

Part III

Perspectives on the evolutionary prerequisites for musical behaviour

OUP UNCORRECTED PROOF – FIRST PROOF, 26/12/2011, Spi

5

Hominid Physiological Evolution and the Emergence of Musical Capacities

Iain Morley

INTRODUCTION

The production and processing of the various elements of musical behaviours relies upon the integration of a number of physiological and neurological capacities. These have evolved into their current form in modern humans from foundations present in our hominid ancestors. Fundamental questions associated with this process concern the original roles of these component capacities, and how and why they developed and integrated in the way that they did, allowing possible musical behaviours of the form with which we are now familiar.

Examination of the fossil record of hominid physiological evolution, of neurological interdependencies, and of primate and developmental studies, can go some way towards addressing these questions, and give some insight into the essential underlying foundations of musical behaviours.

Before the presence of musical instruments in the archaeological record, discussion of the earliest foundations of musical behaviours in the human lineage must make recourse to the features of physiology and neurology that are used in such activities—one is necessarily investigating the origins of the production and processing of complex vocalizations and muscular movements. Vocalization is a fundamental component of musical behaviour, and the production and perception of long sequences of tonally varied and rhythmic sound sequences would be a prerequisite to any behaviour that we would

recognize as musical. This chapter focuses on the physiological (palaeoanthropological), neurological, developmental, and primatological evidence available regarding the emergence of such capabilities.

PHYSIOLOGICAL EVOLUTION—AN OVERVIEW OF PALAEOANTHROPOLOGICAL EVIDENCE

Although there are clearly difficulties with carrying out reconstructions and analyses of fossil specimens, these techniques are becoming increasingly refined, and the evolution of various elements of physiology related to vocalization can be tracked with reasonable confidence in the fossil record. The following constitutes a very brief summary of the findings of a variety of research related to the evolutionary development of these elements of hominid physiology (for a more detailed treatment the reader is directed to Morley 2002, and preferentially to Morley 2003).

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 of the position of the larynx near the top of the throat (Duchin 1990; 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 instead occupies a space in the pharynx further down the throat, opening the pharyngeal resonance cavity 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, but as the child grows up, particularly after the first year of life, the larynx descends towards its adult position. This process is only completed during adolescence (Magriples and Laitman 1987). It seems that a similar process has occurred over the course of human evolution.

The dimensions of the neural canal through the cervical and thoracic vertebrae have been taken as indicative of the level of innervation of musculature involved in the control of the larynx and breathing, respectively (MacLarnon and Hewitt 1999; Frayer and Nicolay 2000). Thoracic vertebral canal dimensions (indicating fine control of breathing and thus over duration and volume of utterances) in *Homo ergaster* are not increased relative to body size (in comparison with modern primates). However, cervical vertebral dimensions (indicating level of laryngeal muscular control over pitch, intonation, and emphasis, for example) are increased in *Homo ergaster*, equivalent to modern humans'. (MacLarnon and Hewitt 1999; Frayer and Nicolay 2000). Australopithecines display no increase in thoracic dimensions relative to body size, in comparison to the great apes.

It seems that *Homo ergaster* had some increased control over pitch, intensity, emphasis, and intonation, as indicated by basicranial flexion and cervical vertebral dimensions, but not over duration of utterances—that is, it would not have been capable of producing finely controlled long sequences of such tonal vocalizations. However, the latter capability appears to have been fully developed by the time of *Homo neanderthalensis* and early modern humans (MacLarnon and Hewitt 1999), who also have modern thoracic vertebral canal dimensions and hyoid bone morphology, suggesting modern levels of control over extended exhalations, and modern laryngeal anatomy, respectively.

This all suggests the subsequent development of the ability to produce extended sequences of controlled vocalizations out of an initial ability to make discrete vocalizations which were first controlled for pitch and tone, the foundations for which were already in place (as evident in higher primates). An initial increase in control over pitch, intensity, and intonation had probably occurred with *Homo ergaster*, followed by the development of control (and planning) of longer sequences of such vocalizations, probably in place in the common ancestor of *Homo neanderthalensis* and *Homo sapiens*.

In addition to the production of sounds, there are also various indicators of the emergence of the perception of sounds. There exists a set of 'natural auditory categories' (Kuhl 1988) or 'articulatory invariants' (Lieberman et al. 1967; Wind 1990) which are universal across all human languages, and these are perceived more effectively than other sounds, even by pre-linguistic infants and non-human mammals. This suggests that the auditory capacity to perceive these

Hominid Physiological Evolution and the Emergence 113

sounds predates the ability to produce them, and that the vocal capacity to produce these sounds (a fundamental component of all human vocalization) evolved in response to this perceptual ability.

