The origin of articulate language revisited: The potential of a semi-aquatic past of human ancestors to explain the origin of human musicality and articulate language

Mario Vaneechoutte

Reprinted from Human Evolution 29: 1-33; 2014.
http://users.ugent.be/~mvaneech/Vaneechoutte.%202014.%20The%20origin%20of%20articulate%20language%20Human%20Evolution.pdf

Mario.Vaneechoutte@UGent.be
Laboratory Bacteriology Research, Faculty of Medicine & Health Sciences, University of Ghent
De Pintelaan 185, 9000 Gent, Belgium
Tel: +32 9 332 3692
Fax: +32 9 332 3659

Abstract

Articulate language depends on very different abilities, such as vocal dexterity, vocal mimicking and the acquirement by children of the very different and arbitrary phonology and grammatical structure of any language.

A vast array of experiments confirm that children acquire grammar by the use of prosodic clues, basically intonation and pitch, in combination with e.g. facial expression and gesture. Prosodic clues, provided by speech, are exaggerated in infant-directed speech (motherese). Moreover, strong overlap between musical and linguistic syntactic abilities in the temporal lobes of the brain has been established. A musical origin of language at the evolutionary level (for the species *Homo sapiens*) and at the ontogenic level (for each newborn) is parsimonious and no longer refutable.

We then should ask why song, i.e. vocal dexterity and vocal learning, was evolved in our species and why it is largely absent from other 'terrestrial' animals, including other primates, but present in disjoint groups such as cetaceans, seals, bats and three orders of birds. We argue that this enigma, together with a long list of other specifically human characteristics, is best understood by assuming that our recent ancestors (from 3 million years ago onwards) adopted a shallow water diving lifestyle. The swimming and diving adaptations of the upper airway (and vocal) tract lead to increased vocal dexterity and song, and to increased finetuning of motoric and mimicking abilities. These are shared by creatures that can freely move in three dimensions (swimming and flying animals) and that can respond instantaneously to behavioural changes of other animals. Increased bodily mimicking together with increased vocal dexterity, both a consequence of a semi-aquatic lifestyle, lead to integrated song and dance, which predisposed to producing and mimicking speech and gesture and to the ability to use prosodic clues to learn the grammar of whichever language.

Key words: swimming, diving, moving in 3 dimensions, mimicking, song, dance, prosody, articulate language, semi-aquatic past

Content

1. Why it is important to explain the origin of language

Articulate language (spoken language, symbolic language) can be regarded as the most important faculty that distinguishes humans from all other animals.

Articulate language, which is considered to be present only in our species, *Homo sapiens*, is enigmatic with regard to how such a peculiar ability could have originated at all, and this question is closely connected to the question why articulate language arose only once, and only in our species.

Although articulate language as such is not necessary for higher mental abilities or for improved hunting skills (see 3.4), it has provided us with a separate reference frame, which enabled reflexive awareness (*i.e.*, consciousness (Vaneechoutte, 2000)) and foresight, which forced us into religiosity (Vaneechoutte, 1993) and brought technology and science.

It is articulate language which, very quickly after its origin (estimted by many at at most 200 000 years ago), changed the appearance and future of planet Earth. Of course, it is important to understand how such a feature could have naturally evolved.

2. What needs to be explained?

First, the possibility to use (produce, reproduce (mimick) and understand/respond to) articulate language depends on many different and independently evolved characteristics (see 2.3). Intriguingly, production by one individual of random utterances (see 3.1) and making sense of these utterances by another, represent independent trajectories as well, which, beyond that, must develop simultaneously. Put otherwise, a conversation requires at least two people – but how could someone invent language at exactly the same time that someone else figured out how to decode it? This requires a more complicated explanation than when accounting for the evolution of the complex eye, which can be achieved by straightforward natural selection of an advantageous trait (better eye = better sight). Moreover, we will argue that language was certainly not very useful at first, and that it is not even all that advantageous when fully developed (see 3.4), which minimizes the likelihood that it was naturally selected.

Second, stating that complex traits such as speech production and speech comprehension, which by nature rely on several other and very different characteristics, was selected because it was advantageous, does not in itself provide the mechanisms via which these traits could have originated and could have been perfectionized. Thus, the explanation must also outline the trajectory via which articulate language could originate.

Third, a useful explanation for the mechanism through which articulate language could have developed in our species, must not only be phylogenetic, *i.e.*, explaining how it could have developed in our species, but also ontogenetic, *i.e.*, explaining how each individual human, from childhood onwards, can develop the ability of producing as well as understanding speech. Ideally, the same kind of mechanism should be able to explain both phylogenetic and ontogenetic processes.

Fourth, a valid hypothesis should also explain why articulate language originated only in our species, and why it is exclusive to our species. For example, better speech for better hunting not only is an improbable explanation, *e.g.*, because chimps are better hunters without language (see 3.4) and because most human hunters keep silent when hunting and rely on sign or click languages), but it also does not provide a mechanism via which language could have developed in our species (no phylogenetic explanation) and it appears improbable that language selected on the basis of its advantage for hunting could be easily picked up by children, hence also lacking ontogenetic explanatory power. Finally, since none of the many predatory species rely on articulate language, the hunting hypothesis fails the exclusivity

criterion as well, as one would expect articulate language to have developed in at least some of the hunting species.
We need more sophisticated explanations, rather than simply claiming that language was advantageous and therefore naturally selected.

2.1. Timing of the origin of language

First, knowing about the appropriate time that language emerged is of importance when evaluating the validity of claims (e.g., Deacon, 1997; Lieberman, 2012; Pinker & Bloom, 1990; Pinker, 1994; Smith & Szathmáry, 1995) that ever better language, starting from primitive proto-language, was sufficiently advantageous to propel natural selection towards fully developed language. However, currently most scholars agree, on the basis of the time of appearance of cultural artefacts from only the last 125,000 years onwards and on the basis of anatomical characteristics (see for example, Lieberman, 2012), that full-fledged language originated only recently. Therefore, the recent origin of language appears to indicate that natural selection propelled by the advantages of language itself is not a valid explanation for its origin, simply because there was insufficient time for natural selection to promote the very different characteristics that are necessary to enable speech. I agree with the view that language originated only recently and that therefore the several independent characteristics that were needed to make language possible (preadaptations) must have been selected for other reasons.
The exact timing of the origin of modern language is also important in order to determine which of the adaptations that can be deduced from fossils and from genetics (see 3.5) can be considered as indicative of language. For an excellent review of the difficulties associated with interpreting the fossil evidence in relation to our vocal abilities, see Wurz (2009) - although she considers the data in the context of our presumed running adaptations.
To summarize, although several fossil characteristics, *e.g.*, the presence and development of Broca's brain area - as can be read from skull endocasts, or indications for a descended larynx, and genetic mutations, *e.g.*, the *foxP2* mutations (Krause *et al.*, 2007), are usually interpreted as indicative of articulate language, it is generally overlooked that these might, just as well, and more parsimoniously, be indicative of the presence of diving (breath holding), running (see Wurz, 2009) and musical (song production) abilities (Maess *et al.*, 2001), and consequently point to evolutionary events that are preadaptions for language, but that are not necessarily indicative of the presence of language (see 3.5).

2.2. Uniqueness: Why only in one species? Why in our species?

Although one cannot exclude that symbolic language exists to some degree in *e.g.*, dolphins, bats, elephants, crows and parrots, and therefore is not entirely unique to our species, we will assume that articulate, symbolic language is a unique characteristic of the species (that names itself) *Homo sapiens*, because it would seem that only our species has made use of this capability to increase its mental possibilities and eventually (during the last few centuries) become a major evolutionary force on Earth. Therefore, any explanation on the origin of language also needs to account for why it evolved in only one species and why in our species. Yet none of the explanations put forward thus far fulfill the uniqueness criterion.
First, the general explanation: 'Articulate language was advantageous and therefore it was naturally selected' (Pinker, 1994; Deacon, 1997) would seem to imply that many, if not most, animals would be using symbolic speech, and therefore that speech should be more ancient than its widely accepted maximum 200,000 years.

Second, the hypothesis that articulate language developed because it improved our hunting abilities also fails to explain its uniqueness, since one would expect that at least several of the many hunting species would be able to communicate by way of symbolic language (see 3.3).
A third hypothesis (Pinker, 1994) considers sexual selection to be the means through which articulate language evolved. Thereby, women chose males on the basis of their oratory abilities, with better speaking males having more offspring and whereby, consequently, the language gene(s) spread. The problem we encounter here is that one must then explain how women were attracted to babbling men, when the women could not understand what the men were saying, unless, though highly improbable, they developed the unrelated ability to understand speech, at exactly the same time as men developed speech. Another problem related to the sexual selection hypothesis is that it also does not explain how women came to speak, even though it is general known that they are the truly talkative gender). Most importantly, this hypothesis fails the uniqueness criterion because many traits in different species have been selected sexually, and therefore speech could have been selected in many species by sexual selection.
On the other hand, I will argue that sexual selection easily explains our musicality (see 4.1) and furthermore that our uniqueness in articulate language can be best understood as the result of the very specific trajectory that the genus *Homo* followed during the last 5-8 million years, including a semi-aquatic phase, which predisposed towards, or strongly improved, our singing capacities, by selecting for voluntary controlled breathing, vocal dexterity and vocal learning. Together with the enlargement of our brains, also best explained by a DHA-rich diet (see 2.3), as can be found in marine/coastal environments, the increased singing and associative capacities in turn predisposed to the more recent emergence of language.
In other words, I claim that the several unrelated preadaptations, necessary for articulate language, are more likely to have been brought about by a swimming/diving lifestyle and by our singing abilities, whereby the latter were also promoted by our semi-aquatic lifestyle.
The elaboration of this peculiar evolutionary route, *i.e.*, a swimming/diving ape that was preadapted for learned song, and consequently for articulate language, not only provides a sufficient explanation as to why articulate language is only observed in one species, thus fulfilling the uniqueness criterium, it may also, at the same time, provide an explanation as to why it is only observed in our species.

2.3. Which adaptations are necessary for articulate language?

Articulate language depends on several very different, unrelated abilities, which can therefore be considered to have originated from different selection pressures and evolutionary pathways. These abilities are best regarded as preadaptations, in other words, traits which evolved from other selection pressures rather than those imposed by the possible advantages of language, but which coincidently developed to a point whence language could emerge.

2.3.1. Vocal flexibility/vocal dexterity

Articulate language depends on vocal flexibility or vocal dexterity to produce the wide range of consonants, vowels and intonations. Vocal flexibility itself depends on respiration (diaphragm, descended larynx), vocalization (vocal chords) and articulation (formants: tongue (most important), and teeth and nose), which are three different abilities, each requiring their own evolutionary explanation. I argue that several elements needed for vocal dexterity were developed for the (shallow water) diving lifestyle of our ancestors (see 5.3), that were further refined for improved singing abilities (see 4.1-4.3).

2.3.2. Vocal learning

Vocal learning or, more precisely, vocal production learning (Janik & Slater, 2000), i.e., the vocal reproduction of sounds that are overheard, is necessary for the reliable imitation of phonemes, rhythms and intonations, as used in song and language. This is well-known in songbirds, the largest group of birds (4000 out of 9000 bird species), parrots and hummingbirds, but is virtually unknown in mammals. Among mammals, only humans, some cetaceans (Janik *et al.*, 1994), some pinnipeds (Ralls *et al.*,1985), some bats (Jones & Ransome, 1993; Boughman, 1998; Bohn *et al.*, 2013), and possibly to a degree elephants (Byrne *et al.*, 2009; Hart *et al.*, 2008), yet no primates (Janik & Slater, 1997), are capable of vocal learning. Humans (and elephants, see 5.4) find themselves in a disjoint group of animals: flying birds and mammals (bats) and swimming/diving mammals.

