocin levels in the blood, once manipulated by the limbic reactions to the music emotional significant cues emulating vocal sounds, can alter the hippocampus[38] and the amygdale functions and morphology. Therefore, music can promote morphological and functional changes in our neural system.

Other authors have posit that music only temporally coordinates emotion-inducing mechanisms such as expectancy —and its fulfilment or violation—, brainstem activation, past-event associations, visual imagery and emotional voice-like acoustic cues (Juslin and Västfjäll, 2008), as a complex emotional experience.

3.3 *Basal ganglia in beat*

Beat perception or regular pulse induction, which marks equally spaced points in time, could function as the ability to encode temporal intervals as multiples or subdivisions of the beat. This ability results in a better reproduction and discrimination of the rhythm, analogous to "chunking" mechanisms, which reduce complex patterns to simpler components.

[38] This regulation of cortisol and oxytocin also alters the birth of new cells in adult hippocampus.

Activations in the premotor cortex (PMC) and supplementary motor areas (SMA), cerebellum and basal ganglia —creating a striato-thalamico-cortical loop network—, are also reported in neuroimaging studies on timing. Experiments involving internal subjective accents —e.g. listening to unaccented isochronous rhythms— show the response of the putamen, caudate and pallidum, as well as PMC and SMA, right and left STG, and right cerebellum.

After contrasting healthy controls and patients with Parkinson's disease, Jessica A. Grahn (2009) concluded that the basal ganglia are strongly linked to the internal generation of the beat (i.e. the pulse). In fact, the activity of basal ganglia is greater when the external cue marking the beat is weak, therefore motivating an internal generation. In Parkinson Disease, a "progressive cell death in the *substantia nigra* that decreases dopamine release by the striatum, affecting excitatory input to the putamen" (Grahn, 2009) leads to an impaired extraction of the beat structure of novel rhythms, which points out that the putamen may encode information about beat timing —facilitating precise movement control for motor areas. Giving more evidence to the role of putamen in beat, a higher activity connecting the putamen to the cortical premotor and supplementary motor areas during rhythmic beat perception has been found in trained musicians, together with an increased connectivity between cortical motor and auditory areas.

SUMMARY

In this part we exposed that music and language are two uniquely-human faculties that are cross-culturally linked to social contexts. While music deals with emotions, language expresses propositional and lexical meanings. Both faculties compile a mosaic of independent traits and components that have gradually evolved and been selected for other purposes. Our ancestors, before speaking, used protolanguages as communicative systems, such as the musical protolanguage proposed by Darwin (1871). It may have given rise to music and language (its phonology). However, both may have appeared recently about 100KYA, together with a symbolic thinking in *Homo sapiens*. As neuroscientific research reveals, some neural mechanisms (hierarchical processing) and perceptual properties (emotional prosody) are shared by these faculties. But music and language also compute domain-specific elements (pitch or rhythm, in the case of music, or semantic meaning, in the case of language). Interestingly, music is found to impact our limbic system, as well as its beat activates basal ganglia.

PART II

PART II: PROTOMUSIC: RHYTHM AND TONALITY ORIGINS

In the previous part we reviewed past and present views on music evolution, as well as some speculations of how music may have gradually evolved in tandem with the emergence of language and grammar. It is undeniable that music, as we know it, has changed from whatever it was in the past, especially after that *Homo sapiens* developed its unique linguistic and symbolic thinking (assuming Tattersall (2013), see fig. 4).

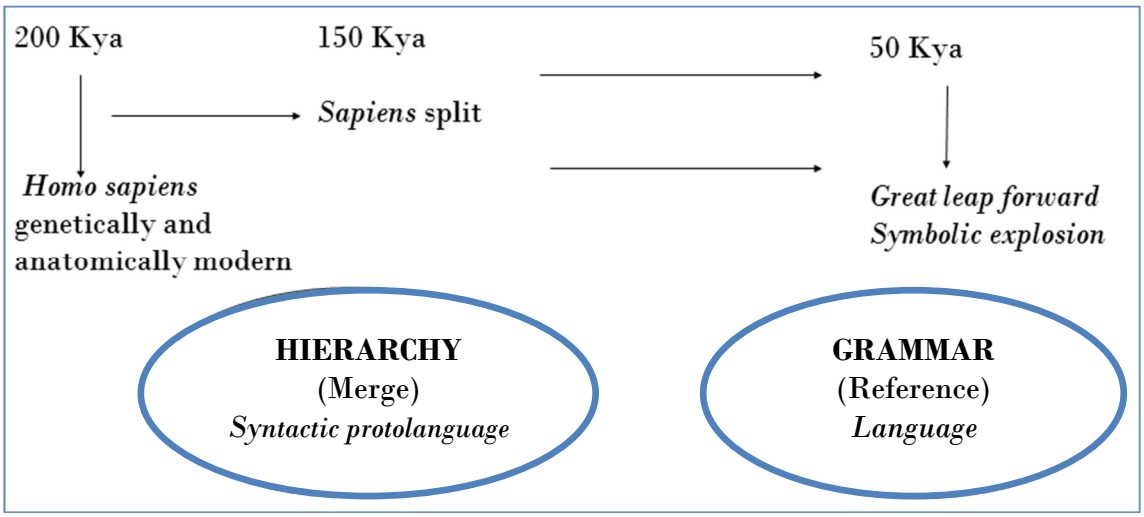


Figure 4

Moreover, music per se has not existed in the past, but only some traits showing *musicality*, which is something that can be selected and found in animals as well. Preceding both language and music, we assumed Darwin's proposal of a musical protolanguage functioning as a communicational system expressing emotional needs and physiological states, which was sexually selected and developed by increased mental powers through different stages. As Fitch's revision of the Darwinian musical protolanguage, we take in account sexual selection (mate choice or pair-bonding mechanism) and kin selection (parental care through *motherese*), although social group cohesion may have been involved as well. From other approaches, we concluded that processing the emotional cues of prosody may have been selected, because it is a trait remaining in both current speech and music, and it could be specifically affected —as it occurs with Alexithymia disease.

Now we will try to unify these views so as to propose two gradual stages for a musical protolanguage, first evolving into a rhythmic protomusic and later into music. At the same time, this musical protolanguage may have given rise to the

phonology of language. We speculated that from a musical protolanguage made up of vocalizations —since we are complex vocal learners sensitive to the pulse—, we developed a rhythmic syllabic protomusic with an underlying beat, which in turn may have changed after the emergence of merge and its linguistic hierarchy: a protomusic with beat and its hierarchical organization (meter) may have appeared. Later, when our symbolic thinking was in place and grammar has arisen, music tonal-harmony appeared as a side-effect of our linguistic reference.

3. FROM A MUSICAL PROTOLANGUAGE TO PROTOMUSIC AND MUSIC

The present thesis supports the position that our ancestors underwent a musical protolanguage stage and that a (musical) rhythmic component was central to it. Given that most complex vocal learners can detect beat (i.e. pulse), and can entrain to it, and given that certain primates (i.e. gorillas, chimpanzees) develop rhythmic behaviours in the wild, it seems highly plausible that song-like human vocalizations (similar to the duets found in singing gibbons) may have manifested an isochronous component as well. We will argue that this isochronous incorporation constitutes the precursor to our music, stemming from a musical protolanguage to protomusic.

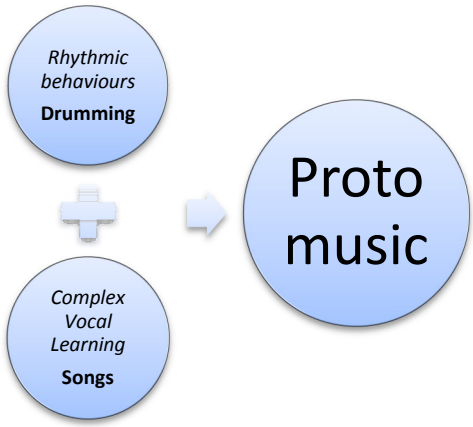


Figure 5

With regard to language phonology, early human vocalizations could have linked together the precursors of consonants and vowels in a musical protolanguage. It would not be the unique case, given that geladas are known to produce proto-syllabic vocalizations during grooming. This syllabic linking may have resulted from different physiological and communicative mechanisms, with their separate neural correlates. These proto-syllabic cycles could be convincingly explained by the content/frame theory developed by Peter MacNeilage (2008). It exposes that our syllabic spoken language could have come from modified closed-open cycles, which, in turn, may have come from the cyclical motion of the jaw and the movements of other articulators, including: the lips, the vocal folds or the tongue.

Notwithstanding, in our thesis, we defend that the rhythmic movement of cyclical syllables, rather than coming from ingestive motor patterns, must have derived from intentional lip-smacking, which is found in other primates as well as in baboons and macaques. In fact, the rhythm of lip-smacking matches the rhythm of human syllables. Moreover, geladas' vocalizations, by showing the simultaneous combination of lip-smacking and phonation, give further support to the non-ingestive theory.

The existence of a musical protolanguage made up of discrete elements seems to be supported by the kind of songs observed in gibbons: duets comprised of discrete elements. In fact, one can go further and consider that this discrete musical protolanguage was already syllabic in the sense of made up of consonant and vowel-like elements. A syllabic musical protolanguage is supported by the innate babbling of human infants and by the existence, even in some monkeys like geladas, of vocalizations akin to vowels and consonants from an acoustic point of view. Hence, we support a musical protolanguage whose vocalizations were discrete syllables.

A discrete, and perhaps syllabic, elementary musical protolanguage equipped with beat could have gained meter through co-opting the linguistic hierarchy and, in turn, provided the basis for the "signifier" part of words —from the saussurean dichotomy "signifier-signified"— when the externalization of language took place.[39] Then, with everything in place for language, a linguistically-mediated thought could have impacted music providing it with a system of reference to quiescent points. In sum, current music incorporates a musical grammar consisting of hierarchically-ordered pitches and chords bearing different functions according to their structural position and their acoustic morphology (harmonic spectra). Equivalently, it was the grammatical organization of music that led to harmonic syntax.

This picture [fig. 6] depicts the evolution from a prosodic protolanguage to our current music. To reach it, it may have crossed two intermediate steps: a musical protolanguage and rhythmic protomusic.

---

[39] We deffend that the externalization of language cognitively changed our mind, because it involved grammar (i.e. reference) and lexicon (indexed concepts via phonology), which were added to an internal mechanism of combining elements from different domains.

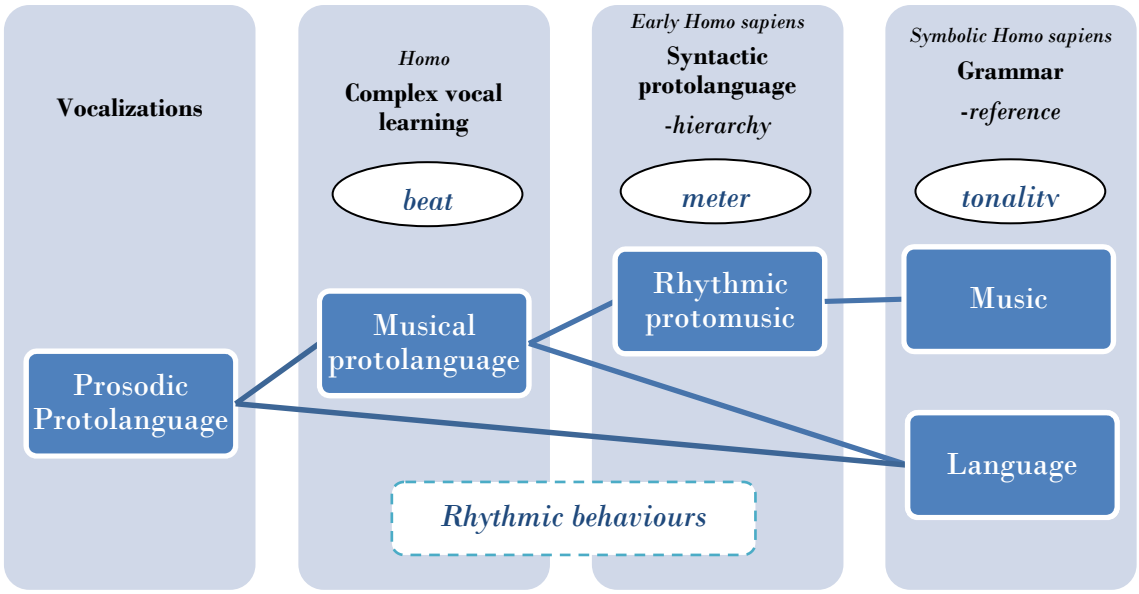


Figure 6

A musical protolanguage or protomusic needs not be built on scales of discrete notes from a small set of elements from the beginning. It was after the emergence of hierarchy that pitches became interrelated and constrained by rules of hierarchical relations, thus leading to a fully-fledged musical grammar. In fact, this step came after the rhythmic protomusic stage, which was only based on beat (i.e regular pulse) and metrical organization.

Rhythmic pulse is very relevant to coordinate group behaviours, as well as meter is crucial for music and dance integration. Perhaps they all moulded our music-ready brain through beat induction and timed motor production. Although musical rhythm was present in the beginning, musical grammar may have come later, hand in hand with linguistic reference. Assuming that, we propose that our current music faculty arises from the interaction of metrical structure (emerging from the hierarchical organization of the beat) and tonal-harmonic structure (emerging from musical grammar). In order to analyse the origins of both structures, the following sections will focus on rhythmic cognition and tonal-harmonic cognition, looking at the animal kingdom and brain studies to support a rhythmic protomusic hypothesis.

4. RHYTHMIC COGNITION

In this section we will review (1) how human cognition categorize rhythm, (2) which processes in our brain allow music beat and meter computation, (3) how they are related to animal musical abilities with respect to rhythm and (4) how they may have evolved to give rise to our unique ability to compute meter. Taking music as an acoustical, psychological and cognitive phenomenon, it is essential to know how its core components have arisen. After focusing on its rhythmic structure, by looking at the interaction between performance and perception, we will analyse the phylogenetic and neural roots of rhythm.

The cognitive process of categorization allows humans to recognize, classify and distinguish objects and events in the world. Similarly, categorical perception is fundamental in rhythmic pattern and timing. As it does not simply map discrete variables from a continuum (which would lose information), categorization functions as "a reference relative to which timing deviations are perceived" (Honing, 2013). Categorical boundaries can be influenced by metrical context because they are not fixed, which allows for variation in rhythm perception and timing. Categorization processing is also affected by top-down cognitive influences, the preceding musical context and the expectations from musical knowledge or earlier exposure.

5.1 *Rhythm and time perception*

While *rhythm* or *grouping* refers to "phenomenal patterns of duration in the world", marking sound onset to sound onset by changes in loudness, timbre, pitch or duration; *meter* refers to an «endogenous sense of rhythmic organization that arises in the perception of periodic stimuli», involving different levels of temporal structure. As London also points out, our musical rhythm[40] perception is active, involving top-down and bottom-up processing in different time scales (from 100ms to 5-7second), as well as concomitant motor behaviour.

Honing (2013) separates rhythm, "any series of sounds or events that has duration", into four basic components: rhythmic pattern, meter, tempo and timing.

[40] Justin London (2012) distinguishes *rhythm* from *time* in music, that is, between groups of durations of acoustical events in the world and the sense of beat cycles in the mind.

1) *Rhythmic pattern* consists of representing a pattern of durations on a discrete symbolic scale, as well as it relates to the process of categorization: "deriving rhythmic categories from a continuous rhythmic signal".

2) *Meter* is a hierarchically organized interpretation of pulse, usually in two or more levels of beat or *tactus* (the induced regular pulse), which yields a metrical framework to assign to the rhythmical signal. In addition, *rhythmic structure* or *grouping* arises from taking figural aspects of the rhythmic signal as a sequential pattern of durational accent, grouped at the surface level.

3) *Tempo* is impression of speed of the sounding pattern, related to the cognitive beat or pulse rate occurring over time.

4) *Timing* relates to the expectancy of sounding events: to the sensation of notes occurring earlier or later. *Expressive timing* is the "deviation from the most frequently heard version of a rhythm" (which depends on memory), rather than its deviation from a canonical integer-related version, notated in scores.

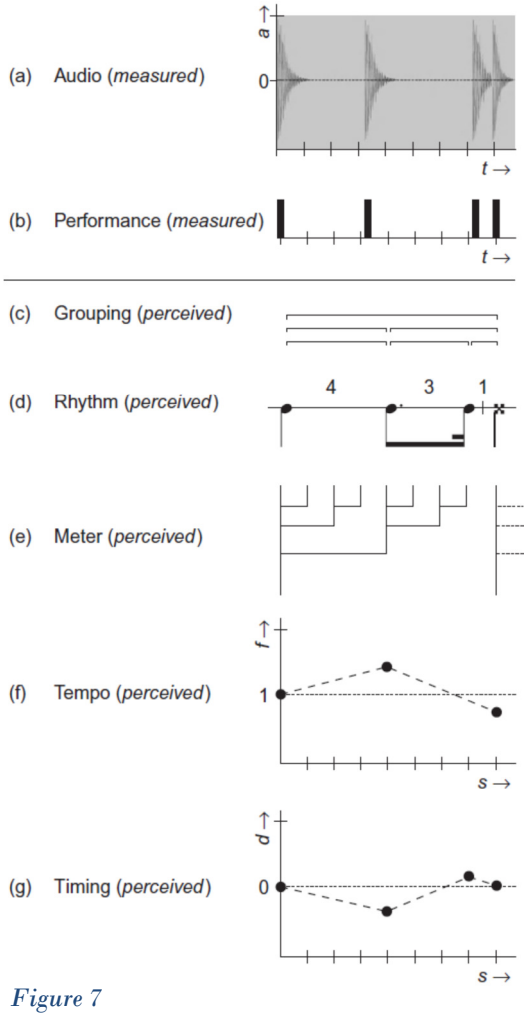


Figure 7

This figure (fig. 7) separates the external acoustic sound (*a* and *b*) from the rhythmically categorized perception of the acoustic beats over time. While *c* and *d* refers to "grouping" the acoustic sounds (minimally irregular in temporal duration) into isochronous patterns yielded from integer-ratio frequencies (1:2, 1:4, 2:3) of the beat or pulse, *e* implies a hierarchy of pulses in strong-weak patterns. Finally, *f* and *g* mean the relative timing perception generated by expectations and their violation or fulfilment. While *f* could be labelled by terms as *allegro* or *lento*, *g* could be indicated as *accelerando* or *ritardando*. With regard to timing and tempo relations, *timing* is tempo-specific in both production and perception, because rhythms are timed differently at different tempi.