On the other hand, aspects of audition have certainly been shaped by vocalization abilities. The stapedius inner-ear muscles regulate the intensity of our perception of our own voice, so that external sounds are not drowned-out by our own vocalizations (Borg and Counter 1989). This facility would presumably have become increasingly important as the range and duration of utterances increased, obscuring more environmental sounds.

Modern-like inner-ear morphology is also first exhibited by *Homo ergaster* (Spoor et al. 1994; Spoor 1996; Spoor and Zonneveld 1998). As with laryngeal anatomy, at least some of the changes in the features of the labyrinthine (inner-ear) anatomy appear to be correlated to changes in the shape of the basicranium. The cranial base accommodates the labyrinthine structure, and with the basicranium shortening and flexing with increased brain size, and the change in position of the foramen magnum (where the spinal cord enters the base of the skull) with full bipedalism, it seems this has had distorting effects upon the labyrinth. Again the first evidence of this is with *Homo ergaster*, and this process of flexion of the cranial base continued to some extent in all subsequent species of hominid up to *Homo heidelbergensis*, as elaborated above.

In summary, the foundations for the control of pitch and contour, as controlled by the cervical vertebral innervation, would have been present in *Homo ergaster*, but the duration of such utterances would have been limited until the time of *Homo heidelbergensis*, or possibly the common ancestor of Neanderthals and modern humans. However, it is with *Homo ergaster* that we see the first evidence for an increase in vocal versatility in the form of laryngeal development. This indicates the beginnings of an increase in tonal versatility and control before increased duration of utterances. It would seem likely that the requirement to control the pitch and contour of vocal sequences, and subsequently their duration, would have increased proportionally with the increase in range of pitch tones producible by the vocal tract as it evolved.

As MacLarnon and Hewitt (1999) point out, many primates vocalize in the form of discrete units of sound created with single air movements, but are limited in the duration of these and the order in which certain sounds can be made in the breathing cycle. They are

also limited in the diversity of such sounds that they can make. This again suggests the subsequent development of long sequences of controlled vocalizations out of an initial ability to make discrete vocalizations which were first controlled for pitch and tone, the foundations for which are already evident in higher primates.

THE FUNCTIONS AND LOCATIONS OF THE NEUROLOGY FOR MUSIC

The fossil record

Any discussion of the evolution of physiology would be incomplete without consideration of neurology, which is, of course, not only itself physiological, but is also intimately related to the control of other physiological capabilities. There are, however, several difficulties associated with making inferences regarding the development of hominid neuroanatomy from the palaeoanthropological 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. Nevertheless, the fossil record can provide some valuable information.

Some light can be shed on the anatomy of early brains by the study of endocasts. These are casts (naturally formed or man-made) of the inside of fossil skulls, which can show the shape and features of the brain which occupied them (e.g. Kochetkova 1978; Falk 1992). However, the extent to which one may extrapolate about brain function from brain physiology is debated. Firstly, an endocast only shows the outer surface of the brain, as it is a cast of the imprint left by the brain on the inner surface of the skull; it can tell us nothing about the internal brain structure and little about connections between different areas (Leakey 1994). Endocasts can, though, show the *shape* of the brain, and this has allowed the identification of development in Broca's area (important in language production) of hominid brains (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, however, with this inference. As discussed further below, language ability is not located exclusively in

Hominid Physiological Evolution and the Emergence 115

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). For example, complex sequences of muscular movements and fine motor muscular co-ordination of oral and facial muscles are controlled in this region (Calvin 1996). This includes musculature necessary for the 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 development of the ability to produce complex sequences and finely controlled vocalizations, and not have any bearing on language ability in the sense of syntax and grammar. In fact, 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 seems to be the nearby Brodmann's area 47 that is responsible for semantic associations. Despite the fact that 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 to this study, as it gives important information regarding capabilities for vocalization and fine vocal control.