Fitch (2006) only considers sound production to be song when it is based on vocal learning and classifies all other animal utterances as stereotypic 'calls'. For example, the limited interindividual variability of dugong 'songs' (Okumura *et al.*, 2007) would be classifed as calls, whereas in manatees, the vocalizations of calves well resemble those of their mothers (Sousa-Lima *et al.*, 2002). Similarly, even gibbons and siamangs, probably the best vocalists among nonhuman primates, use stereotypic calls that are inheritable (Geissman, 2000; Fitch, 2006). Further on (see 5.4), I speculate on why one could expect vocal learning to be present in precisely these disjoint groups.

Notably, for vocal learning to be of use in explaining language, both genders should engage in this activity because in several species (e.g., most song birds, humpback whales) only males display song and vocal learning. In humans, both genders vocalize and have vocal learning - with greater linguistic skills usually attributed to the female gender. In manatees, calves mimic their own mother's song, and bottlenose dolphins and killer whales have vocal learning as well (Filatova *et al.*, 2013). Ralls *et al.* (1985) recorded vocalizations from captive harbor seals. Pups of both sexes vocalized, but females above one year of age rarely vocalized. Two adult males produced sounds that mimicked one or more English words and phrases, which led the authors to speculate that male harbor seals may mimic other males in the wild.

2.3.3. Increased associative capacity

Certainly, increased brain capacity can not be considered disadvantageous for the development of a complex ability such as articulate language and grammar. Still, the importance of a large brain in the ability to acquire symbolic language may be limited (Duchin, 1990). Indeed, linking (auditory) symbol to meaning is for example a capacity that does not need an enlarged human brain, because trained dogs can also recognize at least 200 words. It has also been well demonstrated that apes are able to learn abstract language to some degree, without having a brain the size of humans. It will be argued that the best explanation for the enlarged brain of our species derives from a semi-aquatic past (see 5.1), not forgetting that at present half of the world's human population is still fed by the sea.

2.3.4. Learning of the random syntax (and random phonology) of human languages by young children

Assuming that grammar is not at all universal and also not innate (see 3.1 and 3.2), I have previously advocated (Vaneechoutte & Skoyles, 1999) that children acquire whichever random language using sensory clues offered by the people talking to them, of which the most important clue is the melody of speech, i.e., intonation and rhythm. Although it can be easily

dismissed that children have an innate language acquiring device, they do have a music acquiring device (e.g., Schön *et al.*, 2004). This kind of innateness is much easier to understand from an evolutionary point of view because it is a general characteristic (as opposed to learning specific grammars) and also because it evolved independently in very different evolutionary lineages (unlike linguistic grammar). Numerous studies have shown how, indeed, language and music rely on the same resources (see 4.4) and how children rely largely on prosodic clues when acquiring language (see 4.5).

3. Misconceptions about language and its origin

Before elaborating on what I consider to be the most plausible hypothesis for understanding the many peculiar and unique characteristics of the human species, including articulate language, I would like to deal first with several paradigms which have misguided scientists who are interested in human evolution and the origin of articulate language. There is, of course, the Chomskyan paradigm which claims that linguistic grammar is universal among languages and innate to the human species (see 3.1 and 3.2). Moreover, starting from the notion that language is innate and that we are genetically predisposed for language, there has also been the claim that language was naturally selected for, *i.e.*, that the advantages of language were so huge that language could act as its own selective force: better speaking individuals on average had more offspring than poorly speaking individuals and therefore the language genes could spread throughout the population (see 3.3).

There is also a widespread view of human evolution (the savannah/open plain hypotheses), which claims that we started walking upright for hunting large game on open terrain, endowing us with large brains and a descended larynx, which led to the ability to speak (see 3.4). Although none of these hypotheses have their roots in solid research – as is illustrated by the account of Bender *et al.* (2012: co-authored by Phillip V. Tobias), they have managed to entrench themselves as established truths in the minds of many, if not most, scientists in the field, obtaining the status of irrefutable paradigm.

Another problem is related to the interpretation of how fossils and genetic data provide clues on the origin of language. Most of these data are usually interpreted as indicative of language, whereas I argue that such structures (like the presence and size of Broca's area as seen from skull endocasts) or genetic data (such as the mutations in the FOXP2 gene) may have originated in the function of other and more general abilities, such as voluntary breath control (for diving and later for singing) or singing (see 3.5).

Finally, there is the general overemphasis on the part of linguists of the importance of 'word order' for the syntax of language, whereas I shall argue that word order is not very important in spoken language, compared to prosodic clues (see 3.6).

3.1. Is grammar universal?

A major misconception, which has been the ruling paradigm in linguistics for the last 50 years, is that grammar is universal and innate (Chomsky, 1957). To Chomsky, all humans have a universal grammar, a set of rules that can generate the syntax of every human language; English & Mohawk, for example, are essentially the same languages. However, it has previously been argued (Vaneechoutte & Skoyles, 2000; Kenneally, 2007; Deutscher, 2006) that this is simply not in agreement with fact.

Although it is not my intention to re-open this discussion at length, I would like to provide some examples which clearly demonstrate that grammar is anything but universal, neither has it been particularly stable during the evolution of languages.

3.1.1. Gender variability and instability

Gender is a good example to start with. Languages like Finnish and Hungarian have only one gender, which in fact is the same as having no gender, because they do not distinguish between male and female. Some languages have up to 16 genders and have, e.g., a gender for edible vegetables or for family members (Deutscher, 2006). But gender is not even stable and an incredible (and humorous) example can be seen in the transition that occurred from the Latin female gender and name for the most characteristic female organ, i.e., the vagina, to the male gender in French (a direct descendent of Latin): le vagin. Somehow, the French not only changed the gender of the word, but in addition managed to change the female gender into the male gender. Incidentally, all of the roman languages also lost the neuter gender present in their mother language, Latin. Yet, on the other hand, they acquired the definite article absent in Latin (see below).

English, a modern hybrid language, which was the result of the invasion of the French speaking Normans (= Northern men), Vikings who had conquered 'Normandie' and had adopted the French language, even lost its male and female gender for nonhuman items, which all became neuter, except for 'ship', which is still female (!).

One last example, taken from Deutscher (2010): "Why does the German feminine sun (*die Sonne*) light up the masculine day (*der Tag*), and the masculine moon (*der Mond*) shine in the feminine night (*die Nacht*)? After all, in French, he (*le jour*) is actually illuminated by him (*le soleil*), whereas she (*la nuit*) by her (*la lune*)."

Related to the gender issue is the use of definite articles. Some languages (Russian, Latin) don't have them, some have only female and male (French), some only neuter and non-neuter (Dutch), some have neuter, male and female (German), and again, changes can occur during language evolution (no articles in Latin, yet two in its daughter languages).

3.1.2. Informational content of verbs

Languages vary considerably with regard to the information they provide within words. In English, 'walked' expresses the past tense, but does not supply information about the person. In Dutch, the number of persons (one vs. more) is indicated, and Latin and Arabic also indicate the first, second or third person. Chinese verbs provide no indication of tense nor person. In French, there are special tenses, like *subjonctif* and *passé simple*, with rather complicated rules on when to apply them.

3.1.3. Plural variability

Hawaiian does not have different forms for singular and plural, and also in French both often sound the same (jour, jours) and the difference needs to be indicated by the definite articles 'le' and 'les'. In English, you clearly hear the difference since the final plural 's' is pronounced. German has many plural forms. A servian language distinguishes between the plural of two (*hródaj* = two castles) and the plural of more (*hródy* = many castles).

3.1.4. Absolute positioning vs. relative positioning

This example seems hardly credible. Some languages use(d) absolute positioning when indicating directions. Instead of saying: "your left hand" (relative position of the hand with regard to the owner of the hand), they state "your western hand" (the absolute geographical position) from which it follows that, when you turn around, the same hand will be addressed as "your eastern/northern/southern hand," depending on the direction you are facing (Deutscher, 2010). This means that these people, in order to produce comprehensible speech

and to understand what is meant by the speaker, have to be aware of the absolute geographic position (N, S, E, W) all of the time, also when in the dark or inside. Astonishing, and for the sake of this argument, not at all very universal.

3.1.5. Recursive abilities: Subordination

Chomsky changed terminology and ideas continuously in response to experiments which showed that his theoretical framework was wrong time after time (Kenneally, 2007). Having started with major claims about how language was programmed in our brains, Chomsky put forward subordination as the last resort for innate grammar enthusiasts. Subordination was considered the real hallmark of our language abilities (Hauser *et al.*, 2002).

In fact, in spoken language, we do not make much use of subordination. Because of recursive ability, we could say: "I told Johnny that he should go home to avoid making his wife angry", whereby we use two subordinated sentences or recursion. In reality, we are saying: "I told Johnny. Johnny, I said, go home. Your wife will get angry."

It is important to note that even musical structure is hierarchical, closer to language than previously thought. Thus far, it has only been agreed that music has local syntactic dependencies, i.e., from one note/chord to the next, similar to the AB-AB-AB grammar that also cotton top tamarins can learn. However, the claim of linguists that language is unique because of its recursive hierarchical structure can now be ruled out for the first time with the study by Koelsch *et al.* (2013), showing that our musical faculty uses hierarchical structure as well.

Most languages can subordinate to some degree, but some (like many Australian aboriginal languages and some languages from South America) lack it completely. Deutscher (2006) remarks that ancient languages, such as Hittite, Akkadian and Hebrew are remarkably repetitive and depend on '…and …and…and' concatenated structures of their sentences. Deutscher argues that lack of subordination coincides with simplicity of society. The juxtapositional way of saying things leaves open more room for ambiguity, whereas in complex societies, with complicated rules and relationships, language needs to express accurately the situation at hand. Another possible explanation for the more recent advent of recursion may be that it coincides with the origin of literacy. If it is true that subordination only arose with complex society (10 000-13 000 years ago) or with literacy (only 5000 years ago), subordination is a recently developed ability, and there is little that makes it a hallmark of linguistic capacity or of the human mind.

It is important to note that even musical structure is hierarchical, closer to language than previously thought. Thus far, it has only been agreed that music has local syntactic dependencies, i.e., from one note/chord to the next, similar to the AB-AB-AB grammar that also cotton top tamarins can learn. However, the claim of linguists that language is unique because of its recursive hierarchical structure can now be ruled out for the first time with the study by Koelsch *et al.* (2013), showing that our musical faculty uses hierarchical structure as well.

3.1.6. Phonological variability

Although Chomsky did not claim that phonology is universal, phonology is an important part of the conventions that make up grammar, and shows even more variability. Phonology is also a random convention: different languages use very different words to mean the same thing. The fact that even the phonetically same word may mean many different things, adds to the randomness of phonology.