Rather than perceiving rhythm as an abstract unity or a continuum, rhythm and timing is heard in "clumps": islets on a *chronotopological map* or a *rhythm chart,* based on an abstract mathematical notion which represents a visual space for all possible rhythms in all possible interpretations.

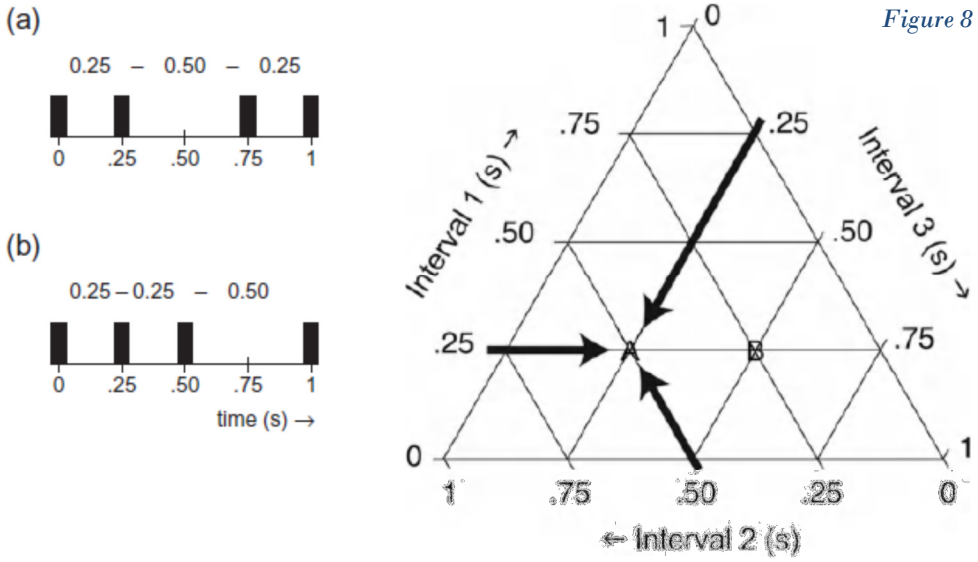


Figure 8

This figure extracted from Honing (2013) shows two sample rhythms (*a* and *b*) and how they are located in a chronotopological map based on three axes. These rhythm charts allow the testing of listeners' perception of slightly altered rhythms, which are mentally categorized as regulars.

Aside from a perceptual phenomenon theory of rhythmic cognition, another proposal is that of embodied cognition. It argues that our physiology and body metrics, as well as our body movement, influence rhythm perception. There are findings coming from babies supporting accentual-beat preferences after being rocked in duple or ternary-timed (i.e. 2/4, 3/4) lullabies. Laurel Trainor (2010) indicates that neural similarities in rhythmic auditory and motor circuits which enable synchronization through movement are also present in other species, such as crickets. Given the discovery of association between sensory- and auditory-motor systems (Zatorre et al., 2007), a hypothesis that metrical interpretation rests upon covert sensorimotor action seems well supported (Repp, 2007). However, another understanding of *meter* (as a musically-specific form of entrainment that allows us to synchronize a periodic aspect of our attention to environmental external rhythms in a perceptual pattern of accentually differentiated beats, i.e. beats as perceptual

abstractions of peaks of attentional energy) regards the *metrical structure* as a "mode of attending" (London, 2012). This is to say that *meter* is a by-product of our attentional cyclical system,[41] rather than a hierarchical musical structure per se.

In summary, rhythmic cognition can be divided into four basic domains: beat, grouping, meter and tempo, which together yield our rhythmic cognitive flexibility, i.e. human ability to "extract structural properties from music and interpret them in multiple contexts" (Ravignani et al., 2014). While *rhythm* is essentially a general structured pattern of temporal change, *beat* is its fundamental element consisting of points in time that occur in a perceptually periodic way (Patel, 2008). In turn, *grouping* corresponds to the organization of the musical stream into motives, phrases, and sections, while *meter* regulates beats in strong and weak patterns. In fact, *grouping* and *meter* can be treated as subsystems of rhythmic organization (Andrea et al. 2013), although they are considered the basic structural components of rhythmic patterns (Lerdahl & Jackendoff, 1983). In relation to their strength, beats are organized hierarchically (building metrical structures), where the level of the primary strong beat is traditionally called the *tactus*. Finally, *tempo* has an important role in the interpretation and perception of rhythms, because it is able to modify the grouping conditions and metrical hierarchy induction in listeners.

5.2 *Rhythmic processing mechanisms*

Tecumseh Fitch (2013) proposes a cognitive and comparative perspective on human rhythmic cognition, distinguishing two fundamental cognitive processes: *pulse extraction* from *meter induction*. While the former consists of converting "a periodic event sequence to an (unaccented) isochronic pulse stream", the latter consists of the "conversion of an event stream or unaccented pulse stream to a hierarchically-grouped metrical tree structure". These cognitive processes are indeed independent, because they can appear separately. *Metrical induction* is present in languages and poetry (but not *isochrony*) and seems unique to humans. Contrarily to what occurs in non-human animals, where *pulse extraction* and synchronized *entrainment* are found,

[41] Despite that this point will be analysed later, we assume that meter is indeed processed as a hierarchical structure of beats, although does square with an attentional explanation.

meter induction is not. Computationally, *pulse* involves detecting periodicity, whereas *meter* involves building hierarchical structures.

Beat induction is a cognitive skill that allows a regular pulse in music to be heard and to synchronize to it, and therefore allows dance and collective musical performance. The *beat* comes from a highly salient, periodic layer of articulation in the musical texture (between 400ms and 800ms) and does not need to be physically present to be perceived. The *meter*, as a cognitive phenomenon, is an emergent temporal structure of at least two levels of pulse that involves our perception and an embodied anticipation of rhythmic patterns (perceived periods of duration present in music). Therefore, perceiving rhythm must be seen as the interaction between the acoustic patterns and the listener projecting meter onto it. Origins of music in beat induction are supported by experiments which show that babies and newborns can detect beat and meter, regardless if they are bounced or rocked in time with the tested stimulus by their parents. Thus, this innate, domain-specific skill might be a "predisposition to extract hierarchically structured regularities from complex rhythmic patterns" (Honing 2013).

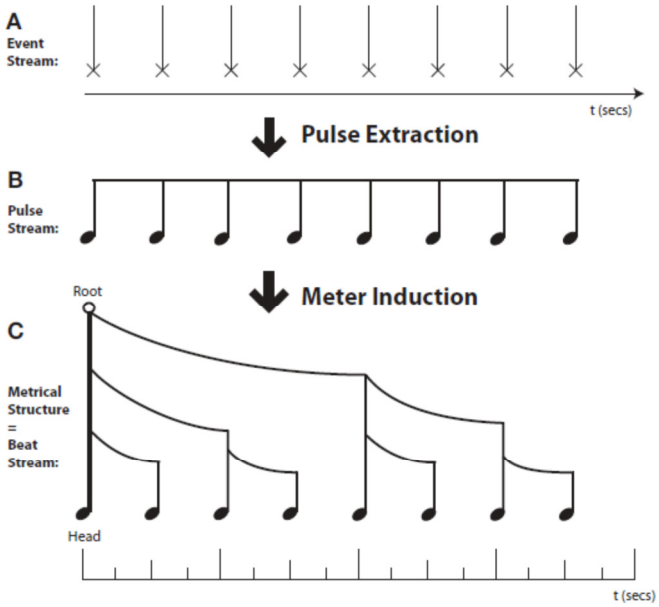


Figure 9

Fitch (2013)'s model of metrical trees equates rhythmic syntax to linguistic syntax considering them as sub-types of hierarchical processing, that are both dominated by a head node. In this way, he shifts the focus from a harmonic syntax (Lerdahl & Jackendoff, 1983) towards a rhythmic syntax. Since *pulse* and *meter* are cognitive constructs (not explicitly present in the raw acoustic signal) which is inferred by the listener, *rhythm* (like *pitch*) becomes a mental construct, which need not be identical to aspects of the signal. Once the pulse frequency is extracted from the incoming events, a downbeat (a prominence) must be located in the stream, thus creating metrical patterns of strongly-weakly accented events around which a hierarchical grouping of sonic events could arise, building a hierarchical structure with downbeats occupying the

head node position. As it occurs in linguistic constituents, the prominence of a musical event depends on its place in the overall metrical hierarchy, and not on its serial location.

Fitch claims that what permits to dance is this metrical structure (not only the pulse), and that «an event's prominence differs depending on the meter assigned by the listener», which create rhythmic and metrical expectancies, whose deviations (i.e. syncopations) or violations (breaking rigid isochrony or meter) conforms effects of surprise in the listener.

Against these views of headed rhythmic hierarchies in metrical cognition (Fitch, 2013; Honing, 2012; Patel, 2014), other approaches (Lerdahl and Jackendoff, 1983; Lerdahl, 2013) deny the existence of a hierarchical head in strong-weak patterns, arguing that this grouping structure does not align with melodic *anacrusis*.[42] Nevertheless, Fitch argues that anacrusis is explained by the interaction of melodic and metrical trees, wherein «pickup notes [are] *melodically* connected to the root of the following metrical tree but are not part of the tree itself». Fitch's Grouping Tree model, though hierarchical, is said not to be necessarily recursive, alleging to the existence of ternary measures or three beat rhythms: triplets.[43]

Periodic beat patterns are basic for every culture's music, permitting entrainment of rhythmic action to sound, as occurs in dance, due to a specific beat perception and synchronization (BPS)[44] mechanism: an «ability to perceive a beat in music and synchronize bodily movement with it». Beat induction "allows us to hear a regular pulse in music, to which we can synchronize [...] to dance and make music together" (H&P, 2012). Recent studies show that beat induction by appearing in young infants as well as in newborns must be innate rather than be the result of learning. While this skill also appears within other species, like some birds (parrots, hummingbirds and songbirds) and mammals (sea lions, Asian elephants...), non-

[42] Anacrusis: 'note(s) preceding the first downbeat in a bar', often configuring the initial melody.

[43] We will discuss the binary implications of ternary meter in section 4.4.2 (also in a foot note).

[44] BPS should be distinguished from simple pulse-based synchronization, since the first involves extracting a regular beat from a complex signal, flexibility in moving tempo and cross-modality rhythmic responses, whereas the latter only implies pulse-extraction from simple pulse trains, limited (if any) flexibility in movement tempo, and modality restriction (Patel, 2009).

human primates seems to lack it —empirical research in chimps demonstrates their limitations in inducing pulse (Fitch, 2013).

Beat induction, although somehow present in language through poetry, could be restricted to music. Meter, instead, shows correspondences between spoken language and music. These links are due to the existence of a few different metrical accent structures among which each language must choose.[45] Different from the perceived linguistic meter, the musical meter shows a stress pattern which maps regularly to the rhythmic tree, suggesting that musical meter has a simpler structure than speech. However, regarding beat induction, cases of "beat deafness" have been tested in people with normal language and normal musical perception (Phillips-Silver et al. 2011), demonstrating its constitutive independence. Assuming that BPS is not a language off-shoot, it may have been selected independently.

After having analysed rhythm categorization in perception and production, as well as *pulse extraction* and *meter induction*, now we turn to beat and meter origins, comparing how both elements are perceived and produced in animals.

5.3 *Animals with rhythms: drumming and songs*

Rhythmic synchronization is very unusual in nature, only appearing in certain anurans, arthropods, birds and mammals —including gibbons— (Bowling, Herbst & Fitch; 2013). Simultaneous acoustic or visual signal production in groups indicates precise patterns of temporal signal interactions, fundamentally based on synchrony or alternation (see Greenfield, 1994). Bowling et al. (2013) point out the important role of isochrony for the development of temporal regularity in vocalizations, because it "makes the behaviour of others predictable", suggesting that musical rhythm origins lie in "cooperative social interaction" facilitating precise temporal group coordination. After testing synchronization skills in human non-isochronic speech, and asking why this isochrony does not appear in a speech coming from a hypothetical musical protolanguage, the authors conclude that sexual selection is not the selective force promoting synchronization (because women and men manifest equal skills), but a more general "cooperative urge" for sharing experiences and emotions. However, they still accept as plausible a synchronous vocal display, i.e.

[45] In fact, these acoustic cues allow rats to detect and distinguish languages only by their rhythm.

chorusing (Merker et al., 2009), enhancing human capacity for isochronous signal production and entrainment.

Patel (2009)'s Beat Perception and Synchronization (BPS) tests (renamed "Pulse Perception and Entrainment" in Fitch (2013)) in non-human animals, together with Schachner et al. (2009) analysis of videos showing dancing animals, demonstrate that birds and mammals can infer pulse from music or visual inputs, as well as can follow and anticipate it through body movements. That is the case for Sulphur-Crested Cockatoos,[46] parrots, budgerigars, an Asian Elephant and a California Sea lion.[47] It is worth to say that BPS or PPE tests applied to non-human primates, concretely to chimpanzee *Pan troglodytes* (Hattori et al., 2013), offer evidence that they lack this ability, which by contrast is present very early in human new-borns. Although both the African Grey Parrot and the Sulphur-Crested Cockatoo maintained a consistent phase matching to the beat, only the latter displayed foot-lifting phase matched with the beat, that is, showing a motor flexibility similar to human highly flexible motor response in entrainment." While synchronization of movement to a musical beat develops spontaneously in humans, it does not occur in most animals. Rhythmic entrainment is distinct from beat perception and synchronization (BPS) because the latter "involves a periodic motor response to complex sound sequences […], can adjust to a broad range of tempi, and is cross-modal" (Patel et al. 2009).

Drumming is closely related to instrumental music —"the use of the limbs or other body parts to produce structured, communicative sound, possibly using additional objects" (Fitch, 2005)— because it involves the use of limbs to hit sounding objects or the own body, and it is developed in our closer relatives, the Great Apes, thus making the non-tonal percussive behaviour of drumming a nice human instrumental music homologue. Apart from gorillas, chimpanzees and bonobos, in which bimanual drumming is used to mark aggressiveness in fighting and hierarchy in social positions, only few other vertebrates, such as palm cockatoos, woodpeckers, kangaroo rats and desert rodents, are found to drum out rhythmic patterns by using either a stick, their own bill, or their hind feet (respectively).

[46] Snowball, sulphured-crested cockatoo video-link: http://www.youtube.com/watch?v=cJOZp2ZftCw
[47] Californian sea lion video-link: http://www.youtube.com/watch?v=6yS6qU_w3JQ

Surprisingly, for this work, bonobos, the most social great apes, are able to "maintain a steady drummed beat for at least 12s" (Fitch, 2005). Curiously, the singing skill of gibbons, i.e. their complex vocal displays in duets, is also developed in absence of experience, and is accompanied by a vigorous movement component (Fitch 2005), suggesting a possible homologue to dance. Hence, this reported behaviour reinforces our protomusic hypothesis by incorporating (together with drumming) complex, rhythmic gestural patterns —linking a common precedent for our current music and dance.

Apart from pulse synchronization and entrainment, a hierarchical metrical structure might also be present in the animal kingdom. Although evoked responses in electro-encephalographic signal (revealing mismatch negativity to metrical structure violations) have been found in human new-borns and adults detecting downbeats omissions, there is a lack of evidence of meter in other non-human animals (macaques, pigeons…). Nevertheless, more experiments should be made to confirm its complete absence. Personally, we have the intuition that metrical hierarchy must be a by-product of our language capacity, even if it is supported and enhanced by our attentional mode of perceiving.

5.2 *Two hypotheses on rhythm origins*

Looking at the basic capacities allowing rhythmic cognitive flexibility, two main hypotheses have arisen after having found beat entrainment and rhythmic behaviours within different species: vocal learning and social convergence; which are strongly related to language and music emergence. On the one hand, the former predicts the appearance of beat entrainment —processing of relative timing of events by expecting their phases or periods, and adjusting these expectations to actual occurrences (Grahn, 2012)— in vocal learning species with vocal mimicry skills, because of the tight connection between motor and auditory brain regions that they present. On the other hand, the latter predicts rhythmic abilities as a social coordination instinct, where group synchronization arises from rhythmic isochrony, which permit cooperation in auditory signal generation (Fitch, 2012).

This vocal learning hypothesis for BPS relies on brain circuitry correspondences: motor-auditory links and overlapping regions as basal ganglia and supplementary

motor areas. However, BPS may also require circuitry for open-ended vocal learning, permitting novel sound patterns imitation throughout life, and the ability to imitate non-verbal movements. Comparative data from animal rhythmic behaviour has demonstrated that pulse extraction and synchronization is not a property unique to humans. Furthermore, not only synchronized behaviour has been attested for audio signalling in insect and frog species but also for visual signalling in fireflies. However, only humans shows a cross-modal capacity to synchronize, and at different tempos.

While data coming from parrots and the Asian elephant seem to support the idea, those studies revealing an absence of entrainment in other vocal learners (songbirds kept in human homes, captive dolphins and orca exposed to music…) challenge this hypothesis, suggesting vocal learning is necessary, but not enough for BPS or PPE. In the case of Californian sea lions, which are otariids and the unique non vocal learners' members within the pinniped family, one can defend that the neuronal connections from a common ancestor shared with walruses and phocids (both vocal learners) still remain in place.

