Evolution of the Physiological and Neurological Capacities for Music

Iain Morley

Neuropsychological and developmental studies suggest human musical ability has a deep evolutionary history; but we do not find evidence of the manufacture and use of instruments, with which musical behaviours have often been assumed to be equated, until 70,000 years after the advent of Homo sapiens*. This anomaly is addressed by examining the evidence from the fossil record for the evolution of the physiology and neurology required for musical behaviours, with the aim of identifying the development of the physiological and neurological capacity to produce and process melody and/or rhythm. Aural and vocal sophistication appear to have developed in tandem, beginning with full bipedalism around 1.75 million years ago, until a vocal apparatus similar to the modern was present in* Homo heidelbergensis *400,000–300,000 years ago. Prosodic and structural aspects of both speech and music production and processing are lateralized in the brain in similar ways suggesting evolutionarily-shared foundations for these mechanisms.*

The last ten years in particular have seen a considerable surge in interest and research into the evolution of the human mind and cognition across a number of disciplines. Work in cognitive archaeology (e.g. Ingold & Gibson 1993; Mellars & Gibson 1996; Mithen 1996), inferring the development and nature of cognitive abilities in prehistoric human ancestors from the fossil and archaeological record, has been paralleled by developments in psychology, neuroscience, linguistics and philosophy (e.g. Barkow *et al.* 1992; Byrne 1995; Deacon 1997; Carter 1998; Dennett 1998; Pinker 1997).

Within this body of research, however, one topic has been severely neglected by all but a very few (e.g. Wallin 1991; Cross 1999a): the evolution of musical abilities. It is only very recently that any unified approach to the issue has been attempted (Wallin *et al.* 2000; Cross in press; Cross & Morley 2002). Neuropsychological and developmental studies suggest that the human musical ability has a deep evolutionary history (e.g. Wallin 1991; Fernald 1992; Papousek 1996; Carter 1998; Dissanayake 2000) but this contrasts with the evidence from the archaeological record (e.g. Scothern 1992; Lawson *et al.* 1998), in which musical instruments and evidence for the possible use of acoustics do not appear until 35,000–40,000 years ago. This is more than 80,000 years after the first anatomically modern humans had appeared. This article aims to address the apparent disparity between the archaeological record and developmental evidence by examining the fossil record of the evolution of the physiology and neurology required for musical behaviour.

Any attempt to extrapolate the nature of any musical ability prior to the presence in the archaeological record of manufactured musical instruments must concentrate on vocal and percussive methods of music production. It may be possible to determine at what point early humans had the physiological and neurological *capacity* to produce and process melody and/or rhythm. To have the capacity is not necessarily to make use of it in that particular way, but it should allow a time to be identified before which musical behaviours would not have been possible. This article presents: 1) an analysis of the fossil evidence of hominids regarding the evolution of the neurology and physiology identified as dedicated to musical ability; 2) a review of the evidence from neuroscience regarding distinct areas of the brain relating to musical ability and their relation to those

Cambridge Archaeological Journal 12:2, 195–216 © 2002 McDonald Institute for Archaeological Research
DOI: 10.1017/S0959774302000100 Printed in the United Kingdom

https://doi.org/10.1017/S0959774302000100 Published online by Cambridge University Press

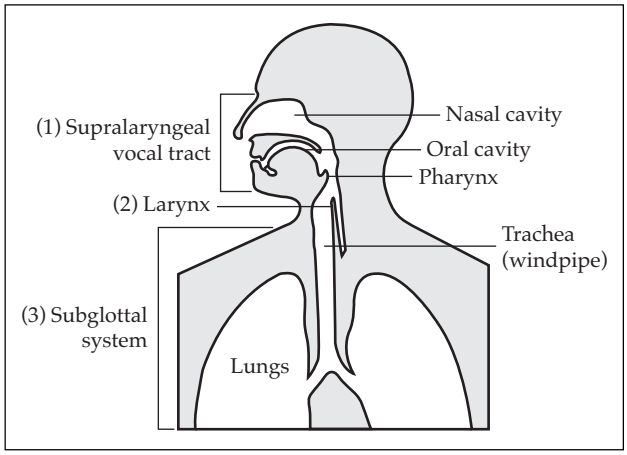


Figure 1. *The components of human vocalization physiology. (After Lieberman 1992, 134.)*

controlling speech; and 3) an examination of the physiology used in the production and processing of sound. Together, these analyses allow conclusions to be drawn regarding when musical capabilities could have begun to emerge, and how such capabilities would have developed over time.

A detailed review and analysis of evidence and rationales for the evolution of musical abilities, relationships between proto-musical and proto-linguistic behaviours, and a proposal for the path of evolution of musical abilities has already been undertaken (Morley 2000). The current article is intentionally and by necessity limited to an examination of the actual physiology and neurology required for such behaviours, and the evolution of these features. Although touched upon here, detailed examination of developmental studies, evolutionary rationales and proposals regarding the possible nature and development of complex vocalizations must await a future article.

Physiology

Sound production: the vocal apparatus and its evolution
The evolution of the physiology necessary for vocalizations has received much attention in the last thirty years, in particular since interest in language evolution has blossomed (e.g. Pinker 1994; Deacon 1997). A consequence of language evolution being the driving force behind most of such research has been that the results have generally been presented in terms of the development of speech capabilities in hominids. Dealing as it does with the fundamental physiological features required for *controlled* and *varied* vocalizations, however, the research is equally valuable as regards hominid capabilities to produce *any* vocalizations, linguistic, melodic or otherwise. Note that vocalizations are *not* equivalent to verbalizations, despite the fact that these terms have been used interchangeably in some literature. Recently, some attention has been given to these features with specific reference to music as well. The evidence for the development of some of the physiological features discussed below is also presented, with greater brevity, by Frayer & Nicolay (2000), to which the reader may refer for additional analysis of some of the aspects of the physiology considered here.

Vocalizations are produced in two stages, in the supralaryngeal tract and the oral cavity (see Fig. 1). The acoustic energy of vocal sound is generated by the larynx, which modulates the airflow in the throat (supralaryngeal tract). This sound is then adjusted by the action of the tongue, lips, teeth and palate in the oral cavity (Lieberman 1992). Unfortunately, the vast majority of the physiology responsible for vocal production (for example, the tongue, larynx and supralaryngeal tract) consists of cartilaginous and soft tissue that is not preserved in the fossil record. Consequently, when considering the evolution of vocal ability in our ancestors, it is necessary to draw conclusions about the nature of these features from the bony structures which do survive. These can provide some clues to the nature of the original oral and laryngeal features with which they were associated, such as the jaw (mandible), bottom of the skull (basicranium) and, in some rare and particularly valuable instances, the hyoid bone.

The larynx and basicranial flexion: The major influences on vocal ability are the position of the larynx in the throat, the size of the oral cavity and the size and control of the tongue, all of which are inextricably linked (Lieberman 1989). We know already that non-human primates (and other mammals) are capable of creating vocalizations, but these are of a limited range, at least in part because the larynx is positioned near the top of the throat (Duchin 1990; Budil 1994). In this position the larynx allows the animal to breath and swallow at the same time without choking, as it separates the upper respiratory and digestive tracts (Laitman 1984; Budil 1994). This also restricts the resonance (and thus variety of vowel sounds) produced by the pharyngeal cavity, the area between the epiglottis and soft palate (Lieberman 1984; Budil 1994). In adult humans, the larynx occupies instead a space in the pharynx further down the throat, opening up the pharyngeal resonance cavity

https://doi.org/10.1017/S0959774302000100 Published online by Cambridge University Press

and allowing the production of the wide range of vowels (such as [a], [u] and [i], the 'vowel triangle') that are universals in all human languages (Lieberman 1984). This capacity to produce a wide variety of sustained vowel sounds is also fundamental to the ability to produce vocal melody.

At birth, the larynx occupies the top of the throat in humans, as in other mammals, with the same important benefit with regard to feeding and breathing. As the child grows up, however, particularly after the first year of life, the larynx descends towards its adult position. This process is only completed during adolescence (Magriples & Laitman 1987). It seems that a similar process has occurred over the course of human evolution. Detailed analysis of chimpanzee vocal anatomy (Laitman & Heimbuch 1982; Duchin 1990) and other non-human mammals including baboons (Laitman & Reidenberg 1988) shows that non-human primates have a high larynx; in contrast humans, and only humans (Duchin 1990, 688), have a larynx low in the throat (see Fig. 2). Thus, if it is possible to reconstruct the position of the larynx in fossil hominids one should be able to draw conclusions as to the extent of their vocal abilities, in comparison with modern humans and non-human primates.

Laitman has asserted that there is a direct relationship between laryngeal position and the degree of curvature of the underside of the base of the skull, or *basicranial flexion* (Laitman *et al*. 1979; Laitman & Reidenberg 1988). This curvature is observable in varying degrees in extant pongids and humans, and is related to the position of the tongue and laryngeal muscular attachments to the base of the skull. Basicranial flexion is taken as an indication of the inclination of the styloid process, a bony protrusion from the temporal bone at the base of the skull, which is rarely preserved (Budil 1994). This, in turn, indicates the position of the hyoid bone and the larynx, as the muscles which support these structures attach to the styloid process (Budil 1994). This allows study of the degree of basicranial flexion in hominid fossils, and thus some reconstruction of the vocal tract can be made.

Laitman and colleagues carried out examinations of Plio-Pleistocene specimens such as KNM–ER 406 (*Paranthropus boisiei*) and Sterkfontein V (*Australopithecus africanus*), both approximately 3 million years old (Laitman & Heimbuch 1982), and found that these specimens show no flexing of the basicranium, as is the case with modern apes. The first evidence for the arching of the basicranium is displayed by *Homo ergaster* skull KNM–ER 3733 (from Koobi Fora, Kenya), approximately 1.75 mya (Laitman *et al.* 1979). A fully flexed, modern-like, basicranium is first seen in *archaic Homo sapiens* (or *Homo heidelbergensis*), 400,000–300,000 years ago (or kya), in specimens such as Steinheim, Broken Hill 1 and Petralona (Laitman 1984). Laitman suggests that *Homo ergaster* would have been able to produce the vowel sounds found in 'boot', 'father' and 'feet', whereas *archaic Homo sapiens* would have been able to produce a full modern repertoire of vowel sounds.

Variations in methods and quality of reconstruction of crania can, however, result in problems with such research. Researchers must use the utmost discretion in their choice of sample. This is illustrated by the initial work carried out by Lieberman & Crelin (1971), who created computer simulations of the vocal tract based on the anatomy of the La Chapelle-aux-Saints Neanderthal specimen. Although ground-breaking at the time, this study has since been severely criticized on a number of fronts. First and foremost, the La Chapelle-aux-Saints cranium is that of an aged and crippled Neanderthal, who was also suffering from the bone-altering disease osteoarthritis, and so is not a good approximation for the species as a whole. Secondly, Lieberman & Crelin's simulation was based partly on the original Boule reconstructions of the skull from 1911–12; Heim's (1985) reconstruction showed that it in fact had a basicranium very similar to that of modern humans (Heim 1985, cited in Scothern 1992).

Despite the criticisms levelled at this particular reconstruction, it does seem that Neanderthal basicrania are, on average, longer and flatter than those of modern humans and most *Homo heidelbergensis* (Laitman *et al.* 1979), and this is often still used as a basis for suggesting incomplete speech production amongst Neanderthals. Other writers (e.g. Arensberg *et al*. 1990; Kay *et al.* 1998), however, now suggest that the basicranium is only one of many complementary features which determine vocalization quality and enunciation ability (see below for discussion of these other factors).

Lieberman has since modified the severity of his conclusions in the light of the above criticisms of the 1971 simulations (Lieberman 1989), shifting his emphasis towards the connections between physiology and neurology. He still maintains that 'classic Neanderthal hominids appear to be deficient with respect to their linguistic and cognitive ability . . . they probably communicated slowly at extremely low rates and were unable to comprehend complex sentences' (Lieberman 1989, 391). He makes this assertion on the basis that there are important neural connections between vocal physiology and the cog-

https://doi.org/10.1017/S0959774302000100 Published online by Cambridge University Press

nitive substrates responsible for language production and perception. He hypothesizes that since the physiology appears to be underdeveloped in Neanderthals, so too was the neurology.