This may be the case through coincidence (e.g., mail vs. male) or as the consequence of different intonation or pitch contour, as in tonal languages, and atonal languages as well (but also see 3.6).
In addition, it has been observed, though with no really good explanation, that some languages use more phonemes than others. When a total of 317 languages were sampled, together a total of 757 phonemes could be produced by their speakers, yet the speakers of each language use only a limited subset of these possibilities. Rotokas from Papua New Guinea has only six distinct consonants (p, t, k, b, d, g), whereas !Kung from Botswana has 47 non-click consonants and another 78 click consonants. With regard to vowels, many Australian languages have only three, Rotokas has five, and English has 12 vowels and 8 diphthongs. The overall number of sounds in Rotokas is therefore 11 (equal to Polynesian Mura: Vaneechoutte & Skoyles, 1998, whereas it amounts to more than 140 in !Kung: Vaneechoutte & Skoyles, 1998; Deutscher, 2010).
The absurdity of Chomskyan claims have been summarized as follows:
"When you hear speech, [Chomskyans claim that] the syntactic module extracts syntax information from the sound wave, the intonation module analyses the pitch variation. After each module has done its job, the information is put together again as language. The grammar part of your brain somehow extracts the grammatical information from the sound waves, but ignores any other information in those waves which might help interpret the sounds. Workings of the language organ are separate from other parts of the brain: separate from context, and with gesture unimportant." (Kenneally, 2007).
Kenneally (2007) continues:
"It is hardly credible that such an obviously counterintuitive hypothesis has ruled the minds of so many. The average person would probably consider context crucial to understanding language. He would count intonation as important, and he would be unlikely to completely separate structure from meaning."
Of course, these clues are important for acquiring language, as has been shown by numerous studies (see 4.5).
The above examples (but see also 3.6 for word order) can be summarized by concluding that whatever variation in grammar can be invented has been invented, and that nothing about grammar is universal. Furthermore, one might add, not only is there nothing universal about grammar but that grammar itself can be considered a random, culturally embedded convention. Still, apparently, children can easily learn whichever random syntax convention and whichever phonological subset of whatever language they happen to be exposed to (see 4.5).

3.2. Is grammar innate?

Chomskyans state that humans possess a language organ which is unique compared to all other animals. Their theory suggests that linguistic ability is hardwired in the human brain and manifests itself without being taught. It is innate.
First, the notion of innate grammar is contradictory to the existence of numerous unrelated grammars with random rules (see 3.1). Moreover, given the instability of grammar during the evolution of languages (see 3.1 and 3.6), it is difficult to comprehend how such a changing grammar could be innate. This is exemplified by the case of irregular verbs. The past tense of the verb 'to make' was regular some 1000 years ago (maked) but has become irregular (made) since then. That such irregularity cannot be innate is clear from the observation that children have to memorize these exceptions, after initial mistakes resulting from their (innate!) ability to generalise.

Second, any evolutionary biologist (but see below) who is confronted for the first time with the notion of 'innate grammar' is perplexed and dismisses it out of hand, assuming that it is a creationist idea. Hence, the fact that such a notion has been the ruling paradigm among linguists for half a century, and still is taken for granted by many today, is even more perplexing (the same is true of the stubbornness of the savannah hypotheses regarding human evolution, see 3.4). Innateness of a complex characteristic such as grammar, simply does not make sense. As Darwin (1871) noted: "Humans don't speak unless they are taught to do so." and "(Language) is certainly not a true instinct, for every language has to be learnt."
However, most perplexing of all is that the view of the innateness of grammar has been defended by some strict neodarwinists (see 3.3).

3.3. Is language advantageous and has it, as such, been naturally selected?

To defend the idea that language was naturally selected because of the obvious advantages it brought, one must suppose that language is genetically encoded, i.e., is an instinct and, as such, can be naturally selected. This is a position held by some linguists (Pinker & Bloom, 1990) who should be accredited with reopening the origin of language debate, after it had been banned from the scientific agenda by linguists for more than a century (Kenneally, 2007). Lieberman (2012) also defends the same position: "Brains and body co-evolved to make human speech possible." Evolutionary biologists Smith & Szathmáry (1995) also adhere to a genetic basis for grammar, because a language or grammar gene can be naturally selected for, so that the origin of language can be explained within the strict neodarwinist corset of gene mutation and the selection of advantageous genes.
However, Darwin (1871) would probably have disapproved of the title of the book, *The Language Instinct* (Pinker, 1994), claiming a Darwinian approach to the problem of the nature and the origin of language. He stated: "(Language) is certainly not a true instinct, for every language has to be learnt. It differs, however, widely from all ordinary arts, for man has an instinctive tendency to speak, as we see in the babble of our young children; whilst no child has an instinctive tendency to brew, bake, or write. ... The sounds uttered by birds offer in several respects the nearest analogy to language, for all the members of the same species utter the same instinctive cries expressive of their emotions; and all the kinds that sing, exert their power instinctively; but the actual song, and even the call-notes, are learnt from their parents or foster-parents".
I maintain that language emerged from a series of preadaptations, which had been naturally selected for other reasons, independent of each other (see 6), and moreover, consider the advantage(s) of language to be too limited to have forced such a series of coincidental evolutionary events. Moreover, any possible advantages probably only emerge when language is fully developed.
I do not consider articulate language to be necessary for successful reproduction, and therefore it brings limited advantage to creatures in the wild. All other animals do well without it. Without articulate language, chimps, our closest relatives, use (like several other species) and produce tools, have higher mental abilities, e.g., self recognition in mirror and theory of mind (Beran *et al.*, 2013), and are probably even better hunters than we are. In fact, they can hunt in a manner which is not only well-organized and planned in advance (and completely silent!), but moreover takes place in a three-dimensional space, taking care not to let the monkeys they are chasing escape through the canopy (http://www.youtube.com/watch?v=A1WBs74W4ik).

Consequently, I suggest that the idea that language was so advantageous that it could provide its own selective force is probably incorrect. Articulate language is not required to carry out many of the higher order mental abilities, or for improved hunting abilities.
Finally, it is generally agreed that its origins are recent (see 1.1), and as such, any eventual advantages of language could not have played a role in the natural selection of the many different capabilities that are needed for articulate language.

3.4. Do we walk upright and have large brains and speak because of running and hunting in the savannah?

The discussion as to whether adaptation to a life of chasing big game on wide open plains can explain why we run on hind legs, are fully upright, naked, sweat, have large brains and can speak, will be addressed in more detail elsewhere in this issue (see also Morgan, 1972, 1997; Vaneechoutte *et al.*, 2011a, 2012). Still, I would briefly like to touch upon it because the idea that our characteristics developed for running on open plains contradicts our thesis that language became possible because of some preadaptations which developed during the last 3 million years of our evolution (predominantly by our ancestor *Homo erectus*) in response to a semi-aquatic life, characterized by shallow water diving and seafood gathering.
The savannah hypotheses, despite being refuted by Phillip V. Tobias, the direct heir of Raymond Dart who framed the hypothesis, are still regarded by most (well-informed as well as ill-informed) as the most plausible approach to explaining the fact that humans walk upright on hind legs. Bender *et al.* (2012) convincingly illustrated how the 'running/hunting over large distances' hypothesis and the savannah hypothesis, are rooted in misconceptions, dating back at least to Lamarck. Unfortunately, when a misconception is repeated many times, it becomes established knowledge, difficult to eradicate, which is also the case with the Chomskyan view of language.
First, australopithecines, considered hominin ancestors, were already bipedal before the cooling of the climate which begat the savannah, which has in fact been known for almost two decades: "The savannah 'hypothesis' of human origins, in which the cooling climate begat the savannah and the savannah begat humanity, is now discredited." (Wood, 1996). Moreover, even older fossils, such as Orrorin, possibly closely related to the latest common ancestor of humans and chimps, where already largely bipedal (Pickford *et al.*, 2002).
Besides the problem of timing related to the savannah hypotheses, walking upright is a very inefficient way of moving around, since most quadrupedals easily outrun us, over short as well as long distances. Walking upright, nakedness and sweating have also been explained as adaptations to cooling in the hot savannah (Wheeler, 1991). Taking into consideration that night time on dry open land is freezing cold, and that equatorial nights last exactly as long as equatorial days, one can but wonder whether the presumed advantage of keeping cool during the day is balanced by the disadvantage of trying to keep warm at night. Also, nakedness, which is accompanied by a subcutaneous–insulating–fat layer, another uniquely human characteristic among primates, does not seem to be a coherent solution to cooling. Finally, sweating for cooling, and producing diluted urine (through multipyramidal kidneys, again absent in other primates: Williams, 2006), unlike land and desert animals who produce concentrated urine and do not sweat (except, to a much more limited degree, camels and horses), is counterintuitive as a strategy to save water, much needed in dry and hot conditions.
This special issue presents several arguments for a semi-aquatic past. Some proponents of the semi-aquatic hypothesis (Niemitz, 2010; Kuliukas, 2011) consider adaptations to wading a sufficient explanation for upright walking. This may be so, but there is much more that needs

to be explained about human morphology and physiology than upright walking alone. For example, wading does not explain other unique human characteristics such as voluntary breath control and nakedness, whereas shallow water diving may.
Because adaptation to dry, open land seems an unlikely explanation for upright walking, upright walking itself is an unlikely explanation for several uniquely human characteristics, such as a descended larynx and voluntary breath control, which are considered necessary (pre)adaptations for the development of our musicality and linguistic abilities. For an opposite viewpoint, discussing how the development of our musical abilities were linked to our presumed running adaptations, see Wurz (2009). As stated before (Vaneechoutte *et al.*, 2011b)(see also 5), most of the preadaptations needed for song and speech, including vocal learning (absent in all terrestrial mammals, except man and the elephant), can be most straightforwardly explained by adopting the view that our past was much wetter than is generally assumed.

3.5. How indicative are fossil and genetic data as markers for the faculty of language?

Many researchers have linked fossil findings, as well as data on the presence and timing of the first occurrence of genetic mutations, to the origin of speech, leaving other possible explanations unexplored, as exemplified by the following statement: "The only apparent selective advantage of the human supralaryngeal vocal tract is that it enhances the robustness of speech communication." (Lieberman, 2012). However, one should be careful with this kind of 'for language' interpretations because the skeletal remains and/ or genetic data could be indicative, not, or not only, for articulate language abilities, but also, or predominantly, for, e.g., voluntary breathing abilities in general (evolved for running or, more likely, for diving and/or singing).
Indeed, there are several problems with the interpretation of fossil data, as summarized by Fitch (2000) and extensively addressed by Wurz (2009).
Another unique characteristic of humans that can be read from fossils, not addressed by Fitch (2000), is not related to the production of vocal signals but to their reception. Apparently, we are endowed with a heightened sensitivity to the midrange frequencies' tones, i.e., 2-4 KHz (Martínez *et al.*, 2009) or 2-5 KHz (Despopoulos & Silbernagl, 2003), compared to other primates. Martínez *et al.* (2009) conclude, through the study of the skeletal anatomy of the middle and outer ear of middle Pleistocene hominids, from the site of the Sima de los Huesos (Spain) and from Neanderthals, that this ability was already present at least 500,000 years ago. Because this frequency is typical of human speech, they conclude that these data provide important clues to the origin and evolution of spoken language. However, this is also the frequency range of the singing voice. Intriguingly, from their data, it appears that the highest sensitivity is situated around 3000 Hz, known as the 'singers' formant, present in the spectra of trained (especially male) singers, but absent in speech. It is this increase in energy at 3000 Hz which allows singers to be heard and understood over an orchestra (http://en.wikipedia.org/wiki/ Music_theory).
Another intriguing change that took place in human evolution regards the size and orientation of the semicircular canals of the labyrinth in the inner ear during hominin evolution, as summarized by Wurz (2009) (see also Morley, 2002, 2003; Spoor *et al.*, 2007):
"The first significant developments of auditory anatomy occur with *Homo ergaster*, 1.5–1.7 million years ago, and also seem to be related to a shift to a fully upright posture. In australopithecines the vestibular structure has an apelike position, and in *Homo habilis* it has a monkey-like configuration. In *H. ergaster*, however, a modern configuration in which an almost