As it has been explained above, Patel's (2006) hypothesis is that vocal learning and rhythmic synchronization are linked: concretely, BPS is a consequence of a selected vocal learning ability. Complex vocal learning (CVL) is the ability of learning to produce complex acoustic communication signals based on imitation (Patel, 2009). These links between BPS and CVL are based on a tight auditory-motor interface integrating auditory perception with rapid and complex vocal gestures promoted by vocal learning, as well as on particular modifications of brain substrates, like vocal learning birds' basal ganglia —structure involved in human beat perception from music. However, CVL is not enough, and BPS needs additional foundations as: open-ended vocal learning, non-vocal movement imitation and complex social group life (Patel, 2009). As neuroanatomical research suggests, homolog brain circuits involving the striatum, thalamus and forebrain, appear in vocal learner birds and mammals, in spite of their divergence 200 million years ago (Jarvis, 2007), thus constituting a case of convergence or deep homology, with similar underlying brain mechanisms. Whereas adult human BPS seems to differ from animal BPS, infant human BPS is found to be closer to animal BPS patterns,

especially in "sporadic synchronization" —limited periods of genuine synchronization to the beat.

Rhythmic entrainment to music is no longer unique to humans, since it is found in several bird species and mammals. This capacity to move the body or the limbs following an external beat is necessary for music playing and dancing cross-culturally. However, it is relevant the absence of this rhythmic entrainment in non-human primates, since "they naturally engage in 'drumming' in the wild" (Fitch, 2009). The fact that gorillas roughly beat their own bodies or objects and chimpanzees drum on rainforest trees with their feet or hands (which generates certain rhythmic signals), plausibly suggest a drumming propensity in our last common ancestor. This tree (rebuilt from Ravignani (2014)) depicts these findings:

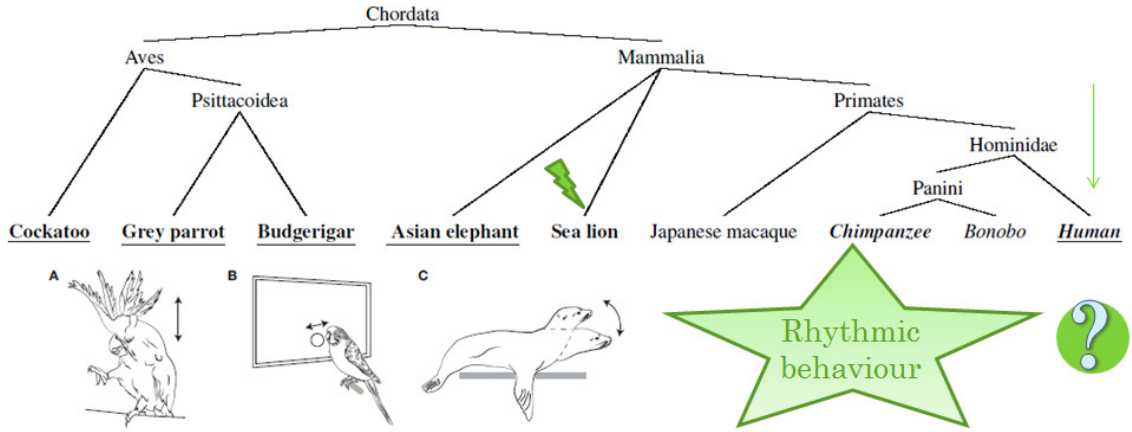


Figure 10. Underlined species are widely accepted vocal learners. Italicized species have rhythmic behaviours.

The vocal learning hypothesis (Patel 2006, 2008), although rightly based in a cross-modal linkage between auditory and motor brain areas, does not explain why certain complex vocal learner species, though possessing "vocal mimicry", lack the ability to entrain. Looking for an explanation, Fitch (2009) proposes that engaging in social action might develop a key role in auditory entrainment, since entrainment is present in group oriented behaviours: such as parrots' vocal "badges" of group membership and children's better entrainment in socially-engaged game-playing contexts. In this line, the correlation of social group behaviour with entrainment to beat and complex vocal learning seems to be coherently related to the Darwinian hypothesis of a greater social interaction pushing human cognition and selecting increased intelligence.

5. VOCAL LEARNING AND OTHER HYPOTHESES

As we have seen, Patel (2006) observed that animals showing BPS were almost all vocal learners. For this reason, an interesting hypothesis linking these two abilities could be established: "selection for vocal learning might lead to a capacity for rhythmic entrainment as a side-effect" (Fitch, 2013). This hypothesis is based on the narrow neuronal connections between auditory and vocal motor systems, which seem to be unusual in vertebrates without the vocal learning ability. Concretely, complex vocal learning may have arisen in both mammalian and avian evolution, and in humans it is tightly linked to speech. In fact, it allows us to learn the socially-shared open-ended vocabulary of spoken languages. Thus, it seems that the selection of complex vocal learning might have promoted more general connections between auditory input and motor behaviour, a linkage that is tested by researchers such as Schachner et al. (2009), Hasegawa et al. (2011), Cook et al. (2013) and Hattori et al. (2013).

6.1 *Vocal behaviour in animals*

Vocal learning is an ability which permits to modify the acoustic and syntactic structure of own species-specific sounds. It is distinct from auditory learning, which is present in most (if not all) vertebrates and consist of forming memories of heard sounds, because vocal learning, although depending on auditory learning, consists of imitating and improvising upon sounds. Vocal learning is more restricted, only found in three avian clades: songbirds, parrots, and hummingbirds; two marine mammal clades: cetaceans (dolphins and whales) and pinnipeds (seals and sea lions); elephants, some bats and humans. In contrast, its presence in non-human primates is dubious, because it would only consist of pitch little changes of innate calls and imitating sounds without using the larynx.

Egnor and Hauser (2004) distinguish three vocal learning behaviours in animals:

1. Vocal comprehension learning: appropriate response to vocalizations
2. Vocal production learning: spectrotemporal features of vocalizations are modified after auditory experience
3. Vocal usage learning: the right use of a call in an adequate social, ecological context

Vocal production learning is obviously present in songbirds, but it does not seem to occur in non-human primates' development.[48] It has been found acoustic variation between social groups (dialects) and acoustic convergence (conspecific vocal behaviour matching) in adult non-human primates' vocalizations, thus supporting that social context affects their vocal production, maintaining and advertising social group membership. While vocal plasticity appears stronger during development in humans and most songbirds, it is hard to detect in the development period of non-human primates. In fact, vocal plasticity in adult non-human primates "consists of a subtle acoustic change on top of an innately determined call structure"[49] (Egnor and Hauser, 2004), which is quite different from human vocal plasticity (found in language acquisition), which involves subcortical and cortical brain structures.

Assuming that we have the "ability to acquire new vocalizations or modify the spectral or temporal structure of existing vocalizations based on environmental cues" (Armador and Margolias, 2011), we have to focus on how the "processing of auditory cues that are memorized" and changed in motor patterns is implemented in the brain. It has been found that non-learning species that produce innate vocalizations only possess midbrain vocal nuclei. Besides, while call production (innately-specified vocalizations) involves the brainstem and the midbrain system, song production (learned vocalizations) recruits the forebrain system.

Vocal mimicry of human speech has been attested in parrots, songbirds and seals. Its selection suggests an auditory-motor linkage, which may have promoted entrainment. In the absence of vocal mimicry, other factors per se, such as phylogenetic proximity to humans, exposure to music, movement imitation and complex social structure, do not entail entrainment to pulse. The evolution of vocal mimicry in avian species is associated with parallel modifications to the basal ganglia, the same mechanisms that support musical beat perception in humans (Schachner et al., 2009). Therefore, vocal mimicry selection has come hand in hand with basal ganglia modifications, which promoted a tight auditory-motor coupling. Since entrainment does not appear in avian species in their natural behaviour, vocal

[48] Although it is found to some extent in primates' adulthood, in both sexes, and in a wide variety of call types: contact and alarm calls and sexual advertisement

[49] This call structure implicates the anterior cingulate cortex, supplemental motor area, motor cortex, cerebellum and subcortical structures (as the periaqueductal grey).

mimicry should be what it might have been selected. The parallel case may have occurred in the human lineage, and, once complex vocal learning was selected, our rhythmic behaviours may somehow participate in our vocalizations as well.

6.2 *The evolution of vocal learner birds' brain*

Although certain elements of call production may have been present in early vertebrates 400 million years ago and they are still well preserved, vocal learning has evolved intermittently and independently in distinct higher vertebrates lineages. Studies on mammals and birds indicate a "strong forebrain regulation of descending motor pathways arising from non-primary auditory forebrain pathways" (Armador and Margoliash, 2011). Given that only the three vocal learning bird groups show seven similar telencephalic brain structures [see the picture below], Jarvis (2006) proposes that these set of seven vocal brain nuclei might have evolved into a complex behaviour in a common ancestor within the past 65 million years, strongly constrained by epigenetics and followed by independent losses.

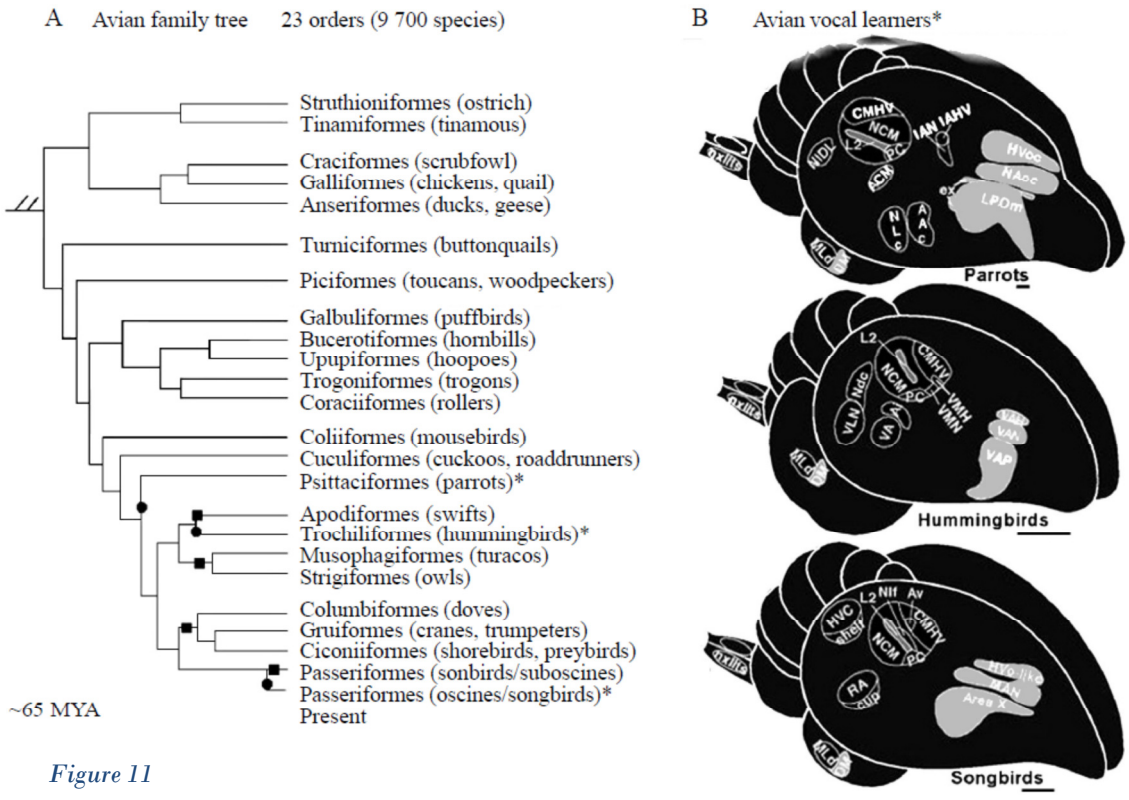


Figure 11

After comparing the seven areas through gene expression analyses (a behavioural molecular mapping detecting the transcription factor ZENK), Jarvis (2006) distinguished hearing-activated areas from vocal-activated areas, the former

creating an auditory pathway similarly located in vocal learning and non-learning birds. The seven vocal-activated areas, can be divided into two groups: (i) a posterior vocal nuclei located away from auditory areas in parrots, adjacent to them in hummingbirds and embedded within them in songbirds, thus forming a posterior vocal pathway for learned vocalizations; as well as (ii) an anterior vocal nuclei, within the forebrain, forming a loop connecting the cerebrum and the thalamus. Jarvis points out that in songbirds, "the anterior vocal pathway may be responsible for vocal learning and some as yet undefined role in the social context of singing, as well as song syntax". It applies to parrots as well, although their differences arise from the interactions between posterior and anterior vocal pathways.

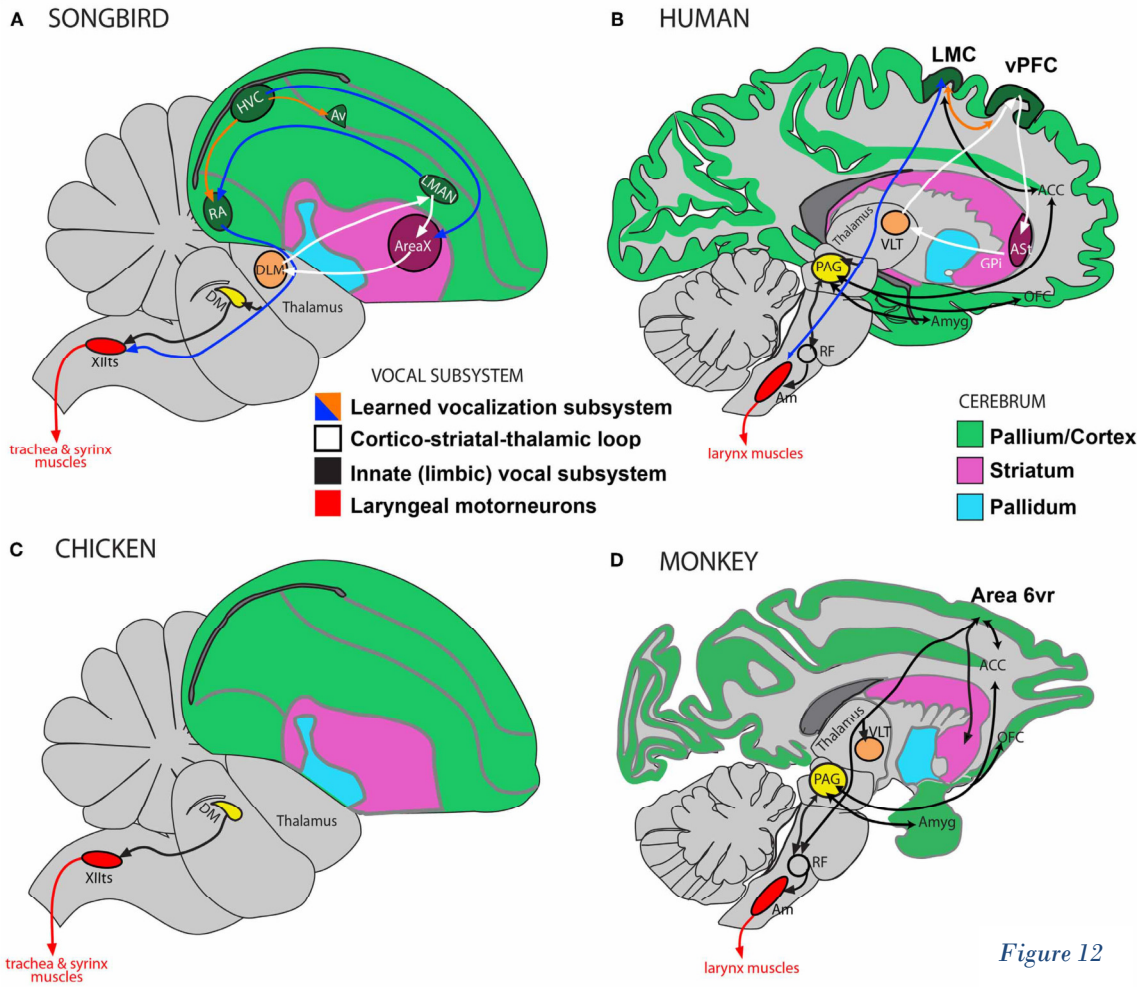


Figure 12

If the cerebral nuclei for vocal learning are divided into a posterior vocal pathway (PVP) for the production of learned vocalizations and an anterior vocal pathway (APV) forming a loop for the control of vocal learning, similarities with the circuitry in mammals arise, because PVP projects to motor neurons (as it occurs in mammalian brain motor cortex projections) and APV loop resembles mammalian

cortical-basal ganglia-thalamic-cortical loop. As such, this avian system of organization is comparable to the mammalian six-layered cortex connected to basal ganglia, suggesting a deep homology between birds and mammals' vocal learners (see the picture above [fig. 12] extracted from Jarvis and Petkov (2012)).

In short, as Jarvis (2006), we defend that a common ancestor of birds possessed vocal learning and the seven cerebral nuclei. However, this trait could not be manifested in certain orders because of epigenetic constraints imposed by the environment and the animal morphology: survival cost or predation danger, syringeal and respiratory system, and so on. Furthermore, assuming Jarvis (2006), the vocal learning system could be a universal brain structure, even for mammals, perhaps inherited from a common reptilian ancestor with avian, similar to the auditory pathway in vertebrate groups, permitting auditory learning.

M.A. Arbib and A. Iriki (Arbib, 2013) summarize Jarvis' vocal learning findings relating language and music evolution to birdsong evolution, in the following points:

1. All vocal learning species, even those with evolutionarily quite distant lineages, share neuroanatomical circuitry that is topologically similar, even when the concrete neural structures comprising each component in each species may be different.
2. A common group of genes, including the ones that guide axonal connections, are commonly, but specifically, expressed in these circuits in vocal learning species, but not in closely related vocal non-learning species.
3. These shared patterns of neuroanatomical circuitry and gene expressions necessary for vocal learning may be coded via common (but still not evident) sequences of genes that are language-related in humans and which start functioning upon environmental demand at different evolutionary lineages (a deep homology to subserve convergent evolution).

