

JMM - *The Journal of Music and Meaning*, vol.2, 2004, section 2 - printout.

Original article URL:

<http://www.musicandmeaning.net/issues/showArticle.php?artID=2.2>

[Refer to this article by section numbers, not by page number]

**Jakob Christensen-Dalsgaard**

## **Music and the Origin of Speeches**

Jakob Christensen-Dalsgaard  
Center for Sound Communication,  
Institute of Biology  
University of Southern Denmark  
Campusvej 55  
DK-5230 Odense M, Denmark  
e-mail [JCD@biology.sdu.dk](mailto:JCD@biology.sdu.dk)

### **Abstract**

As far as we know, music is a uniquely human trait. Theories of the origin of music are very speculative, since the adaptive significance of music is questionable, and it is very difficult to date the origin of music. I review the biological findings on animal communication and human music and language processing and the available theories on the origin of music. I propose that the close link between music and human prosody (the non-semantic, slowly varying pitch contours and rhythms of speech), which is also corroborated by recent neurophysiological experiments could reflect a common origin, and that an adaptive value of music could be related to rhetoric, which is an effective way of controlling or manipulating the emotions of a group. The close connections between music, rhetoric and prosody have been recognized since Plato. I argue that rhetoric also would have been important in early human societies, suggested by the fact that we are (still) surprisingly easy to manipulate using the right combinations of speech and music.

## 2.1. Introduction

It is evident that our perception of music is dependent on our auditory, cognitive and affective abilities, which in turn are shaped by physiological adaptations throughout our evolutionary history. Also, it is evident that our auditory and cognitive abilities are part of the 'instrument' on which a composer plays his or her tunes, since composers are craftsmen who cunningly have used their knowledge of human audition to write effective pieces of music. In the context of music perception, the physiological and cognitive mechanisms that underlie music perception and even music enjoyment can be characterized in biological terms, as the result of investigations, aided by new technologies such as non-invasive brain scanning. Thus, it is clear that bio-musicology (for a current review of the field, see Zatorre and Peretz, 2001) is a new and growing discipline that supplements rather than supersedes traditional musicological studies, and it is to be hoped that the future will see a fruitful collaboration between biological and other approaches to music. In a biological approach to 'music and meaning' it should always be borne in mind, however, that the definition of meaning is very different in a biological compared with an individual context: Biologically, the 'meaning' of a trait such as musical ability is its evolutionary drive, and its evolutionary relevance is based on the increase in individual fitness (number of surviving offspring) it confers to an organism. This is a very different definition than our usual one when we refer to the meaning of a piece of music or a text, and the meaning as defined by biology may very well be one that we do not like! For the same reason, it is unlikely that aesthetic judgements will be informed to any large extent by biological findings such as the demonstration that certain preferences are natural or inborn. Should it be shown, for example, that the natural harmonic series was robustly encoded in the central nervous system and therefore in some sense 'natural', as suggested by Helmholtz (1885), it would only demonstrate that listeners and performers would recognize this series with greater ease and explain why music tended to use harmonically related sounds, but not explain why those kinds of sounds should be aesthetically favorable. Also, we cannot - even in the future - expect biological studies to do much more than to localize the music processing centers and describe their physiology. Thus, the subjective musical experience, which is the really important part of the meaning of music, remains inaccessible to this kind of approach.

Biology can and must offer evolutionary speculations regarding the origin of music and its evolutionary meaning, since 'nothing in biology makes sense except in the light of evolution' (Dobzhansky, 1973). However, as it will be evident from the following, none of the theories of the evolutionary origin of music are much more than speculative at present.

In the present paper, I will try to outline current theories of the origin of music based on evolutionary theory (for a recent volume on music and evolution, see Wallin, Merker and Brown, 2000). I will also present an alternative hypothesis assuming a close link between music and the non-semantic (prosodic) component of language, the speech melody that consists of slow variations in pitch and conveys the emotional, non-lexical meaning of speech. The basis of the age-old understanding of the rhetorical power of music is that music is closely related not to speech, but to speeches. Speeches do more than convey semantic meanings in language to an audience; rather, they manipulate the listeners' understanding effectively by adding non-semantic emotional and gestural components. The close connection between music and rhetoric is an idea leading back to Plato's discussion of prosody in the *Republic*, 399a, where the inherent

qualities of Greek melodic modes are discussed in relation to their rendering of word and prosody as these would be employed by brave, weak, and moderate men. Plato clearly believed that the different melodic modes corresponded to different types of speech and could influence behaviour directly: The Lydian and Mixolydian modes were fit for lamentations, the Ionian and Lydian for banquets and the Dorian and Phrygian for imitating the speech of brave and moderate men. According to a modern view of modes as somewhat comparable to ancient versions of major and minor scales, such statements are puzzling; the Greek modes, however, may also have indicated specific rhythms and melodic progressions. In that case, it is not too difficult to imagine that the modes would have generated very different melodies that simulated different prosodies (and, interestingly enough, we would probably all have an idea about what a 'lamenting' melody or a melody imitating the speech of brave and moderate men would be like). If music and rhetoric are closely related, one could suppose that music and rhetoric also have had a common origin and a common evolutionary rationale, namely, to communicate emotions effectively in order to manipulate a group. This suggestion is the main thesis of the present paper.