90-degree rotation has occurred is found. This indicates that *H. ergaster* was fully bipedal and that complex movements, such as running and jumping, could have been performed. The two vertical canals, the anterior and posterior canals, of humans are enlarged relative to the horizontal canal. The enlarged vertical canals may be linked to monitoring accurately fast vertical body movement in bipedal running on an irregular substrate."
Of course, this is another example of a possibly biased interpretation because these changes may be linked just as well to a more aquatic lifestyle of *H. ergaster*. At least they contradict the claims of the same classical palaeo-anthropologists, that australopithecines were already fully bipedal.
Genetic data as well have been interpreted as unambiguous markers for the presence of linguistic abilities. *FoxP2* has been dubbed a grammar gene (Pinker, 2001), or a language gene. It is among the 5% of the most conserved proteins among mammals, but two amino acid substitutions have fixed in the human lineage since our split from our common ancestor with the chimpanzee. The date of the emergence of these genetic changes was originally estimated to be at around 200,000 years, on the basis of only extant human diversity data (Enard *et al.*, 2002), i.e., at around the earliest possible date for modern language to have originated. However, Krause *et al.* (2007) showed that the Neandertals carried a FOXP2 protein that was identical to that of present-day humans in the only two positions that differ between the human and the chimpanzee. The most plausible explanation establishes that these changes were already present in the common ancestor of modern humans and Neandertals, i.e., before 300,000-400,000 years ago. This older estimate seems to confirm that FOXP2 is not language specific.
More importantly, animal studies indicate a much broader function and more conserved roles of this gene in patterning and plasticity of neural circuits, including those involved in integrating incoming sensory information and outgoing motor behaviors. It has been linked to motor skills in mice and to vocal production learning in songbirds (Fisher & Ridley, 2013), and the rapid evolution of this gene has been observed in bats as well (Jones *et al.*, 2013). Since language is about vocal dexterity (i.e., motoric activity), it can be expected that mutations in this gene will have consequences for language, but these consequences are pleiomorphic and will include deficits other than specifically linguistic ones. Fisher & Ridley (2013) conclude, as we did before (Vaneechoutte & Skoyles, 1998), that it is unlikely that FOXP2 triggered the appearance of spoken language in a nonspeaking ancestor. Also, because singing and dancing may be influenced by this kind of gene, it can be concluded that this gene is linked with vocal dexterity in general, but not specifically with language.

3.6. How important is word order for syntax?

Many linguists have been, and still are, obsessed with the importance of word order for syntax and they consider word order to be the major grammatical characteristic and clue by which people structure their sentences. This conviction can best be explained by assuming that they have studied written language, which lacks many of the syntactic and semantic markers that are continuously used in spoken language. Even Deutscher (2006), otherwise well-informed, states (p. 214): "Me-Tarzan protolanguage relied solely on a single strategy: the ordering of its words."
First, as I have illustrated above for other characteristics that there is nothing very universal about grammar (see 3.1), there seems to be nothing very universal about word order either.

All six possibilities of word order for Subject (S), Object (O) and Verb (V) are used, with a predominance for SOV and SVO. In some languages, almost whichever word order is possible (Deutscher, 2006).

Word order	Example	Language
SVO	Cows eat grass	English, Chinese, Swahili
SOV	Cows grass eat	Turkish, Hindu, Japanese
VSO	Eat cows grass	Welsh, classic Arabic, Samoan
VOS	Eat grass cows	Malagasy (Madagascar), Tzotzil (Mayan)
OSV	Grass cows eat	Kabardian (Northern Caucasus)
OVS	Grass eat cows	Hixkaryana (Brazil)
Whichever		Finnish/Hungarian - Greek

Some languages, like German, use SVO for main sentences, but SOV for some subordinate sentences. In fact, English speakers used to change word order as well, in subordinate sentences, but switched (after hybridizing with French) to one word order only (Deutscher, 2006):

Medieval English: The hye god, whan he hadde Adam maked (Canterbury Tales, 1390).
Modern English: The high god, when he had made Adam.

Turkish and Finnish speak in reverse order. Or is it the Indo-European languages which speak in reverse?

Most interesting for the case I want to defend here is that sentences with exactly the same word order can have a very different meaning, depending on the intonation (the overall pitch of the sentence), the pitch contour (rising, lowering pitch) and the rhythm of the sentence, together known as prosody (the melody of speech).

For example, the following four words: 'he', 'knows', 'she' and 'knows', in exactly the same order, can express very different meanings in spoken language, depending on prosody. Just some (out of many) examples below:

Written language	Spoken language*
He knows and she knows.	He knows. PAUSE. She knows.
They sure both know!	He knows! She knows!
He knows that she knows.	He knows she knows.
Does he know that she knows?	He knows? she knows.
He knows, but does she know?	He knows. She? knows?

*Using written notation, it is impossible to fully represent the prosodic subtleties of spoken language. Which is, of course, an important reason why word order is more important in written language. And, on the other hand, the power of prosody results in the limited importance of word order in spoken language. A simple sentence, such as "You go to school" can be pronounced (intonated) in at least 20 different manners, with, as a result, 20 different meanings (examples not presented).

In tonal languages (like Mandarin), word intonation also changes the meaning of a single phonetic word. However, this happens quite often in an 'atonal' language, such as English, as well. Take the word 'great', which can have at least four different meanings depending on (context, facial expression and) prosody. The first two utterances are exactly each others opposite:

Greát! (rising pitch) = Congratulations! How magnificent that you achieved that.
Greàt! (lowering pitch) = Damn! Now you've ruined it.
Gréat = answer to the question "How are you?"
Great = answer to whether you want something small or great (large).

'Hungry' can mean the opposite of 'Hungry'
Hungry? = Do you want some food? I'll give you some.
HUNGRY!! = I am very hungry! I'd very much appreciate you bringing me some food.

These examples illustrate how in language pitch is modulated in specific ways to convey different communicative meanings, such as: word stress, focus and sentence type (Xu *et al.*, 2005).
I will advocate below, increasingly supported by experimental data (see 4.5), that it are such environmental clues, such as prosody (intonation, pitch, pitch contour, loudness, loudness changes, rhythm), facial expression, eye gaze, gesture and body language, that can be used by children to learn the syntax of any language.

4. The importance of music and gesture for the origin of language

4.1. Song predates language

Music, like language, is an acoustically based form of communication with a set of rules (syntax) for combining a limited number of sounds in an infinite number of ways (Lehrdahl & Jackendorf, 1993; Tramo, 2001; Patel, 2003; Berwick *et al.*, 2011; Sammler *et al.*, 2013). Language is generally considered a very recent phenomenon (no older than 200,000 years), whereas vocal musicality has originated independently in several unrelated warm blooded taxa (in 3 orders of birds, in cetaceans, pinnipeds, manatees, bats and humans).
Imaging studies of trained singers indicate that singing involves the specialized contribution of auditory cortical regions, along with somatosensory and motor-related structures, suggesting that singing makes particular demands on auditory-vocal integration mechanisms related to the high level of pitch accuracy required for singing in tune (Zatorre & Baum, 2012). In music, encoding takes place as part of the entire pitch pattern. It can be hypothesized that our pitch differentiation abilities, that have been developed through our musicality (whereby a semitone may be of importance), have been partially re-used for conveying meaning in language, where pitch differences are one octave, and as such are more salient than in music. Our musical pitch differentiation abilities could easily be - partially - used in language to convey, and/or add, meaning of phonemes, to recognize phrases, to develop and use syntax, and to recognize sentence structure. Below are some examples, out of a vast array of experimental data, which confirm this assumption.
The existence of a rich diversity of phonemes and intonations, i.e., a large palette of sounds and tones, is strongly indicative of an origin of vocal dexterity other than language because every language uses only a very limited subset of the vocal variation that can be produced by our species (see 3.1). Fitch (2000a) agrees that our production and perception capacities go beyond what would be necessary for a very high level of vocal communication, although, at the same time, he was reluctant to recognise the role of music in language origins (Fitch, 2000b). This well-developed vocal dexterity can most easily be interpreted as though it served a more general (and vocally more demanding) ability. Our extremely well-developed singing ability is the most likely candidate.
Given the many analogies between bird song and human song (e.g., Darwin, 1871; Doupe & Kuhl, 1999; Berwick *et al.*, 2011; Lipkind *et al.*, 2013), it is most parsimonious to conclude that

this specific trait evolved for improved singing abilities in both humans and song birds, instead of assuming that it specifically developed in humans for language, out of the blue.

4.2. Song and dance play pivotal roles in human societies

For several authors, the role of music in human life and evolution is not clear. Pinker (1994) considers music as 'auditory cheesecake' and Fitch (2006) questioned the utility of human music: "For human music, an activity whose current utility is quite obscure, …" They overlook that we live in societies that are only remotely reminiscent of our small hunter-gatherer bands, only 10,000 years ago. They also overlook the importance of music in our daily lives. The music business is an important industry, festivals with tens of thousands of people are among the most important mass events, there are thousands of radio stations broadcasting 24 hours per day, the deepest emotions are brought to the fore by music, singing contests are still very popular, new pair formation is still frequently initiated by song and dance, male musicians are highly attractive to females (see below). To name but a few examples.

In original societies, singing, dancing and trancing played (and still play in extant tribes) pivotal roles in pair bonding, community bonding, territorial defence, transition rites, as my co-author and I have asserted previously (Vaneechoutte & Skoyles, 1998; see also Merker, 1999).

It is also easy to see how musicality, unlike language (see 3.3), could have been selected for, also by sexual selection. Indeed, the strongest pair bonding species are those where both genders sing. All primates that sing are monogamous and tropical song birds, where both genders engage in duets, are the most monogamous among song birds (Vaneechoutte & Skoyles, 1998).

The problem of using sexual selection to explain the origin of language is not posed by sexual selection for males gifted with musical abilities. Sexual selection for song, through a female preference for excellent male singers, whereby better singing capacities improve a male's chances of reproduction, has occurred many different times in unrelated taxa. Sexual attraction to male singers only requires some general emotional influence on female minds, without the need for the simultaneous uncovering of meaning, conveyed by randomly convened sounds.

Importantly, song can develop in both genders, as can be seen in some primates (most prominently in siamangs), tropical song birds (and probably some bats and cetaceans), and is used as a bonding mechanism. Merker (1999) has hypothesized that human song originated from synchronous chorusing by males to attract females and, as such, played a role in our divergence from the chimp. Female choice acted both between and within groups of cooperatively chorusing males.

4.3. Musicality improves the general mental and motor capacities of individuals

Song and dance are not only important as social tools; they not only improve our ability to distinguish pitch in music and language (see 4.4), but music training, for example, also improves our mathematical performance and increases fine tuning of motoric activity (Hannon & Trainor, 2007). This is again indicative of the more ancient and profound influence of music on our minds.

Cross (1999) hypothesized that early musical abilities could act to reinforce the fundamental integrative processes of learning and cognition and that proto-music promoted our cross-domain, associative capacity, regarded as the most innovative characteristic of the human brain (see also Bidelman *et al.*, 2009).

4.4. Song and language use common resources

4.4.1. Similarities between music and language

Music and language syntax show intriguing similarities. Musical structure is complex, consisting of a small set of elements that combine to form hierarchical levels of pitch and temporal structure, according to grammatical rules (Hannon & Trainor, 2007).

Individuals with significant impairment in music (amusics) are impaired in linguistic abilities too. Although this has been questioned (Peretz & Coltheart, 2003), other studies indicate that amusical individuals do have difficulties discriminating pitch glides that mimic the prosody of speech (Patel *et al.*, 2005) and that, vice versa, individuals with language-specific disorders have impaired perception and production of musical rhythms (Alcock *et al.*, 2000).