There are still, however, several major problems with these assertions. Kuhl (1988) concludes from her studies of the perceptual mechanisms and abilities of non-human mammals and pre-linguistic infants (see below p. 201) that the mechanisms relied upon to process speech sounds were in place long before humans began to use speech. She concludes instead that speech was shaped by our perceptual abilities. This means that Neanderthals could have had the neurological mechanisms necessary to perceive speech sounds even if they *were* unable to produce those sounds themselves.

Fidelholtz (1991) suggests that the brain capacity for language production was present in the 'generalized Neanderthals' such as Swanscombe and Steinheim (around 200,000 years ago, now usually classified with *Homo heidelbergensis*, although considered to be of a type ancestral to Neanderthals), and that the subsequent developments in the vocal anatomy (including the loss of jaw prognathism) were responses to the use of the speech ability.

Furthermore, research into the vocal abilities of individuals with Down's syndrome, who may possess abnormally high laryngeal position and large tongue, as well as lower than average intellectual capacities (whose linguistic ability Lieberman originally used as an analogy for Neanderthals'), suggests that such individuals can produce virtually all of the sounds present in modern speech, and have a well-developed singing ability (Scothern 1987, cited in Scothern 1992). It would seem that these anatomical features are not such a barrier to vocal production as has sometimes been asserted.

Other supralaryngeal reconstructions have been attempted. Budil (1994) compared the basicrania of 109 modern adult humans (from disparate ethnic groups) with those of 103 non-human primates (from 18 genera), 28 fossil hominids and the measurements from a further 16 fossil hominids as quoted in Laitman's publications. He used the evidence from all of the modern specimens (human and primate) to devise a formula relating basicranium, styloid process, hyoid, and laryngeal position. Based on this, he reconstructed the supralaryngeal tract of the Petralona specimen (*Homo heidelbergensis*: approximately 200,000 years old). His results indicate that a modern-like supralaryngeal resonance cavity was present in the Petralona specimen, and others with similarly flexed basicrania, including other *Homo heidelbergensis* such as Broken Hill and Steinheim, and *Homo sapiens* such as Taforalt 12, Cro-Magnon 1, Predmosti 3 and 4, Afalou 5 and Skhul 5. His conclusions thus agree with those of the earlier work by Laitman & Reidenberg (1988) and Arensberg *et al*. (1990, see below), that an upper respiratory tract similar to that of modern humans (and thus similar vocal ability) first appeared 400–300 kya (thousand years ago), in *Homo heidelbergensis*, and that the laryngeal physiology of the Australopithecines resembled that of modern apes. The results for Neanderthal specimens again fall outside the range of any modern humans, despite post-dating the other specimens by many millennia, agreeing with Lieberman's analysis of the physiology. To Budil, this finding 'remains enigmatic'.

It may not be quite as enigmatic as it first seems if *Homo neanderthalensis* lay on a different evolutionary path from the cited *Homo heidelbergensis* specimens of 300,000–400,000 years ago. It may be that those *Homo heidelbergensis* reconstructed are ancestral only to anatomically modern humans, not Neanderthals, and that *Homo heidelbergensis* specimens ancestral to Neanderthals would exhibit similar basicranial features to the Neanderthals. If the specimens of *Homo antecessor* (dating to about 800 kya) from Atapuerca are ancestral to both *Homo neanderthalensis* and *Homo sapiens* as the authors assert (Arsuaga *et al*. 1999), then the degree of basicranial flexion they exhibit should help resolve this issue. Despite thirty years of debate and many reinterpretations of the evidence, the precise nature of the linguistic and vocal ability of *Homo neanderthalensis* remains unresolved.

The hyoid bone and mandible: Although the flexing of the basicranium can indicate the position of the larynx in the throat, laryngeal position alone may not be enough to draw accurate conclusions about the extent of vocal ability possessed by a specimen. Fortunately, other physiological features can be considered. A rare find from Mount Carmel, Israel, of the intact hyoid bone of a hominid dated to 60 kya has allowed more detailed reconstructions of the vocal apparatus. This find was unique in its completeness (though Frayer & Nicolay 2000, make mention of a second, less complete one) until the recent find of three more hyoids at Atapuerca, Spain (Steven Mithen pers. comm.) which, sadly, are so far unpublished. The analysis of these is eagerly awaited, as it could shed important light on the vocal abilities of West European *Homo heidelbergensis* compared with those from Israel. Any differences may relate to the

https://doi.org/10.1017/S0959774302000100 Published online by Cambridge University Press

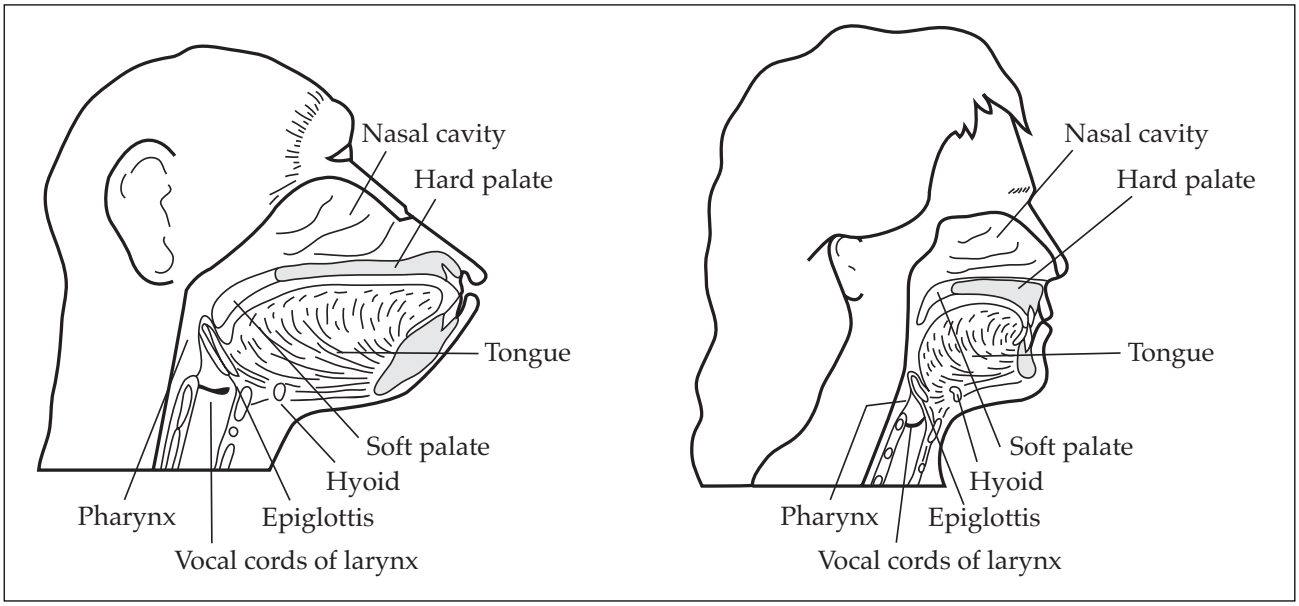


Figure 2. *Chimpanzee and human adult speech anatomy, showing position of larynx, pharynx, tongue and hyoid bone. (After Lieberman 1992, 135.)*

ancestry of these different *Homo heidelbergensis* to Neanderthals and modern humans.

The cranium of the Mount Carmel specimen (Kebara 2) does not survive, but the postcranial skeleton and age indicates that it is of a very robust Neanderthal-type hominid (Arensberg 1989). Arensberg and colleagues (1990) carried out a reconstruction of the vocal tract of the hominid using evidence from the muscular attachments on the hyoid, spine and mandible, and compared the hyoid bone with that of modern apes and humans. They also cite evidence (Solow 1966, cited in Arensberg *et al.* 1990) that in fossil hominids a lack of basicranial flexion is accompanied by a compensatory increase in maxillary width and posterior height. This has the consequence of there being no change in the overall nasopharyngeal volume as a result of the lack of flexion; thus the size of the supralaryngeal space is unaffected.

It would also seem that parallels drawn between the flat basicranium of apes and Neanderthals are mistaken. In apes and humans, the hyoid and larynx rest in the same position relative to the vertebrae of the neck, but owing to the forward posture (kyphosis of the vertebral column) of apes the hyoid and larynx are much nearer the cranial base. In hominids with a fully upright posture the hyoid and larynx, although in the same position relative to the vertebrae, rest further from the cranial base. Consequently, in fully upright hominids, the lack of basicranial flexion is of far less significance with regard to the supralaryngeal volume and range of vocalizations than it is in apes.

Arensberg and colleagues (1990) conclude that features of the mandible are a far better indicator than basicranial flexion of the position of the hyoid (and thus the larynx) and consequent vocal ability. The position of the muscular systems supporting the hyoid bone and larynx are indicated by the angle of a groove on the inside of the mandibles, called the *mylohyoid groove* (see Fig. 2 for the position of the hyoid bone). Evidence from Middle Palaeolithic hominids, including Neanderthals and early moderns (Kebara 2, Amud 1, Qafzeh 9, Skhul V, Tabun I and II, Regourdou and Krapina J and C) show that this is identical in position and orientation to that of modern humans 'suggesting a suite of muscular relations similar to modern humans must have been present' (Arensberg *et al*. 1990, 141). The conclusion drawn from all this is that, regardless of basicranial flexion, the hyoid and larynx of Neanderthals must have been low relative to the cranial base; they should thus have had supralaryngeal spaces approximately the same size as those of modern humans, and no significant difficulties in articulation.

The hyperglossal canal: Recently another physiological feature has been used as a source of information regarding the nature of the vocal soft tissue of human ancestors: the *hyperglossal canal* (Kay *et al.* 1998). The hyperglossal canal transmits the nerve that sup-

https://doi.org/10.1017/S0959774302000100 Published online by Cambridge University Press

plies the muscles of the tongue, and the authors hypothesize that the relative size of this canal (and accordingly the hyperglossal nerve that occupies it in life) may reflect the degree of motor co-ordination of the tongue present. They found that the relative size of the canal in *Australopithecus afarensis* and possibly *Homo habilis* (based on three skulls from Sterkfontein) is within the range of modern *Pan troglodytes* and *Gorilla gorilla*, whereas those of *Homo heidelbergensis* (Kabwe and Swanscombe), *Homo neanderthalensis* (La Chapelle-aux-Saints and La Ferrassie 1) and early *Homo sapiens* (Skhul 5) are well within the range of modern humans and significantly larger than *Pan troglodytes*. Given the estimated age of Swanscombe and Kabwe, Kay and colleagues suggest that a modern human pattern of tongue motor innervation and vocalization abilities had evolved by around 300,000 years ago, in *Homo heidelbergensis*, and was present in subsequent species including Neanderthals.

Another factor in the ability to produce controlled vocalizations is the fine control of the rib (intercostal) muscles responsible for breathing. Frayer & Nicolay (2000) cover this in some depth, with particular reference to the shape and muscular control of the ribcage of WT–15000, the 'Nariokotome Boy' *Homo ergaster* skeleton. Without reproducing the discussion in full here, their conclusion is that, *contra* MacLarnon (1993) and Walker (1993), there is no reason to believe that *Homo ergaster* did not have a modern ability to control its breathing and vocalization. This is in line with their interpretation of the evidence for other physiological features too, as they conclude from their analyses that by 1.5 mya, with *Homo ergaster*, 'both the articulatory capacity to form vowels and the respiratory capacity to maintain high-volume airflow were present' (Frayer & Nicolay 2000, 233). Whilst the latter may well be the case, they base their conclusions regarding articulatory capacity on the evidence from the hyoid bone and basicranial flexion. The sole hyoid evidence comes from the Kebara specimen, dating to 60 kya (see above), so conclusions based on this cannot be extended to *Homo ergaster*. The basicranium, whilst showing the first evidence of flexion in *Homo ergaster*, does not achieve modern proportions until *Homo heidelbergensis*, 400–300 kya (see above). It should be possible to attribute to *Homo ergaster* some articulatory versatility, but greater value in developmental terms is gained by identifying that it was with *Homo ergaster* that the first major steps were made towards modern vocalization abilities.