## 2.2. An Evolution Primer

Since much of the following will be based on the evolutionary principles of biological science, it may be useful to clarify some of the principles of evolution, since the theory is widely misunderstood outside the biological sciences. Firstly, it should be appreciated that the theory of evolution (Darwin, 1859; for a recent review, see Freeman and Herron, 2004) actually consists of two parts: 1) A historical theory stating that all organisms are related and have evolved from ancestral species and 2) A mechanism explaining evolutionary change as caused by random events (genetic drift) or by a mechanism such as adaptation by natural selection. Natural selection means that the organism is shaped throughout its evolutionary history by the differential reproductive success of individuals - i.e. individuals that most efficiently utilize the resources in the environment produce most offspring and therefore their genes end up dominating the gene pool. This very simple mechanism is surprisingly powerful and produces adaptation of a population within a few generations, if the selection pressure (i.e. the differential reproductive success) is large enough. It should be noted that selection influences the immediate reproductive success of the organism, so the organism cannot suffer a reduction of fitness in order to achieve an increased fitness in future generations, and, moreover, there is no direction of evolution. Also, it should be realized that chance events (such as natural disasters) can change the outcome of evolution.

The origin of music within the context of sexual selection was first proposed by Darwin (1871), so it may be useful to review the concept of sexual selection briefly here. Darwin was puzzled by some morphological or behavioral traits that were seemingly non-adaptive and thus unexplainable by 'normal' natural selection - for example, structures such as the peacock's tail. To Darwin, music was also such a mysterious trait. The mechanism invoked was sexual selection: Individuals of one of the sexes choose a partner of the other sex based on these accessory or secondary sexual characteristics. Usually, sexual selection is found where the reproductive investments, i.e., the resources the individuals spend on reproducing, are unevenly distributed between the two sexes. Generally, in tetrapods (land-living vertebrates, i.e., amphibians, reptiles, birds and mammals) the female sex has the highest reproductive investment, especially in a viviparous species such as mammals - compare the resources spent on carrying the foe-

tus by the pregnant female with the resources used to generate male sperm cells. Of the mammals, the relative reproductive investment of females is probably highest in humans because of the extended period of child care. In such systems, the theory is that the 'choosy' sex should act to decrease the possibility of mating with a low-status partner, since there are far more dire consequences for the choosy sex if it squanders the opportunity to reproduce on such a partner. Animals of the 'chosen' sex can develop extreme secondary sexual structures, of which the tail of the peacock is a well-known and striking example, but even more extreme examples are found in some bird species, where many males are aggregated in an arena or 'lek,' and perform intricate behavioural displays for the selective females. An important fact is that sexual selection always leads to sexual dimorphism, as witnessed in the peafowl: The choosy sex does not develop the extreme secondary sexual structures.

Four different and not mutually exclusive mechanisms have been proposed for sexual selection. One is the handicap theory, where the partner shows that he can survive with large, expensive and unnecessary structures (tail, antlers) or make complicated courtship displays. The sensory exploitation theory states that animals choose the partners that are most conspicuous and therefore easiest to find (for example, animals emitting the loudest or most directional calls), so the secondary sexual characteristics exploit an inherent bias in the sensory system of the choosy sex. The good genes theory asserts that the structure preferred by the female is an indication that the male has high-quality genes, for example conferring low susceptibility to pathogens and finally, the theory of runaway sexual selection states that secondary characteristics escalate, simply based on the fact that the genes for the trait itself (from the chosen sex) and the genes for the selectivity for it (from the choosy sex) are united in the offspring. Both the trait and the preference for it will therefore be 'amplified' in the offspring.

### **2.3. Animal Communication.**

In biological investigations of animal communication a baseline assumption is that animals do not make calls for fun, since animals spend energy on calling and make themselves conspicuous and vulnerable to predation. Rather, the drive for evolution of communication systems is always thought to be based on some selection advantage that the sender achieves by communicating. (Note that this is a narrower definition of communication than would be used by some writers; see Bradbury and Vehrencamp, 1998.) An important point to note is that because of evolutionary adaptation, animal communication is different from the sort of communication dealt with in 'rational' communication theory in the sense that the aim is not to transfer information as efficiently as possible, but rather to manipulate the receiver as deftly as possible. Thus, it cannot generally be assumed that animal communication is honestly signalling the emotional state of the sender: If that were the case, the sender would be easy to manipulate, and deceitful signals, which are commonly seen, would be impossible. For example, some bird species will use alarm calls, which are normally used as warnings against predators, to chase competitors for food away (Munn, 1986), and a similar use of alarm calls has been described on the part of vervet monkeys (Cheney and Seyfarth, 1990; see especially p. 184-203 for an extended review of deception in animals). Rather, if the signalling should confer any advantage to the sender, communication signals would be a controlled and 'filtered' version of the animal's emotions and are thus often designed to deceive or manipulate the listener. Models from game theory can be used to predict, for example, whether it 'pays' evolutionarily to make honest signals or to use deceitful communica-

tion (see Bradbury and Vehrencamp, 1998) . In general, the outcome of the models depends on the likelihood that a deceiver meets the 'victim' again, and therefore depends also on the social structure.