4.4.2. Shared syntactic resources for music and language in the brain

Darwin (1871) already pointed out compelling parallels between human language and birdsong (see also Doupe & Kuhl, 1999; Patel, 2003; Koelsch *et al.*, 2005; Berwick *et al.*, 2011). Studies of syntax in language and music based on neuroimaging indicate overlap, whereas some neuropsychological approaches indicate dissociation. Patel (2003) and Patel *et al.* (2005) suggested that both domains share syntactic processes, located in overlapping frontal brain areas, but with applications to different, domain specific, syntactic representations in posterior brain regions.

Although this view has been questioned (Fedorenko *et al.*, 2011; Rogalsky *et al.*, 2011), it has been largely confirmed by neuro-imaging research by Koelsch *et al.* (2005) and Steinbeis & Koelsch (2007), and more recently by Sammler *et al.* (2013). The left planum temporale, the pride of musicians with perfect pitch, is also involved in language processing (Tramo, 2001).

Musical structure creates expectations which, when not fulfilled, are experienced as violations. This implies that musical structure creates predictability, a feature also of importance in language for the listener to be able to keep pace with the production of sounds and meaning by the speaker. Indeed, violations of linguistic and musical syntax activate similar networks in the temporal lobes of the brain (Maess *et al.*, 2001; Koelsch & Siebel, 2005; Koelsch *et al.*, 2005; Sammler *et al.* 2013).

The fact that music is processed in Broca's area (Maess *et al.*, 2001) also sheds more light on claims that an increase of the Broca's area in fossils is a direct indication of improved linguistic abilities (see 3.5).

Finally, it has been demonstrated for the first time that humans, when listening to music, apply cognitive processes that are capable of dealing with long-distance dependencies, resulting from hierarchically organized syntactic structures, and that a brain mechanism fundamental for syntactic processing is engaged during the perception of music. Thus, processing of hierarchical structure with nested nonlocal dependencies is not just a key component of human language, but a multidomain capacity of human cognition (Koelsch *et al.*, 2013), probably evolved for the purpose of song.

4.4.3. Pitch processing in music and language

Training in one domain results in the refinement of brainstem enhancement in the other domain, to the extent that musicians show better encoding of linguistic tone while tone-language speakers show enhancement of musical tones (Bidelman *et al.*, 2009; Wong *et al.*, 2007). Also Schön *et al.* (2004) showed that music training facilitates pitch processing in both music and language.

Liu *et al.* (2010) also showed that pitch processing is not domain specific, but is shared by language and music. Patients with congenital amusia, a neuro-developmental disorder of musical perception, also had impaired speech intonation processing, and showed impaired discrimination, identification and imitation of statements/questions that are characterized by pitch direction differences in the final word. This intonation-processing deficit in amusia was largely associated with a pitch direction discrimination deficit. Intonation perception deficits in amusics are readily observed when testing for musical abilities, because pitch differences are subtle. When testing for language, these deficits may be overlooked because i) pitch differences are usually larger in the test sentences used, and/or ii) the amusics may rely on additional clues (including lexical, learned syntactical) to deduce the meaning of the sentence/word. This may explain the discrepant conclusions, reached by neuroimaging and neuropsychology, as outlined by Patel *et al.* (2005).

In addition, there are numerous reports on the application of Music Intonation Therapy (Albert *et al.*, 1973) to treat language disorders, as is exemplified by the quotes below:

"In order to develop a useful communication system, a 3-year-old, non-verbal autistic boy was treated for 1 year with a Simultaneous Communication Method involving signed and verbal language. As this procedure proved not useful in this case, an adaptation of Melodic Intonation Therapy (signing plus an intoned rather than spoken verbal stimulus) was tried. With this experimental language treatment, the patient produced trained, imitative, and finally, spontaneous intoned verbalizations which generalized to a variety of situations." (Miller & Toca, 1979).

"We examined mechanisms of recovery from aphasia in seven nonfluent aphasic patients, who were successfully treated with Melodic Intonation Therapy (MIT) after a lengthy absence of spontaneous recovery." (Belin *et al.*, 1996).

"In patients with brain lesion, a preverbal, emotionally focussed tonal language almost invariably is capable of reaching the still healthy sections of the person. Hence, it is possible for music therapy to both establish contact with the seemingly non-responsive patient and re-stimulate the person's fundamental communication competencies and experience at the emotional, social and cognitive levels." (Jochims, 1994).

Together, these findings clearly indicate that pitch processing in language and music involves shared mechanisms, leading to the suggestion that our linguistic syntactic ability relies on our hypothesized musical past, whereby our ability to recognize subtle differences in intonation and pitch in melodies is applied to recognize structure (syntax) in language (see 4.5).

4.5. Children acquire whatever random grammar by relying on its prosodic properties

4.5.1. The importance of melody contour (in both music and language) for memorizing auditory stimuli

Numerous studies have shown that human fetuses are able to memorize auditory stimuli from the external world by the last trimester of pregnancy, with a particular sensitivity of newborns to melody contour in both music and language (DeCasper *et al.*, 1986; Sansavini *et al.*, 1997; Granier-Deferre *et al.*, 1998; Kisilevsky *et al.*, 2004). Their perceptual preference for the surrounding language (Mehler *et al.*, 1988; Moon *et al.*, 1993; Mehler & Dupoux, 1994) and their ability to distinguish between prosodically different languages (Mehler & Christophe, 1995; Nazzi *et al.*, 1998; Ramus *et al.*, 2000; Friederici *et al.*, 2007) and pitch changes (Carral *et al.*, 2005) are based on prosodic information, primarily melody (Mampe *et al.*, 2009). Newborns use adult-like processing of pitch intervals to appreciate musical melodies and emotional and linguistic prosody (Stefanics *et al.*, 2009).

4.5.2. Infants prefer consonance over dissonance

Although many species differentiate between consonance and dissonance, the aesthetical preference for consonance over dissonance seems to be uniquely human among primates (Trainor *et al.*, 2002; Masataka, 2006). Nonhuman primates seem to dislike music in general, opting for silence over music (McDermott & Hauser, 2006). By 4 months of age, babies prefer consonant musical intervals (major and minor thirds) to dissonant intervals (minor seconds): Zentner & Kagan (1998). This ability may be re-used to learn the prosodic features of any language and to develop expectations and consequently experience violations, when non-grammatical prosodic features are presented. Children, like most animals, use context dependent clues, such as eye movement, facial expression, body language and (manual) gestures, but, on top of these, can add strongly refined interpretation of pitch, pitch contour, intonation and rhythm, i.e., rely on the melody of speech, probably as the major clue.

This is best exemplified and corroborated by the use of motherese when addressing infants. The power of prosody in motherese to convey meaning to infant listeners has been long recognized (e.g., Fernald, 1989). Kuhl (2004) reported on recordings of women speaking English, Russian or Swedish while they spoke to their young infants. Acoustic analyses showed that the vowel sounds (the /i/ in 'see', the /a/ in 'saw' and the /u/ in 'Sue') in infant-directed speech were more clearly articulated. Women from all three countries exaggerated the acoustic components of vowels ('stretching' the formant frequencies). This acoustic stretching makes the vowels contained in motherese more distinct. Infants might benefit from the exaggeration of the sounds in motherese, because the clarity with which individual mothers spoke, was related to her infant's skill in distinguishing the phonetic units of speech. Mothers who stretched the vowels to a greater degree had infants who were better able to hear the subtle distinctions in speech.

4.5.3. Children rely predominantly on the prosodic clues of a language to extract the syntactic rules of any language

The importance of prosody for grasping the syntax of the language to which a child happens to be exposed, has been formulated as the 'prosodic bootstrapping hypothesis'. Soderstrom *et al.* (2003) explored infants' use of prosodic clues coincident with phrases in processing fluent speech. After familiarization with two versions of the same word sequence, both 6- and 9-month-olds showed a preference for a passage containing the sequence as a noun phrase, over a passage with the same sequence as a syntactic non-unit. However, this result was found only in the group exposed to a stronger prosodic difference between the syntactic and non-syntactic sequences. Six-month-olds were tested in the same way on passages containing verb phrases. In this case, both groups preferred the passage with the verb phrase, to the passage with the same word sequence as a syntactic non-unit. These results provide evidence that infants as young as 6 months old are sensitive to prosodic markers of syntactic units smaller than the clause, and, in addition, that they use this sensitivity to recognize phrasal units, both noun and verb phrases, in fluent speech.

4.5.4. Vocal learning by infants is influenced by the surrounding speech prosody

Mampe *et al.* (2009) analyzed the crying patterns of 30 French and 30 German newborns with regard to their melody and intensity contours. The French group preferentially produced cries with a rising melody contour, whereas the German group preferentially produced falling

contours. The data show an influence of the surrounding speech prosody on newborns' cry melody, possibly via vocal learning based on biological predispositions (Mampe *et al.*, 2009).
In fact, we are such perfect imitators that each village has a different dialect, which is difficult to imitate by grown ups, even by neighbours from nearby villages with related dialects (see also the 'Password hypothesis' as a possible explanation for vocal mimicking in song: Feekes, 1977; Fitch, 2006), but of which every detail is perfectly mimicked by infants, quite often posing problems later on when trying to speak the standard common language.
In summary, there is a vast and evergrowing amount of evidence that children rely heavily on prosodic clues, basically pitch and intonation, to learn the syntax of language and to guide vocal imitation.

5. How do language and music relate to a swimming/diving past?

5.1. Seafood and large brains

A semi-aquatic past seems to be the best explanation for why we have such large brains. Cunnane (2005), Cunnane *et al.* (2007) and Crawford *et al.* (2013) showed that some of the major brain nutrients, such as docosahexaenoic acid (DHA), iodine and selenium, are only abundantly present in seafood, although Langdon (2006) has questioned the importance of an aquatic diet for hominin brain evolution. Regardless, the influence of an aquatic lifestyle - whether or not in combination with a carnivorous (fish, shellfish) diet - on brain size increase becomes clear from the following summary, taken largely from Eisert *et al.* (2013): "Marine mammals are of particular interest in comparative studies of mammalian encephalization because they encompass the upper mammalian size range and because most species (especially odontocetes) have relatively large brains.
Pinnipeds generally have relatively larger brains than fissipeds, or terrestrial carnivores (Kruska, 2005). Even among fissiped carnivores, an aquatic lifestyle correlates with increased brain size compared with fully terrestrial species (Kruska, 2005). When taking brain mass at birth, expressed as a proportion of adult brain mass, as a measure of the degree of neonatal maturity, or relative precociality, the neonates of seals and cetaceans are morphologically precocial, especially in the case of the Phocidae, and would be predicted to have brains that have achieved a large proportion of adult brain mass at birth. While body mass typically increases by a factor of 5 to 25 from birth to adulthood in pinnipeds and cetaceans, brain mass increases only by a factor of 1.5 to 5 from neonate to adult (Kruska, 2005). Thus, the brain represents a much greater proportion of body mass in neonates and juvenile animals than in fully grown animals. This is best exemplified by the Weddell seal, whereby the brains of Weddell seals appear to be unusually well-developed at birth, both in terms of mass as a proportion of adult brain mass and in terms of neurologic function. Weddell seal pups not only have very large brains (almost adult size) but are also very precocial" (this information has not been quoted literally).
Needless to say, the large brain size and precocity of these marine mammals is reminiscent of newborn humans and unlike other primates or terrestrial animals, representing just one more argument, out of a long list (see this issue; Vaneechoutte *et al.*, 2011a), supporting the notion that our ancestors must have been semi-aquatic.
If large brain size is considered elementary to developing articulate language, a semi-aquatic past is the most straightforward explanation for its threefold increase in humans compared to our closest primate relatives.