In summary, the debate concerning Neanderthal physiology notwithstanding, all the above research reaches a consensus with regard to the earliest occurrence of modern-like vocal physiology and ability. It would seem that from *Homo ergaster* (*c.* 1.73 mya) to *Homo heidelbergensis* (300–400 kya) there was a gradual co-evolution of the various features of the vocal apparatus, allowing a progressively greater range of sounds to be produced. By archaic *Homo sapiens*, or *Homo heidelbergensis*, a set of vocal apparatus that was essentially fully modern was in place.

It is worth noting at this stage that the potential individual cost of the lowering of the larynx in the throat (namely, that the individual is increasingly subject to choking on liquid) is considerable; for the progressive lowering of the larynx to have occurred over the course of hominid evolution, the selective advantage must have outweighed the potential cost.

Sound production: percussive sound and rhythm

In terms of percussive sound production we know already, from the evidence of stone-tool manufacture (e.g. Wynn 1993; Johanson & Edgar 1996), that very early humans had the co-ordination and dexterity necessary to strike one item with another precisely. This would result, intentionally or unintentionally, in percussive sound. Recent preliminary research by Cross (1999b) has investigated the potential use of lithics themselves for intentional sound-making. Even without co-ordination to knock two objects together, percussive sequences could be created by foot-stamping, an activity physiologically possible long before the advent of the genus *Homo*. The question remaining is at what point *Homo* could have had the mental capacity for the forethought and planning that would have been necessary to create these sounds in deliberate complex sequences necessary to produce a metre or pulse.

The other question is at what point *Homo* would have actually had the *incentive* to produce percussive rhythm for its own sake (rather than as a by-product of tool use). Several potential evolutionary rationales for the use of music and rhythm have been proposed (Wallin *et al.* 2000; Morley 2000). McNeill (1995) reports how the use of rhythm alone may be significant in inducing group cohesiveness. He observes that when people move together in dance (or military drill), this elicits a physiological arousal in the participants, and a great sense of emotional solidarity. He calls this 'muscular bonding', and proposes that such effects would have important selective benefits to any hominid group that practised them. On the basis that all surviving human cultures make use of music and dance, he sug-

https://doi.org/10.1017/S0959774302000100 Published online by Cambridge University Press

gests that those groups that did not do it, did not survive. Those that had practices of moving together rhythmically developed stronger communities and became better at co-operating in other spheres, particularly hunting.

Whilst this is speculative, the principle of 'muscular bonding' has a strong foundation. It is used in military circumstances the world over, and ethnographic evidence for the action of dance in evoking group solidarity is very widespread (e.g. Nettl 1992; Locke 1996; Johnston 1989). Further, the very close association of the areas of the brain responsible for complex vocalizations and complex sequences of muscular movements, which seem to have evolved in tandem (see below), may go some way to explaining the association of rhythm, dance and music in almost all cultures.

Sound perception: the ear and its evolution

The production and perception of sound are inextricably linked. In most mammalian species the frequency range of their vocalizations (mainly determined by their laryngeal system) largely coincides with that of their auditory system (Wind 1990). A number of important changes appear to have occurred during the evolution of the human auditory system relating specifically to its use for perceiving vocalizations. It would seem that sometimes the development of the auditory system has been spurred by developments in the vocalization capabilities, at other points vice versa.

Liberman *et al*. (1967) identified certain *articulatory invariants* across all human cultures, regardless of language. Articulatory invariants are the sounds produced by certain mechanisms of articulation that are human universals (Liberman *et al*. 1967; Wind 1990); because the articulatory mechanisms are universal, so are the sounds they produce. This is not to say that the *meaning* attributed to these sounds is universal, only the sounds themselves. The conventional view has been that the auditory system in humans is a special perceptual mechanism which evolved to detect these articulatory invariants.

Laryngeal and respiratory differentiation in mammals made more variation in pitch and amplitude possible, as well as more frequent vocalizations, while increasing auditory sophistication enabled a greater variety of sounds to be perceived and analyzed. As Wind puts it, 'Mammalian vocalizations are mainly used for intraspecific communication. Therefore auditory function must have been shaped by selective pressures gearing them to vocalization capacities' (Wind 1990, 185).

Kuhl (1988), however, suggests that the evolutionary process may have been the reverse of this; that the acoustical properties of speech were selected for as a consequence of the particular invariant properties that are detected by the auditory system. Kuhl found that non-human animals and pre-linguistic infants (neither of whom are capable of producing articulate speech) show a particular tendency to perceive speech sounds. She found that the categories of sound perceived most effectively (which she calls 'natural auditory categories': Kuhl 1988) conform to the phonetic categories of speech. This set of phonetic sounds is universal across all human languages. Despite the great variability of the sounds that it is possible to produce with the human vocal tract, a very limited set of acoustic forms are used in the world's languages.

That these 'natural auditory categories' are most effectively perceived by non-human mammals and pre-linguistic humans suggests that the mechanisms relied on for the perception of speech sounds is a very ancient one, pre-dating humans by millions of years. As hominids developed the ability to control their vocalizations in order to communicate, there would have been strong selective pressure to use those sounds that are most easily perceived by others. So it would seem that audition, specifically the existence of 'natural auditory categories', was initially responsible for the formation of speech categories, and that the mechanisms for perceiving such categories were in place in our hominid ancestors before they were capable of actually producing articulate speech. This has important implications for the debate concerning Neanderthals' information-processing abilities with respect to their supposed vocal abilities (*contra* Lieberman 1989).

No matter in which order the major evolutionary developments occurred, it appears that the vocal and auditory systems must have evolved in response to each others' capabilities. A further piece of evidence indicating that the inner-ear systems are intimately linked with the development of vocalization is that the stapedius inner-ear muscles of humans (which regulate intensity of the operation of the stapes bones on the eardrum) involuntarily contract before the onset of vocalization. This reduces the intensity of the sound of our own voice and prevents our own speech from masking other sounds from the surrounding environment, allowing us to hear other sounds even while we speak (Borg & Counter 1989). It suggests that the auditory system, over the course of its evolution in human ancestors, also faced significant selective pressure as a consequence of

https://doi.org/10.1017/S0959774302000100 Published online by Cambridge University Press

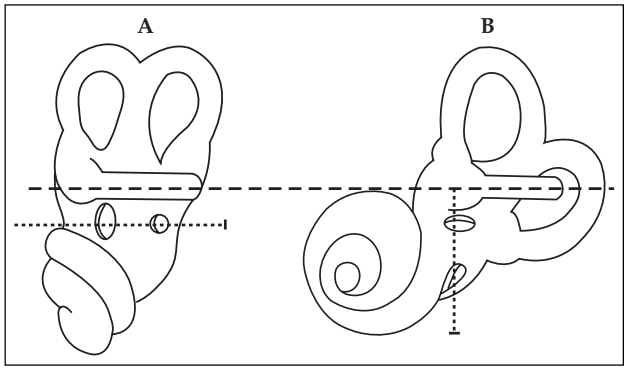


Figure 3. *The left labyrinths of the inner ear of (A) a quadraped and (B)* Homo sapiens*. The horizontal plane is represented by the dashed line. The whole arrangement has rotated through 90 degrees as a consequence of bipedalism. (After Daniel 1990, 260.)*

our vocalization abilities.

Another major factor in the development of the human auditory system was bipedalism (Daniel 1990; Spoor *et al*. 1994). The labyrinthine structure of the inner ear is important for the regulation of balance, as well as hearing. Daniel (1990) and Spoor and colleagues (1994) link the anatomy of this structure with bipedal posture. In all quadrupedal mammals, the labyrinthine capsule of the inner ear (see Fig. 3) is situated vertically, whereas in anatomically modern humans it is rotated through almost 90 degrees (Daniel 1990). Fortunately, the bony labyrinth is often preserved inside fossil hominid skulls. This has allowed Spoor and colleagues (1994) to carry out detailed computer reconstructions of the labyrinthine system of various fossil hominids, based on x-rays of the skulls. They compared the rotation and dimensions of the labyrinths of the fossils with those of modern humans and extant great apes. The labyrinths of *Australopithecus* and *Paranthropus* closely resemble those of modern great apes, whereas fully modern morphology is first exhibited by *Homo ergaster* (SK 847, *c.* 1.5 mya), implying that the Australopithecines and Paranthropines were only partially bipedal, whereas *Homo ergaster* had full bipedalism.

Daniel relates this rotation to various other neurological changes. In humans, rotation of the labyrinth from vertical to horizontal occurs ontogenically too, during the development from foetus to adult. This reinforces the idea that bipedal locomotion is a major cause of the rotation (Daniel 1990). It would seem that these adjustments to the labyrinthine system as a consequence of bipedalism had important neurological knock-on effects on language and learning. The vestibular neurological system associated with the labyrinth is also phylogenically and ontogenically the first to exhibit lateralization tendencies. Daniel (1990) proposes from this, and clinical connections between vestibular dysfunction and language and learning deficits, that the foundations of linguistic functions are first evidenced by *Homo ergaster*. According to Spoor and colleagues (1994), *Homo ergaster* (1.5 mya) is the first hominid to have modern anatomy in this respect, providing further evidence that *Homo ergaster* is the earliest hominid to exhibit signs of vocal and linguistic ability.

Ultimately, the conclusion is that regardless of when language actually emerged in humans, the neurological predisposition for it (and for other complex vocalizations such as singing) arose with bipedalism and the associated labyrinthine-vestibular changes at least 1.5 mya in *Homo ergaster*.

Neurology

Having the physiological mechanisms does not necessarily imply the cognitive capacity to use them for the production or perception of rhythm and melody. It is important to examine the evidence provided by neurological studies and music psychology to identify the areas of the brain responsible for the production and perception of vocalizations, rhythm and melody. It may then be possible to identify evidence of the evolution and development of these areas in palaeoanthropological evidence and comparative studies.

Locations of the production and processing of music and language

As singing and speech make use of the same physiological mechanisms for their production and perception (namely, the vocal chords and ear) one might expect that the two are closely linked in evolutionary terms and the neurological mechanisms they use. Neurologically, however, the parallels between music and language function are less clear-cut. Research relating musical functions (and deficits) to neurology is reviewed comprehensively by Marin & Perry (1999), and earlier by Borchgrevink (1982). Such research has a long history (almost half of Marin & Perry's 290 references pre-date 1980) and this body of research continues to grow. Dealing as it does predominantly with the western musical tradition, much of this research is concerned with the ability to play instruments, to name pieces of music and to read and write notation. Here we are concerned with musical behaviours at a more fundamental level,

https://doi.org/10.1017/S0959774302000100 Published online by Cambridge University Press

however, that pre-date such elements of musicality as written notation and the naming of tunes. The neuropsychological literature may allow us to relate this information to such evidence as we have of neurological development in the fossil record. We may also attempt to identify interrelationships and differences between the neurology responsible for musical and linguistic functions, in attempting to propose a path for their development. Consequently, of greatest interest are those investigations exploring prosodic, tonal, rhythmic and lexical-semantic comprehension and production in music and language.

Vocalizations vs verbalizations: evidence from pathology

As Borchgrevink (1982) and Sloboda (1985) observe, conventional belief for many years was that language function is controlled almost exclusively by the left hemisphere, and music function by the right. Detailed neurological studies using a variety of techniques, however, have shown that this division is not accurate. Observations of brain pathologies which result in aphasias (loss of language production/processing abilities) and/or amusias (loss of music production/processing abilities) are illustrated; some functions are shared and some are lateralized (Borchgrevink 1982; Schweiger 1985; Marin & Perry 1999). The process of identifying the precise locations of neurological areas implicated in different tasks is complicated by the fact that pathologies are rarely neatly localized, and cerebral anaesthetization is very difficult to achieve selectively. Scanning techniques provide some insight, but are complicated by the necessity of 'factoring out' other cerebral processes that are inevitably occurring at the same time as those under observation.

Schweiger (1985) describes several cases of lesions or strokes in both hemispheres in musicians and composers; surprisingly, there appears to be little consistency in terms of which musical functions suffer as a consequence of lesions in a given hemisphere (performance, composition or recognition, for example). In some cases amusia exists without aphasia, sometimes vice versa and sometimes both together. For example, Broca's aphasia, which can result in severe disruption of semantic speech performance, can leave affective-tonal vocal performance quite intact. Marin & Perry (1999) describe one of their own clinical cases of Broca's aphasia, which resulted in severe linguistic expressive disorder, but no impairment of the performance of non-lyrical vocal melodies, highlighting the distinction between vocal control and linguistic function.