If we turn to some of the species that use sound communication extensively and have a large repertoire, bird and especially song bird sound communication have been compared to music in our species, since birdsong can almost rival (simpler) human music in complexity. The repertoire of most bird species is relatively stereotyped and learnt during adolescence, however, after which it is more or less fixed. It is usually only produced by males and is used for maintaining territories as well as being a product of sexual selection. Song learning in birds has a clear neurophysiological basis and has formed one of the model systems in neurobiology during the recent years (Brainard and Doupe, 2002).

In mammals, one of the more well-known communication systems is that involving the long and complex calls produced by humpback whales. Humpback whales migrate over distances of several thousand kilometres to specific breeding grounds. The male humpback whales produce varied and extremely complicated calls, lasting from 5 to 35 minutes and with an extraordinary variation in pitch as well as timing (Payne, 2000). While this system is not easy to study, since it is hard to make direct observation of the calling and responding animals, recent work suggests that the calling males aggregate in a lek and that females probably select males based on their calls. Thus, the humpback calls would be another example of call complexity generated by sexual selection.

Our closest relatives, the primates, all use vocal communication (Geissman, 2000; Hauser, 2000; Hauser and McDermott, 2003). Many species have a relatively large repertoire of calls, ranging from alarm calls to social calls that can be very long and powerful. For example, the loud gibbon calls can last for over an hour and may function in male-female pair bonding (Geissman, 2000). Previously it was assumed that the calls were reflections of the caller's emotional state, but more recent research has showed referential components in the calls. For example, vervet monkeys use alarm calls to warn other monkeys of approaching predators (Cheney and Seyfarth, 2000). There are, however, different alarm calls dependent on the type of predators (snake, eagle or leopard). Call diversity has adaptive value, since different behavioral responses are appropriate in each case: If the predator is a snake, the monkeys should stand erect and scan the ground; if a leopard, they should climb up a tree; if the predator is an eagle, the appropriate response is to seek cover on the ground. The important point is that in this case the calls are not emotional, but referential. One may argue, however, that the distinction between emotional and referential calls presupposes that emotional calls are more or less precise communications about the emotional state of the animal. This would be very unlikely, since there is no general adaptive value for an animal to communicate its emotive state precisely, as stated above. Rather, it must be assumed that all calls - both emotional and referential - serve a purpose.

#### **2.4. Human Sound Communication**

It is natural to view human language as a major evolutionary adaptation of our species (Pinker, 1993), and human language shows several unique characteristics: It is clearly of adaptive value, and there is a universal syntactical structure underlying all human languages. The major difference between human language and the signals used in animal communication is that human language has a unique combinatorial structure (generative

grammar; see Chomsky, 1957), which permits an infinite number of sentences. Another unique aspect of human language is the symbolic aspect: the ability to refer to objects and events in the future and the past, i.e. without a direct reference (note that whereas some primates may use referential calls, they still refer to present objects). Human sound communication is not only language (or music), however; we use many non-verbal signals, usually denoting some kind of emotion. Furthermore, language contains at least two different components that it may be important to distinguish between. In the following, I will use the term 'semantics' for the 'lexical' meaning of language such as would be communicated by a typed transcript. In contrast, I use the term 'prosody' (as 'speech-melody') for slow variations in pitch contour and rhythm, carrying information (for example) about large-scale sentence structure and emotional content (note that this use of the term prosody is close to Plato's original (*Republic*, 399a)). Prosody is the component that a skilled orator would manipulate to enhance the effect of a speech and the main component in non-verbal sound communication or parent-infant communication. It is evident that some rhetorically effective spoken performances or declamations come close to musical (sung) performances; in fact, slowed-down speeches with more sound energy put into the vowel sounds (thus making the speech audible at longer distances) would almost be heard as song. Physiologically, the semantic-prosodic distinction makes sense, since the two components - following Zatorre et al. (2002), they could also be called 'slow temporally varied, tonal' (prosody) and 'fast temporally varied less tonal' (semantic) components - are processed by different brain centers, as outlined below. Also, patients with defects in the first brain center will be deaf to the prosodic and emotional aspects of language, whereas patients with defects in the other brain center will be aphasic, but with full sensitivity to prosody and emotion. A problem in much of the recent literature on language and music is that the semantic/prosodic division in language is not addressed.

Humans possess a large number of species-specific adaptations for sound production which enable us to produce an unrivalled number of different sounds. For example, monkeys and apes cannot produce the large number of speech sounds - most notably the majority of the consonants - that humans produce. The basis of the diversity of human speech sounds is found in the morphological arrangement of the vocal system. The basic structure and mechanism are similar to other mammals, but a notable difference is that the larynx has descended into the throat, leaving the tongue free to move in two dimensions in the vocal tract (Fitch, 2000). Most importantly, however, is that the human ability to produce a large variety of sounds is accompanied by an unusual ability for vocal learning and imitation, which is not found in any of our closest relatives (Fitch, 2000). Many of the adaptations for human sound production are soft structures that do not show up in fossils, so it is difficult to date the origin of human sound communication conclusively. Three features that are found in fossil material may be related to human sound production (Frayer and Nicolay, 2000; Morley, 2002), however, though none of them can be linked conclusively to human speech (Fitch, 2000). The hyoid bone (tongue bone) has a special structure in humans, and of course the mobility of the tongue is a prerequisite for speech production in recent humans. Also, the human rib cage has a specialized, barrel-like structure, which is not found in apes or monkeys, and which probably is important for controlling the air stream to the vocal cords. Finally, the protruding nose of humans is probably important in the production of speech sounds (Frayer and Nicolay, 2000). These adaptations are found in early humans dating back 1.5 million years, but probably were first fully developed 400,000 years ago (Morley, 2002). Within

the last two million years the evolution of the human brain involved rapid enlargement, i.e. a tripling of volume. During this development, the specialized centers for language processing probably appeared, as did the strong lateralization of the language-related centers.