In contrast, Rizzolati and Arbib (1998) and Corballis (2002) hypothesize that vocal learning might have arisen in the hominid lineage from a gestural system — therefore differing from bird vocal learning origin— without considering the role of audition in human vocal learning. In the same line, Arbib and Iriki (2013)'s hypothesis[50] supports a gestural origin that is based on a mirror neuron system. This

[50] Arbib and Iriki (2013) defend a the notion of a language-ready brain arising from a niche construction as a bridge between biological and cultural evolution, a process based on altering the relation to the environment, thus changing the adaptive pressure constraining species evolution, as well as altering the cultural niche in which human evolve so as to construct new intentional and neuronal niches, wherein new behaviours remodel the brain by social contact.

system would function for imitation, intention attribution and language, but it would be activated in monkeys[51] and primates as well, during action recognition, manual dexterity and grasping, and communicative gestures. Given that vocal learning is lacking in non-human primates, and it has not led to language in vocal learner species, other mechanisms must be also implied in language emergence, such as a communicative gestural system, social community living, orofacial gestures, and intentional behaviour. In fact, these other systems also promoting language could be seen as necessary ingredients of the increased "mental powers" proposed by Darwin, which enriched human cognition and allowed *Homo sapiens* to pass from musical protolanguage to language.

6.3 *An audiomotor hypothesis for beat evolution*

Looking at our closest phylogenetic relatives (i.e. primates), an audiomotor evolutionary hypothesis is proposed by Merchant and Honing (2014). It decomposes the neurocognitive mechanisms underlying interval-based timing and rhythmic entrainment, so as to suggest their gradual emergence. They also claim that humans and other primates share interval-based timing, but that rhythmic entrainment ability is only partially shared.

Human rhythmic entrainment implies two features: tempo or period matching, "the period of movement equals the musical beat period", and phase matching, "rhythmic movements occur near the onset times of musical beats" (Merchant and Honing, 2014); both based on temporal anticipation (Repp, 2005). Moreover, this cognitively complex auditory-motor interaction shows flexibility to synchronize to broad range of tempi, as well as integer rates of fractions and multiples of the basic beat (Honing, 2013), suggesting a human mind access to distinct levels of periodicity, selected as the beat at every case (Drake et al., 2000).

Merchant and Honing (2014) challenge the vocal learning hypothesis (Hasegawa et al. 2011, Patel et al. 2009; Schachner et al. 2009) arguing that the studied sulphur-crested cockatoo showed only occasional periods of synchronization and that Cook et

[51] Arbib and Iriki (2013)'s mirror system hypothesis links macaques F5 brain region to Broca's area, as well as its connection to vocal folds found in squirrel monkeys, to some extent explaining a protospeech emergence coming from controlling a protosign.

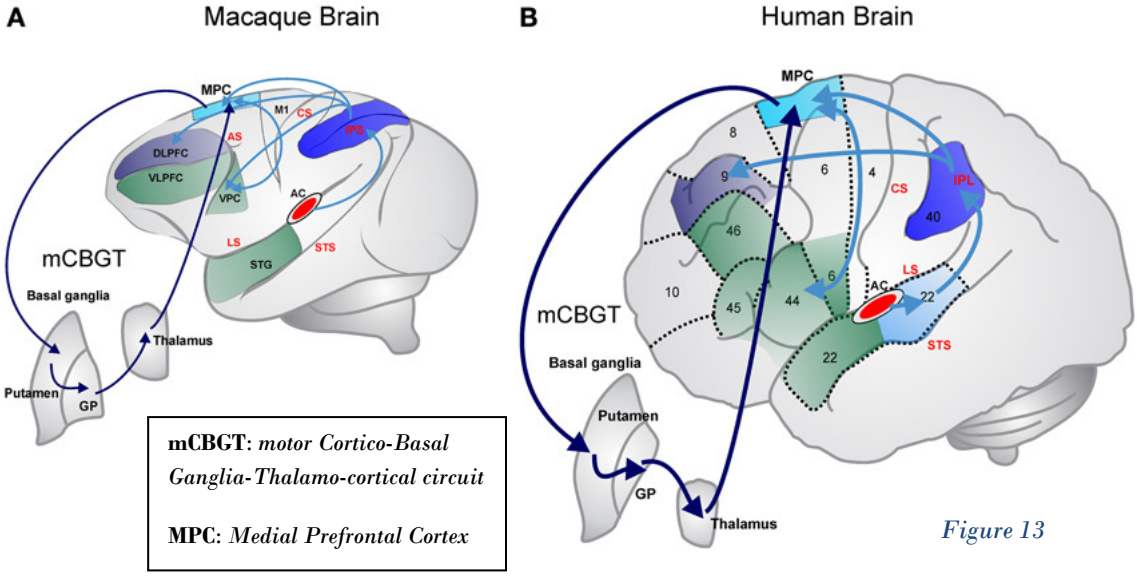
al.(2013)'s Californian sea lion is not considered a mimic vocal learner. The gradual evolution of complex vocal learning proposed by Petkov and Jarvis (2012) would shift beat entrainment to a gradual development of auditory-motor skills, part of them already found in non-human primates. Hence, rhythmic entrainment would be gradually developed in primates across evolution, selecting some different constitutive properties. In monkeys, for instance, there is a preference for visuomotor integration, manifested in behavioural imitation during socially coordinated actions with some level of rhythmic entrainment.

Merchant and Honing (2014) hypothesize that the similar timing performance for single intervals found in primates and the rhythmic entrainment gradually increased in anthropoids may depend on the neural system defining "the nested hierarchical properties of sequential and temporal behaviour", computing single sensorimotor associations, simple action chunks and superordinate action chunks. They report that macaques' performance of single interval tasks —such as interval production, categorization and interception, i.e. rhythmic grouping— is comparable to human skills. However, their multiple interval tasks —such as rhythmic entrainment, synchronization and continuation— differ from them. It may be due to a strong coupling absence between the auditory and motor systems, which is otherwise found in complex vocal learners.

Rhythmic behaviours, such as music and dance, engages a motor cortico-basal ganglia-thalamo-cortical circuit [henceforth, mCBGT], also used in sequential and temporal processing, which controls voluntary skeletomotor movements including SMA and the putamen (Coull et al. 2011). Studies in monkeys also reveal the engagement of a mCBGT circuit for perceptual and motor aspects of timing and control of movement sequences.

However, while mCBGT circuit in humans shows different loops responsible for the concatenation of sequential auditory information (or formation of chunks) and for temporal chunking of sensory information, starting in the anterior part of Broca's area and its right homologue, "the anterior prefrontal CBGT and the mCBGT circuits in monkeys might be less viable to multiple interval structures, such as regular beat", perhaps due to monkeys partial development of Broca's area

and its association with basal ganglia and premotor areas. In addition, human direct connections between medial and ventral premotor areas and Broca's area are reduced to a smaller tract in macaques. We can compare macaque and human brains and pathways, as they are shown by Merchant and Honing (2014).



mCBGT: *motor Cortico-Basal Ganglia-Thalamo-cortical circuit*

MPC: *Medial Prefrontal Cortex*

Figure 13

All these findings suggest that the similarities among primates in executing and perceiving single interval timing may depend on the conserved functional-architecture of medial and ventral premotor areas and putamen forming the skeletomotor mCBGT loop, which may permits an abstract neural representation of time during rhythmic behaviours in primate lineage.

6. A RHYTHMIC BRAIN

We have compared rhythmic behaviours in animals following two main hypotheses: the complex vocal learning hypothesis permitting entrainment to the beat and the social convergence hypothesis of rhythmic behaviours in primates, which were, to some extent, linked together in the audiomotor theory. We will now focus on how beat and meter are processed by humans through attentional fluctuations that follow the environmental stimuli, via neuronal rhythmic oscillations that engage their firing synchronously with the rhythmic beat.

7.1 *The Dynamic Attending Theory on beat and meter*

The Dynamic Attending Theory (DAT) (Jones and Boltz, 1989; Large and Jones, 1999) focuses on "how attention is directed in time", and considers the metrical

structure as an active listening strategy,[52] rather than a simple rhythmic parsing mechanism. In other words, meter's dynamic structure permits "to facilitate future oriented attending, to direct perception and to coordinate behaviour with external events" (Bolger et al. 2013).[53] Thus, attentional dynamics (Large & Jones, 1999) aims to explain the listeners' response to time-varying events, proposing that internal oscillations or attending rhythms are able to entrain to external events and targeting attentional energy to expected points in time. This theory therefore postulates a coordinated relationship between external rhythms,[54] created by distal events, and internal rhythms, actively generating temporal expectancies.

Given that the brain must represent and process beat and meter periodicities, Nozaradan et al. (2011) provide electroencephalogram (EEG) evidence of neural entrainment to beat and meter showing that "beat elicits a sustained periodic EEG response tuned to the beat frequency" as well as "meter imagery elicits an additional frequency tuned to the corresponding metric interpretation of this beat". In fact, Nozaradan et al. (2011)'s support the resonance theory for beat and meter perception[55] (Large and Kolen, 1994), where the emergence of beat perception comes from the entrainment of neuronal populations resonating at the frequency of the beat, and where meter perception comes from higher-order resonance of subharmonics of beat frequency. Nozaradan et al. (2011)'s experimental results show that beat perception from a complex auditory signal elicits a periodic response in the EEG spectrum, appearing as a steady-state beat evoked potential (EP) at the beat frequency, as well as the voluntary binary- or ternary-metric interpretation of the

---

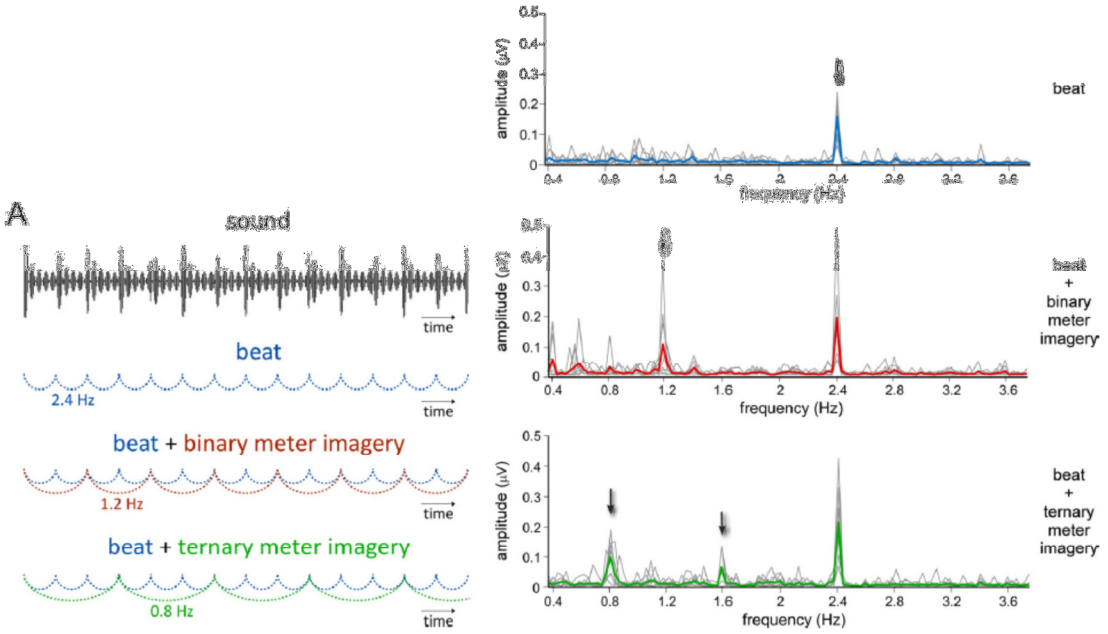
[52] A modulation of attentional resources over time occurs in correspondence with the induced meter, and the temporal events coinciding with the strong beats are highly anticipated.

[53] Their research support the role of meter in generating temporal expectations, in orienting attention, and in affecting pitch accuracy judgements and temporal differences.

[54] Large & Jones (1999)'s idea of rhythms broadly involves non-isochronous and isochronous time structures —such as the time patterns found in language or in music—, considering external rhythm as "a sequence of temporally localized onsets, defining a sequence of time intervals that are projected into the flow by some external event".

[55] As Nozaradan et al. (2011) state, "beats can be organized in meters, corresponding to subharmonics (i.e. integer ratios) of the beat frequency". However, beat perception could also consist of perceiving periodicities from not necessarily periodic sounds.

beat induces an additional periodic signal in the EEG at the corresponding subharmonic of beat frequency: *f*/2 or *f*/3, respectively.



These pictures have been extracted from Nozaradan et al. (2011) to report Evoked Potentials of different amplitude peaks depending on perceiving (1st) a steady beat at the frequency of 2.4Hz; (2nd) the steady beat categorized through a binary meter with a secondary peak marking the strong beat of the binary pattern (1.2Hz); and (3rd) the steady beat categorized through a ternary meter showing two secondary peaks: one marking the strong beat of ternary pattern (0.8Hz) and other marking the second weak beat —thus creating a binary subdivision of this ternary pattern.

Bolger et al. (2013) experimentally revealed that meter-driven orienting of attention over time is cross-modal:[56] processing visual and auditory targets equally. Their results indicate that the cross-modal meter effect is not restricted to isochronous stimuli, but that it also applies to general structured rhythms as well as highly variable rhythmic patterns.[57] Moreover, they also support an attractor hypothesis where highly expected positions within the meter structure create an anticipatory effect that presents greater attentional energy and leads to a secondary periodic oscillation over time interacting with the main metrical structure.

[56] Bochard, Tassin, Zagar (2013) also demonstrate that oscillatory attention tapped into cognitive processes combining visual stimuli, auditory rhythm and language. When a correctly-divided word syllable is presented on-beat, its visual recognition is quicker than when it is presented off-beat, showing that auditory rhythmic attention influence word recognition beyond the auditory modality.

[57] For instance, music complex rhythms are found to engage more oscillators whose coupling builds more stable beat periods, as well as their internal fluctuation facilitates the metrical structure access.

In order to achieve attentional synchrony, anticipatory attending is needed, in other words, "a temporal shift of attention that anticipates the onset time of a sound" (Jones, Moynihan, MasKenzie and Puente, 2002). While the beat is a psychological phenomenon relating to the subjective emphasis of certain events equally spaced in time, the emergent property of meter is characterized by multiple, hierarchically-related periodicities over time scales as well as is based on beats, perceived with different salience in a metrical structure of stronger and weaker events. The figure (fig.14) shows how the attentional energy fluctuates according to the metrical position of each beat, that is, as a function of the metrical salience of temporal positions.

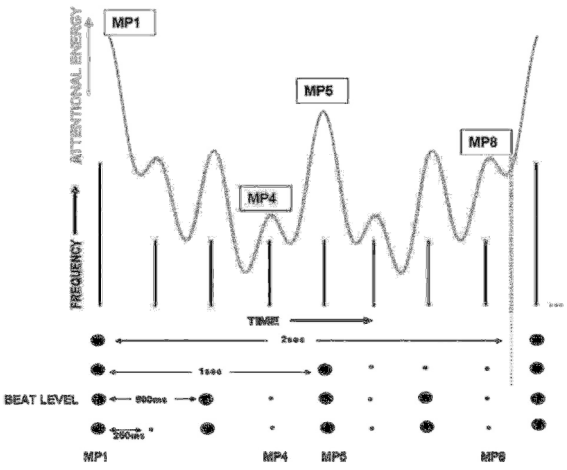


Figure 14

A top-down structure onto rhythmic experience is usually imposed by listeners, grouping isolated sound sensations into temporally-arranged system of ideas, describing a metrical hierarchy: "repeating intervals of equal duration, which are further subdivided into equal intervals" (Motz et al. 2013). Nested metrical levels indeed constraint the represented rhythmical structures: metrical patterns are preferentially treated, because rhythms occurring at equally-spaced subdivisions of a repeating cycle have perceptual advantages and are represented more accurately.[58] The "top-down imposition of metrical levels on [...](rhythmic) patterns is the dynamic result of oscillators resonating at integer multiples of the duration between beats" (Motz et al. 2013), in other words, neural oscillators synchronize with the external event which endogenously deploy focused attention to expected upcoming sounds. When a non-metrical sequence is perceived, this non-integer ratio sequence is systematically regularized, i.e. temporally distorted, shifted toward the nearest integer ratio subdivision, attracted to a stable metrical pattern. Therefore, individuals employ effective cortical representations of non-integer ratio sequences regularized towards the nearest expected metrical structure.

[58] A lot of evidence support that people are better remembering, reproducing, synchronizing with, detecting changes and making perceptual judgements when sound events sequences occur within a repeating time period equally subdivided into integer ratio relationships or harmonics.

The following table compiles the views of beat and meter from the DAT:

BEAT: Although acoustic features —such as loudness modulations, timbre variations, and melodic or harmonic accents— normally induce musical beats; prior musical experience, periodicity expectation and periodic motions' generation also generate mental representations inducing the musical beat. The *Dynamic Attending Theory* (George and Boltz, 1989; Large and Jones, 1999) considers beat perception as the synchronization of the beat periodic structure with the listener's attention, which leads to a periodic modulation of expectancy as a function of time. As it is suggested in primate studies, the neuronal beat-induced periodic EEG response varies according to the phase of the beat-induced cycle, which can elicit a cyclic fluctuation of the responding neuronal population excitability, thus modulating their amplitude.

METER: Although accents or periodic physical changes in beat —such as changes in duration, loudness, timbre or pitch— usually induce musical meter, its mental representation can also emerge (voluntarily or involuntarily) in cases of absence of accentual cues. Given that metric structure introduces additional periodicities based on beat frequency integer ratios (a natural human tendency in timing perception and production) suggests that metric interpretation enhances subharmonics within the neuronal network entrained by the beat. There is a human natural bias for binary structures in timing processing (Repp, 2005), even in ternary meter. It explains, for instance, the existence of musical "amiolias" —when the ternary meter briefly changes to binary meter— because of secondary attentional peaks.

7.2 *Brain oscillations in synchronzationi and anticipatory attending*

Predictive action representing temporal information is required to move in synchrony (playing an instrument, dancing...) with an auditory rhythm. Through the low-level cortical oscillations underlying sensory predictions, the brain creates an internal model of the world, inferring and predicting *what* and *when* is going to happen in the sensory environment (Giraud et al., 2012).