Certain sub-functions of music and language seem to be shared, whereas functional lateralization seems to obtain in others (e.g. Borchgrevink 1982; Schweiger 1985; Marin & Perry 1999). As Marin & Perry (1999) observe, where both speech and music are affected by avocalias (impairment of vocalization ability), this does 'suggest the possibility of a common neural substrate at a lower level in the hierarchy of vocal expressive functions' (Marin & Perry 1999, 665). They go on to comment, however, that 'dissociations in performance between verbal and melodic vocal tasks give support for a contralateral localization of crucial neural processing at higher levels' (Marin & Perry 1999, 665).

An important distinction to note here is the difference between the terms 'vocal' and 'verbal', and 'speech' and 'language'; *verbal* necessarily implies linguistic lexical content, whereas *vocal* does not, encompassing non-lexical utterances too. This distinction between *vocal* and *verbal* is an important one, made appropriately by Marin & Perry, but often neglected in other literature where the terms have sometimes been used interchangeably. Likewise, for the purposes of this discussion, *language* will be used to refer specifically only to lexical-semantic elements of *speech*, which also encompasses tonal elements of communication.

Marin & Perry (1999, 665) go on to state that examples of aphasia without amusia 'decisively contradict the hypothesis that language and music share common neural substrates'. They are using the term 'language' to refer only to verbal, *lexical* language functions, and not the affective tonal elements of speech. Such cases of aphasia without amusia only indicate that music and the *lexical* elements of speech do not share common neural substrates.

Marin & Perry (1999) again refer to one of their own observed cases, a patient with severe bilateral cortical atrophy. The patient had no spontaneous language at all, and an inability to sing lyrics. Her spontaneous vocalizations, however, contained appropriate sound contours expressing emotion and intention, consisting of iterative sounds in sentence-like sequences, demonstrating that the prosodic element of speech was intact; preserved along with this was perfect rhythm, intonation and prosody in singing, lacking only lyrical content. That aphasics often continue to be able to produce and comprehend the *prosodic* components of language, its emotional tonal content, confirms the distinction between the linguistic and prosodic elements of speech function.

The areas of the brain responsible for controlling affective (emotional) processing seem to be

https://doi.org/10.1017/S0959774302000100 Published online by Cambridge University Press

evolutionarily ancient. Karow *et al*. (2001) compared the abilities of cortical brain-damaged subjects with or without concomitant sub-cortical basal ganglia damage to process affective speech prosody, emotional facial expression and linguistically-based information. Those subjects with only cortical damage performed all tasks without significant difficulty, irrespective of the location of the lesion in the right or left hemisphere. Amongst those with sub-cortical damage, however, there was a distinct lateralizational bias in their abilities: those with left-hemisphere sub-cortical damage had difficulty processing the linguistic information, whilst those with right-hemisphere sub-cortical damage had the greatest difficulty processing prosodic-emotional information and facial expression. As well as reinforcing the distinction between these elements of vocal content (linguistic-semantic and affective), the integration of neurologically ancient sub-cortical systems into the processing of emotive-vocal information suggests that the evolutionary roots for these abilities are ancient themselves.

The comprehension and production of the prosodic elements of language often exist intact with the comprehension and production of the prosodic elements of music, and when one of these functions is damaged, so too is the other (Mazzucchi *et al*. 1982; Peretz 1993; Peretz *et al*. 1994; Patel *et al*. 1998). This indicates that there *are* shared common substrates between elements of speech and musical functions, namely, those concerned with emotive expression and comprehension. Snow goes even further, with regard specifically to the intonation of utterances, concluding that 'linguistic analysis suggests that even intonation patterns traditionally described as nonemotional have their underpinnings in the speaker's emotions' (Snow 2000, 1). He goes on to observe that pathological studies 'indicate that intonation patterns described as either linguistic or emotional are mediated by right-hemisphere substrate specialized for emotional experience' (Snow 2000, 1). This suggests, then, that whatever the content of speech, its production, specifically its intonation, unavoidably involves some input from neurological systems concerned with affective function.

Lesions isolating speech areas in the left hemisphere can also result in the condition of echolalia, whereby the patient is able faithfully to repeat any linguistic utterance they have heard, but have no understanding of the linguistic content of the utterance (Geschwind *et al.* 1965). That the phonological loop (allowing repetition of heard utterances) can function in isolation from the linguistic cognitive processes indicates that the systems controlling complex vocalization and language comprehension are distinct. It appears that the neurology responsible for language comprehension and production is quite distinct from that responsible for carrying out complex vocalizations and, furthermore, from that responsible for the emotive prosodic content of such complex vocalizations, be it in speech or in song.

So, the evidence from pathologies resulting in aphasias and amusias suggests that we have a common neurology for musical and linguistic vocalizations at a 'lower' emotive prosodic level, and contralateral localization of neural processing at 'higher' linguistic semantic levels. This would suggest that the specialized human neurology dedicated to the higher linguistic and musical functions emerged later out of a set of common neurological substrates concerned with vocal emotional expression and comprehension. Indeed, following their review of a large body of research, Marin & Perry (1999) propose that:

> The close correspondence between the networks of regions involved in singing and speaking suggests that speech may have evolved from an already-complex system for the voluntary control of vocalization. Their divergences suggest that the later evolving aspects of these two uniquely human abilities are essentially hemispheric specializations (Marin & Perry 1999, 692).

Prosody vs rhythm: evidence from brain activity

These specializations are further informed by research involving observations of brain activity. Buchanan *et al*. (2000) used functional magnetic resonance imaging (fMRI) to observe brain activity during tasks involving discrimination of the emotional tone and the phonemic characteristics of heard spoken words. They found that in both the phoneme analysis and the emotional tone analysis tasks, significant bilateral activity was observed in comparison to baseline activity, but that between the two tasks, there was hemispheric specialization. The task that involved analysis of emotional characteristics resulted in significant activity in the right inferior frontal lobe, whereas the phonemic characteristic analysis led to greater activation in the left inferior frontal lobe. Again this provides evidence of shared neurology for some elements of both linguistic and tonal-affective content analysis, other elements of those functions being undertaken by left and right cerebral hemisphere specializations.

Some specifics of these specializations are suggested by Benson (1985), who observed PET (positron emission tomography) scans of the brain activity

https://doi.org/10.1017/S0959774302000100 Published online by Cambridge University Press

of subjects while they carried out various tasks relating to the production and perception of speech. His results show that the left hemisphere appears to be dominant for semantic verbal meaning and syntactic sequencing and relationships, whilst the right hemisphere is dominant in prosodic melody and semantic visual images (Benson 1985, table 3, 201). For the other aspects of gestural, prosodic, semantic and syntactic language, both hemispheres participate.

This accords well with the findings of Borchgrevink (1980; 1982) on rhythm and melody production and perception. Borchgrevink selectively anaesthetized either the right or the left hemisphere of his subjects by administering a local anaesthetic (sodium Amytal) via the right and left carotid artery. Anaesthetization of the right hemisphere resulted in loss of control of pitch, whilst left-hemisphere anaesthetization arrested both speech and singing (Borchgrevink 1980). Subsequently he observed the performance of the subjects at various tasks relating to the production and perception of rhythm and melody, with one or other hemisphere anaesthetized. He found that the right hemisphere seemed to be responsible for controlling pitch and tonality of singing (cf. prosodic melody in speech, above), whilst musical rhythm appeared to be controlled by the left hemisphere (cf. syntactic sequencing, above) (Borchgrevink 1982). These links between the systems of linguistic and musical prosody and the systems of linguistic and musical rhythm are further reinforced by Peretz *et al.* (1994), whose patients with bilateral lesions of the superior temporal cortex exhibited simultaneously impaired perception of prosody in speech and perception of melody in music, whilst having preserved perception of the rhythmic elements of speech and music.

This general distinction was also found in later experiments by Borchgrevink (1991), in the majority of subjects. Some individuals have speech control bilaterally represented, but in the vast majority of cases examined (88 per cent) it is unilaterally represented. In these cases, regardless of the laterality of the speech function, pitch control was always represented in the right hemisphere. Where speech control was bilateral, pitch control was more likely to be, with 7 per cent of all the subjects showing bilateral pitch control.

So there appears to be a consistency in the location of pitch control and processing in the brain. Zatorre (1984) notes, however, that both hemispheres are required to be functional for normal performance of singing when encompassing all elements of performance rather than solely pitch control.

Analysis of sound

The degree of involvement of the left hemisphere in the processing of tones seems to be dependent upon the level of analysis required in the processing, with greater left hemisphere involvement when greater analysis of the sound is required. Peretz & Morais (1980; 1983) report that non-musicians show right-ear (left hemisphere) advantage when required *analytically* to process dichotic tones, and a left-ear (right hemisphere) advantage when required to *holistically* to process the tones. Shanon (1982) and Schweiger & Maltzman (1985) also found that the degree of tendency for right-ear/left-hemisphere dominance depended upon the complexity of the analytic (vs holistic) task; greater analytical demands require greater left-hemisphere participation.

When no conscious processing of the sound is required by the situation, it seems that there is little left-hemisphere involvement. Reporting on PET scan data during dichotic tests, Bogen (1985) suggests that there may be a link between human voice processing and the processing of other sounds with a rich timbre. Melodies without much timbre or harmonic content (played on a wooden recorder) showed little if any left-ear advantage; in contrast, an organ chord produced a distinct left-ear advantage. Bogen (and van Lancker & Canter 1982) observes that voice recognition is detrimentally affected by right-lesion damage, and proposes that the advantage for the right hemisphere in timbre processing might be related to voice recognition. That human voice recognition requires little or none of the analytical input of the left hemisphere is interesting, as it implies that the mechanisms for the recognition of voices are quite separate from those of linguistic analysis and comprehension. The mechanisms for voice recognition are almost certainly far older evolutionarily than those for linguistic processing; the fact that timbre-rich musical sounds are processed exclusively by these mechanisms could suggest that musical processing also predates linguistic processing, *or at least that the processing of tonal content predates the processing of semantic content*. This conclusion is given extra depth by findings by Watt & Ash (1998) that music has an action on the mind similar to that of interaction with a person.

Distinctions between the processing required for linguistic, musical and environmental sounds are also observed as a result of neural pathologies. In cases of auditory agnosia (disruption of the perception of speech, animal noises, music and other sounds), when the brain lesion is in the left hemisphere only, the perception of non-speech sounds

https://doi.org/10.1017/S0959774302000100 Published online by Cambridge University Press

(including music) tends to be largely intact, whilst speech perception is disrupted (Marin & Perry 1999). In contrast, there are also cases of unilateral right-hemisphere lesions in which normal speech perception is preserved but the perception of non-verbal sounds (including music) is disrupted (Fujii *et al*. 1990). In the latter case, and in generalized auditory agnosia (usually the result of bilateral damage), the perception of timbre and complex sounds, as well as rhythm and temporal information in the stimuli, is often severely disrupted (Marin & Perry 1999).

As far as localization of these functions is concerned, defects in music perception, discrimination of pitch between two tones and disturbance in singing seem to result from damage to the temporal regions of the right hemisphere. The temporal lobes also seem to be implicated in perception of tonality; Peretz (1993) describes two cases in which loss of sensitivity to tonality in melody resulted from bilateral temporal lobe stroke; perception of rhythm, speech and environmental sounds, however, was intact. Marin & Perry (1999) suggest that areas in the left hemisphere appear to be responsible for control of both speech and sung vocalization, whereas the right hemisphere appears to include areas responsible for the modulation of vocal pitch and the processing of pitch information.

In summary, it seems that both hemispheres are required in the production and processing of both music and language; the functions sometimes overlap, using the same neurological mechanisms, and sometimes they appear to be independent. Musical functions as a whole do not seem to be clearly localized, but tasks relating to pitch and pitch discrimination do seem to be right-hemisphere dominated. Linguistic functions seem to be most detrimentally affected by left-hemisphere lesions, whereas most musical functions seem to be impaired in some respect by damage to either hemisphere. Combinations of aphasia and amusia tend to result from left-hemisphere damage, whilst cases of sensory amusia without aphasia tend to be related to right-hemisphere lesions.