To summarize, the uniqueness of human language is based on three different properties: 1) The diversity of speech sounds, 2) The combinatorial property of language, and 3) The symbolic structure, enabling reference to past and future objects (Donald, 1991). These components need not have originated at the same time. It would have been possible, for example, to have a simple language that still had syntax, but not necessarily the full complement of speech sounds. Conversely, one could imagine a protolanguage with the full diversity of speech sounds, but no syntax. Such a language would probably be comparable to bird songs and could have served the same function - sexual selection of the male with the most varied repertoire (Fitch, 2000) - though present-day language does not show any of the sexual dimorphism that always accompanies sexual selection in other animals. However, it is as likely that a diverse, asyntactical protolanguage would have served important social functions, as in other primates (Cheney and Seyfarth, 1990; Hauser, 1996) or in parent-infant communication (Trevathan, 1979).

If language evolution has been a major selective force within our species, it has probably also shaped the evolution of the brain, especially with regard to the special computations needed in human syntactical language. One of the consequences of the drastically enlarged brain of our species is that human babies have very large heads (pushing the birth canal to its limits) and that their brains take a long time to develop full cognitive powers. Infants must therefore be protected until a relatively advanced age, which promotes a social structure with well-knit relations between members. Thus, advanced social structure is probably one of the prerequisites for human survival during the course of evolution. Also, language and social structure are interdependent; language is less useful outside a reasonably solid social structure (since the vocabulary is learned and shared between the group), and language can be expected to promote social structure by allowing social complexity such as gossiping and tall tales (Bickerton, 2000). It follows that communication within larger groups as well as parent-infant communication becomes a very important adaptive trait in humans.

## 2.5. Music and Human Hearing

If music is a specific human trait, which the auditory system evolved long before music arose, then our processing of music is using parts of the auditory pathway already in place and shaped by selection for non-musical functions of hearing. To understand the evolution of human hearing it is important to know which functions hearing is serving and their selection (survival) value. Three major functionalities of human hearing may be identified quite easily (Christensen-Dalsgaard, 2002). One is a reflexive response to loud, sudden sounds, which is probably a primitive response reflecting the fact that loud, sudden sounds usually are 'danger signals,' since such sounds are warnings about mechanical events in the vicinity of the observer. A second functionality is the ability of our auditory system to assign sound components to sound sources or acoustic objects (Yost, 1993). This is not a trivial task, given that sound components are mixed at the two eardrums, and is probably a complicated computation comparing parameters such as

onsets, frequencies, location, simultaneous amplitude modulation etc. for the different sound components (Bregman, 1990). Finally, an important function of hearing must be that of analyzing human language sounds, i.e. to translate the sounds to their symbolical equivalent and process the syntactical structure. Language works perfectly well with non-acoustic - for example, visual - signals (sign language), so the final processing must be in some general (i.e. not auditory) symbolic center. Mammals generally have very good hearing, and among mammals humans excel in their frequency discrimination abilities (Long, 1994), which may be an adaptation for processing the spectral fine structure of speech sounds, but also is very important for our auditory streaming abilities. Language processing in humans is largely lateralized with several dedicated centers in the auditory cortex that are specialized in semantic (areas 22 and 39 - Wernicke's area, left superior temporal gyrus), prosodic aspects of speech (right superior temporal gyrus), and syntactic processing (Broca's area, left frontal lobe). None of these centers is unique to humans (Hauser, 2000; Hauser and McDermott, 2003), but they are hypertrophied in humans compared to other primates. It has been shown that the cortical brain centers processing music and language generally overlap, but that music centers are generally placed in the right hemisphere. In the case of trained musicians, however, it is processing by the left hemisphere in particular that becomes increasingly important. This may partly be due to increasingly 'verbalized' processing of music, but there are also recent data showing that 'syntactic' elements in music (harmonic progressions) are analyzed in Broca's area (Maess et al., 2001), which is regarded as a language center, but may be a more general symbolic or syntax-processing center.