In some way related to BPS, Fujioka, Trainor, Large and Ross (2012) have found that "the periodic modulation of beta activity following fast-paced regular auditory stimuli could aid the initiation of movement", which supports an auditory-motor facilitation. Gamma (28-48Hz) and beta (15-20Hz) oscillatory patterns detect violations of expectations during the perception of an isochronous sequence of tones, with a first larger gamma-band response followed by an increased beta rebound. Passive listening to isochronous sound stimuli modulates beta oscillations, which reveal an initial rapid beta decrease following the stimulus onset and the subsequent rebound —probably representing the internalized interval—, and show a temporally

correlated beta modulation in auditory- and motor-related cortical and subcortical areas, as well as a neural synchrony measured as cortico-cortical phase coherence at beta frequencies modulated with the sound rhythm (Fujioka et al. 2012).

Although neural processing for timing at the frequency range of 1-3Hz involves basal ganglia and cerebellum, reported in musical tempo prediction, beta-band activity (around 20Hz) modulation, reflecting changes in an active status of sensorimotor functions, provides a mechanism for maintaining predictive timing and coordinating auditory and motor systems. Following the sound stimulus tempo, sensorimotor cortex, inferior frontal gyrus, supplementary motor area and cerebellum —as well as the thalamus and the posterior parietal cortex— are activated, which allows us to anticipate acoustic events through predictive temporal representation of stimulus rate, spanning motor and auditory brain areas.

7.3 *Rhythm and meter in language*

Paralleling the metrical structure of strong-weak beat perception found in music, a similar patterning of strong and weak elements also occur in speech, where stressed and unstressed syllables offer relevant prosodic information. Although the same degree of temporal regularity does not appear in speech compared to music, listeners seem to perceive stressed speech events isochronously, thus yielding regularity, as well as it occurs with the repetitive —hence, predictable— prosodic information. In speech, metrical stress patterns have a key role facilitating higher-order semantic processing. Cason and Schön (2012) apply the *Dynamic Attending Theory* framework to speech perception, claiming that "attentional resources are preferentially allocated to locations at which stressed syllables are predicted to occur". According to that, dynamic attending may enhance speech sounds processing, because the stressed syllables timing expected by listeners contributes to speech perception.

Cason and Schön (2012)'s experiment reveals that a music-like rhythmic prime when is matched to the speech's prosodic features enhances spoken words' phonological processing. This is due to the beat and metrical structure of the prime, which permits the generation of temporal expectations. Moreover, in word comprehension, predictive mechanisms seem to involve segmental rather than

lexical predictions (Gagnepain et al. 2012), which may give evidence that auditory cortex samples speech into segments, making them predictable in time.

For speech rehabilitation, the use of rhythmic therapies enhancing phonological processing has been found very useful. For instance, Rhythmic Speech Cuing (Thaut, 2005), which paces speech production by using "patterned" cues placing a beat on salient syllables, could benefit non-fluent aphasics production, as well as the use of metrical structure for practising bisyllabic and trisyllabic word patterns in languages showing these metrical feet. Metrical stress therapies are also applied to Cochlear Implanted children, since their speech acquisition priming effect; and rhythmic regular temporal structure aiding learning and memory improve word recall in Multiple Sclerosis patients.

Before starting this new section, we will summarize what has been said on rhythmic cognition and our rhythmic brain in order to refresh certain properties which seem to be common with pitch and tonality origins. Humans cognitively categorize acoustic events by regularizing their duration to integer ratios of pulse so as to group them and extract the underlying beat and meter, as well as to experience its tempo and tempo's fluctuation. Then, we highlighted two key processes, pulse extraction and meter induction, which permit to yield music metrical structure. While the latter seems absent in non-human animals, the former process has been found in certain animals entraining to the pulse through body movements, creating expectancies about the phase of the beat. Furthermore, these animals perceiving and entraining the pulse are found to be complex vocal learners, which entails the existence of tights connection between auditory and motor regions in the brains, as the study of songbirds' brain has revealed. However, rhythmic social behaviour (i.e. primate's drumming), as well as macaques' grouping during rhythm perception, should be considered as important contributors to the human protomusic stage as well, because of their phyolgenetically proximity. On the other hand, beat and meter perception in humans could, alternatively, be explained from a Dynamic Attentional Theory in which neuronal populations —especially those firing in beta rhythms— synchronize with the (visual or auditory) stimulus' frequency, creating expectancies or certain salient points over time.

7. PITCH AND TONAL-HARMONIC COGNITION

Even though several neural network studies have focused on both production and perception of music, pointing to a general purpose system able to learn complex harmonic structures, they do not specify why human auditory system easily seeks tonality (pitches or sounds containing integer harmonics), that is to say, why our hearing mechanism is specialized to the harmonic spectra that only appears in animal vocalizations when most inorganic sounds occurring in nature are indeed not tonal. However, it should not be surprising, because this specialization of our ear to analyze harmonic series might be a simple ancient adaptation to the characteristics of our voice, paralleling other animal ability in identifying their own species-specific communicative sounds, vocalizations, calls or cries.

8.1 *The spectral origins of pitch*

Kammraan Z Gill and Dale Purves (2009) realized that, although humans can distinguish between 240 pitches over an octave in the mid-range of hearing, the most widely used scales cross-culturally comprise five to seven tones dividing octaves into specific intervals. Concretely, these intervals are indeed "those with the greatest overall spectral similarity to a harmonic series", meaning that all the simultaneous overtones or secondary frequencies accompanying a harmonic acoustic sound. These harmonics are integer ratios of the fundamental pitch, i.e. following proportions as 1:2, 1:3 or 2:5. This picture geometrically expresses how frequencies are divided into integer ratios, illustrating harmonics and subharmonics.

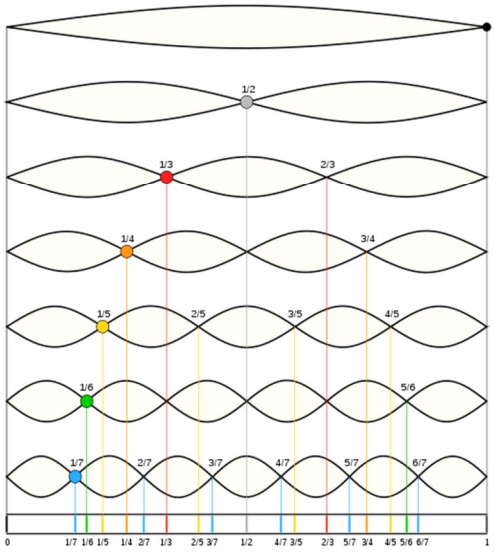


Figure 15

The following musical scale expresses the (approx.) harmonic series of C_0 (16.35Hz):

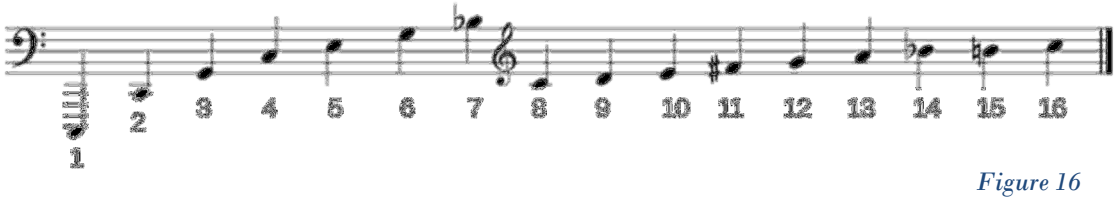
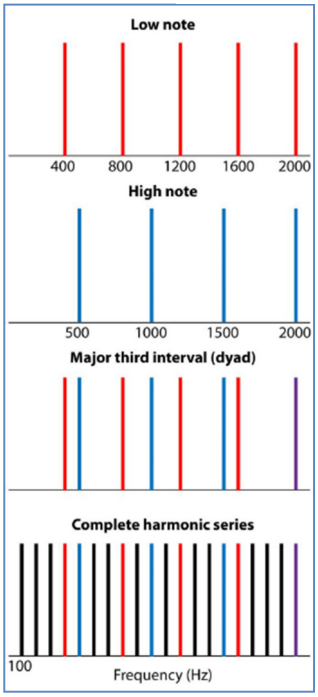


Figure 16

However, there is a lack of general agreement on the harmonic series' role in scales. Ball (2008), for instance, does not find any reason to base scales on it. Nevertheless, he accepts that it is undeniable that they are not arbitrary, since "most have between four and seven notes arranged asymmetrically within the octave", and that all show unequal intervallic steps. In fact, it is this asymmetry which indicates a tonal centre to the listener, together with the tones' structural position.

Previous approaches to scale structure and its origins, such as (i) consonance curves made up of integer ratios, (ii) musical patterns defining a musical grammar, (iii) competing preferences for small integer ratios and equal intervals, and (iv) multiple-tones scales (see Gill and Purves, 2009), have failed to explain the human preference for 5-to-7 tones' scales and its biological rationale.

Figure 17



For this reason, Gill and Purves (2009) join Helmholtz's (1877) view of the relative consonance deriving from harmonic relations of two tones with Bernstein's (1976) consideration of the scale structure determined by the appeal of lower harmonics in naturally-generated harmonic series. Then, they compare the harmonic structure of every interval in any scale to the general harmonic series structure (rather than the intervals between fundamental frequencies and individual harmonics). Finally, they evaluate their degrees of similarity by making an average of all the scale-intervals contrasted to the harmonic series intervals [see fig. 17].[59]

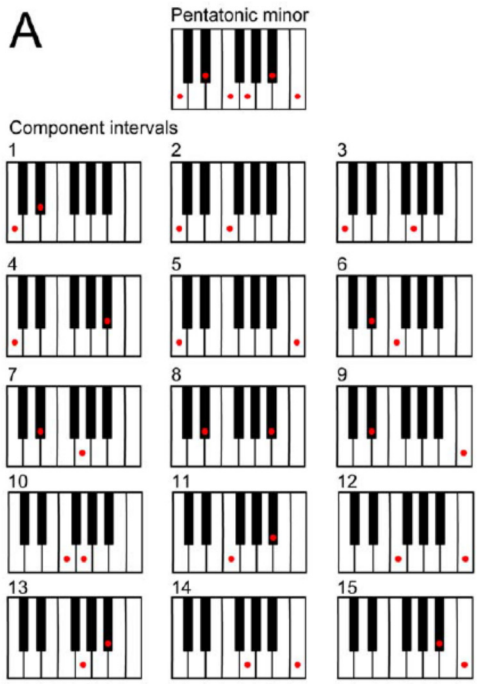
It is found that "many of the relatively small number of scales [...] comprise intervals that, when considered as a set, are maximally similar to a harmonic series", and are more similar to harmonic series when the number of discrete tones is decreased. The number of tones in musical scales seems to be delimited by the difficulty to sing larger intervals (requiring greater neuromuscular energy for

[59] Figure 17, taken from Gill and Purves (2009), shows the mechanism used for obtaining the average of similarity between scales (the internal intervallic relations of two notes of a scale and their harmonics) and the natural harmonic series , in particular, it illustrates the Major third interval.

coordinating production) and the optimal minimum of enough variety available for comfortable intervallic combinations.

The picture placed below (fig. 18) shows, as a sample, the case of comparing different pairs of notes from the minor pentatonic scale, with all the possible dyadic relations (A) and its percentage of similarity (B), when the intervals are compared to the harmonic series.

A



B

Figure 18

Interval	Name	Ratio	Percentage similarity
1	Minor third	6:5	33.33
2	Perfect fourth	4:3	50.00
3	Perfect fifth	3:2	66.67
4	Minor seventh	9:5	28.89
5	Octave	2:1	100.00
6	Major second	10:9	20.00
7	Major third	5:4	40.00
8	Perfect fifth	3:2	66.67
9	Major sixth	5:3	46.67
10	Major second	9:8	22.22
11	Perfect fourth	4:3	50.00
12	Perfect fifth	3:2	66.67
13	Minor third	6:5	33.33
14	Perfect fourth	4:3	50.00
15	Major second	9:8	22.22
Mean percentage similarity:			46.44

Gill and Purves (2009), applying this method, have found that human cross-cultural pentatonic and heptatonic scales occupy the ranking top position, showing the highest possible average made of integer ratios close to the harmonic series ratios. For instance, the pentatonic minor scale, whose interval relations were analysed above [fig. 18], occupies the top position of mean percentage similarity [see fig. 19] because its integer ratios are the most similar to ratios found within the harmonic series.

Scale	Scale degrees	Mean percentage similarity
Minor	3:2, 4:3, 6:5, 16:9	46.44
Ritusen	3:2, 4:3, 5:3, 10:9	46.44
Candrika todi	3:2, 4:3, 6:5, 8:5	44.28
Asa-gaudi	3:2, 4:3, 5:3, 5:4	44.09

Figure 19

8.2 *Harmony from pitch*

Computing pitch relations is unique and critical for music processing. But it does not mean that it has evolved for this purpose, because pitch information processing

allows human to distinguish environmental sounds, which only some of them show naturally occurring periodic sounds with relevant pitch information. Neurally, the right-hemisphere auditory cortex specialization for pitch processing may have emerged for the way of processing the environmental sounds: quickly and roughly, or slowly and accurately (Zatorre, 2005).

A main point of Gill and Purves (2009)'s statistical results is that human clearly prefer particular characteristics of harmonic series in musical scales, probably because "human ability to perceive tonal (i.e. periodically repeating) sound stimuli has presumably evolved because of its biological utility", that is, because of the presence of harmonic resonances in nature, mostly produced by animal species producing periodic sound for socialization and reproduction. Although the harmonic stimuli are present in stridulating insects' sounds, songbirds' songs and mammals' vocalizations, human vocalizations might have been the most biologically relevant and frequently experienced. Since primates are specifically attracted to conspecific vocalizations and human auditory systems have specialized for processing vocal sounds (harmonic series depending on vocal fold vibrations permitting voiced speech and vowels), it is plausible that musical scales resulted from a preference for dyads resembling maximally to harmonic series, in other words, human vocalizations.

However, not only harmonicity matters, but also frequency ranges, timbres and prosodic fluctuations coming from human vocalizations are influential in musical preferences, as affective responses of nonhuman primates to music similar to their vocalizations frequencies and prosody (Snowdon and Teie, 2009) strongly suggest. Hence, while scale preferences seem based on harmonic series coming from vocal fold vibrations, other musical aspects (embellishments, microtonal intervals, glissandos...) may come from additional features of the human voice.

8. ORIGINS OF TONAL-HARMONY

In music, hierarchy not only occurs in metrical structure, but also in tones. Cross-cultural tonal hierarchies convert certain tones into reference pitches, which show stability, frequent repetition and rhythmic emphasis, as well as occupy structural important positions. Therefore, more prominent and stable structurally significant musical tones yield a hierarchical ordering of them into different levels. Cross-

cultural musical styles express the notion of tone centrality, i.e. "one central tone anchors a subset of hierarchically related tones". While an acoustic approach looks at the harmonic series of complex periodic sounds to explain the formation of musical scales and chords, a cognitive approach takes into consideration the role of cultural experience in musical learning and perception.

9.1 *Tonal hierarchies*

Krumhansl and Cuddy's ([henceforth K&C], 2010) theory of tonal hierarchies rests upon three interrelated propositions:

1. Tonal hierarchies have psychological reality, that is, their cognitive representations play a central role in how musical sequences are perceived, organized and remembered, as well as how expectations are formed.
2. Tonal hierarchies are musical facts, made evident in the musical surface and characterizing diverse styles and genres.
3. Tonal hierarchies are abstracted by the listeners by using statistical frequency patterns of tones distributions and tones combinations.

While (1) reflects the psychological, internal status of tonal hierarchies organizing prominent, stable, and structurally significant tones, which affects memory, sense of stability, phrasing and generation of expectations, (2) refers to the musical, external status of tonal hierarchies, i.e. the tones salience on the surface of music, emphasized by frequency and duration. Then, (3) describes the relationship between the subjective and objective descriptions of tonal hierarchies, claiming that sensitivity to tones' distributions enables the listener to abstract the tonal hierarchy.

K&C (2010) propose two basic cognitive principles underlying tonal hierarchies' structure: the existence of cognitive referential points and sensitivity to statistical regularities. *Cognitive reference points* —"to which other category members [other tones and chords] are encoded, described and remembered" efficiently (K&C, 2010)— provide and economical description of the domain by guiding perception and cognition. These reference points show processing priority and memory stability. Distinct from other domains, they are not independently defined from the category, i.e. they do not have invariant cross-contextual inherent qualities, because the function of a tone depends on the musical context, relying on a listener's relational processing (relative pitch) rather than on fixed labels (absolute pitch).

This stable hierarchical tonal framework is isolable at the neural level and can be selectively affected in brain pathologies, tightly related to musical memory. With respect to the second principle, sensitivity to environmental regularities consists of an extracting mechanism which processes statistical properties of musical surface, just beginning with the statistical learning occurring early in human development.

K&C (2010) claim that tone centrality arises from musical style regularities, such as "repetition of tones and tone sequences, melodic and rhythmic emphasis, durational and metric stress, and positioning of central tones at or near beginnings and endings of phrases". Then, a mental representation is developed by the listener after exposure to music, coupling tone distributions and style-specific knowledge, which permits to generate expectations and remember musical patterns. According to Huron (2006), generating expectancies through statistical learning (i.e. anticipating frequently occurring events) has adaptive value in evolution, because knowing "what, when and where something is likely to occur speeds perception, action and evaluating the consequence of alternative actions" (K&C, 2010).

A tonal hierarchy gives a stable, abstract frame of reference, which does not contain information about specific pitch heights, but reference to pitch classes, whose relative stability does not depend on tones' position in music. Thus, tonal hierarchies are cognitive representations of the implicit knowledge of abstract musical structure within a culture or a genre, while *event hierarchies* describe the relative prominence of musical events occurring within particular sequences of specific pieces of music. Both hierarchies are complementary and provide the musical structure to the listener, as well as they give patterns of stability and instability. For that reason, Bharucha and Krumhansl (1990) propose three principles of tonal stability in terms of psychological distance and memory: contextual identity, contextual distance and contextual asymmetry.