The locations of the production and processing of rhythm

The location and the development of the areas of the brain responsible for the production and processing of rhythm and metre, have also been considered in a number of important recent pieces of work. Petsche *et al*. (1991) report several pathological cases illustrating the importance of the left frontal hemisphere in rhythm perception and production. Their own work using EEG scans illustrated that many areas in both hemispheres are stimulated when processing rhythm, and that the two halves of the brain are functionally closely connected whilst doing so, more so than during any other sensory function (Petsche *et al*. 1991, 326).

Although both hemispheres are involved, the dominance of the left hemisphere for the ordering of temporal information, which must be fundamental to rhythm production and perception, is a consistent finding. Noffsinger (1985) also reports this amongst the functions of the left hemisphere in dichotic listening tests (where different information is presented simultaneously to both ears and the subject is required selectively to attend to one). He found that words, nonsense syllables, formant transitions, processing and ordering of temporal information and difficult rhythms generate a right-ear/left-hemisphere advantage. In contrast, pitch perception, musical chords, emotional sounds and hummed melodies produce a left-ear/right-hemisphere advantage (in normally lateralized subjects).

That melody and rhythm are not processed simultaneously in the brain by the same mechanisms (Noffsinger 1985; Peretz & Kolinsky 1993) would suggest that they have separate functional and behavioural roots and have come to be used together in musical performance subsequently.

Origins of rhythmic behaviour

Several components of human neurology seem to contribute to the development and execution of rhythmic functions, and recent theories as to the roots and development of co-ordinated rhythmic behaviour are considerably varied. Merker (1999) suggests that co-ordinated synchronous rhythmic and melodic group behaviour could be derived from synchronous chorusing activities carried out to attract female hominids from neighbouring territories. This is on the premise that, like chimpanzees, Lower and Middle Palaeolithic populations may have been exogamously organized (females leaving their natal population group and joining another where they settle to raise young). In this situation, males can make distance-calls to attract prospective mates from neighbouring groups. Merker points out that if this activity was carried out by several individuals at once, synchronizing their calls (as is the case with several insect species such as crickets and, visually, glow-worms), then the area reached by the sound increases proportionally. Having attracted females from neighbouring groups in this way, it would then be in the interest of individual members of the calling group to have an individual distinct calling ac-

https://doi.org/10.1017/S0959774302000100 Published online by Cambridge University Press

tivity to attract females to themselves rather than to the rival males in their group.

Distance-calls amongst higher primates in the wild are almost invariably accompanied by synchronized body movements such as branch-waving and beating of the ground. Merker proposes that such behaviours, in a group-synchronized form, were the root of musical behaviours in hominids. This is an interesting idea which in many respects makes sense. It remains the case, however, that there is no known precedent for this type of co-operative behaviour amongst primates, it is principally known amongst insects. This is not to say that there is any reason to believe that it *could not* have developed amongst primates; simply that no precedent is yet known to reinforce the idea. Furthermore, it implies two conflicting manifestations of the same behaviour, firstly co-operative, then competitive. Initially there would have to be a group selective advantage to carrying out the calling behaviour in synchrony, followed by an individual selective advantage in creative variation of the vocal behaviour to distinguish oneself from ones' rivals, according to Merker's proposal; i.e. the selective pressure for synchronous chorusing would by necessity impose an equal selective pressure for *asynchronous* calling, if this were indeed the behaviour used to woo prospective mates. Further evidence from primatology may prove useful in testing this idea.

As for neurological and developmental studies, there is much evidence to link rhythmic behaviours and motor and emotional systems. Trevarthen (1999) expounds upon the evidence for what he calls the human 'Intrinsic Motive Pulse', or IMP. Existing as the brain stem, basal ganglia and limbic structures of the Emotional Motor System, the Intrinsic Motor Formation (IMF) is responsible for integrating functions of attention, learning, and physiological actions of expression, including the synchronization and co-ordination of sequences of movements. The physiological and functional parameters of the IMF and IMP ('The body-moving rhythmic and emotionally modulated system is the Intrinsic Motive Pulse, or IMP': Trevarthen 1999, 160) are not clear from Trevarthen's exposition, but there does seem to be a close interrelationship between the emotional-controlling elements of the limbic system and the areas responsible for the co-ordination of motor sequences and posture. In Trevarthen's own words:

> The movement-creating reticular networks and nuclei are intricately combined with the neurochemical systems of emotion. The same activating neurones that select movements and control their energy and smoothness also cause changes in the emotions felt and the intensity and 'colour' of consciousness (Trevarthen 1999, 161).

It is common knowledge that our posture and movements can express a great deal about our emotional state, intentionally and unintentionally. Musicality and rhythmic movement, Trevarthen points out, involve deliberate control and sequencing of this system. This can result in a self-directed feedback from movement into emotional state and, importantly, feedback and interaction between individuals, in synchrony of movement and emotional state. This system is active automatically in all types of interpersonal interaction, but is deliberately used in musical activities. Trevarthen believes that this system has been able to reach its current state of versatility and expressiveness as a consequence of our bipedal posture, which has freed up large proportions of the body for spontaneous movement that would otherwise be occupied by locomotion.

That this co-ordination of emotional communication and of rhythmic behaviour is innate is evidenced by studies of infant–parent interactions. It seems that infants can accompany the emotive-tonal prosodic infant-directed speech known as 'motherese' (or more accurately, 'parentese'), or in other circumstances singing, with rhythmic body movements which co-ordinate with the rhythm of the vocalization (Trevarthen 1999). This can occur even in congenitally blind infants, who cannot possibly be imitating the movement of the parent.

Evidence for an inherited capacity for the perception of melody and rhythm

If a sense of rhythm and a sense of melody exist in neonates and infants, this would indicate that perception of these phenomena (inherent in music) is not simply a product of cultural influences, but instead has some basis in hereditary factors. Further, if pre-linguistic (and specifically, pre-lexical) infants can be shown to understand some content from rhythm and tone alone, this would provide some confirmation of the ability of non-verbal (rhythmic and melodic) vocal utterances to communicate important emotional information. Fernald (1989; 1992) showed that this is exactly the case. Four-month-olds heard utterances of approval and of prohibition in four languages, and displayed significantly different affective (emotional) response to the two types of utterances, regardless of language, and regardless of whether they had heard that language before. The test was also carried out with nonsense-syllables, with the same result, and in all instances intensity of

https://doi.org/10.1017/S0959774302000100 Published online by Cambridge University Press

utterance was controlled for, so that the only variations were in rhythm and tone. This inherited capacity to perceive melody and rhythm and to gain affective information from them does not necessarily mean that these capacities were selected for by music, but it does suggest that the ability to comprehend non-linguistic emotional utterances was selectively important.

Dissanayake (2000), like Papousek (1996), proposes that the musical perceptual abilities of infants were selected for because they are used in the non-verbal vocal communication between mother and infant. These utterances rely on tone, inflection and intonation, and have a wide vocal range; they communicate information about well-being, emotional status, and needs, and are expressive of approval and intention. Dissanayake hypothesizes that this mother–infant behaviour emerged as hominid infants were born progressively earlier, and thus more helpless, as brain size increased, particularly around the time of archaic *Homo sapiens* (*Homo heidelbergensis*). She proposes that these vocalizations then came to be adopted increasingly into daily life and ritual, eventually becoming the basis for musical behaviour, because they 'were found by evolving human groups to be emotionally affecting and functionally effective . . . to emotionally conjoin and enculturate the participants' (Dissanayake 2000, 401).

The benefit of such social-emotive interaction between individuals is well-argued, and the connection between musical behaviours and affective vocalization is an important one, but it is a little difficult to envisage hominids beginning to use a behaviour in other situations which was originally evolved as a 'baby-talk', because they realized it 'promoted affiliation and congruence in adult social life' (Dissanayake 2000, 401). An alternative, and perhaps more parsimonious, explanation might be that this was a form of communication that came to be used *throughout* the social group at a much earlier time. Originally both adult–adult and infant–adult, it is now perpetuated only in adult–infant interactions. Additionally, the fact that the vocal tract of *Homo heidelbergensis* was effectively fully modern (see p. 198) would suggest that selective pressure for its use for wide-ranged expression had come much earlier. This form of communication might gradually have evolved into music as Dissanayake suggests, but could also have been the basis for language amongst a whole population (cf. Aiello & Dunbar 1993; p. 210 below). It may be that the persistence of such emotive utterance for communication between (pre-linguistic) infants and adults today is a vestigial use of this earliest form of social communication; after all, it can still be selectively useful, at least until the child gains full linguistic abilities.

In summary, it seems likely that the production and processing of rhythm and metre in speech and of rhythm and timing in music use related neurological mechanisms, in the left hemisphere and in the reticular and emotional limbic systems. On the other hand, the prosody of speech, pitch perception, melodies and emotional sounds seem to be controlled and processed by neural substrates that are less lateralized, although much evidence suggests that important elements of these functions are predominantly controlled by the right hemisphere; these mechanisms probably predate those in the left hemisphere. The greater the analytical and temporal ordering demands of either music or language, the greater the involvement of the left hemisphere, and the greater the demands in terms of melody, tonal changes and prosody, the greater the involvement of the right hemisphere.

Pathologies resulting in aphasias and amusias suggest that we have a common neurology for musical and linguistic vocalizations at an emotive prosodic level, and contralateral localization of neural processing at linguistic semantic levels. This suggests that the specialized human neurology dedicated to the more analytical 'higher' linguistic and musical functions emerged later out of a set of common neurological substrates concerned with vocal emotional expression and comprehension. It could be that the emotional limbic system and the IMF proposed by Trevarthen (1999) constitute the uniting neurology between the tonal emotive vocalization elements and the rhythmic-motor elements of interaction and musicality.

Studies of infant capabilities and their development show that the capacity to perceive melody and rhythm and tonal vocalization and to gain affective information from them is inherited and innate at birth. It seems likely that these capabilities provide the foundations for the more specialized, culturally determined acquisition of specific language and musical abilities later.

Proto-music, proto-language and social vocalization

We have seen that human ancestors gradually evolved the apparatus necessary to make complex vocalizations over a period of around 1.5 million years, between *Homo ergaster* and *Homo heidelbergensis*. By 300,000–400,000 years ago, *Homo heidelbergensis* had a brain of 1220 cc (Petralona 1: Johanson & Edgar 1996, 200) and a vocal apparatus that was essentially

https://doi.org/10.1017/S0959774302000100 Published online by Cambridge University Press

modern. It is not generally suggested that *Homo* had modern language by 400,000 years ago, however, in terms of developed lexical complexity and syntax; in fact, there is little evidence for fully modern language until the appearance of cultural artefacts produced by anatomically modern humans (Mithen 1998). If they did not have modern language or symbolic thought, what were they using their large brains and fully-developed vocal tracts for? What was the selective pressure responsible that led to development of vocal ability perhaps 360,000 years before it was used for fully-developed language?

Vaneechoutte & Skoyles (1998) observe that modern language does not *need* to use all of the vocal tract (although it usually *does* use it all). With syntax and grammar to disambiguate meaning, one can speak in a monotone without fear that the meaning will be misunderstood. The utterance would have no emotive content through tone, but even this could be substituted, to a large degree, by careful choice of words. On the other hand, the use of the voice as an instrument — singing — *does* make use of the entire vocal tract. Vaneechoutte & Skoyles (1998) conclude that this suggests that the vocal tract evolved to the extent that it has in order to support (animal-like) song rather than speech. In some animals, song is used to declare territorial information, or readiness to mate. Amongst other animals (specifically, tropical song birds, whales, porpoises, wolves and gibbons), song has evolved as a means of establishing pair- and group-bonding. Vaneechoutte & Skoyles (1998) assert that this is an older requirement than linguistic speech and hypothesize that the ability to sing evolved in humans for the same reasons that it did in other animals. Consequently, they suggest that language now uses mechanisms originally selected for by song, concluding that 'the ability to sing provided the physical and neural respirational control that is now used by speech' (Vaneechoutte & Skoyles 1998, 1).