Before about 2 Mya, fossil hominid brains show no development in the area of endocasts corresponding to 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 in this area, in comparison to the rest of the brain's growth (Kochetkova 1978). While there is a general increase in brain size of around 50 per cent between *Homo habilis* OH7 (at around 674cc; Johanson and Edgar 1996) and *Homo ergaster* KNM-WT 15000 (estimated adult cranial capacity 909 cc; Johanson and Edgar 1996), there is an 'especially intense' growth of the lateral tuba, corresponding to Broca's area (Kochetkova 1978, p. 200). This suggests that there was a significant development of fine motor control of vocalization as part of the speciation of these hominids. By *Homo neanderthalensis* (e.g. Djebel Ihroud I, 1,305 cc; Holloway 1981), this area is 'very protuberant' (Holloway 1981 p. 387).

Neurological foundations for production and perception of speech, melody, and rhythm

Other methods must be used to investigate connections between different areas of the brain, as these are unlikely to show up in endocasts. In addition to primatological and developmental studies, explored further below, clues relating to the evolutionary emergence of vocal complexity can be gained from studies of the interdependencies of neurological systems relating to such behaviours in modern humans.

Musical and linguistic vocalization capabilities seem to share neurological overlaps at a fundamental level concerned with the instigation and control of vocalization, but are dissociated at higher levels concerned with the specific production and processing of *content* in melody and speech (Marin and Perry 1999). The neurology responsible for complex vocal production is quite distinct from that responsible for lexical linguistic production and comprehension—pathologies show that it is quite possible to maintain full ability to produce complex linguistic sentences, whilst having lost all linguistic comprehension (Geschwind, Quadfasel, and Segarra 1965); that is, the *structure and rhythm of speech* can be perfectly preserved, along with *melodic contour and prosody*, whilst all lexical linguistic ability is lost.

Intonation patterns in speech, even if apparently non-emotional, seem to have their underpinnings in the speaker's emotions, and are mediated by right-hemisphere substrates specialized for emotional experience (Snow 2000). These substrates seem to be used in the comprehension and production of both the prosodic intonation of speech and the melodic elements of music (Mazzucchi et al. 1982; Peretz 1993; Peretz et al. 1994; Patel et al. 1998)—when one is damaged, so is the other, indicating common substrates for those elements of speech and musical functions concerned with prosodic intonation and melodic contour. This ability to process emotional information in vocalizations seems to be integrated with the ability to process emotional information in other media such as facial expression (Karow et al. 2001), which are evolutionarily ancient sub-cortical systems. They are separate from the ability to process linguistic content, and almost certainly far older. This suggests, then, that whatever the content of vocalization at 'higher' linguistic or melodic levels, its production and perception unavoidably involve some input from neurological systems concerned with affective function.

Hominid Physiological Evolution and the Emergence 117

Brain scanning studies reveal similar patterns of function distribution: lateral biases between hemispheres occur for various auditory processing tasks. Emotional tone analysis, whilst resulting in bilateral activation, involves significant activity in the right inferior frontal lobe, whereas phoneme analysis of words results in activation of the left inferior frontal lobe (Buchanan et al. 2000). Both involve bilateral activation and clearly share some neurology, but neurological specialization is also required. In the production and perception of speech sounds, prosodic melody and semantic visual images also rely on right-hemisphere activity, whereas semantic meaning and syntactic sequencing and relationships result in left-hemisphere activation. The right hemisphere also seems to be responsible for the control of pitch and tonality in singing, whilst musical rhythm seems to be controlled by the left hemisphere (Borchgrevink 1980, 1982, 1991).

Further reinforcement of the association between linguistic and rhythmic function, and the distinction between melodic and rhythmic function, comes from observations by Alcock et al. (2000), who studied a family with an inherited developmental speech and language disorder. The family members were tested on various tasks related to the production and perception of melody and rhythm. It was found that, whilst the affected family members were not deficient in the production or perception of pitch, be it in the form of melodies or single notes, they were deficient in the production and the perception of rhythm, both vocally and manually. Alcock et al. conclude that the oral and praxic defects of the family's condition cannot be at the root of the impairment in timing ability, but that the reverse must be the case; that is, rhythmic capacity forms an important component of oral/praxic ability. It seems that the capacity to perform planned sequences of complex muscular movements of rhythmic behaviour, both orally and manually, predate oral/praxic abilities, and that oral/praxic abilities are built upon muscular timing and sequencing capabilities.