5.2. Why are humans musical apes?

We then should ask, why, among all apes, primates, and terrestrial mammals, are we the most musical ones (see 5.3), and the only terrestrial mammals (except elephants) to develop the ability for vocal production learning (see 5.4)? The combination of vocal dexterity and vocal learning is present in three orders of birds, in flying mammals (bats), in aquatic mammals (some of the sirenians, pinnipeds and cetaceans), in elephants, and in man, but not in primates, our closest relatives. I argue below that humans and elephants, the odd 'terrestrial' ones among this already disjoint group of animals, are vocal learners because of a semi-aquatic past, which occurred only recently in our genus/species.

5.3. Diving, the oral cavity and vocal dexterity

Although the importance of the laryngeal descent in human language has been questioned (Duchin, 1990), its descent freed space for the tongue and consequently contributed (indirectly) to vocal dexterity (Fitch, 2006). The tongue can be moved vertically and horizontally, can close the oral cavity, and can touch every part of it, thus being the major formant in vocal production. Jeffrey Laitman thinks the larynx descended to enable quick intakes of breath (Jeffrey Laitman, pers. comm., 2013). In his view, this was for sudden runs and not for diving, because he considers the naso-oral cavity as "too leaky" to be suited for diving (Jeffrey Laitman, pers. comm., 2013). However, most animals can produce sudden runs without descended larynx, and, as for the opposite, our naso-oral cavity can be perfectly closed by several different mechanisms.

Our nostrils are muscled and some people can close their nostrils to block out water when jumping or diving feet first into water. For some, this is a response learned spontaneously in childhood and used whenever entering water (Johnny Weyand, pers. comm., 2013). In addition, many people can raise their closed lips to touch the nose, and in fact, the philtrum (again probably unique to humans) matches perfectly the nasal septum, improving closure of the nose with the upper lip (Morgan, 1997). The mouth itself can be perfectly closed by muscled, fleshy lips, the teeth form a parabolic closed row, and the globular tongue can move vertically and horizontally and secure the oral cavity, all unlike the abilities of chimps. In other words, the naso-oral cavity is perfectly designed to be fully lockable and together with the assumption that the larynx descended to enable quick and large intakes of air, this fits with swimming and diving adaptations of our ancestors (and, still partially, with ourselves), see also Vaneechoutte *et al.* (2011b). Vocal dexterity can be perfectly explained as resulting from adaptations to a semi-aquatic, shallow water diving past, also because voluntary breath control is another prerequisite for human song and speech (Vaneechoutte & Skoyles, 1998; Skoyles, 2000), a characteristic largely absent in terrestrial mammals. Humans combine both autonomous breathing and voluntary breathing, whereas full aquatics only have the capacity for voluntary breathing. See also Wurz (2009) for an explanation of vocal dexterity in the context of the running hypothesis.

5.4. Full-fledged 3-D experience of bodily movement may explain the ability for mimicking and predicting gestures, sounds and intonations: vocal production learning

Also with regard to vocal production learning (see Janik & Slater, 1997, for appropriate terminology), we may wonder why only our species among primates can mimic song (and spoken sentences).

In humans, the primary motor cortex, and a connection from it to the nucleus ambiguus, are necessary for the production of learned vocalizations, such as speech or the humming of tunes,

but not for the production of vocalizations like crying or laughing (Groswasser *et al.*, 1988). Nonhuman primates lack the direct connection between the primary motor cortex and the nucleus ambiguus. To date, there is no evidence of phonatory production learning in nonhuman primates.

Fitch (2000a) correctly noticed that we find ourselves among 'disjoint' groups, such as aquatic mammals and song birds, which are excellent vocal learners as well. Vocal learning in human infants and two species of song birds has been shown to be rather comparable. Lipkind *et al.* (2013) found a comparable stepwise acquisition of vocal combinatory capacity in songbirds and human infants. Besides, this cumbersome achievement of production may explain the well-known gap between the early well-developed abilities for perception of new pairwise vocal transitions, and the long time span before we can produce these vocal combinations.

Fitch (2000a) forgot to mention another disjoint group, besides singing birds, singing aquatic mammals and singing humans, that has the capacity for vocal learning, namely bats. These mammals make human speech look simple: In a behavior called echolocation, a bat must coordinate its nose, mouth, ears, and larynx to emit and receive calls, all the while executing flight maneuvers guided in part by these signals (Jones & Ransome, 1993; Boughman, 1998; Jones *et al.*, 2013). Moreover, elephants, though probably to a lesser extent, can also mimick sounds (Hart *et al.*, 2008; Byrne *et al.*, 2009). This adds to the disjointedness of these groups, but upon closer inspection may also hold the answer to this enigma.

What do all these groups have in common? What follows is purely speculative but, once more, with regard to vocal learning, an aquatic past is the most plausible explanation for the uniqueness of humans among primates and among most terrestrial mammals.

Besides being warm blooded and, of course, gifted with vocal dexterity in the first place (also – for our species and aquatic mammals – possibly attributable to aquatic adaptations (see 5.3)), it can be noted that aquatic mammals and flying mammals and birds share a full-fledged 3-D life, whereby instantaneous mimicking of body movements in 3 dimensions (of species' mates or of prey to pursue or of predators to escape from) is an intrinsic part of their lives. The speed, the almost instantaneousness, with which these creatures can mimick the movements of others is amazing. And although other (cold blooded) aquatic vertebrates (e.g., fish) and invertebrates (e.g., squid) are equally amazing with regard to their immediate mimicking of the movements of other individuals (and – even more amazing – the mimicking of complex skin colour patterns, by squids), they never developed communication by sound, probably because they remained aquatic.

Briefly then, mimicking movement in 3 directions and doing so almost instantaneously, is something that all these species can do, birds and bats in the air, cetaceans and pinnipeds in the water. Not only do they need very flexible and fast responding musculatures and nervous systems, but they must also be able to predict the movements of others, to enable simultaneous response. Again, the ability to predict is essential for music and language (see, for example, expectation violation studies: Sammler *et al.*, 2013).

I speculate, not based on literature, that our vocal learning is best explained by an aquatic past, which provided us with improved capabilities for mimicking bodily movements (later used in dance, and in linguistic gesturing), which in turn enabled the mimicking of sounds and intonations. If this speculation could be corroborated by experimental data, the presence of vocal production learning in humans may become one of the strongest arguments for an aquatic past in our species. Because of its absence in terrestrial animals, it seems unlikely that vocal learning can be explained as resulting from adaptations to upright walking and running (see Wurz, 2009). In fact, the presence of vocal learning in elephants may unexpectedly support

the aquatic theory. It may seem odd, but Gaeth *et al.* (1999) convincingly argued in favour of the aquatic ancestry of elephants, and, of course, the fully aquatic Sirenians (dugongs and manatees) are their closest relatives.
Also, the characteristic human vestibulum/labyrinth may support an aquatic explanation. For fluent movement in 3 dimensions (as in water), a well-developed vestibular apparatus is of utmost importance. I have argued above (see 3.5) how our vestibular system (semicircular canals) may have changed its orientation as an adaptation to a semi-aquatic lifestyle, rather than for upright running (Wurz, 2009). A relationship between vestibular function and language development has been suggested (Magrun *et al.*, 1981). Bailey (1978) described in detail the anatomical and neurological relationships between the vestibular system and speech centers. Interestingly, hearing a rhythm evokes physical movement and the resulting vestibular stimulation also influences the auditory interpretation of the rhythm (Trainor *et al.*, 2009). Concurrently, song and dance are intrinsically linked. Song/music evokes moving and dancing, which is reflected by the fact that in some languages the notions 'song' and 'dance' are synonymous.
This close link between movement and speech, via mimicking and song, may also explain the importance of gestures in language, whereby gesture (bodily movement) is an intrinsic part of linguistic expression, although it can only explain some aspects of articulate language.

6. A plausible scenario

Deutscher (2005) stated: "Failing the discovery of a camcorder left behind by careless aliens on a previous visit, it is thus difficult to see how the first emergence of speech in hominids can even be much more than the stuff of fantasy." He thereby re-enforces the view that was already held by the Académie Française more than a century ago and which led to a ban on publishing all research on the origin of language, because it was impossible to substantiate anyway. This attitude is illustrative of the size of the difficulties encountered when trying to explain the origin of language. However, by bringing together insights on human evolution and studies on the link between language and music, we can now be more optimistic.
Deutscher's more pessimistic point of view may stem from his conviction (which he shares with most linguists) that when first developing language all our ancestors had was 'word order'. However, an ever-growing amount of experimental data on similarities between human language and bird song, on how children acquire language (think of motherese), and on how language and music processing are controlled by strongly overlapping domains in our brains, make it clear that our non-speaking ancestors in the first place had melody, rhythm and dance with which to conveyed symbolic meaning to the many sounds and tones they could produce.
I have argued that the most straightforward and parsimonious evolutionary approach to explaining our vocal dexterity, our extremely well-developed ability to handle melody and rhythm, and our ability to perfectly imitate sound and intonation (vocal learning), is to assume that we started as musical apes (Vaneechoutte & Skoyles, 1998), an idea put forward by Darwin and re-discovered independently several times since then. The parallels with song birds (and probably several cetaceans, pinnipeds and bats) have been pointed out.
Song and dance are equally intrinsically linked, supporting the idea that well-developed mimicking motoric abilities, for moving freely in 3 dimensions (in air or in water, but not in trees), developed for diving/swimming. This ability further evolved for dancing/singing, and predisposed for spoken/gestural language.

I therefore suggest that the intrinsic links between song, rhythm, dance and movement, which predisposed us to articulate language, are best understood as having evolved in an aquatic environment.

Acknowledgments

I would like to thank Daniela Sammler (Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany) for her critical reading and for providing up-to-date neurolinguistic information and Claire Stypulkowski for careful editing.

References

Albert, M.L., Sparks, R.W., & Helm, N.A. (1973). Melodic Intonation Therapy for Aphasia. *Arch. Neurol.*, 29:130–131.

Alcock, K.J., Passingham, R.E., Watkins, K., & Vargha-Khadem, F. (2000). Pitch and Timing Abilities in Inherited Speech and Language Impairment. *Brain Lang.*, 75:34–46.

Bailey, D.M. (1978). The Effects of Vestibular Stimulation on Verbalization in Chronic Schizophrenics. *Am. J. Occup. Ther.,* 33:445–450.

Belin, P., Van Eeckhout, P., Zilbovicius, M., Remy, P., Francois, C., Guillaume, S., Chain, F., Rancurel, G., & Samson, Y. (1996). Recovery from Nonfluent Aphasia after Melodic Intonation Therapy: A PET Study. *Neurol.*, 47:1504–1511.

Beran, M.J., Smith, J.D., & Perdue, B.M. (2013). Language-Trained Chimpanzees (*Pan troglodytes*) Name What They Have Seen but Look First at What They Have Not Seen. *Psychol. Sci.*, 24:660.

Berwick, R.C., Okanoya, K., Beckers, G.J.L., & Bolhuis, J.J. (2011). Songs to Syntax: The Linguistics of Birdsong. *Trends Cogn. Sci.*, 15:113–121.

Bender, R., Tobias, P.V., & Bender, N. (2012). The Savannah Hypotheses: Origin, Reception and Impact on Paleoanthropology. *Hist. Phil. Life Sci.*, 34:147–184.

Bidelman, G.M., Gandour, J.T., & Krishnan, A. (2009). Cross-Domain Effects of Music and Language Experience on the Representation of Pitch in the Human Auditory Brainstem. *J. Cogn. Neurosci*., 23:425–434.