Of these three functionalities, it may be assumed that only the last two would be really important in music processing by human hearing, since loud, impulse-like 'startle' sounds probably are processed by a specialized neural pathway due to the reflex-like response. On the other hand, the ability to separate sounds into different 'streams' is very important for our music perception, not only for identification of instruments in an ensemble, but probably also for fundamental aspects such as consonance and dissonance, since tones in different streams do not really create dissonance (Bregman, 1990). The ability to process language has provided us with a series of exquisite tools to analyse pitch, pitch contours and rhythm, features that are central to music perception as well. Indeed, one theory for the origin of music is that there is no special selection pressure for music - it is just a non-adaptive 'game' created and exploited for pleasure by the (idle) brain of our ancestors and its ability to process acoustic patterns (the so-called 'auditory cheesecake' theory, Pinker 1997) by using neural pathways already in place to analyze language. Recent evidence indicates, however, that there might be specialized centers for music. Peretz et al. (2002) have studied individuals with *amusia* (i.e. a selective disability for recognizing musical elements such as melody and rhythm), either of the congenital sort or the kind caused by stroke, and they present evidence that these individuals have normal language processing (at least concerning the semantic component of language). The authors do not ascribe the music processing to a well-defined center, but rather too many distributed neural circuits. Also, the functional explanation for the amusia is that the individuals lack the ability to track fine-scale pitch contours (whereas the pitch changes in speech - typically much larger - are detected without problems). Another study has shown that an individual with amusia (caused by a stroke affecting the right frontoparietal cortex) was also impaired with respect to prosodic discrimination, for example that of intonation differences in language (Nicholson et al., 2003). Furthermore, another study on patients with amusia showed



that they had normal rhythmical processing, indicating a dissociation of melody and rhythm processing (Hyde and Peretz, 2004; Morley, 2002). Interestingly, Zatorre et al. (2002) after studying music and speech processing in the auditory cortex have suggested that the functional asymmetry - that music is processed primarily in the right, speech primarily in the left Heschl's gyrus - reflects a functional specialization of the auditory cortex: one area (right) is specialized in accurate pitch processing with low time resolution (i.e., processing slowly varying pitch contours), and the other (left) is specialized in low pitch resolution and high temporal resolution. We naturally associate these two kinds of processing with music and (semantic) speech processing; one really interesting aspect, however, is that the cortical specialization could have predated music and speech, since Zatorre et al. report that identical divisions are also found in other mammals. It might be, then, that speech (and music) co-opted existing structures, and that the speech and music signals during evolution were modified so as to be easy to analyze.

The search for brain centers or networks dedicated to music processing is important in the context of evolution, since such centers would show that there was sufficiently strong selection pressures associated with music to lead to dedicated centers. On the basis of current knowledge it is difficult to maintain, however, that there is large selection pressures associated with music in humans. Also, it is evident that one of the features of persons with amusia(s) is that they generally are individuals who function very well in other respects. So even if there is an evolutionary benefit associated with musicality, it may not be large enough to drive the evolution of large, dedicated brain structures.

## 2.6. Music and Emotion

Everybody agrees that music produces an emotional response in listeners and music has even been called the 'language of emotions'. This would assume that emotions could be rather precisely encoded and transmitted by music. The link between music and language has been stated most explicitly by Cooke (1959), who proposed that the music of the last six centuries is a coherent emotional language, where certain melodic formulae have well-defined emotional meaning. Cooke based his argument on examples from Western music over the last 600 years. It may still be questioned, however, whether the effects of music on the emotional state of the listener are universal or whether they depends on shared cultural background, verbal description, context - and liner notes. I think that it is reasonably safe to assume that there is a shared 'emotional codex' in classical and romantic music, but I think it is much more tenuous to associate well-defined emotional states with melodic formulae in music of other cultures, including (from my personal experience) medieval and renaissance Western music. If there were such a well-defined 'emotional codex' it could also be assumed that music theorists and composers like Johann Mattheson (1739) or even more recent composers such as Aaron Copland (1939) or Carl Nielsen (1925) would have mentioned these associations, but that is not the case - instead, most composers seem to be as unspecific as the 13<sup>th</sup> century theorist Franco of Cologne, stating that 'he who wishes to write a *conductus* [a medieval polyphonic form, JC-D] ought first to invent as beautiful a melody as he can' (translated in Strunk, 1965). Mattheson (1739), while stressing the importance of music emulating different emotions, is not more specific than that he states that joyous melodies should have large intervals and sad melodies narrow intervals. Also, it would be nice to see Cooke's ideas subjected to experimental testing, but in the only case where that has been tried, the results did not show any clear association between emotion and well-

defined melodic patterns (Gabriel, 1978; for criticism of the experiments, see Sloboda, 1985). In other words, I do not believe that a well-defined 'emotional language' is a musical universal.

Rather, I think that the true musical universals - that slow, low-pitched sounds are 'sad' and fast, higher pitched sounds joyful (or aggressive) - would also be prosodic universals. In that sense, music is not the language of emotions - but prosody is, and as far as music emulates prosody, it can also encode emotions. Another powerful emotion associated with music is 'the chills' experienced by listeners. In measurements of cerebral blood flow in subjects listening to their favourite pieces of music the chills have been shown to be associated with pleasure and reward centers such as ventral striatum, midbrain and amygdala in the brain, and the response comparable to responses to other pleasurable stimuli such as drugs and arousal (Blood and Zatorre, 2001; see also review by Panksepp and Bernatzky, 2002). In my experience, the chills are not caused by any special kind of music, but are a personal (extremely subjective) feeling of the quality of the music (of being moved by the music). I would like to note that such a general feeling of quality and well-being could be a very powerful rhetorical tool and may provide a partial explanation of the persuasive powers of music.