That the establishment of pair- and group-bonding is an older requirement than full linguistic speech is evident from many animal behaviours. It is important to note, however, that language also aids this behaviour. In fact, the social-bonding property is an important common feature of linguistic and musical behaviour. There is plenty of evidence for the importance, or even dominance, of the social function of language. Dunbar *et al*. (1997) and Emler (1992) found through analysis of natural conversations that 60–70 per cent of all conversation time is devoted to social information exchange.

Ethnographic evidence from a wide diversity of hunter-gatherer and other 'traditional' cultures parallels this finding for music; many of the most important roles of music amongst Plains Indians (Nettl 1992; McAllester 1996), African Pygmies (Kisliuk 1991; Turino 1992; Ichikawa 1999), Yupik and Inuit Eskimos (Nettiez 1983; Johnston 1989) and Australian Aborigines (Breen 1994; Myers 1999) are social, interactive and integrative, and the performers themselves often see these as the most important consequences of the activity. Blacking (1995) also observes, from his studies of the Venda peoples, that performance of a musical pattern 'may announce social situations, recall certain feelings and even reinforce social values' (Blacking 1995, 39). Blacking believed that social bonding was not the limit of the function of music, but that there is a relationship between music and social structure, with particular types and elements of music being used by different groups in an integrated culture.

There is thus an alternative, and possibly more parsimonious, explanation to Vaneechoutte and Skoyles' suggestion. Language can function today without full tonal use of the vocal chords because lexical complexity and syntax have removed ambiguity from our vocal utterances. This would not have been the case in a proto-language without modern lexical complexity and syntax (similar to that proposed, for example, by Sperber (1996), Aiello & Dunbar (1993), or Brown (2000); see below), and would not be applicable to melody. These *would* rely on the full use of the vocal tract.

It may be that the vocal tract developed to the extent that it did, not to support song before language, as Vaneechoutte & Skoyles (1998) suggest, but instead to support tone-dominated social-bonding utterances that were a precursor to both language and melodic music. In this case, it is only since complex lexicon and syntax have developed that language has been able to be used without recourse to the entire human vocal range.

Scherer, too, suggests that the earliest forms of language and music might have had emotive (affective) communication as a common foundation:

> . . . one might reasonably speculate that both proto-speech and proto-music might have used affect vocalizations as building blocks . . . It may well be, then, that the externalization of affect or emotion via vocalization is at the very basis of music and speech (Scherer 1991, 147).

The finding of Watt & Ash (1998), that music has an action on the mind similar to the action of interacting with a person, further complements the idea that melody in music could have been developed from

https://doi.org/10.1017/S0959774302000100 Published online by Cambridge University Press

the use of a social proto-language.

Thus the central issue, that the capacity for complex *vocalization* existed long before complex *language*, is not a problem if we accept the hypothesis that proto-language was based on tone-dominated emotional-social utterances. Aiello & Dunbar (1993; Dunbar 1998) propose that early linguistic utterances were first used to facilitate and then to replace social grooming. Such utterances may have been emotive vocalizations, and these would seem most likely to be effective in facilitating social grooming.

Amongst primates such as chimpanzees, grooming has a very important function in maintaining an individual's social network within a group; such a network is important in providing support in power, mate or food resource contests. Beyond a certain group size, however, it would become impossible to maintain a social network effectively through manual grooming alone, as the time taken would be too great (Dunbar 1998). Dunbar suggests that physical grooming began to be supplemented by utterances that increased its efficiency, and that as group size increased there would be a selective pressure for the 'grooming' utterances to become progressively more efficient and expressive. This could then feed back allowing group size to increase further, and so on.

Aiello & Dunbar (1993) also note that there is a proportional relationship amongst the higher primates and humans between degree of encephalization (cortical development) and group size; as group size increases, so does neocortical development. Treating this as a straight-line relationship, they calculated the probable group sizes of the various Palaeolithic hominids on the basis of their brain-size.

This proposed relationship between encephalization and group size is the possible weak link in Aiello & Dunbar's (1993) 'grooming' hypothesis. It may be the case that this relationship is not a straight line at all, and that it is impossible to accurately predict the group size of earlier hominids from it, in which case it would not be possible to assert at what point physical grooming would have become impractical. Future fossil evidence will add to the record of data regarding the group size of hominids; the 'pit of bones' site at Atapuerca, Spain, contains the remains of more than 32 individuals apparently part of the same community (Arsuaga 1997) so may eventually provide clues as to the group size of *Homo heidelbergensis*. In any case, the evidence that the use of language to establish and maintain social networks far outweighs its use for 'higher' discussion provides important support for the idea that sociality was one of the fundamental foundations of language.

In this case, users would have derived considerable selective benefit from the development of increasing vocal agility, because as the variety of sounds they could make increased, so too would the variety of sound sequences (and thus expressiveness). With greater expressiveness, social grooming would have become more efficient, and the individual's social network could become larger and more cohesive. An individual able to maintain a large cohesive social network would have considerable selective advantages over its peers.

In summary, it seems plausible that the earliest form of both language and music could have been non-verbal vocal emotive utterances which were made to increase the efficiency of physical grooming activities, and which gradually came to be used in broader spheres of behaviour. Exactly these utterances could also be used to accompany rhythmic percussion and dance, thus increasing the emotive response and social bonding experienced as a consequence of the activity. Emotive vocables are used in exactly this way by Plains Indians (Nettl 1989; 1992; McAllester 1996) and African Pygmies (Locke 1996).

A common ancestral 'musilanguage' has recently been proposed by Brown (2000), who suggests that music and language are 'reciprocal specializations of a dual-natured referential emotive communicative precursor, whereby music emphasizes sound as emotive meaning and language emphasizes sound as referential meaning' (Brown 2000, 271).

If the social hypothesis is correct, then human use of non-verbal song would seem to have more in common with the soothing and cooing sounds made by and to infants than with the birdsong, which is territorially or reproductively orientated rather than expressive and socially bonding (Slater 2000).

The evolution of the neurology

How far back in our evolutionary history do we have to go to see evidence of these systems? Several difficulties impede the making of inferences regarding the development of hominid neuroanatomy from the archaeological evidence. Not least of these is that there is no *direct* fossil evidence for the nature of early human brains, because brains do not fossilize.

Some light can be shed on the anatomy of early brains by the study of endocasts (e.g. Kochetkova 1978; Falk 1992). The extent to which one may extrapolate about brain function from brain physiology, however, is debated. First, an endocast only

https://doi.org/10.1017/S0959774302000100 Published online by Cambridge University Press

shows the outer surface of the brain, and can tell us nothing about the internal brain structure and little about connections between different areas (Leakey 1994). Endocasts can nonetheless show the *shape* of the brain, and this has allowed the identification of development in Broca's area (important in language production; Fig. 4) (Holloway 1983). This identification has led to the assertion that *Homo habilis* (skull KNM–ER 1470, from East Turkana) had some language abilities, as it shows evidence of development in that area of the brain (Holloway 1983).

There are some problems with this inference. Language ability is not located exclusively in Broca's area, but in many other areas too (e.g. Borchgrevink 1982; Benson 1985). Furthermore, areas around Broca's area are also used for processes other than language (Calvin 1996). Complex sequences of muscular movements and fine motor muscular coordination of oral and facial muscles are controlled in this region (Calvin 1996). These include the musculature necessary for production of facial expressions and vocalizations, as well as sequences of movements for rhythm production. Note that *vocalizations* are not the same thing as *language*. Thus development in Broca's area and the surrounding areas may only indicate the ability to produce complex sequences and finely-controlled vocalizations; it need have no bearing on language ability in the sense of syntax and grammar. Peterson *et al*. (1988) conclude from PET studies of the brain that the classic Broca's area (Brodmann's areas 44 and 45) is not concerned with semantic associations of words at all, but exclusively with motor functions. It is the nearby Brodmann's area 47 that is responsible for semantic associations. Uncertainty about the precise functions of Broca's area may have led to the attribution of language abilities to hominids who probably had none. The identification of Broca's area in endocasts is still very useful, as it gives important information regarding capabilities for vocalization and fine vocal control.

Before 2 mya, endocasts of fossil hominid brains show no development in Broca's area (Falk 1992). The first evidence for any growth in this area this comes from KNM-ER 1470, *Homo habilis* (Falk 1992). Between *Homo habilis* and *Homo ergaster* there is a disproportionate level of growth here, in comparison to the rest of the brain (Kochetkova 1978). While there is a general increase in brain size of around 50 per cent between *Homo habilis* OH7 (at around 674 cc: Johanson & Edgar 1996) and *Homo ergaster* KNM-WT 15000 (estimated adult cranial capacity 909 cc: Johanson & Edgar 1996), there is an 'especially intense' growth of the lateral tuba, corresponding to Broca's area (Kochetkova 1978, 200). This suggests that there was a significant development of fine motor control of vocalization as part of the speciation of these genera. By *Homo neanderthalensis* (Djebel Ihroud I, 1305 cc: Holloway 1981), this area is 'very protuberant' (Holloway 1981, 387).

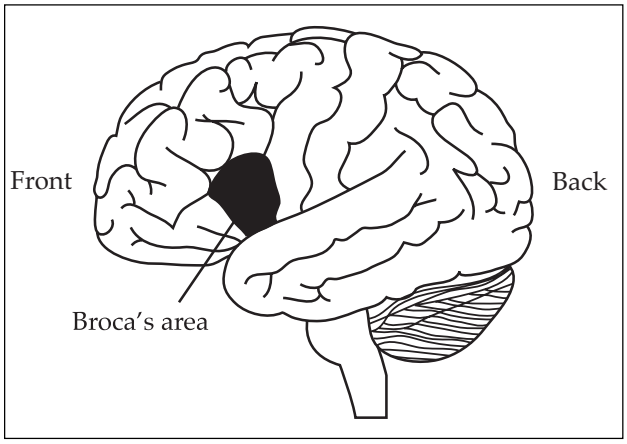


Figure 4. *The location of Broca's area in the human brain. (After Carlson 1994.)*

Other methods must be used to investigate connections between different areas of the brain, as these are unlikely to show up in endocasts. Neuropathological studies have identified the anterior limbic cortex as being responsible for the production of voluntary emotional vocal expressions in all primates (Jurgens 1992). Destruction of this area in humans does not affect the ability to produce involuntary vocal reactions to external stimuli but does result in an inability to *voluntarily* produce joyful exclamations, angry curses or pain outbursts. Without the anterior limbic cortex, voluntary vocal utterances which would normally be highly emotionally communicative sound more or less monotonous, with very flat intonation (Jurgens & von Cramon 1982, reported in Jurgens 1992). This is an evolutionarily ancient area of the brain, fulfilling the same function in Rhesus monkeys and probably in all primates since, with the same effects when destroyed (Aitken 1981; Kirzinger & Jurgens 1982, reported in Jurgens 1992). It is thus highly likely that the ability to perform voluntary emotional vocal expression was present in all members of the *Homo* genus and probably in all primates on the lineage between rhesus monkeys and *Homo*.

What separates us from other primates, however, is the vocal behaviour which involves voluntary control over the *structure* and *complexity* of vocal

https://doi.org/10.1017/S0959774302000100 Published online by Cambridge University Press

utterances. Jurgens reports that modern humans have a direct connection between the primary motor cortex (responsible for planned motor control) and the site of the laryngeal motoneurons, the nucleus ambiguus (Jurgens 1976, in Jurgens 1992). This is directly adjacent to Broca's area (Carlson 1994). The connection is not present in monkeys, nor probably in any non-human mammals (Jurgens 1992), and would seem to be what allows humans to carry out planned structured vocal utterances.

In summary, the first evidence for any development of the neurology required for fine, planned, vocal control is observed in *Homo habilis*, and this subsequently developed in prominence, disproportionately to the rest of the brain in *Homo ergaster*. There was a gradual increase in the variety of sounds that could be produced with the dropping of the larynx in the throat, from *Homo ergaster* until modern capabilities were reached by *Homo heidelbergensis*. It seems likely that these two phenomena are connected, and that there was a 'bootstrapping' (self-perpetuating) effect between them. As the vocal anatomy evolved, allowing an increasing range and diversity of utterances, the neurology also developed allowing increasingly complex planned vocal muscular sequences.