The following points are reported from experiments evidencing a tonal hierarchy:

1. A tone is expected to resolve (or lead up) to a tone of greater stability in the hierarchy.
2. Absolute pitch listeners name tones faster and more precise when they pertain to higher hierarchical positions in C major.
3. If a melody ends with an out-of-scale tone, a larger P300 component appears in ERP.
4. Tonal melodies are easier to recognize, as well as to detect pitch alterations.

Even though K&C (2010) defend the cognitive origin of tonal hierarchies from a psychoacoustic explanation based on the harmonic structure of complex tones, thus emphasizing the role of experience in music internalization, a recent experiment contrasting the harmonic relations among tones of all world musical scales reveals a human specialization for detecting the harmonic series spectrum, suggesting a conspecific vocalization specialization in human perceptual systems. While in a learning approach tonal hierarchy arises after an extensive exposure internalizing music, likely through a statistical input processing of frequent tones and their combinations; in a non-learning approach hierarchy reflects acoustic properties of tones depending on complex tones harmonics.

9.2 *Acquisition and loss of tonal hierarchies*

Internalized as cognitive resources, tonal hierarchies require a mature memory, and it is plausible that they emerge in development later than the basic perceptual sensitivities building them. While statistical learning occurs in the one-year-old infant brain, discriminating melodic contours, frequencies, harmonic ratios, phrasing and some pitch-scale patterns; tonal hierarchies apprehension and representation do not occur until the age of 5-6 years (for detecting stable tonal centres), whereas the 7-year olds and adults perfectly detect harmonic and key implications and changes. Sensitivity to statistical regularities seems to depend on the maturation of our memory system, which would make it able to deal with hierarchical processing.

Failures and loss of tonal hierarchy are accompanied by musical memory failures. This link has been corroborated by studies on neurological disorders and dissociations: normal language and intellectual functioning contrasting with musical functioning failures. Studies of acquired *amusia*, a clinical disorder affecting musical abilities after brain damage, support the linkage between tonality and recognition memory for familiar melodies, suggesting a role of tonal encoding of pitch in accessing to stored memory representations. Hence, impairment of acquired cognitive references leads to severe failures of melody recognition. The same is found in congenital amusia, a developmental disorder due to a neurogenetic anomaly, where the inability to detect out-of-key notes and recognize familiar tunes is also

found. Conversely, Alzheimer Disease manifests the alternative dissociation: musical memory is preserved as well as tonal encoding of pitch.

9.3 *Tonality as the musical grammar*

Perceiving a tonality depends on psychological mechanisms involving harmonic interactions of frequencies creating virtual pitches, memory traces of immediate contexts and internalized regularities of harmonic sequences. "Tones of greater surface salience in the sequence [… act] as reference points or anchors for other tones of lesser surface salience" (K&C, 2010), and, as Smith and Schmuckler (2004) have found, structural salient cues rely on total duration rather than frequency of occurrence. Hence, a distributional emphasis of tones establishes and maintains the listener's sense of tonal reference points, correlating subjective and objective musical properties in all the cultures. For that reason, any listener can use tone distributions to perceive the main anchoring tones of tonality in different musical styles or genres.

Paralleling some linguistic experiments on teaching invented languages with parametric rules applying to words according to their lineal position instead of their constituent position, the twentieth century western music 12-serialism created a musical grammar abolishing the tonal hierarchy, where the twelve chromatic tones were strictly ordered in series or tone rows, so as to avoid giving salience to any tone. Despite the stylistic intention of avoiding tonal implications, listeners are found to be influenced by local tonal implications, without internalizing the ordered sequence of tones in the series. Musical organization mechanisms are therefore not avoided.

Reviewing what has been seen in this section on the evolution of pitch and tonality, first we claim that discrete pitches configuring cross-cultural scales are determined by the similarity among their intervallic relations and the distances of the harmonic series overtones, which divide a note frequency into integer ratios forming simultaneous harmonics. The brain specialization in processing acoustic harmonic sounds is proposed to come from human conspecific vocalizations' preference over non-human or non-harmonic acoustic sounds. With respect to the tonal-harmony evolution, it is claimed that it depends to some extent on the emergence of hierarchy, organizing the scale pitches or notes in hierarchical levels, and essentially on grammar, yielding a system of internal reference to quiescent

points, whose salience over the others in turn depends on (i) their position in the metrical structure, (ii) their acoustic properties and (iii) the hierarchical scale level determining their function in each moment. Finally, the interaction between tonal-harmonic structure and the metrical structure gives rise to the musical ebb-and-flow.

SUMMARY

In this part we proposed a protomusic stage, between a musical protolanguage and our current music, in which *meter* was in place due to the influence of a syntactic protolanguage (200-150KYA) and cognitive mechanism *merge* yielding hierarchies. A rhythmic protomusic with a hierarchical organization of the beat may have arisen. Given that some animals (certain complex vocal learners) can perceive beat and entrain to it, and that certain primates show rhythmic behaviour (drumming) and grouping categorization, we claim that humans were able to externalize the beat via metrical structures. In fact, this meter is essential for music and dance, and may have impacted our phonology (accented syllables) and poetry as well. Although our brain shows proficiency in processing a steady beat, it also has developed a specialization for processing pitch and harmonic relations. We claim that it has come from selecting the processing of our conspecific vocalizations, made of harmonic spectra. Then, once hierarchy and grammar emerged, a musical grammar also appeared, showing musical scales organizing pitches and referential quiescent points yielding tonalities. Thus, music and language evolution are deeply linked.

CONCLUSIONS

9. GENERAL OVERVIEW

The faculty of music parallels the faculty of language in many ways: both are governed by structural principles that are easily and unconsciously learned through environmental culture exposure, and both are used in social interaction to communicate intentional meaning. Their internal computing mechanisms are also genetically developed, neurally grounded and physiologically restricted. This, in turn, allows the processing of learned rule systems (grammars) which should be distinguished from the external perceivable output, which is culturally driven through musical idioms and languages. Moreover, music and language faculties are indeed bimodal, either externalized through acoustic sounds (music and spoken languages) or by gestural movement (dance and sign languages), generally implying the co-occurrence of both (instrumental music, co-speech gestures). While both possess grammatical systems, their expressed meanings differ: music evokes emotions by structurally marking sound qualities and their internal relations, whereas language expresses lexical concepts by referring to entities (objects or events) and propositions. And this fundamental distinction entails very different communicative usages within determined social contexts.

Despite making use of certain linguistic-related neural mechanisms —such as Broca's area to compute harmonic relations—, music is found to activate brain regions specific both to music perception and to structural processing (discrete pitch interrelations and isochronous rhythm grouping). In fact, some properties including spectral frequencies and the organization of time do not appear to contribute to language in a significant way. In addition, although the prosodic cues of speech also activate emotional brain areas, music seems to be even more tightly linked to the limbic system and emotions than language is. Music even yields (to some extent) a modification of subcortical brain morphology throughout a lifespan, thus affecting hormonal production, mood, memory functions and other computational skills positively. This is the reason why music is used as a clinical tool. Despite the fact that specific music disorders are found in some patients, music therapy encourages the improvement of certain pathological symptoms, enhancing motor control, and even promoting communicational behaviours.

As language and music have a strong genetic basis and both are deeply grounded in the brain and developed effortlessly within every culture, natural selection must have underpinned them in some way. More concretely, they must have been selected, not as a whole, but rather through the selection of their different constitutive components, which may have had their own evolutionary history for other purposes. The emergence of linguistic hierarchy and grammar [see fig.4], which only occurred in *Homo sapiens*, may have cognitively impacted our minds, completely changing our interaction with the world and giving rise to our cultural and uniquely-human behaviour. Thus, a grammar providing lexical items (indexed concepts via phonology) with reference (including deixis) was added to *merge*, an internal mechanism that was able to combine elements from different domains (protoconcepts). At that point, our current language/symbolic thinking was in place.

Assuming the existence of ancestral communicative systems (i.e. protolanguages that preceded our linguistic capacity), and also accepting Darwin's proposal of a musical protolanguage, the emergence of symbolic thinking may have impacted this musical protolanguage. Consequently, it may have split into music and spoken language. The prosodic components of speech and vocal (as well as instrumental) music support this common origin, which is further backed up by the fact they share the same neural substrate. Furthermore, (i) phylogenetic studies on emotive processing of animal vocalizations, (ii) the different neural pathways affording calls and songs, and (iii) the existence of a vocal learning capacity in humans (as a vocal memorizing and manipulative mechanism), come together to suggest that there is a common rudimentary vocal communicative system among primates and other mammals, whose components are partially shared with even further removed taxa (e.g. some birds). Even if this may have led to a protolanguage that was more prosodic than musical —hence showing more emotional cues than rhythmical or pitch-discrete elements—, after different ingredients showing "musicality" were selected, a musical protolanguage may have evolved into protomusic, with rhythmic components participating as well [see fig.1]. The existence of other protolanguages, however, must not be rejected. A multimodel in which protolanguages interact with each other may bring different components of language (intentionality, creativity, structure…) together.

Rhythmic cognition and pitch (tonal-)harmonic cognition also show cognitively rooted components within the brain, which indicate an evolutionary process of selecting these mechanisms and predispositions. Essentially, the selection of complex vocalizations may have, in turn, promoted beat extraction and spectral perception of pitch which enabled humans to perceive meter and create tonality (obviously, once hierarchy and reference had emerged). While beat and meter may have led to musical structure, reference applied to organized pitches may have yielded a musical harmonic grammar [see fig.6].

Although rhythm has been overlooked until just a few decades ago, complex rhythmic mechanisms seem to be a promising clue to explain musical structure. Beat extraction and metrical induction enable a chunking mechanism to process and remember groups of sounds categorized to follow integer ratios (normally binary subdivisions) of an isochronous pulse. This, in turn, becomes hierarchically organized in strong and weak patterns, leading to a metrical structure. While beat extraction is found in animals (with complex vocal learning species displaying entrainment to the beat) [see fig.10], meter has not yet been found in any animal.

In contrast, humans innately detect meter from birth. For this reason, the vocal learning hypothesis for beat extraction and entrainment proposes that, after having selected this complex ability, the tight relation of motor and auditory brain areas may have triggered the motor ability to entrain pulse. That would not only create phases of expectation, but also suggest that metrical structure —which permits music and dance— may not have been present from the very beginning, as it implies hierarchical structure processing.

Phylogenetically, human externalization of rhythm seems to parallel the rhythmic behaviour found in primates (i.e.drumming), which indicates social position within the group and serves to intimidate both advertencies and intruders. Moreover, macaques and other primates show a half-developed perception system of grouping rhythms. Aside from animal research, cognitive studies on attention also corroborate a metrical organization of music. According to DAT, meter could be explained by attentional energy fluctuations. It would be carried out by beta-band activations within the brain that synchronize with the input stimulus and create

cyclic expectancies. These rhythmic neuronal activations can also explain the existence of ternary meters, in contrast to the simple binary meter, by appealing to secondary attentional peaks that interact at different frequencies.

Briefly regarding pitch and (tonal-)harmonic cognition, the acoustic manner of processing music —if we consider music as a faculty which is externalized bimodally—, it must be highlighted that discrete pitches (sounds showing harmonic spectra) are interrelated to each other following cultural scales (learned intervallic steps separating successive pitches within an octave). Moreover, these pitches are also hierarchically ordered through giving predominance to certain notes over the rest by positioning them at important structural positions and modifying their acoustical properties.

Our human capacity to perceive the harmonic spectra of different pitches and their spectral relations (harmony) seems to be promoted by conspecific vocalization specialization. This is furthermore supported by the cross-cultural use of 5 to 7 notes within scales which is constrained by the (limits of) comfortable production of vocal folds and by an optimal combinability. Furthermore, tonal hierarchies give functions to pitches, considering them as stable or unstable acoustic anchors, i.e. referential points over time. This creates the ebb-and-flow of music, derived from the flux between tension and release.

Given the existence of percussive drumming, we should see music as having a core structural element, rhythm that structures pitches in a temporal stream. A secondary acoustical structure arises from pitch-interrelations, which yields melody and its internal harmonic functional relations: ([(rhythm) pitch] melody...harmony).

In summary, although music and language may have had a common origin in the past —a musical protolanguage sharing melodic contours and prosody—, it may have split into music and language by incorporating linguistic-specific properties and music-specific properties. A protomusic stage made of (1) syllabic vocalizations, (2) discrete-pitch signals and (3) externalized rhythmic synchronizing behaviours which follow pulse and are accompanied by movement, would therefore be expected. We defend that new properties were gradually added to the musical protolanguage.

Assuming that language and music may have impacted each other, we propose that one of these interactions may have been hierarchy. We argue that hierarchy, as a cognitive and computational process, originates from linguistic merge but it was, in turn, co-opted in meter. Furthermore, the ability to generate hierarchy renders a (musical) grammar which makes "reference to a quiescent point" over time and frequency, i.e. tonality. As such, the products of hierarchy (meter) and reference (pointing to quiescent points, tonality) yield the general ebb-and-flow of music. Conversely, meter has also impacted language, at least the syllabic stress of words, and perhaps prosodic rhythm, since current rhythmic therapies enhances the fluency and the recovery of speech.

CONCLUDING REMARKS

Music and language, unique to humans and cross-culturally found, are based on a common original system of communication, a musical protolanguage, which explains the sharing of prosody, rhythm and neural mechanisms among the current faculties.

Vocalizations may have been fundamental for both core aspects of music: rhythm and pitch-based tonal harmony, as well as for speech, phonology and the language acquisition of an open-ended lexicon. In fact, the evolutionary selection of complex vocal learning has had an important role in our species, promoting a brain circuitry for processing harmonic spectra, internal pulse generation and coordinating movement, which led to the two essential components of our current music: rhythm (following beat and meter) and pitches (within a hierarchical tonal system).

Animal evidence of pulse extraction in vocal learners, as well as the rhythmic social behaviour of primates, clearly points to a human protomusic made up of vocalizations and rhythmic behaviour, an ancestral system of music and dance. Neural and clinical studies corroborate that our brain is specialized in harmonic spectra and rhythmic meter processing, which supports the role of vocalizations in founding our uniquely-human music. In turn, this specialization has also influenced our linguistic prosody and phonology.

10. FURTHER ISSUES

Looking at animals' rhythmic categorization will reveal which mechanisms are used by our brain to categorize acoustic sound patterns of different durations, and will indicate which processes are common and shared with other animals, as well as which are unique to humans, and (perhaps) driven by language and our linguistic thinking. In contrast, from the assumptions of this thesis, it is not expected to find any tonal-harmonic processing in animals, given that it comes from our grammar. Otherwise, the hypothesis presented here should be revised or rejected immediately.

Being speculative, while rhythmic cognition may have appeared simultaneous to linguistic hierarchy emergence (200-150KYA), tonal-harmonic cognition may have appeared afterwards, at the same time of the emergence of grammar. Perhaps it was as a consequence of a simplified referential device, strongly related to the language externalization in culturally-modern humans with symbolic minds. More studies interrelating grammar emergence and the symbolic figurative production of *Homo sapiens* should be done, in order to precisely make chronologies of our language and music evolution. Although arguable, assuming our dissociated hypothesis, perhaps grammatical deficits appear separated from linguistic deficits (i.e. from the basic mechanism merge and its hierarchy). Future investigations should clarify this point.

"Music"[60] should be seen as a bimodal capacity, which links auditory (and visual) perception and production to motor activities. As such, new studies should consider the implications of linking vocal or instrumental music and dance (as two possible ways to express the cognitive experience of music), given that both are based on rhythm as the structural component of "music". In contrast, the harmonically interconnected pitches and the accurately linked gestures give to "music" its evoking counterpart. Thus, it should be re-explored how to use "music" clinically, through music therapy and dance therapy in order to recover or improve damaged and affected (motor) skills in certain diseases, as well as to enhance children's cognitive and social abilities. Furthermore, the relation between *beat* and repetitive rhythmic behaviours in autism or schizophrenia should be deeply analysed in future research.

[60] Here we use the term "music" referring to an internal capacity, which is structurally-driven by rhythms and grammatically-externalized or -perceived through evoking emotions. For instance, a rhythmic play of coloured lights could also be included in this reinvented-term.

REFERENCES

Amador, A. & Margoliash, D., 2011. Auditory Memories and Feedback Processing for Vocal Learning. *The Auditory Cortex*, 1, pp. 561-575.

Arbib, M, 2005. From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics. *Behavioral and brain sciences*, 28, 105-167.

Arbib, M., 2005. The mirror system hypothesis: How did protolanguage evolve? In Tallermann (Hg.): *Language Origins. Perspectives on Evolution*. Oxford, *Oxford University Press*.

Arbib, M., 2013. Language, Music, and the Brain: A Mysterious Relationship, MIT Press, Cambridge.

Arbib, M. & Iriki, A., 2013. Evolving the language-and music-ready brain. In *Language, music, and the brain: A mysterious Relationship,* MIT Press, Cambridge.

Arbib, M.A., Liebal, K. & Pika, S., 2008. Primate Vocalization, Gesture, and the Evolution of Human Language. *Current Anthropology*, 49(6), pp.1053–1076.

Aubé, W., Peretz, I. & Armony, J., 2013. The effects of emotion on memory for music and vocalisations. *Memory*.

Ball, P., 2008. Facing the music. *Nature*, 453 (May), pp.160–162.

Belin, P., Campanella, S. & Ethofer, T., 2013. *Integrating face and voice in person perception*. New York

Belin, P., Zilbovicius, M. & Crozier, S., 1998. Lateralization of speech and auditory temporal processing. *Journal of cognitive Neuroscience*.

Bernstein, L., 1976. The unanswered question: Six talks at Harvard.

Berwick, R.C. et al., 2011. Songs to syntax: the linguistics of birdsong. *Trends in cognitive sciences*, 15(3), pp.113–21.

Bharucha, J., 1996. Melodic anchoring. *Music Perception*.

Bickerton, D., 1995. Language and human behavior.