Bohn, K.M., Smarsh, G.C., & Smotherman, M. (2013). Social Context Evokes Rapid Changes in Bat Song Syntax. *Animal Behav.*, 85:1485–1491.

Boughman, J.W. (1998). Vocal Learning by Greater Spear-Nosed Bats. *Proc. R. Soc. Lond. B*, 265:227– 233.

Byrne, R.W., Bates, L., & Moss, C.J. (2009). Elephant Cognition in Primate Perspective. *Comp. Cogn. Behav. Rev.,* 4:65–79.

Carral, V., Huotilainen, M., Ruusuvirta, T., Fellman, V., Naätänen, R., & Escera, C. (2005). A Kind of Auditory 'Primitive Intelligence' Already Present at Birth. *Eur. J. Neurosci.,* 21:3201–3204.

Chomsky. N. (1957). *Syntactic Structures*. Mouton, The Hague.

Crawford, M.A., Broadhurst, C.L., Guest, M., Nagar, A., Wang, Y., Ghebremeskel, K., & Schmidt, W.F. (2013). A Quantum Theory for the Irreplaceable Role of Docosahexaenoic Acid in Neural Cell Signalling Throughout Evolution. *Prostaglandins, Leukotrienes, Essential Fatty Acids*, 88:5–13.

Cross, I. (1999). Is Music the Most Important Thing We Ever Did? Music, Development and Evolution. In S.W. Yi (Ed.), *Music, Mind and Science.* Seoul, South Korea: Seoul Nat. Univ. Press

Cunnane, S.C. (2005). *Survival Of The Fattest: The Key to Human Brain Evolution.* 368 pp. ISBN: 978-981-256-191-6.

Cunnane, S.C., Plourde, M., Stewart, K., & Crawford, M.A. (2007). Docosahexaenoic Acid and Shore- Based Diets in Hominin Encephalization: A Rebuttal. *Am. J. Human Biol.*, 19:578–581.

Darwin, C. (1871). *The Descent of Man, and Selection in Relation to Sex*. Concise Editon and Commentary by Carl Zimmer (2007), Plume, Penguin Books. ISBN 978-0-452-28888-1.

Deacon, T.W. (1997). *The Symbolic Species: The Co-Evolution of Language and the Brain*. New York: W.W. Norton.

DeCasper, A.J., & Spence, M.J. (1986). Prenatal Maternal Speech Influences Newborns' Perception of Speech Sounds. *Infant Behav. Dev.*, 9:133–150.

Despopoulos, A., & Silbernagl, S. (2003). *Color Atlas of Physiology*. 5th Edition, p. 362. New York, Stuttgart: Thieme. ISBN 3-13-545005-8.

Deutscher, G. (2006). *The Unfolding of Language: An Evolutionary Tour of Mankind's Greatest Invention*. New York: Holt. ISBN-13: 978-0-8050-8013-4.

Deutscher, G. (2010). *Through the Language Glass: Why the World Looks Different in Other Languages.* New York: Holt. ISBN978-0-312-61049-4.

Doupe, A.J., & Kuhl, P.K. (1999). Birdsong and Human Speech: Common Themes and Mechanisms. *Annu. Rev. Neurosci.*, 22:567–631.

Duchin, L.E. (1990). The Evolution of Articulate Speech: Comparative Anatomy of the Oral Cavity in *Pan* and *Homo*. *J. Hum. Evol.*, 19:687-697.

Eisert, R., Potter, C.W., & Oftedal, O.T. (2013). Brain Size in Neonatal and Adult Weddell Seals: Costs and Consequences of Having a Large Brain. *Marine Mamm. Sci.*, 30:184-205.

Enard, W., Przeworski, M., Fisher, S.E., Lai, C.S., Wiebe, V., Kitano, T., Monaco, A.P., & Pääbo, S. (2002). Molecular Evolution of FOXP2, a Gene Involved in Speech and Language. *Nature*, 418:869–872.

Fedorenko, E., Behr, M.K., & Kanwisher, N. (2011). Functional Specificity for High-Level Linguistic Processing in the Human Brain. *Proc. Natl. Acad. Sci USA*, 108:16428–16433.

Feekes, F. (1977). Colony Specific Song in *Cacicus cela* (Icteridae, Aves): The Password Hypothesis. *Ardea*, 65:197–202.

Fernald, A. (1989). Intonation and Communicative Intent in Mothers' Speech to Infants: Is the Melody the Message? *Child Developm.*, 60:1497–1510.

Filatova, O.A., Burdin, A.M., & Hoyt, E. (2013). Is Killer Whale Dialect Evolution Random? *Behav. Proc.*, 99:34–41.

Fisher, S.E., & Ridley, M. (2013). Culture, Genes, and the Human Revolution. *Science*, 340:929–930.

Fitch, W.T. (2000a). The Evolution of Speech: a Comparative Review. *Trends Cogn. Sci.*, 4:258–267.

Fitch, W.T. (2000b). Without Breath and Without Song. Reply to Skoyles. *Trends Cogn. Sci.*, 4:405– 406.

Fitch, W.T. (2006). The Biology and Evolution of Music: A Comparative Perspective. *Cognition,* 100:173–215.

Friederici, A.D., Friedrich, M., & Christophe, A. (2007). Brain Responses in 4-Month-Old Infants Are Already Language Specific. *Curr. Biol.*, 17:1208–1211.

Gaeth, A.P., Short, R.V., & Renfree, M.B. (1999). The Developing Renal, Reproductive, and Respiratory Systems of the African Elephant Suggest an Aquatic Ancestry. *Proc. Natl. Acad. Sci. USA*, 96:5555–5558.

Geissmann, T. (2000). Gibbon Songs and Human Music from an Evolutionary Perspective. In N.L. Wallin, B. Merker and S. Brown (Eds.), *The Origins of Music (pp. 103-123).* Cambridge, MA: MIT Press. ISBN 0-262-23206-5.

Granier-Deferre, C., Bassereau, S., Jacquet, A.Y., & Lecanuet, J.P. (1998). Fetal and Neonatal Cardiac Orienting Response to Music in Quiet Sleep. *Dev. Psychobiol.*, 33:372.

Groswasser, Z., Korn, C., Groswasser-Reider, J., & Solzi, P. (1988). Mutism Associated with Buccofacial Apraxia and Bihemispheric Lesions. *Brain and Language*, 34:157–168.

Hannon, E.E., & Trainor, L.J. (2007). Music Acquisition: Effects of Enculturation and Formal Training on Development. *Trends Cogn. Sci.*, 11:466–472.

Hart, B.L., Hart, L.A., Pinter-Wollman, N. (2008). Large Brains and Cognition: Where Do Elephants Fit In? *Neurosci. Biobehav. Rev.*, 32:86–98.

Hauser, M.D., Chomsky, N., & Fitch, W.T. (2002). The Faculty of Language: What Is It, Who Has It, and How Did It Evolve? *Science,* 298:1569–1579.

Janik, V.M., Dehnhardt, G., & Todt, D. (1994). Signature Whistle Variations in a Bottlenose Dolphin, *Tursiops truncatus*. *Behav. Ecol. Sociobiol.*, 35:243–248.

Janik, V.M., & Slater, P.J.B. (1997). Vocal Learning in Mammals. *Adv. Study Behav.*, 26:59–99.

Janik, V.M., & Slater, P.J.B. (2000). The Different Roles of Social Learning in Vocal Communication. *Animal Behav.*, 60:1–11.

Jochims, S. (1994). Establishing Contact in the Early Stage of Severe Craniocerebral Trauma: Sound as the Bridge to Mute Patients. *Rehab.*, 33:8–13.

Jones, G., & Ransome, R.D. (1993). Echolocation Calls of Bats are Influenced by Maternal Effects and Change over a Lifetime. *Proc. R. Soc. London, Series B*, 252:125–128.

Jones, G., Teeling, E.C, & Rossiter, S.J. (2013). From the Ultrasonic to the Infrared: Molecular Evolution and the Sensory Biology of Bats. *Frontiers Physiol.*, 4:117.

Kenneally, C. (2007). *The First Word. The Search for the Origins of Language*. London, UK: Penguin Books. ISBN978-0-14-311374-4.

Kisilevsky, S., Hains, S.M., Jacquet, A.-Y., Granier-Deferre, C., & Lecanuet, J.P. (2004). Maturation of Fetal Responses to Music. *Dev. Sci.*, 7:550–559.

Koelsch, S., & Siebel, W.A. (2005). Towards a Neural Basis of Music Perception. *Trends Cogn. Sci.*, 9:578–584.

Koelsch, S., Gunter, T.C., Wittfoth, M., & Sammler, D. (2005). Interaction between Syntax Processing in Language and in Music: An ERP Study. *J Cogn. Neurosci.*, 17:1565–1577.

Koelsch, S., Rohrmeier, M., Torrecuso, R., & Jentschke, S. (2013). Processing of Hierarchical Syntactic Structure in Music. *Proc. Natl. Acad. Sci. USA*, 110:15443-15448.

Krause, J., Lalueza-Fox, C., Orlando, L., Enard, W., Green, R.E., Burbano, H.A., Hublin, J.-J., Hänni, C., Fortea, J., de la Rasilla, M., Bertranpetit, J., Rosas, A., Pääbo, S. (2007). The Derived FOXP2 Variant of Modern Humans Was Shared with Neandertals. *Curr. Biol.*, 17:1908–1912.

Kruska, D.C.T. (2005). On the Evolutionary Significance of Encephalization in Some Eutherian Mammals: Effects of Adaptive Radiation, Domestication, and Feralization. *Brain, Behavior Evol.*, 65:73– 108.

Kuhl, P.K. (2004). Early Language Acquisition: Cracking the Speech Code. *Nature Rev.*, 5:831–843.

Kuliukas, A.V. (2011). A Wading Component in the Origin of Hominin Bipedalism. In M. Vaneechoutte, A. Kuliukas, and M. Verhaegen (Eds.), *Was Man More Aquatic in the Past? Fifty Years After Alister Hardy - Waterside Hypotheses of Human Evolution (pp. 173–180). Oak Park, IL:* Bentham eBooks, Bentham Science Publishers. eISBN:978-1-60805-244-8, 2011.

Laitman, J.T., & Heimbuch, R.C. (1982). The Basicranium of Plio-Pleistocene Hominids as an Indicator of their Upper Respiratory Systems. *Am. J. Phys. Anthropol.*, 59:323–343.

Langdon, J.H. (2006). Has an Aquatic Diet Been Necessary for Hominin Brain Evolution and Functional Development? *Brit. J. Nutrition*, 96:7–17.

Lehrdahl , F., & Jackendorf, R. (1993). *A Generative Theory of Tonal Music*. Cambridge: MIT Press.

Lieberman, P. (2012). Vocal Tract Anatomy and the Neural Bases of Talking. *J. Phonetics*, 40: 608–622.

Lipkind, D., Marcus, G.F., Bemis, D.K, Sasahara, K., Jacoby, N., Takahasi, M., Suzuki, K., Feher, O., Ravbar, P., Okanoya, K., & Tchernichovski, O. (2013). Stepwise Acquisition of Vocal Combinatory Capacity in Songbirds and Human Infants. *Nature*, 498:104–109.

Liu, F., Patel, A.D., Foucin, A., & Stewart, L. (2010). Intonation Processing in Congenital Amusia: Discrimination, Identification and Imitation. *Brain*, 133:1682–1693.

Magrun, W.M., Ottenbacher, K., McCue, S., & Keefe, R. (1981). Effects of Vestibular Stimulation on Spontaneous Use of Verbal Language in Developmentally Delayed Children. *Am. J. Occup. Ther.,* 35:101–104.