Conclusions

Homo ergaster, 1.75 mya, exhibits the first incidence of many changes in physiology and neurology important to the production and perception of music and language. This reports the first incidence of a fully modern aural anatomy and probably also the associated neurological labyrinth. There is also at this time the first major development in Broca's area, important for the planning and performance of complex sequences of fine muscular control, in particular of the vocal apparatus. This development would also have had significant benefits for the ability to plan and execute complex sequences of manual movements, important for rhythm production. The first evidence for development in the vocal apparatus itself is also displayed by *Homo ergaster*. Considerable concurrence in a wide variety of physiological and neurological evidence leads to the conclusion that we can, with some confidence, attribute modern vocal and perceptual versatility and complexity to *Homo heidelbergensis*, 400,000–300,000 years ago. This would have allowed it to produce complex planned vocal sequences with wide variation in pitch, and a high emotive content. The laryngeal, auditory and vestibular systems would seem to have been evolving in tandem since *Homo ergaster*, 1.75 mya.

It is likely that from *Homo ergaster* onwards, the ability to plan increasingly complex rhythmic and vocal sequences developed along with the development of the larynx. By *Homo heidelbergensis*, the larynx was essentially modern, with a full range of sounds, the production of which would have been controlled by areas of the brain responsible for voice recognition, emotional sounds, and pitch and melody functions. Given the potential selective costs of this development of the larynx (namely, choking), the selective *benefits* of the development of this laryngeal arrangement must have been very great.

Use of rhythm, music and emotive vocalizations seem to have important benefits for social bonding and group cohesiveness. Whether or not this is itself selectively beneficial, and it could convincingly be argued to be so if group selection theory is well-grounded, it also betrays a fundamental social-emotional element to music-use. Indeed, human musical vocalizations seem to have more in common with emotive communicative vocalizations than with territorial and mating song in other animals. Aiello & Dunbar (1993) propose social use as the primary selective pressure for the development of complex vocalizations and eventually full language, and there is much evidence to suggest that music has similar social and emotional bases.

In modern humans, music and language share functions lateralized in the brain in similar ways. The left hemisphere is implicated in 'higher' functions relating to syntactic sequences and relationships, the ordering of temporal information, difficult rhythms and semantic verbal meaning. The greater the analytical requirements of a task, the greater the involvement of the left hemisphere. The right hemisphere is more involved in functions relating to emotional sounds, voice recognition, prosodic melody of speech, pitch perception, melodies and chords. The greater the prosodic, harmonic or timbre content of the function, the greater the involvement of the right hemisphere. These functions are less strictly unilateral than those of the left hemisphere, and are almost certainly older than those of the left hemisphere. These similarities in the functions subserved by each hemisphere suggest common foundations for music and language functions, with those relating to voice recognition, emotional sounds, pitch and melody perception being evolutionarily older than those relating to the processing of semantic, lexical content, structure and verbal meaning.

Sense of rhythm and tonal perception seem to be innate, and are used in parent–infant vocal inter-

https://doi.org/10.1017/S0959774302000100 Published online by Cambridge University Press

actions (so-called 'motherese', or 'parentese') which relies on a wide tonal range and pitch changes, and has a high emotive content. This mode of communication has much in common with song; it is proposed here that it, and other non-linguistic emotive vocal utterances which we still use, may be the vestiges of a more widespread interpersonal communication which preceded lexical language, and was a common ancestor to both musical behaviours and modern language.

Acknowledgements

Sincere thanks are due to Ian Cross, Steven Mithen and Preston Miracle for advice and input at various stages in the production of this manuscript. Thanks also to the *Cambridge Archaeological Journal* referees for their comments and suggestions. The author's research is currently funded by the AHRB (2001–2003), and has hitherto been contributed to by a scholarship from Trinity Hall, Cambridge (2000–2001), a scholarship donation from the Wingate Foundation (summer 2001) and an AHRB scholarship (1999–2000).

Iain Morley
Department of Archaeology
University of Cambridge
Downing Street
Cambridge
CB2 3DZ
UK
Email: irm28@cam.ac.uk

References

Aiello, L.C. & R.I.M. Dunbar, 1993. Neocortex size, group size, and the evolution of language. *Current Anthropology* 34(2), 184–93.

Aitken, P.G., 1981. Cortical control of conditioned and spontaneous vocal behavior in rhesus-monkeys. *Brain and Language* 13(1), 171–84.

Arensberg, B., 1989. Anatomy of middle Palaeolithic populations in the middle east, in Mellars & Stringer (eds.), 165–71.

Arensberg, P., L.A. Schepartz, A.M. Tillier, B. van der Meersch & Y. Rak, 1990. A reappraisal of the anatomical basis for speech in middle Palaeolithic hominids. *American Journal of Physical Anthropology* 83, 137–46.

Arsuaga, J.L., 1997. Sima de los huesos (Sierra de Atapuerca, Spain): the site. *Journal of Human Evolution* 33, 109–27.

Arsuaga, J.L., I. Martinez, C. Lorenzo, A. Gracia, A. Muñoz, O. Alonso & J. Gallego, 1999. The human cranial remains from Gran Dolina lower Pleistocene site (Sierra de Atapuerca, Spain). *Journal of Human Evolution* 37, 431–57.

Barkow, J.H., L. Cosmides & J. Tooby (eds.), 1992. *The Adapted Mind*. Oxford: Oxford University Press.

Benson, D.F., 1985. Language in the left hemisphere, in Benson & Zaidel (eds.), 193–203.

Benson, D.F. & E. Zaidel (eds.), 1985. *The Dual Brain*. New York (NY): Guildford Press.

Blacking, J., 1995. *Music, Culture and Experience*. Chicago (IL): University of Chicago Press.

Bogen, J.E., 1985. The dual brain: some historical and methodological aspects, in Benson & Zaidel (eds.), 27–43.

Borchgrevink, H.M., 1980. Cerebral lateralisation of speech and singing after intracarotid Amytal injection, in *Aphasia: Assessment and Treatment*, eds. M. Taylor Sarno & O. Hooks. Stockholm: Almkvist & Wiksell, 186–91.

Borchgrevink, H.M., 1982. Prosody and musical rhythm are controlled by the speech hemisphere, in *Music, Mind and Brain*, ed. M. Clynes. New York (NY): Plenum Press, 151–7.

Borchgrevink, H.M., 1991. Prosody, musical rhythm, tone pitch and response initiation during Amytal hemisphere anaesthesia, in *Music, Language, Speech and Brain: Proceedings of an International Symposium at the Wenner-Gren Center, Stockholm, 1990*. Basingstoke: MacMillan Academic, 327–43.

Borg, E. & S.A. Counter, 1989. The middle-ear muscles. *Scientific American* 261(2) (August), 63–8.

Breen, M., 1994. I have a dreamtime: aboriginal music and black rights in Australia, in *World Music: the Rough Guide*, eds. S. Broughton, M. Ellingham & R. Trillo. London: Rough Guides Limited, 655–62.

Brown, S., 2000. The 'musilanguage' model of music evolution, in Wallin *et al.* (eds.), 271–300.

Buchanan, T.W., K. Lutz, S. Mirzazade, K. Specht, N.J. Shah, K. Zilles, & L. Jancke, 2000. Recognition of emotional prosody and verbal components of spoken language: an fMRI study. *Cognitive Brain Research* 9, 227–38.

Budil, I., 1994. A functional reconstruction of the supralaryngeal vocal tract of the fossil hominid from Petralona, in Wind *et al.* (eds.), vol. 3, 1–19.

Byrne, R., 1995. *The Thinking Ape: Evolutionary Origins of Intelligence*. Oxford: Oxford University Press.

Calvin, W.H., 1996. *How Brains Think*. London: Weidenfeld & Nicholson.

Carlson, N.R., 1994. *The Physiology of Behavior*. Boston (MA): Allyn & Bacon.

Carter, R., 1998. *Mapping the Mind*. London: Weidenfeld & Nicholson.

Cross, I., 1999a. Is music the most important thing we ever did? Music, development and evolution, in *Music, Mind and Science*, ed. Suk Won Yi. Seoul: Seoul National University Press, 10–39.

Cross, I., 1999b. Lithoacoustics – Music in Stone. Unpublished preliminary study reported on web page, http://www.mus.cam.ac.uk/~ic108/lithoacoustics/.

https://doi.org/10.1017/S0959774302000100 Published online by Cambridge University Press

Cross, I., in press. Music and biocultural evolution, in *The Cultural Study of Music: a Critical Introduction*, eds. M. Clayton, T. Herbert & R. Middleton.

Cross, I. & I. Morley, 2002. Music and evolution: the nature of the evidence, in *Proceedings of the 7th International Conference on Music Perception and Cognition*, eds. C. Stevens, D. Burnham, G. McPherson, E. Schubert & J. Renwick. Adelaide: Causal Productions, 416–19.

Daniel, H.J., 1990. The vestibular system and language evolution, in Wind *et al.* (eds.), vol. 1, 257–71.

Deacon, T., 1997. *The Symbolic Species*. London: Allen Lane.

Dennett, D.C., 1998. *Brainchildren: Essays on Designing Minds*. London: Penguin.

Dissanayake, E., 2000. Antecedents of the modern arts in early mother-infant interaction, in Wallin *et al.* (eds.), 389–410.

Duchin, L.E., 1990. The evolution of articulate speech: comparative anatomy of *Pan* and *Homo*. *Journal of Human Evolution* 19, 687–97.

Dunbar, R., 1998. Theory of mind and the evolution of language, in *Approaches to the Evolution of Language,* eds. J.R. Hurford, M. Studdert-Kennedy & C. Knight. Cambridge: Cambridge University Press, 92–110.

Dunbar, R., N. Duncan & A. Marriott, 1997. Human conversational behaviour. *Human Nature* 8, 231–46.

Emler, N., 1992. The truth about gossip. *Social Psychology Newsletter* 27, 23–37.

Falk, D., 1992. *Braindance: New Discoveries About Human Origins and Brain Evolution.* New York (NY): Henry Holt & Co.

Fernald, A., 1989. Emotion and meaning in mothers' speech to infants. Paper presented to The Society for Research in Child Development, Kansas City, April.

Fernald, A., 1992. Meaningful melodies in mothers' speech to infants, in Papousek *et al.* (eds.), 262–82.

Fidelholtz, J.L., 1991. On dating the origin of the modern form of language, in *Studies in Language Origins,* vol. 2, eds. W. Von Raffler-Engel, J. Wind & A. Jonker. Amsterdam: Benjamins, 99–113.

Frayer, D.W. & C. Nicolay, 2000. Fossil evidence for the origin of speech sounds, in Wallin *et al.* (eds.), 217–34.

Fujii, T., R. Fukatsu, S. Watabe, A. Ohnuma, T. Teramura, I. Kimura, S. Saso & K. Kogure, 1990. Auditory sound agnosia without aphasia following right temporal lobe lesion. *Cortex* 26, 263–8.

Geschwind, N., F. Quadfasel, & J. Segarra, 1965. Isolation of the speech area. *Neuropsychologia* 6, 327–40.

Heim, J.L., 1985. L'apport de l'ontogenése à la phylogénie des Néanderthaliens, in *Second International Congress Human Palaeontology.* Turin, 184.

Holloway, R., 1981. Volumetric and asymmetry determinations on recent hominid endocasts: Spy I and II, Djebel Ihroud I, and the Salè *Homo erectus* specimens, with some notes on Neanderthal brain size. *American Journal of Physical Anthropology* 55, 385–93.

Holloway, R., 1983. Human palaeontological evidence relevant to language behavior. *Human Neurobiology* 2, 105–14.

Ichikawa, M., 1999. The Mbuti of northern Congo, in Lee & Daly (eds.), 210–14.

Ingold, T. & K. Gibson, 1993. *Tools, Language and Cognition in Human Evolution*. Cambridge: Cambridge University Press.

Johanson, D. & B. Edgar, 1996. *From Lucy to Language*. London: Weidenfeld & Nicolson.

Johnston, T.F., 1989. Song categories and musical style of the Yupik Eskimo. *Anthropos* 84, 423–31.