## 2.7. Origin of Music

A fundamental and unanswerable question at present is the question of when music arose during human evolution. The finding of a putative musical instrument (edge flute) from a Neanderthal settlement from approx. 50000 BC suggests that even advanced music making might be ancient (Kunej and Turk, 2000; note that it is still disputed whether the flute is a human artefact), but of course music making without instruments, e.g., singing, would have been possible long before that, and the fossil record for singing is as inconclusive as the fossil evidence for speech reviewed above. Four points can be noted:

1) Music depends on a social structure with a high degree of protection, since it is dangerous to make sounds that may attract predators

2) Music can use the same sound-producing features as speech -singing is basically slowed-down speech where much more sound energy can be packed into the vowel sounds

3) Music has some similarities to the non-semantic part of language, e.g. in the prominence of pitch contours for delineating phrase structure and emotional content

4) Music is a human universal; there are no human cultures that do not produce music. Indeed, one current theory is that music is non-adaptive and just uses the vocal production and analysis apparatus already present due to an intense selection for language (the 'auditory cheesecake' theory described above, Pinker, 1997). In that case, it may be difficult to explain the universal appearance of music in human societies. Furthermore, there are indications that music could be adaptive. For one thing, music has special functions with regard to social bonding, for example between parent and infant (Trevorthen, 1979; Trehub, 2001), and the predispositions for melodic as well as rhythmic interactions between parent and infant may be the basis of musical ability (Disanayake, 2000), but music also has the feature that it can coordinate the behavior of large groups where language may not be as useful. Another explanation for the origin of music has been that music originates through sexual selection (Darwin, 1871; Miller, 2000), where females should prefer males with a large repertoire. This is in complete analogy with the songbird communication system. In the case of songbirds, however, the

vocal communication shows the expected sexual dimorphism, i.e., the females do not use the large variety of songs found in the males; only the brains of male birds show large nuclei dedicated to song learning and song production. Now, if this theory applied to humans, we should see a robust sexual dimorphism in the abilities for music production in males and females. This sexual dimorphism is simply not found in humans - there is no clear difference in musical abilities between the sexes (accounting for social and historical bias). Furthermore, even though males may serenade there is no clear and robust behaviour associated with music-making and courtship. Therefore, I do not believe that sexual selection in the form known in other animals is applicable to music, nor, in my opinion, is it applicable to other human art forms. I do believe, however, that there is sexual selection for traits in humans, but that is probably for more obvious traits such as male power.

Another viewpoint is that music and language have a common origin (the 'musilanguage' hypothesis, Brown, 2000). It is likely that our ancestors had some kind of vocal communication before language originated. We may get an impression of the types of pre-lingual vocal communication from our non-verbal acoustic signals ('aah', 'eech', 'mmm,' etc.) that are simple, usually tonal, and generally emotional. This vocal communication would probably be similar to vocal communication in other primates, and therefore closer to the emotional, prosodic component of language, although as described previously, monkey vocalizations also can be referential, and that the border line between referential and emotional communication is not clearly defined. The new development of human semantic language would then entail co-opting brain centers analyzing fast time-varying pitch contours, according to the theory of Zatorre et al. (2002), but retaining the 'older' centers for processing of prosody - and of music.

## 2.8. Music and Rhetoric

While I think it very likely that music serves a function in other aspects of social behavior, I would like to advance a hypothesis related to the musilanguage model: Music could have its origin not in speech, but in speeches. It has been well-known since Plato, and especially in the writing of theorists from the Middle Ages until the Baroque, that there is a close link between music and rhetoric (reviewed in Unger, 1941), where rhetoric is understood as the ability to impress listeners with ones viewpoints or even to manipulate them. Much of the music-rhetorical literature contains rather detailed analogies between rhetorical figures in speech and in music, and while such ideas undoubtedly have been important for renaissance and baroque composers (and therefore important from a musicological viewpoint), they may seem to be purely theoretical constructions today. It should not be overlooked, however, that there always was an underlying, highly practical rationale, namely, that music is a very powerful rhetorical element. It was formulated quite explicitly by renaissance theorists; for example, Jacopo Sadoletto in 1533 states: '...By themselves the words have no mean influence upon the mind, whether to persuade or restrain. Accommodated to rhythm and metre they penetrate much more deeply. If in addition they are given a melodic setting they take possession of the inner feelings and of the whole man.' (translated in Palisca, 1985). It is also well-known that music still can be used with great rhetorical effect to manipulate listeners, one of the most infamous examples probably being the orations of Hitler at the mass rallies in Nürnberg (Storr, 1992). If speeches were important in early human societies, and they must have been, since they would have been a way of coordinating responses of the whole group, music might have its origin in the rhetorical prosody of speeches.

Note that this could well be an example of an evolutionary 'meaning' of which we do not approve, since it suggests that we might be much more susceptible to rhetorical persuasion than we would like to admit.