Bickerton, D., 2007. Language evolution: A brief guide for linguists. *Lingua*.

Boeckx, C.; Leivada, E.; Martínez-Alvárez, A.; Martins, P.T.; Rosselló, J. On the need of a new concept: syntactic protolanguage. Comunicación. *Ways to Protolanguage 3*. Wroclaw, 25-26, 05, 2013.

Bolger, D., Trost, W. & Schön, D., 2013. Rhythm implicitly affects temporal orienting of attention across modalities. *Acta psychologica*.

Bolhuis, J. & Everaert, M., 2013. Birdsong, speech, and language: exploring the evolution of mind and brain.

Bowling, D.L., Herbst, C.T. & Fitch, W.T., 2013. Social Origins of Rhythm ? Synchrony and Temporal Regularity in Human Vocalization. , 8(11).

Brown, S., 2000. The “musilanguage” model of music evolution.

Cason, N. & Schön, D., 2012. Rhythmic priming enhances the phonological processing of speech. *Neuropsychologia*.

Christensen-Dalsgaard, J., 2004. Music and the origin of speeches. *JMM: The Journal of Music and Meaning* 2, section 2m pp.1-16.

Colwell, R., 2006. *MENC Handbook of Musical Cognition and Development* R. Colwell, ed., Oxford University Press.

Cook, P. & Rouse, A., 2013. A California sea lion (< em> Zalophus californianus) can keep the beat: Motor entrainment to rhythmic auditory stimuli in a non vocal mimic. *Journal of Comparative Psychology* 10, pp.1-16.

Corballis, M., 2002. *From hand to mouth: The origins of language*. Princeton, Princeton Univ. Press.

Cross, I., 2007. Music and cognitive evolution. In Dunbar & Barrett *Handbook of evolutionary psychology* , Oxford, oxford university Press.

Cross, I., 2009. The evolutionary nature of musical meaning. *Musicae scientiae,* special issue 179-200.

Darwin, C., 1871. The descent ofman. *Great books ofthe western world*, volume 49, Chicago.

Deutsch, D., 2010. Speaking in tones. Issue of *Scientific American Mind*, 21, pp.36-42.

Deutsch, D. ed., 2013. *The Psychology of Music*, 3rd Edition, Oxford Academic Press.

Dissanayake, E., 2008. If music is the food of love, what about survival and reproductive success? *Musicae Scientiae.*

Drake, C., Jones, M. & Baruch, C., 2000. The development of rhythmic attending in auditory sequences: attunement, referent period, focal attending. *Cognition* 77.

Duncan, J., 2001. An adaptive coding model of neural function in prefrontal cortex. *Nature Reviews Neuroscience*, volume 14, No. 4, 173.

Duncan, J., 2010. The multiple-demand (MD) system of the primate brain: mental programs for intelligent behaviour. *Trends in cognitive sciences.*

Egnor, S. & Hauser, M., 2004. A paradox in the evolution of primate vocal learning. *Trends in neurosciences* 27(11), pp.649-654.

Fedorenko, E., 2011. Functional specificity for high-level linguistic processing in the human brain. *Proceedings of the National Academy of Sciences,* 108(39), 16428-33.

Fedorenko, E., 2012. Sensitivity to musical structure in the human brain. *Journal of Neurophysiology*, 108.

Fenk-oczlon, G. & Fenk, A., 2009. Some parallels between language and music from a cognitive and evolutionary perspective. *Musica Scientiae*, pp.1–26.

Fenk-Oczlon, G. & Fenk, A., 2009. Some parallels between language and music from a cognitive and evolutionary perspective. *Musicae Scientiae* 13(2).

Fitch, W.T., 2006. The biology and evolution of music: a comparative perspective. *Cognition*, 100(1), pp.173–215.

Fitch, W.T., 2009. Biology of music: another one bites the dust. *Current Biology*, volume 19, issue 10, pp.403-404.

Fitch, W.T., 2009. Musical Protolanguage: Darwin's Theory of Language Evolution Revisited, *Language Learning and Development* 7: 253–262.

Fitch, W., 2010. *The evolution of language,* Cambridge, Cambridge University Press.

Fitch, W., 2013. Rhythmic cognition in humans and animals: distinguishing meter and pulse perception. *Frontiers in systems neuroscience*, 7.

Fitch, W., Hauser, M. & Chomsky, N., 2005. The evolution of the language faculty: clarifications and implications. *Cognition* 97, pp.179-210.

Fitch, W. & Rebuschat, P., 2012. The biology and evolution of rhythm: unravelling a paradox. In Rebuschat P, Rohmeier M, Hawkins JA, Cross I, editors. *Language and music as cognitive systems,* Oxford university Press.

Fitch, W.T., Grau, J.W. & Texas, A., 2013. Rhythmic cognition in humans and animals : distinguishing meter and pulse perception. , 7(October), pp.1–16.

Foote, A.D. et al., 2006. Killer whales are capable of vocal learning. , (1991), pp.1–4.

Friederici, A. & Fiebach, C., 2006. Processing linguistic complexity and grammaticality in the left frontal cortex. *Cerebral Cortex*, 13(2), 170-177.

Fujioka, T. & Trainor, L., 2012. Internalized timing of isochronous sounds is represented in neuromagnetic beta oscillations. *The Journal of Neuroscience*, 32(5), pp. 1791-1802,

Gagnepain, P., Henson, R. & Davis, M., 2012. Temporal predictive codes for spoken words in auditory cortex. *Current Biology*.

Geiser, E. et al., 2008. The neural correlate of speech rhythm as evidenced by metrical speech processing. *Journal of Cognitive Neuroscience,* 21 (7), 1255-68.

Gill, K.Z. & Purves, D., 2009. A Biological Rationale for Musical Scales. , 4(12).

Ghazanfar, A. a et al., 2012. Cineradiography of monkey lip-smacking reveals putative precursors of speech dynamics. *Current biology : CB*, 22(13), pp.1176–82.

Giraud, A. & Poeppel, D., 2012. Speech perception from a neurophysiological perspective. *The human auditory cortex*.

Goerlich, K., Aleman, A. & Martens, S., 2012. The sound of feelings: electrophysiological responses to emotional speech in alexithymia. *PloS one*.

Grahn, J., 2012. Neural mechanisms of rhythm perception: current findings and future perspectives. *Topics in cognitive science.*

Grahn, J., 2009. The role of the basal ganglia in beat perception. *Annals of the New York Academy of Sciences*.

Grahn, J. & Brett, M., 2009. Impairment of beat-based rhythm discrimination in Parkinson's disease. *Cortex*.

Grahn, J. & McAuley, J., 2009. Neural bases of individual differences in beat perception. *Neuroimage*.

Grahn, J. a & Rowe, J.B., 2013. Finding and feeling the musical beat: striatal dissociations between detection and prediction of regularity. *Cerebral cortex (New York : 1991)*, 23(4), pp.913–21.

Greenfield, M., 1994. Synchronous and alternating choruses in insects and anurans: common mechanisms and diverse functions. *American Zoologist, 34*.

Ha, P., 2012. Rhesus Monkeys (Macaca mulatta) Detect Rhythmic Groups in Music , but Not the Beat. PLoS ONE, 7(12).

Hannon, E.E. & Johnson, S.P., 2005. Infants use meter to categorize rhythms and melodies: implications for musical structure learning. *Cognitive psychology*, 50(4), pp.354–77.

Hardy, M.W. & Lagasse, A.B., 2013. Rhythm , movement , and autism : using rhythmic rehabilitation research as a model for autism. , 7(March), pp.1–9.

Hasegawa, A. et al., 2011. Rhythmic synchronization tapping to an audio-visual metronome in budgerigars. *Scientific reports* 1, 120.

Hattori, Y., Tomonaga, M. & Matsuzawa, T., 2013. Spontaneous synchronized tapping to an auditory rhythm in a chimpanzee. *Scientific reports*, 3.

Hauser, M.D. & McDermott, J., 2003. The evolution of the music faculty: a comparative perspective. *Nature neuroscience*, 6(7), pp.663–8.

Hauser, M.D. et al., 2014. The mystery of language evolution. Frontiers in Psychology, 5: 401.

Helmholtz, H., 1954. On the sensations of tone, 1877. *Trans. AJ Ellis, Dover, New York*. Hewes, G., 1973. An Explicit Formulation of the Relationship Between Tool-Using, Tool-Making, and the Emergence of Language. *Visible Language*.

Hillert, D., 2014. *The Nature of Language*, New York, NY: Cambridge University Press.

Honing, H., 2011. *The Illiterate Listener: On Music Cognition, Musicality and Methodology*, Amsterdam, Amsterdam University Press.

Honing, H., 2012a. Cognition and the Evolution of Music: Pitfalls and Prospects. *Topics in Cognitive Science*, pp.1–12.

Honing, H., 2012b. Without it no music: beat induction as a fundamental musical trait. *Annals of the New York Academy of Sciences*, 1252, pp.85–91.

Honing, H., 2013a. The structure and interpretation of rhythm in music. In D. Deutsch, *Psychology of Music*.

Honing, H. et al., 2012. Rhesus monkeys (Macaca mulatta) detect rhythmic groups in music, but not the beat. *PLoS ONE*, 7 (12).

Honing, H. & Ploeger, A., 2012. Cognition and the evolution of music: Pitfalls and prospects. *Topics in cognitive science* p.1-12.

Huron, D., 2006. *Sweet anticipation: Music and the psychology of expectation*. Cambridge, MIT Press.

Jackendoff, R. & Lerdahl, F., 2006. The capacity for music: what is it, and what's special about it? *Cognition*, 100(1), pp.33–72.

Jarvis, E., 2006. Selection for and against vocal learning in birds and mammals. *Ornithological Science*, 5, pp. 5-14.

Jarvis, E.D., 2006. Evolution of brain structures for vocal learning in birds : a synopsis 1 Introduction 2 Cerebral vocal nuclei of vocal- learning birds. , 52, pp.85–89.

Jarvis, E., 2007. Neural systems for vocal learning in birds and humans: a synopsis. *Journal of Ornithology* 143, S35-44.

Jentschke, S. & Koelsch, S., 2009. Musical training modulates the development of syntax processing in children. *Neuroimage* 47, pp. 735-744.

Jespersen, O., 2013. *Language: its nature and development*, Hamlin Press.

Jolij, J. & Meurs, M., 2011. Music Alters Visual Perception. Plos ONE 6(4).

Jones, M. & Boltz, M., 1989. Dynamic attending and responses to time. *Psychological review*, 96(3), pp.459-491.

Jones, M. & Moynihan, H., 2002. Temporal aspects of stimulus-driven attending in dynamic arrays. *Psychological Science*, 13.

Juslin, P. & Västfjäll, D., 2008. Emotional responses to music: The need to consider underlying mechanisms. *Behavioral and brain sciences*, 31, 559-621.

Katz, J. et al., 2011. The Identity Thesis for Language and Music. Unpublished manuscript. http://ling.auf.net/lingBuzz/000959

Katz, J. & Mit, D.P., 2009. The Recursive Syntax and Prosody of Tonal Music 1 . A consensus about music and language 2 . Against the consensus : the Identity Thesis for Language and Music 3 . Prolongational Reduction : Musical Merge. , (May), pp.1–10.

Khalfa, S. & Bella, S., 2003. Effects of relaxing music on salivary cortisol level after psychological stress. *Annals of the New York* Academy of Sciences, volume 999, pp. 374-376.

Kivy, A.P., 2007. Moods in the Music and the Man : A Response. MIT Press.

Koelsch, S., 2011a. Toward a neural basis of music perception–a review and updated model. *Frontiers in psychology*,2, 110.

Koelsch, S., 2010. Towards a neural basis of music-evoked emotions. *Trends in cognitive sciences*, volume 14, pp. 131-137.

Koelsch, S., 2011b. Towards a neural basis of processing musical semantics. *Physics of life reviews* 8.

Koelsch, S., Gunter, T. & Zysset, S., Lohman G, Friederici, A, 2002. Bach speaks: a cortical "language-network" serves the processing of music. *Neuroimage* 17, 959-966.

Konec, V.J., 2013. Music , Affect , Method , Data : Reflections on the Carroll Versus Kivy Debate. , 126(2), pp.179–195.

Krizman, J. & Marian, V., 2012. Subcortical encoding of sound is enhanced in bilinguals and relates to executive function advantages. PNAS 10.

Krizman, J., Skoe, E. & Kraus, N., 2012. Sex differences in auditory subcortical function. *Clinical Neurophysiology* 123, 590-597.

Krumhansl, C., 1990. *Cognitive foundations of musical pitch*. New York, Oxford University Press.

Krumhansl, C.L. & Cuddy, L.L., 2010. 3- A Theory of Tonal Hierarchies in Music. In M. R. J. et al. (eds.), ed. *Music Perception*. Springer Science+Business Media, pp. 51–87.

Ladinig, O. & Honing, H., 2009. Newborn infants detect the beat in music, PNAS, 106 (7), pp.1–4.

Large, E. & Jones, M., 1999. The dynamics of attending: How people track time-varying events. *Psychological review*, Volume 106,(1), pp.119-159.

Large, E. & Kolen, J., 1994. Resonance and the perception of musical meter. *Connection science*, 6, pp.177-208.

Lerdahl, F., 2001. *Tonal pitch space*. New York, Oxford University Press.

Lerdahl, F., 2013. Musical syntax and its relation to linguistic syntax. In M. A. Arbib (ed.), *Language, Music and the Brain: A mysterious relationship*.

Lerdahl, F. & Jackendoff, R., 1983. An overview of hierarchical structure in music. *Music Perception* 1, no. 2, pp.229-52.

Livingstone, F., 1973. Did the Australopithecines sing? *Current Anthropology*, volume 14, pp 329-360.

London, J., 2012. Three Things Linguists Need to Know About Rhythm and Time in Music. *Empirical Musicology Review*, 7(1), pp.5–11.

London, J., 2012. Hearing in time: *Psychological aspects of musical meter* (2nd ed.). Oxford, UK: Oxford University Press, pp 256.

MacNeilage, P., 2008. *The origin of speech*, Oxford University Press, Oxford.

Manning, F. & Schutz, M., 2013. "Moving to the beat" improves timing perception. *Psychonomic bulletin & review*, 20.

McDermott, J., 2008. The evolution of music. *Nature*, 453(7193), pp.287–8.

McDermott, J., Lehr, A. & Oxenham, A., 2010. Individual differences reveal the basis of consonance. *Current Biology*, 28(1), 175-184.

Menon, V. & Levitin, D., 2005. The rewards of music listening: response and physiological connectivity of the mesolimbic system. *Neuroimage*.

Merchant, H. & Honing, H., 2013. Are non-human primates capable of rhythmic entrainment? Evidence for the gradual audiomotor evolution hypothesis. *Frontiers in neuroscience* 7(274), 1-8.

Merker, B., 2000. Synchronous chorusing and the origins of music. *Musicae Scientiae*, Special Issue pp. 59-73.

Merker, B., Madison, G. & Eckerdal, P., 2009. On the role and origin of isochrony in human rhythmic entrainment. *Cortex*.

Miller, G. F. (2000). Evolution of human music through sexual selection. In N. L. Wallin, B. Merker, & S. Brown (Eds.), The origins of music, MIT Press, pp. 329-360.

Mithen, S., 2005. *The singing Neanderthals: The origins of music, language, mind, and body*. Cambridge, Harvard University Press (2006).

Montinaro, A., 2010. The musical brain: myth and science. *World neurosurgery*, 73(5), pp.442–53.

Motz, B., Erickson, M. & Hetrick, W., 2013. To the beat of your own drum: Cortical regularization of non-integer ratio rhythms toward metrical patterns. *Brain and cognition* 81(3), 329-336.

Nozaradan, S., Palmer, C. & Peretz, I., 2011. Born to dance but beat deaf: A new form of congenital amusia. *Neuropsychologia*, 49(5), 961-9.

Nozaradan, S. & Peretz, I., 2011. Tagging the neuronal entrainment to beat and meter. *The Journal of Neuroscience*, 31(28), 10234-10240.

Nozaradan, S., Peretz, I. & Mouraux, A., 2012. Selective neuronal entrainment to the beat and meter embedded in a musical rhythm. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 32(49), pp.17572–81.

Oatley, K. & Johnson-Laird, P., 2014. Cognitive approaches to emotions. *Trends in cognitive sciences*, volume 178, Issue 3, pp.134-140.

Owren, M., Amoss, R. & Rendall, D., 2011. Two organizing principles of vocal production: Implications for nonhuman and human primates. *American journal of Primatology*, 73, 530-544.

Patel, A.D., 2003. Language, music, syntax and the brain. *Nature neuroscience*, 6, 674-681.

Patel, A.D., 2008. Talk of the tone. *Nature*, 453 (June), pp.726–727.

Patel, A.D., 2008, Music , Language and the Brain . Oxford : Oxford University Press, Organization. *Empirical Musicology Review*, 3(4), pp.218–222.

Patel, A.D., 2010. Music, biological evolution, and the brain. In M. B. (Ed.), ed. *Emerging Disciplines*. Houston: TX: Rice University Press, pp. 1–37.

Patel, A.D., 2014. The evolutionary biology of musical rhythm: was darwin wrong? *PLoS biology*.

Patel, A.D. & Hopkins, J.J., 2008. Rhythm in speech and music. , 3(2006).

Patel, A.D. & Iversen, J., 2009. Studying synchronization to a musical beat in nonhuman animals. *Annals of the New York Academy of Science,* 1169; 459-669.

Patel, A. et al., 2009b. Experimental evidence for synchronization to a musical beat in a nonhuman animal. *Current Biology*, 7.

Pearce, M. & Rohrmeier, M., 2012. Music cognition and the cognitive sciences. *Topics in cognitive science*, 4(4), pp.468–84.

Perani, D. & Saccuman, M., 2010. Functional specializations for music processing in the human newborn brain. PNAS, Pproceedings of the National Academy of Sciences 107(10), 4758-4763.

Peretz, I., 2010. The amusic brain: Lost in music, but not in space. *PloS ONE*, 5(4).