Jurgens, U., 1992. On the neurobiology of vocal communication, in Papousek *et al.* (eds.), 31–42.

Karow, C.M., T.P. Marquardt & R.C. Marshall, 2001. Affective processing in left and right hemisphere brain-damaged subjects with and without subcortical involvement. *Aphasiology* 15, 715–29.

Kay, R.F., M. Cartmill & M. Balow, 1998. The hyperglossal canal and the origin of human vocal behaviour. *Proceedings of the National Academy of Science of the USA* 95, 5417–19.

Kisliuk, M., 1991. Confronting the Quintessential: Singing, Dancing and Everyday Life Among the Biaka Pygmies (Central African Republic). Unpublished PhD dissertation, New York University.

Kochetkova, V.I., 1978. *Paleoneurology*. Washington (DC): Winston & Sons.

Kuhl, P.K., 1988. Auditory perception and the evolution of speech. *Human Evolution* 3, 19–43.

Laitman, J.T., 1984. The anatomy of human speech. *Natural History* 93(8) (August), 20–27.

Laitman, J.T. & R.C. Heimbuch, 1982. The basicranium of Plio-Pleistocene hominids as an indicator of their upper respiratory systems. *American Journal of Physical Anthropology* 59, 323–44.

Laitman, J.T. & J.S. Reidenberg, 1988. Advances in understanding the relationship between the skull base and larynx with comments on the origins of speech. *Human Evolution* 3, 99–109.

Laitman, J.T., R.C. Heimbuch & E.S. Crelin, 1979. The basicranium of fossil hominids as an indicator of their upper respiratory systems. *American Journal of Physical Anthropology* 51, 13–34.

Lawson, G., C. Scarre, I. Cross & C. Hills, 1998. Mounds, megaliths, music and mind: some acoustical properties and purposes of archaeological spaces. *Archaeological Review from Cambridge* 15:1, 111–34.

Leakey, R., 1994. *The Origin of Humankind*. London: Weidenfeld & Nicolson.

Lee, R.B. & R. Daly (eds.), 1999. *The Cambridge Encyclopedia of Hunters and Gatherers*. Cambridge: Cambridge University Press.

Liberman, A.M., F.S. Cooper, D.P. Shankweiler & M. Studdert-Kennedy, 1967. Perception of the speech code. *Psychological Review* 74, 431–61.

Lieberman, P., 1984. *The Biology and Evolution of Language.* Cambridge (MA): Harvard University Press.

Lieberman, P., 1989. The origins of some aspects of human language and cognition, in Mellars & Stringer (eds.), 391–414.

Lieberman, P., 1992. Human speech and language, in *The*

https://doi.org/10.1017/S0959774302000100 Published online by Cambridge University Press

Cambridge Encyclopedia of Human Evolution, eds. S. Jones, R. Martin & D. Pilbeam. Cambridge: Cambridge University Press, 134–7.
Lieberman, P. & E.S. Crelin, 1971. On the speech of Neanderthal man. *Linguistic Enquiry* 11, 203–22.
Locke, D., 1996. Africa: Ewe, Mande, Dagbamba, Shona and BaAka, in Titon (ed.), 78–143.
McAllester, D.P., 1996. North America/Native America, in Titon (ed.), 16–66.
MacLarnon, A., 1993. The vertebral canal, in Walker & Leakey (eds.), 359–90.
McNeill, W.H., 1995. *Keeping Together in Time*. Cambridge (MA): Harvard University Press.
Magriples, U. & J.T. Laitman, 1987. Developmental change in the position of the fetal human larynx. *American Journal of Physical Anthropology* 72, 463–72.
Marin, O.S.M. & D.W. Perry, 1999. Neurological aspects of music perception and performance, in *The Psychology of Music*, ed. D. Deutsch. 2nd edition. New York (NY): Academic Press, 653–724.
Mazzucchi, A., C. Marchini, R. Budai & M. Parma, 1982. A case of receptive amusia with prominent timbre perception defect. *Journal of Neurology, Neurosurgery, and Psychiatry* 45, 644–7.
Mellars, P. & K. Gibson, 1996. *Modelling the Early Human Mind*. (McDonald Institute Monograph.) Cambridge: McDonald Institute for Archaeological Research.
Mellars, P. & C. Stringer (eds.), 1989. *The Human Revolution: Behavioural and Biological Perspectives in the Origins of Modern Humans*. Edinburgh: Edinburgh University Press.
Merker, B., 1999. Synchronous chorusing and the origins of music. *Musicae Scientiae* Special Issue, 59–73.
Mithen, S.J., 1996. *The Prehistory of the Mind*. London: Phoenix Press.
Mithen, S.J., 1998. A creative explosion?, in *Creativity in Human Evolution and Prehistory*, ed. S. Mithen. London: Routledge, 165–92.
Morley, I.R., 2000. The Origins and Evolution of the Human Capacity for Music. Unpublished MA Thesis, University of Reading. (Hard copy available at Reading; electronic copy available from author.)
Myers, F.R., 1999. Pintupi-speaking Aboriginals of the Western Desert, in Lee & Daly (eds.), 348–57.
Nettiez, J.J., 1983. Some aspects of Inuit vocal games. *Ethnomusicology* 27, 457–75.
Nettl, B., 1989. *Blackfoot Musical Thought: Comparative Perspectives*. Kent (OH): Kent State University Press.
Nettl, B., 1992. North American Indian music, in Nettl *et al.* (eds.), 260–77.
Nettl, B., C. Capwell, P. Bohlman, I. Wong & T. Turino (eds.), 1992. *Excursions in World Music*. Englewood Cliffs (NJ): Prentice Hall.
Noffsinger, D., 1985. Dichotic-listening techniques in the study of hemispheric asymmetries, in Benson & Zaidel (eds.), 127–41.
Papousek, H., U. Jurgens & M. Papousek (eds.), 1992. *Nonverbal Vocal Communication*. Cambridge: Cambridge University Press.
Papousek, M., 1996. Intuitive parenting: a hidden source of musical stimulation in infancy, in *Musical Beginnings: Origins and Development of Musical Competence*, eds. I. Deliege & J. Sloboda. Oxford: Oxford University Press, 88–114.
Patel, A.D., I. Peretz, M. Tramo & R. Labreque, 1998. Processing prosodic and musical patterns: a neuropsychological investigation. *Brain and Language* 61, 123–44.
Peretz, I., 1993. Auditory atonalia for melodies. *Cognitive Neuropsychology* 10, 21–56.
Peretz, I. & R. Kolinsky, 1993. Boundaries of seperability between melody and rhythm in music discrimination: a neurological perspective. *The Quarterly Journal of Experimental Psychology* 46A(2), 301–25.
Peretz, I. & J. Morais, 1980. Modes of processing melodies and ear asymmetry in non-musicians. *Neuropsychologia* 18, 477–89.
Peretz, I. & J. Morais, 1983. Task determinants of ear differences in melody processing. *Brain and Cognition* 2, 288–92.
Peretz, I., R. Kolinsky, M. Tramo, R. Labrecque, C. Hublet, G. DeMeurisse & S. Belleville, 1994. Functional dissociations following bilateral lesions of auditory-cortex. *Brain* 117, 1283–301.
Peterson, S.E., P.T. Fox, M.I. Posner, M. Mintum & M.E. Raichle, 1988. Positron Emission Tomographic studies of the cortical anatomy of single-word processing. *Nature* 331, 585–9.
Petsche, H., P. Rappelsberger, O. Filz & G.H. Gruber, 1991. EEG studies in the perception of simple and complex rhythms, in Sundberg *et al.* (eds.), 318–26.
Pinker, S., 1994. *The Language Instinct: How The Mind Creates Language*. New York (NY): William Morrow.
Pinker, S., 1997. *How The Mind Works*. London: Allen Lane.
Scherer, K.R., 1991. Emotion expression in speech and music, in Sundberg *et al.* (eds.), 146–56.
Schweiger, A., 1985. Harmony of the spheres and the hemispheres: the arts and hemispheric specialisation, in Benson & Zaidel (eds.), 359–73.
Schweiger, A. & I. Maltzman, 1985. Behavioural and electrodermal measures of lateralisation for music perception in musicians and nonmusicians. *Biological Psychology* 20, 129–45.
Scothern, P.M.T., 1992. The Music-Archaeology of the Palaeolithic within its Cultural Setting. Unpublished PhD thesis, University of Cambridge.
Shanon, B., 1982. Lateralisation effects in music decision tasks. *Neuropsychologia* 18, 21–31.
Slater, P.J.B., 2000. Birdsong repertoires: their origin and use, in Wallin *et al.* (eds.), 49–63.
Sloboda, J.A., 1985. *The Musical Mind.* Oxford: Oxford University Press.
Snow, D., 2000. The emotional basis of linguistic and nonlinguistic intonation: implications for hemispheric specialization. *Developmental Neuropsychology* 17, 1–28.
Solow, B., 1966. The pattern of craniofacial associations. *Acta Oolontologica Scandinavia* 24, Supplementum 46, 125.

https://doi.org/10.1017/S0959774302000100 Published online by Cambridge University Press

Sperber, D., 1996. *Explaining Culture*. Oxford: Blackwell.
Spoor, F., B. Wood & F. Zonneveld, 1994. Implications of hominid labyrinthine morphology for evolution of human bipedal locomotion. *Nature* 369, 645–8.
Sundberg, J., L. Nord & R. Carlson (eds.), 1991. *Music, Language, Speech and Brain*. Basingstoke: MacMillan Academic.
Titon, J.T. (ed.), 1996. *Worlds of Music: an Introduction to the Music of the World's People*. 3rd edition. New York (NY): Schirmer.
Trevarthen, C., 1999. Musicality and the intrinsic motive pulse: evidence from human psychobiology and infant communication. *Musicae Scientiae* Special Issue, 155–215.
Turino, T., 1992. The music of Sub-Saharan Africa, in Nettl *et al.* (eds.), 165–95.
Van Lancker, D.R. & G.J. Canter, 1982. Impairment of voice and face recognition in patients with hemispheric damage. *Brain and Cognition* 1, 185–95.
Vaneechoutte, M. & J.R. Skoyles, 1998. The memetic origin of language: modern humans as musical primates. *Journal of Memetics – Evolutionary Models of Information Transmission*, 2. Also available online at http://www.mmu.ac.uk/jom-emit/1998/vol2/vameechoutte_m&skoyles_jr.html
Walker, A., 1993. Perspective on the Nariokotome discovery, in Walker & Leakey (eds.), 411–32.
Walker, A. & R. Leakey (eds.), 1993. *The Nariokotome* Homo erectus *Skeleton*. Cambridge (MA): Harvard University Press.
Wallin, N., 1991. *Biomusicology: Neurophysiological, Neuropsychological and Evolutionary Perspectives on the Origins and Purpose of Music.* Stuyvesant (NY): Pendragon Press.
Wallin, N., B. Merker, & S. Brown (eds.), 2000. *The Origins of Music*. London: MIT Press.
Watt, R.J. & R.L. Ash, 1998. A psychological investigation of meaning in music. *Musicae Scientiae* 2(1), 33–54.
Wind, J., 1990. The evolutionary history of the human speech organs, in Wind *et al.* (eds.), vol. 1, 173–97.
Wind, J., A. Jonker, R. Allot & L. Rolfe (eds.), 1990. *Studies in Language Origins*. Amsterdam: Benjamins.
Wynn, T., 1993. Two developments in the mind of early *Homo*. *Journal of Anthropological Archaeology* 12, 299–322.
Zatorre, R.J., 1984. Musical perception and cerebral function: a critical review. *Musical Perception* 2, 196–221.

Author biography

Iain Morley is carrying out doctoral research into the origins and evolution of the human capacity for music in the Department of Archaeology at the University of Cambridge, and is a member of Trinity Hall College. He studied Psychology as an undergraduate, graduating in 1998, and completed his MA in Cognitive Evolution (incorporating Palaeolithic archaeology, linguistics and primatology) in 2000. Other research interests include the origins of religious thought and ritual behaviour, symbolism, creative thought, play and the process of neotenization in humans.

https://doi.org/10.1017/S0959774302000100 Published online by Cambridge University Press