Oliver Sacks (1985) recounts a thrilling example in which aphasic patients with damage to the left temporal lobe were compared with Emily D., a patient with a tumor in the right temporal lobe. The aphasic patients did not recognize the semantic aspects of speech, but were still sensitive to its emotional (prosodic) content. In contrast, Emily D. was *only* sensitive to the semantic aspect of language. When these patients were listening to a speech by an American president, noted for his use (or misuse) of rhetorical tricks and emotional appeal, they all sensed that it was, in some way, wrong or untrue. The aphasic patients sensed that intonation and rhythm were misplaced; Emily D. sensed that the words were used in a non-standard, incorrect way. Oliver Sacks' point is that, interestingly, we as normal listeners are deceived by the combination of prosody and semantics - whereas the patients, insensitive to one or the other parameter are not deceived. This raises a fundamental question: why has evolution left us powerless against such rhetorical manipulation - perhaps because it has been the prize of coordinating a group, which has been essential for the survival, and hence for the fitness of all the members?

The theory linking the origin of music to the origin of speeches that I have sketched briefly here may seem to be pure speculation; it does accord, however, with the available neurobiological facts which show a close relationship between prosody and music perception. It would be useful to test this relationship in further experiments, for example making detailed analyses of musical compositions and the prosody of the composer's native tongue. It should also be possible to investigate directly the relationship between rhetoric and music in controlled experiments - for example to investigate how appropriate music affects our value judgements in relation to speech.

A major problem in biological studies of music is that it is not very clearly understood how music is distinguished from non-musical sounds. I have not addressed this topic in the present paper, but a theory of music as a biological adaptation clearly implies that music perception is not synonymous with sound perception, and that there should be some defining, universal characteristics of music (across cultures and independent of conventions and context). It would be highly valuable to investigate in the future whether such universal features of music can be delineated.

### Acknowledgements

I thank the two anonymous referees and the editorial staff of this journal for their perceptive and constructive criticism. The work has been supported by the Danish National Research Foundation.

### References.

Bickerton, D. (2000). Can biomusicology learn from language evolution studies? In N.L. Wallin, B. Merker & S. Brown (Eds.) *The origin of music* (pp. 153-163). Cambridge: MIT Press.

Blood, A.J., & Zatorre, R.J. (2001). Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. *Proceedings of the National Academy of Science* 98: 11818-11823

- Bradbury J.W. & Vehrencamp SL (1998). *Principles of animal communication*. Sunderland: Sinauer Associates.
- Brainard, M.S., & Doupe, A.J. (2002). What songbirds teach us about learning. *Nature* 417: 351-358.
- Bregman, A.S. (1990). *Auditory scene analysis*. Cambridge: MIT Press.
- Brown, S. (2000). The 'Musilanguage' model of music evolution. In Wallin, N.L., Merker, B., & Brown, S. (Eds.), *op. cit.* (pp. 271-300).
- Chomsky, N. (1957) *Syntactic structures*. The Hague: Mouton & Co.
- Christensen-Dalsgaard, J. (2002). Towards a neuroethology of human hearing. In: Tra-  
nebjerg, L., Christensen-Dalsgaard, J., Andersen, T., & Poulsen, T. (Eds.) *Genetics and  
the function of the auditory system*. Proceedings of the 19<sup>th</sup> Danavox symposium (pp.  
170-181). Denmark: Holmens trykkeri.
- Cheney, D.L., & Seyfarth, R.M. (1990). *How monkeys see the world*. Chicago: University  
of Chicago Press:
- Cooke, D. (1959). *The language of music*, repr. 2001. Oxford: Oxford University Press.
- Copland, A. (1939). *What to listen for in music*. New York: McGraw Hill Book Company.
- Darwin, C. (1859). *The origin of species by means of natural selection, or the preserva-  
tion of favoured races in the struggle for life*. London: Murray.
- Darwin, C. (1871). *The descent of Man, and selection in relation to sex*. London: Murray.
- Dissanayake, E. (2000). Antecedents of the temporal arts in early mother-infant interac-  
tion. In Wallin, N.L., Merker, B., & Brown, S. (Eds.), *op. cit.* (pp. 389-410).
- Dobzhansky, T. (1973). Nothing in biology makes sense except in the light of evolution.  
*American Biology Teacher* 35: 125-129
- Donald, M. (1991). *Origins of the modern mind: Three stages in the evolution of  
culture and cognition*. Cambridge: Harvard University Press.
- Fitch, T. (2000). The evolution of speech: A comparative review. *Trends in the Cognitive  
Sciences* 4: 258-267.
- Fruyer, D.W., & Nicolay, C. (2000). Fossil evidence for the origin of speech sounds. In  
Wallin, N.L., Merker, B., & Brown, S. (Eds.) *op. cit.* (pp. 217-234).
- Freeman, S., & Herron, J.C. (2004). *Evolutionary analysis*. New Jersey: Pearson Prentice  
Hall.

Gabriel, C. (1978). An experimental study of Deryck Cooke's theory of music and meaning. *Psychology of Music* 6:13-20

Geissmann, T. (2000). Gibbon songs and human music from an evolutionary perspective. In Wallin, N.L., Merker, B., & Brown, S. (Eds.), *op. cit.* (pp. 103-123).

Hauser, M. (1996). *The evolution of communication*. Cambridge: MIT Press.

Hauser M (2000). The sound and the fury: Primate vocalizations as reflections of emotion and thought. In Wallin, N.L., Merker, B., & Brown, S. (Eds.), *op. cit.* (pp. 77-102).