Peretz, I., 2006. The nature of music from a biological perspective. *Cognition* 100.

Peretz, Isabelle; Zatorre, Robert J. (ed.). 2003. *The Cognitive Neuroscience of Music*, Oxford University Press.

Perlovsky, L., 2013. Cognitive function , origin , and evolution of musical emotions. *Cybernetics and informatics volume* **11** *- number* **9***, 11(*9), pp.1–8.

Petkov, C. & Jarvis, E., 2012. Birds, primates, and spoken language origins: behavioral phenotypes and neurobiological substrates. *Frontiers in evolutionary neuroscience* 4(1), 12.

Petkov, C.I., Logothetis, N.K. & Obleser, J., 2014. Neuroscientist Where Are the Human Speech and Voice Regions, and Do Other Animals Have Anything Like Them?*Journal of neuroscience*, 34(7)

Phillips-Silver, J., Toiviainen, P. & Gosselin, N., 2011. Born to dance but beat deaf: a new form of congenital amusia. *Neuropsychologia*, 49(5).

Pinker, S., 1997. *How the mind works*. NewYork: Norton.

Pinker, S., 2007. *The stuff of thought: Language as a window into human nature*, Penguin Group, Harvard University.

Ravignani, A., 2014. The evolutionary origins of rhythm : A top-down / bottom-up approach to temporal patterning in music and language. *Procedia - Social and Behavioral Sciences*, 126, pp.113–114.

Ravignani, A.; Gingras, B.; Asano, R.; Sonnweber, R; Matellán, V, & Fitch, T. 2013 The Evolution of Rhythmic Cognition : New Perspectives and Technologies in Comparative Research. pp.1199–1204.

Repp, B., 2005. Sensorimotor synchronization: a review of the tapping literature. *Psychonomic bulletin & review* 12(6), 969-992.

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

Richman, B., 1993. On the evolution of speech: Singing as the middle term. *Current Anthropology*. Rizzolatti, G. & Arbib, M., 1998. Language within our grasp. *Trends in neurosciences*.

Rogalsky, C. & Rong, F., 2011. Functional anatomy of language and music perception: temporal and structural factors investigated using functional magnetic resonance imaging. *The Journal of Neuroscience,* 31(10).

Rohrmeier, M., 2007. A generative grammar approach to diatonic harmonic structure. *Proceedings SMC'07, 4th Sound andMusic Computing Conference*, (July), pp.11–13.

Rohrmeier, M. et al., Music Cognition : Learning and Processing Moderators : , pp.41–42.

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

Rosselló, J. 2013 A musical protolanguage consonant with human syllables. Ways to Protolanguage 3. Comunicación. Wroclaw.

Rosselló, J., 2014. On the separate origin of vowels and consonants. In *EvoLang X*.

Salimpoor, V. & Bosch, I. van den, 2013. Interactions between the nucleus accumbens and auditory cortices predict music reward value. *Science,* 340, pp.216-219.

Salimpoor, V. & Zatorre, R., 2013. Neural interactions that give rise to musical pleasure. *Psychology of* Aesthetics, Creativity, *and the* Arts, 7(1)62-75.

Sammler, D., Koelsch, S. & Friederici, A.D., 2011. Are left fronto-temporal brain areas a prerequisite for normal music-syntactic processing? *Cortex; a journal devoted to the study of the nervous system and behavior*, 47(6), pp.659–73.

Särkämö, T., Tervaniemi, M. & Laitinen, S., 2008. Music listening enhances cognitive recovery and mood after middle cerebral artery stroke. *Brain*, 131 (3), 866-876.

Schachner, A. et al., 2009. Spontaneous motor entrainment to music in multiple vocal mimicking species. *Current Biology*, Volume 19, Issue 10, pp.831–836.

Schlaug, G., Marchina, S. & Norton, A., 2009. Evidence for Plasticity in White-Matter Tracts of Patients with Chronic Broca's Aphasia Undergoing Intense Intonation-based Speech Therapy. *Annals of the New York Academy of Sciences*, 1169, 385494.

Scholar, V., 2011. Music . Cognitive Function , Origin , And Evolution Of Musical Emotions Music . Cognitive Function , Origin , And Evolution Of Musical Emotions Theories of Musical Emotions and Music Origins. , 2(2), pp.1–15.

Slocombe, K., Waller, B. & Liebal, K., 2011. The language void: the need for multimodality in primate communication research. *Animal Behaviour*, 86(2), pp.483-488.

Smith, N. & Schmuckler, M., 2004. The perception of tonal structure through the differentiation and organization of pitches. *Journal of experimental psychology: Human perception and performance,* 30, 268.282.

Snowdon, C. & Teie, D., 2009. Affective responses in tamarins elicited by species-specific music. *Biology letters* 6, 3042.

Sparks, R., Helm, N. & Albert, M., 1974. Aphasia rehabilitation resulting from melodic intonation therapy. *Cortex* 10(4), pp 303-316.

Spencer, H., 1950. The Origin and Function of Music, (1857). *Literary Style and Music,* London: Watts and Co.

Stahl, B. et al., 2011. Rhythm in disguise: why singing may not hold the key to recovery from aphasia. *Brain*.

Steinbeis, N. & Koelsch, S., 2008. Shared neural resources between music and language indicate semantic processing of musical tension-resolution patterns. *Cerebral cortex (New York, N.Y. : 1991)*, 18(5), pp.1169–78.*

Taglialatela, J.P. et al., 2009. Visualizing vocal perception in the chimpanzee brain. *Cerebral cortex (New York, N.Y. : 1991)*, 19(5), pp.1151–7.

Tallerman, M., 2007. Did our ancestors speak a holistic protolanguage? *Lingua 117*, pp. 579-604, Oxford.

Tallerman, Maggie; Gibson, K.R., 2011. *The Oxford Handbook of Language Evolution* Oxford University, New York.

Tallerman, M., 2013. Join the dots: A musical interlude in the evolution of language? *Journal of Linguistics*.

Tattersall, Ian, 2013. "An evolutionary context for the emergence of language." *Language Sciences* 46, pp 199-206. (Actually published in 2014)

Thaut, M., 2005. Rhythm, music, and the brain: Scientific foundations and clinical applications, New York, Routledge, Taylor & Francis Group.

Trainor, L., 2008. ESSAY The neural roots of music. *Nature*, 453(May), pp.598–599.

Trehub, S.E. & Hannon, E.E., 2006. Infant music perception: domain-general or domain-specific mechanisms? *Cognition*, 100(1), pp.73–99.

Trollinger, V.L., 2010. The Brain in Singing and Language. *General Music Today*, 23(2), pp.20–23.

Vaux, B. & Myler, N., 2011. Metre is music : A reply to Fabb and Halle. pp.43–50.

Vuust, P., Roepstorff, A. & Wallentin, M., 2006. It don't mean a thing: Keeping the rhythm during polyrhythmic tension, activates language areas (BA47). *Neuroimage*.

White, S.A., 2009. Genes and vocal learning. *Brain and Language* 2010; 115: 21-28.

Wray, A., 2000. Holistic utterances in protolanguage: the link from primates to humans, In Chris Knight, Michael Studdert-Kennedy, and James R. Hurford (eds.) *The evolutionary emergence of language: Social function and the origins of linguistic form*. Cambridge: Cambridge University Press.

Wu, D., Li, C. & Yao, D., 2009. Scale-Free Music of the Brain. , 4(6), pp.4–11.

Yu-sha, L., Yu-hong, J. & Fei, L., 2014. Exploration of the Facilitative Effects of Music Acquiring Device on Language Acquisition. *Sino-US English Teaching*.

Zatorre, R., 2005. Music, the food of neuroscience? *Nature*, 434 (March), pp.312–315.

Zatorre, R., Belin, P. & Penhune, V., 2002. Structure and function of auditory cortex: music and speech. *Trends in cognitive sciences*.

Zatorre, R., Chen, J. & Penhune, V., 2007. When the brain plays music: auditory–motor interactions in music perception and production. *Nature Reviews Neuroscience*, 8(7), pp.547–58.

Zimmerman, M., Pan, J. & Hetherington, H. 2008. Hippocampal neurochemistry, neuromorphometry, and verbal memory in nondemented older adults. *Neurology*.

ANNEX

11. *Musical diseases and music therapies*

About 4% of the general population suffers from amusia, a disorder associated with structural differences in temporal and frontal cortices distinct from aphasic linguistic deficits. Congenital tone deafness is considered to be a developmental disorder arising from failures in fine-grained pitch encoding, related to general psychoacoustic difficulty in fine pitch resolution. Individuals affected by amusia show a reduced quantity of white matter in the right IFG but a larger amount of grey matter, showing a cortical malformation, due to an altered pattern of neuronal migration. Amusicians also may have affected the left frontotemporal auditory-motor network.

Other musical diseases include hallucinations and epilepsy. Musical hallucinations, which may be self-generated, disordered impulses within the secondary auditory areas, activate all the musical listening regions except for the primary auditory cortex, as it lacks an external stimulus. Musicogenic epilepsy is due to an anomalous activation of temporal-limbic structures associated with the emotional response to music. An ictal hyperfusion in the right temporal lobe, insula, amygdale and hippocampus head, seem to be involved in this diseases.

Music therapy aids in a vast range of diseases and disorders, because it improves fine-movement precision, posture control and walking, affective states and mood (Montinaro, 2010). Music therapy, with dance and rhythmic games, is used for neuromotor rehabilitation in stroke patients, Parkinson and Alzheimer diseases, multiple sclerosis, ataxia and spasticity; for social communication enhancing in children with autism, for neurotrophin modulation and mood restitution in hypertensive patients, and depression, anxiety and stress diagnostic cases. Therefore, it is found that moving with a musical beat alleviates symptoms in movement disorders, as Parkinson Disease (Fitch, 2009).

Another clinical observation of music comes from cases of aphasics recovering verbal fluency in which music promotes structural changes in patients' brains after having undergone a melodic intonational therapy (Albert et al, 1973; Schlaug et al., 2009). The main change is an increase in the thickness of the right arcuate fasciculus,

(although non-significant), related to the degree of verbal fluency improvement, as it is shown through greater right hemisphere activation in speaking.

Stahl et al. (2011) highlights the clinical importance of rhythm (rather than melody) in music therapy for aphasic patients, due to the role of the supplementary motor cortex and the basal ganglia in human beat perception (Grahn, 2009). Moreover, it has been said that simply tapping to a beat enhances our auditory time perception abilities (Manning & Schutz, 2013). Given that perceiving and producing hierarchical structures in language are usually attributed to Broca's area —BA 44 and 45— (Friederici et al., 2006), together with findings of activations of Broca's right homolog in harmonic syntax tasks (Koelsch et al, 2002), we can point to the human expanded Broca's region connecting to posterior associative and auditory areas to place rhythmic cognition, thus playing a role in building hierarchical structures in metrical perception (Vuust et al., 2006; Geiser et al., 2008).

12. *Alexithymia (an emotional disease)*

Alexithymia is recognized as a risk factor of psychiatric and medical disorders, such as somatisation, anxiety, depression, hypertension, and chronic pain; and exhibits high comorbidity with disorders of the Autism Spectrum. Affected individuals are described as cold and distant, interpersonally indifferent, and show paucity of facial emotional expressions and stiff wooden posture. In addition to that, the impairment of facial emotional recognition and words connoting emotional meanings lead to social communication problems. This neurobiological dysfunction could be attributed to right hemisphere hypoactivity and left hemisphere hyperactivity, as well as to some interhemispheric communication deficit. A deficiency in detecting the emotional qualities of prosody —cues of others' emotional state and intention in social communication— has also been reported for this trait. Goerlich, Aleman and Martens (2012)'s experiments on women —because of the gender behavioural and electrophysiological differentiation in perceiving prosody— show reduced sensitivity during the perception of mismatches in the emotional cues of speech and music.

Since alexithymia affects how the brain processes emotional speech qualities —equally in attended and unattended processing—, it could be inferred a link from this ability to an evolutionary selected trait implied in the musical protolanguage. In

fact, the two alexithymia dimensions reveal a dissociable impact on emotional processing, with a left-hemisphere bias during early stages of unattended processing (for the cognitive dimension, rather related to the linguistic feelings) and with an additionally sensitive to the intensity of emotional speech at later processing stages (for the affective dimension, rather related to prosody, music or emotions).

13. *Adaptationist vs. Non-adaptationist models for music evolution*

Darwin (1871) observed that music, despite being a human universal carrying a physiological cost and playing an important role in society, does not show any obvious function. For that reason, music would be better seen as a fossil remaining from a former adaptation, that is, a communicational system used by earlier hominids whose core original function is now developed by language.

Several authors has taken his idea about musical protolanguage (Jespersen, 1922; Livingstone, 1973; Richman, 1993; Brown, 2000; Mithen, 2005; Fitch, 2006), i.e. a common origin of music and language (more concretely, speech), and are now investigating the cognitive, neural and genetic mechanisms underlying both faculties in modern humans, as well as comparing our human vocalizing abilities in this domains to non-human animal communicational systems with complex learned or innate vocalizations.

Tecumseh Fitch (2006) proposes music to be "an instinct to learn, fuelled by certain proclivities and channelled by various constraints", in which cultural and biological aspects intertwine. An example of a proclivity would be the innate template of young birds to pay attention to and imitate their conspecific vocalizations, that is, to distinguish their species-specific song from others which do not fit their template. Hence, their adult normal song are neither innate, nor entirely learned, but channelled by a species-specific set of proclivities and constraints, such as their vocal learning ability or their innate templates. A parallel case for human music might be applied, following Fitch (2006), where music is constituted by basic learning and imitative abilities, as well as *proclivities* "for tonal and rhythmic sounds arranged in interesting structures, with particular favoured frequencies and tempos", and *constraints* on "repetition rates, frequency limens, number of notes in a scale, basic consonance and dissonance judgements". The adaptative ability to acquire complex

novel aspects from the environment is the authentic responsible for the great diversity in our music, language and culture.

From a non-adaptationist approach, musical abilities have not been naturally selected; they are simply by-products and peculiarities of our nervous system, which links them to pleasure likely due to an accidental brain-circuitry wiring (Spencer, 1857; James, 1890; Pinker, 1997). In this position, Patel (2010) enumerates Pinker's non-musical foundational elements building on music, without being selected specially for it:

> 1) A prosodic component of language: music has prosody-like properties, and the brain rewards the analysis of prosodic signals (patterns of linguistic rhythm and intonation) because prosody is an important component of language
>
> 2) An auditory scene analysis: music is rich in harmonic sounds (sounds in which frequency components are integer multiples of some fundamental frequency), and the brain rewards the analysis of such sounds because harmonicity is an acoustic cue used to identify sound sources, an important part of auditory scene analysis
>
> 3) Emotional calls: music can evoke strong emotions because it contains pitch and rhythm patterns that resemble our species' emotional calls,
>
> 4) Habitat selection: because it contains sound patterns reminiscent of evocative environmental sounds (e.g. "safe" or "unsafe" sounds such as thunder, wind, or growls)
>
> 5) Motor control: musical rhythm engenders rhythmic movement (e.g., in dance), and such movement is rewarded by the brain because rhythmic motor patterns are associated with biologically meaningful behaviours, such as walking, running, or digging.

Similarly, but different in some sense, Patel's Transformative Technology of the Mind (TTM) theory maintains that "music is a human invention that can have lasting effects on such non-musical brain functions as language, attention, and executive function, and is concerned with explaining the biological mechanisms underlying these effects". Furthermore, Patel proposes that complex and universal human traits can originate as inventions, instead of biological adaptations, indicating parallel cases, such as *reading* or *fire-making*. Thus, within TTM framework, functional specializations in brain simply come as an experience-dependent neural plasticity product in the individual lifetime. Showing that music

cognition is rooted in non-musical human brain functions that is also shared with other species, would imply that some musical aspects are not shaped by natural selection for music; like tonality processing or synchronization of movement to a musical beat.

14. *The Darwinian musical protolanguage*

Stage (i) could be linked to the genus homo (*Homo habilis*), or even the genus Australopithecus, as social intelligence and technological-ecological intelligence played a key role in these early societies. Looking at (ii), the aesthetical use of vocalizations may have selected and evolved complex vocal learning abilities. From that, one can infer the idea that some phonological and syntactical aspects may have preceded the ability of speech to convey propositional meanings —which fits with cross-species findings of complex vocal learning evolution without propositional meaning. However, the role of sexual selection can be challenged by two facts of modern language (Fitch, 2013a): "it is equally developed in males and females" (if not better in females), and "it is expressed very early in the ontogeny, essentially at birth" (although even in the womb, when tuning phonology); contrasting to the normal expression of sexual traits in the competitive sex at sexual maturity. Solving this possible incongruence, Fitch (2013a) proposed that two different forces selected a musical protolanguage: the sexual selection of mature males' song and the kin selection of mother-infant communication. The latter may have occurred during the evolution of propositional semantic meaning in the musical protolanguage, and may have been supported by the current child-care context of *motherese*, and by the fact that both male and female infants participate in parents-offspring communication during the extended childhood, which enhances the survival of human small reproductive outcome.In addition, current research has recently demonstrated that sexual selection can often induce female birdsong, and that "pair-bonding" mechanisms can also make both sexes choosy, allowing the competition for thigh-quality mates, and in turn better offspring. In contrast, others could defend that it is quite possible that sexual selection did not take any part in selecting a musical protolanguage, but rather kin selection only, given the current functions of mother infant music, as lullabies, and infants' preferences to song over speech.

Stage (iii) presents a big challenge to Darwin, because his model explains lexical semantics but not phrasal semantics, thus missing the origin of functional words and grammatical morphemes. Otto Jespersen's (1922) hypothesis of a holistic protolanguage —rediscovered and supported with evidence by Alison Wray (2000) and Michael Arbib (2005)— squares with Darwin's musical protolanguage model and fills the phrasal semantics gap by suggesting that a cognitive analytical process may have slowly divided the entire sung phrases (with whole propositional meanings) into isolated musical chunks, which in turn were associated to individual meaningful components from a precursor of our conceptual system.