Maess, B., Koelsch, S., Gunter, T.C., & Friederici, A.D. (2001). Musical Syntax is Processed in Broca's Area: An MEG Study. *Nat. Neurosci.*, 4:540–545.

Mampe, B., Friederici, A.D., Christophe, A., & Wermke, K. (2009) Newborns' Cry Melody Is Shaped by Their Native Language. *Curr. Biol.*, 19:1994–1997.

Martínez, I., Quam, R., Arsuaga, J.L., Lorenzo, C., Gràcia, A., Carretro, J.M., Rosa, M., & Jarabo, P. (2009). Approche Paléontologique de l'Évolution du Langage: Un État des Lieux. *L'Anthropol.*, 113:255–264.

Masataka, N. (2006). Preference for Consonance over Dissonance by Hearing Newborns of Deaf Parents and Hearing Parents. *Dev. Sci.*, 9:46–50.

McDermott, J., & Hauser, M.D. (2006). Nonhuman Primates Prefer Slow Tempos but Dislike Music Overall. *Cognition*, 105:654–668.

Mehler, J., Jusczyk, P.W., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A Precursor of Language Acquisition in Young Infants. *Cognition*, 29:143–178.

Mehler, J., & Dupoux, E. (1994). *What Infants Know*. Oxford, UK: Blackwell.

Mehler, J., & Christophe, A. (1995). Maturation and Learning of Language in the First Year of Life. In M.S. Gazzaniga (Ed.), *The Cognitive Neurosciences: A Handbook for the Field (pp. 943–954).* Cambridge, MA: MIT Press.

Merker, B. (1999). Synchronous Chorusing and Human Origins. In N.L. Wallin, B. Merker and S. Brown (Eds.), *The Origins of Music*. Cambridge, MA: MIT Press.

Miller, S.B., & Toca, J.M. (1979). Adapted Melodic Intonation Therapy: A Case Study of an Experimental Language Program for an Autistic Child. *J. Clin. Psych.*, 40:201–203.

Moon, C., Cooper, R.P., & Fifer, W.P. (1993). Two-Day-Olds Prefer their Native Language. *Infant Behav. Dev.,* 16:495–500.

Morgan, E. (1972). *The Descent of Woman*. London, UK: Souvenir Press.

Morgan, E. (1997). *The Aquatic Ape Hypothesis - The Most Credible Theory of Human Evolution*. London, UK: Souvenir Press.

Morley, I. (2002). Evolution of the Physiological and Neurological Capacities for Music. *Cambridge Archaeol. J.*, 12:195–216.

Morley, I. (2003). *The Evolutionary Origins and Archaeology of Music: An Investigation into the Prehistory of Human Musical Capacities and Behaviours*. PhD thesis, Univ. Cambridge, UK.

Munro, S., & M. Verhaegen. (2011). Pachyosteosclerosis in Archaic *Homo*: Heavy Skulls for Diving, Heavy Legs for Wading? Pp. 91–116. In M. Vaneechoutte, A. Kuliukas and M. Verhaegen (Eds.), *Was Man More Aquatic in the Past? Fifty Years After Alister Hardy. Waterside Hypotheses of Human Evolution*. Bentham eBooks (2011). eISBN: 19781608052448.

Nazzi, T., Floccia, C., & Bertoncini, J. (1998). Discrimination of Pitch Contours by Neonates. *Infant Behav. Dev.,* 21:779–784.

Niemitz, C. (2010). The Evolution of the Upright Posture and Gait - a Review and a New Synthesis. *Naturwissenschaften*, 97:241–263.

Okumura, N., Ichikawa, K., Akamatsu, T., Arai, N., Shinke, T., Hara, T., & Adulyanukosol, K. (2007). Stability of Call Sequence in Dugongs' Vocalization. *IEEE,* 2007:1–4.

Patel, A.D., Foxton, J.M., & Griffiths, T.D. (2005). Musically Tone-Deaf Individuals Have Difficulty Discriminating Intonation Contours Extracted from Speech. *Brain Cogn.*, 59:310–333.

Patel, A.P. (2003). Language, Music, Syntax and the Brain. *Nature Neurosci.*, 6: 674-681.

Peretz, I., & Coltheart, M. (2003). Modularity of Music Processing. *Nat. Neurosci.*, 6: 688–691.

Pickford, M., Senut, B., Gommery, D., & Treil, J. (2002). Bipedalism in *Orrorin tugenensis* Revealed by its Femora. *Comptes Rendus Palevol.*, 1:191–203.

Pinker, S., & Bloom, P. (1990). Natural Language and Natural Selection. *Behav. Brain Sci.*, 13:707–784.

Pinker, S. (1994). *The Language Instinct: How the Mind Creates Language*. New York: Morrow.

Pinker, S. (2001). Talk of Genetics and Vice Versa. *Nature*, 413:465–466.

Ralls, K., Fiorelli, P., & Gish S. (1985). Vocalizations and Vocal Mimicry in Captive Harbor Seals, *Phoca vitulina. Can. J. Zool.*, 63:1050–1056.

Ramus, F., Hauser, M.D., Miller, C., Morris, D., & Mehler, J. (2000). Language Discrimination by Human Newborns and by Cotton-Top Tamarin Monkeys. *Science*, 288: 349–351.

Rogalsky, C. Rong, F. Saberi, K, & Hickok, G. (2011). Functional Anatomy of Language and Music Perception: Temporal and Structural Factors Investigated Using Functional Magnetic Resonance Imaging. *J. Neurosci.*, 31:3843–3852.

Sammler, D., Koelsch, S., Ball, T., Brandt, A., Grigutsch, M., Huppertz, H.-J., Knösche, T.R., Wellmer, J., Widman, G., Elger, C.E., Friederici, A.D., & Schulze-Bonhage, A. (2013). Co-Localizing Linguistic and Musical Syntax with Intracranial EEG. *Neuroimage*, 64:134–146.

Sansavini, A., Bertoncini, J., & Giovanelli, G. (1997). Newborns Discriminate the Rhythm of Multisyllabic Stressed Words. *Dev. Psychol.*, 33:3–11.

Schön, D., Magne, C., & Besson, M. (2004). The Music of Speech: Music Training Facilitates Pitch Processing in Both Music and Language. *Psychophysiol.*, 41:341–349.

Skoyles, J. (2000). Without Breath and Without Song. *Trends Cogn. Sci.*, 4:405.

Smith, J.S., & Szathmáry, E. (1995). *The Major Transitions in Evolution*. Oxford, UK: Oxford University Press. ISBN 0-19-850294-X.

Soderstrom, M., Seidl, A., Kemler Nelson, D.G., & Jusczyk, P.W. (2003) The Prosodic Bootstrapping of Phrases: Evidence from Prelinguistic Infants. *J. Memory Language*, 49:249–267.

Sousa-Lima, R.S., Paglia, A.P., & Da Fonseca, G.A.B. (2002). Signature Information and Individual Recognition in the Isolation Calls of Amazonian Manatees, *Trichechus inunguis* (Mammalia: Sirenia). *Animal Behav.*, 63:301–310.

Spoor, F., Garland, T., Krovitz, G., Ryan, T. M., Silcox, M.T., & Walker, A. (2007). The Primate Semicircular Canal System and Locomotion. *Proc. Natl. Acad. Sci. USA*, 104:10808–10812.

Stefanics, G., Háden, G.P., Sziller, I., Balázs, L., Beke, A., & Winkler, I. (2009). Newborn Infants Process Pitch Intervals. *Clin. Neurophysiol.*, 120:304–308.

Steinbeis, N., & Koelsch, S. (2008). Shared Neural Resources between Music and Language Indicate Semantic Processing of Musical Tension-Resolution Patterns. *Cereb. Cortex*, 18:1169–1178.

Trainor, L.J., Tsang, C.D., & Cheung, V.H.W. (2002). Preference for Sensory Consonance in 2- and 4-Month-Old Infants. *Music Percept.*, 20:187–194.

Trainor, L.J., Gao, X., Lei, J., Lehtovaara, K., & Harris, L.R. (2009). The Primal Role of the Vestibular System in Determining Musical Rhythm. *Cortex*, 45:35–43.

Tramo, J.M. (2001). Music of the Hemispheres. *Science*, 291:54–56.

Vaneechoutte, M. (1993). The Memetic Basis for Religion. *Nature*, 365:290.

Vaneechoutte, M., & Skoyles, J.R. (1998). The Memetic Origin of Language: Humans as Musical Primates. *J. Memetics*, 2. Accessible at: http://users.ugent.be/~mvaneech/ORILA.FIN.html

Vaneechoutte, M. (2000). Experience, Awareness and Consciousness: Suggestions for Definitions as Offered by an Evolutionary Approach. *Found. Sci.*, 5:429–456.

Vaneechoutte, M., A. Kuliukas, & M. Verhaegen. (2011a). *Was Man More Aquatic in the Past? Fifty Years After Alister Hardy. Waterside Hypotheses of Human Evolution (p.244)*. Bentham eBooks. eISBN: 19781608052448.

Vaneechoutte, M., Munro, S., & Verhaegen. M. (2011b). Seafood, Diving, Song and Speech. In M. Vaneechoutte, A. Kuliukas, and M. Verhaegen (Eds.), *Was Man More Aquatic in the Past? Fifty Years After Alister Hardy. Waterside Hypotheses of Human Evolution (pp. 181-189).* Bentham eBooks. eISBN: 1 9781608052448.

Verhaegen, M., Munro, S., Puech, P.-F., &.Vaneechoutte, M. (2011). Early Hominoids: Orthograde Aquarboreals in Flooded Forests? In M. Vaneechoutte, A. Kuliukas, and M. Verhaegen (Eds.), *Was Man More Aquatic in the Past? Fifty Years After Alister Hardy. Waterside Hypotheses of Human Evolution (p*p. 67-81). Bentham eBooks. eISBN: 19781608052448.

Vaneechoutte, M., Munro, S., & Verhaegen, M. (2012). Reply to John Langdon's Review of the eBook: Was Man More Aquatic in the Past? Fifty Years after Alister Hardy. Waterside Hypotheses of Human Evolution. *J. Comp. Biol.*, 63: 496–503.

Wheeler, P.E. (1991). The Thermoregulatory Advantages of Hominid Bipedalism in Open Equatorial Environments: The Contribution of Increased Convective Heat Loss and Cutaneous Evaporative Cooling. *J. Hum. Evol.*, 21:107–115.

Williams, M.F. (2006). Morphological Evidence of Marine Adaptations in Human Kidneys. *Med. Hypoth.*, 66:247–257.

Wong, P., Skoe, E., Russo, N., Dees, T., & Kraus, N. (2007). Musical Experience Shapes Human Brain- Stem Encoding of Linguistic Pitch Patterns. *Nature Neurosci.*, 10: 420–422.

Wood, B.A. (1996). Apocalypse of Our Own Making. *Nature*, 379:687.

Wurz, S. (2009). Interpreting the Fossil Evidence for the Evolutionary Origins of Music. *Southern African Humanities*, 21:395–417.

Xu, Y. (2005). Speech Melody as Articulatorily Implemented Communicative Functions. *Speech Comm.*, 46:220–251.

Zatorre, R.J., & Baum, S.R. (2012). Musical Melody and Speech Intonation: Singing a Different Tune? *PloS Biol.*, 10:e1001372.

Zentner, M.R., & Kagan, J. (1998). Infants Perception of Consonance and Dissonance in Music. *Infant Behav. Dev.*, 21:483–492.