Hauser, M., & McDermott, J. (2003). The evolution of the music faculty: A comparative perspective. *Nature Neuroscience* 6: 663-668

Helmholtz, H. (1885, repr. 1954). *On the sensations of tone* (A.J. Ellis, Trans.). New York: Dover,

Huron, D. (2001). Is music an evolutionary adaptation? *Annals of the New York Academy of Science* 930: 43-61

Hyde, K.L., & Peretz, I. (2004). Brains that are out of tune but in time. *Psychological Science* 15: 360

Kunej, D., & Turk, I. (2000). New perspectives of the beginnings of music: Archaeological and musicological analysis of a middle Palaeolithic bone 'flute'. In Wallin, N.L., Merker, B., & Brown, S. (Eds.) *op. cit.* (pp. 235-268).

Long, G.R. (1994). Psychoacoustics. In Fay, R.R. & Popper, A.N. (Eds.), *Comparative hearing: Mammals. Springer Handbook of Auditory Research, vol. 4* (pp. 18-56). New York: Springer.

Maess, B., Koelsch, S., Gunter, T.C., & Friederici, A.D. (2001). Musical syntax is processed in Broca's area: An MEG study. *Nature Neuroscience* 4: 540-545

Mattheson, J. (1739) *Der Volkommene Capelmeister*. Hamburg: Christian Herold.

Miller, G. (2000). Evolution of human music through sexual selection. In Wallin, N.L., Merker, B., & Brown, S. (Eds.), *op. cit.* (pp. 329-360).

Morley, I. (2002). Evolution of the physiological and neurological capacities for music. *Cambridge Archaeological Journal* 12: 195-216.

Munn, C.A. (1986). The deceptive use of alarm calls by sentinel species in mixed species flocks of neotropical birds. In Mitchell, R.W. & Thompson, N.S. (Eds.), *Deception: Perspectives on human and nonhuman deceit* (pp. 169-175). Albany: State University of New York Press.

Nicholson, K.G., Baum, S., Kilgour, A., Koh, C.K., Munhall, K.G., & Cuddy, L.L. (2003). Impaired processing of prosodic and musical patterns after right hemisphere damage. *Brain and Cognition* 52: 382-389

Nielsen, C. (1925) *Levende musik* [in Danish]. Copenhagen: Martins forlag.

Palisca, C.V. (1985). The Artusi-Monteverdi controversy. In Fortune, N & Arnold, D. (Eds.), *The new Monteverdi companion*, (pp 127-158). London: Faber and Faber,

Panksepp, J., & Bernatzky, G. (2002). Emotional sounds and the brain: The neuro-affective foundations of musical appreciation. *Behavioral Processes* 60: 133-155

Payne, K. (2000). The progressively changing songs of humpback whales: A window on the creative process in a wild animal. In Wallin, N.L., Merker, B., & Brown, S. (Eds.), *op. cit.* (pp. 135-150).

Peretz, I. (2002). Brain specializations for music. *The neuroscientist* 8:374-382

Peretz, I., Ayotte, J., Zatorre, R.J., Mehler, J., Ahad, P., Penhune, V.B., & Jutras, B. (2002). Congenital amusia: A disorder of fine-grained pitch discrimination. *Neuron* 33: 185-191

Pinker, S. (1993). *The language instinct*. New York: Morrow.

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

Plato (ca. 400 BC). *The republic* (B. Jowett Trans., 3<sup>rd</sup> ed. 1888). Oxford: Clarendon Press.

Rauschecker, J.P. (1998). Cortical processing of complex sounds. *Current Opinion in Neurobiology* 8: 516-521.

Sacks, O. (1985). *The man who mistook his wife for a hat*. New York: HarperCollins..

Sloboda, J.A. (1985). *The musical mind*. Oxford: Oxford University Press.

Storr, A. (1992). *Music and the mind*. London: HarperCollins,

Strunk, O. (1965). *Source readings in music history. Vol. 1. Antiquity and the middle ages*. New York: W.W. Norton..

Trehub, S.E. (2001). Musical predispositions in infancy. *Annals of the New York Academy of Science* 930:1-16

Trevarthen, C. (1979). Communication and cooperation in early infancy: A description of primary intersubjectivity. In M. Bullowa (Ed.), *Before speech: The beginning of human communication* (pp. 321-347). London: Cambridge University Press.

Unger, H.-H. (1941). Die Beziehungen zwischen Musik und Rhetorik im 16.-18. Jahrhundert. Hildesheim: Georg Olms Verlag, (Reprinted 1969)

Wallin, N.L., Merker, B., & Brown, S. (Eds., 2000). *The origins of music*. Cambridge: MIT press.

Yost, W.A., & Sheft, S. (1993). Auditory perception. In Yost, W.A., Popper, A.N., & Fay, R.R. (Eds.), *Human psychophysics. Springer Handbook of Auditory Research, vol. 3* (pp.193-236). New York: Springer.

Zatorre, R.J., Belin, P., Penhune, V.B. (2002). Structure and function of auditory cortex: Music and speech. *Trends in the Cognitive Sciences 6*: 37-45

Zatorre, R.J., & Peretz, I. (Eds., 2001). The biological foundations of music. *Annals of the New York Academy of Science, 930*.