So, rhythmic ability appears to form a foundation for linguistic ability, and rhythmic planning, execution, and processing; and linguistic semantic and structural processing seem to be handled largely by left-hemisphere structures. Prosodic elements of speech, melody, pitch, and the emotional content of sound production and perception seem to be handled predominantly by related right-hemisphere structures. It also seems that analysis of sounds with a greater amount of timbral quality (such as organ chords or the human voice, as opposed, for example, to the flute) requires greater right-hemisphere involvement

than sounds with more neutral, or less rich, timbre. Human voice recognition also relies on right-hemisphere neurology, and little or no input from left-hemisphere analytical structures; it may be that voice recognition and timbre processing rely on related or overlapping structures (Bogen, 1985). The mechanisms for voice recognition are almost certainly evolutionarily far older 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 and Ash (1998) that music seems to have an action on the mind similar to the action of interacting with a person. The above are just a few of many studies from pathologies, fMRI, PET, EEG and other scans to demonstrate overlaps between prosodic, melodic, tonal, and emotional sound content production and perception on the one hand, and structural, rhythmic, syntactic, and sequencing content production and perception on the other (Noffsinger 1985; Marin and Perry 1999; Fujii et al. 1990).

In summary, the right hemisphere appears to be responsible for processing and production, in both melody and speech vocalization, of prosodic melody, pitch control, tonality of singing, timbre processing, and voice recognition. The left hemisphere appears to be implicated in production and processing of semantic verbal meaning and syntactic sequences, as well as rhythmic and analytical functions. Thus, both hemispheres are involved in the production and processing of both music and language; some of the fundamental elements of music and language production and perception are shared, using the same neurological mechanisms, and some have subsequently become specialized. Musical functions as a whole are less clearly lateralized than language functions, 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; most musical functions seem to be impaired in some respect by damage to either hemisphere.

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 higher linguistic and musical functions emerged later out of a set of

Hominid Physiological Evolution and the Emergence 119

common neurological substrates concerned with (complex) vocal emotional expression and comprehension. Indeed, following their review of a large body of research, Marin and Perry (1999) propose that ‘The close correspondence between the networks of regions involved in singing and [linguistic] speaking suggests that [linguistic] 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 specialisations’ (p. 692).

PHYLOGENY AND INNATENESS OF CAPACITIES
UNDERLYING MUSICAL BEHAVIOURS

Use of rhythm and music both seem to have important benefits with regard to interpersonal social bonding and group cohesiveness. Whether or not this is itself selectively beneficial, it betrays a fundamental social-emotional element to music-use. Indeed, human musical vocalizations seem to have far more in common with emotive communicative vocalizations than with territorial and mating song in other animals (Slater 2000).

Primate evidence for neurological structures

Studies of primate vocalization neurology can inform as to the phylogenetic origins of such neurological systems. Neuropathological studies have identified the anterior limbic cortex as being responsible for the production of voluntary emotional vocal expressions in all primates (Jürgens 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 use of the anterior limbic cortex, voluntary vocal utterances, which would normally be highly emotionally communicative, sound more or less monotonous, with very flat intonation (Jürgens and von Cramon 1982; Jürgens 1992). This is an evolutionarily ancient area of the brain, fulfilling the same function in rhesus monkeys (and probably at least all primate species

since), with the same effects when destroyed (Aitken 1981; Kirzinger and Jürgens 1982, reported in Jürgens 1992).

On the basis of many years researching the neurological roots of vocalization, Jürgens (1998) considers that ‘Squirrel monkey vocalization can be considered as a suitable model for the study in humans of the biological basis of non-verbal emotional vocal utterances, such as laughing, crying and groaning’ (p. 376), and much of his research has been carried out on these primates. Critical in the production of non-verbal vocal emotional utterances is the periaqueductal grey matter (PAG) of the mid-brain, which acts as a critical relay station for such utterances (Jürgens and Zwirner 1986). Interestingly, recent findings show that the PAG is only implicated in the origination of the limbic *non-verbal emotional* vocal utterances, and not *learned* utterances with a neocortical origin; a second neurological pathway not involving the PAG seems to be responsible for learned vocal patterns (Jürgens and Zwirner 1986), although these would also be integrated at the nucleus ambiguus where control of the vocal folds occurs (Vanderhorst et al. 2001).

Muscle control for both voiced ‘vowel’ (laryngeal) and voiceless ‘consonant’ (orofacial) vocal sounds seems to be generated in the lateral column of the mid-brain PAG (Davis et al. 1996). Laryngeal muscle control is thought to be mediated by a neurological pathway from the PAG via the nucleus retroambiguus to the nucleus ambiguus, where monosynaptic connections exist to the laryngeal motoneurons (Vanderhorst et al. 2001, on the basis of the study of rhesus monkeys). This pathway is critical to the production of vocalizations; as well as being responsible for integrating vocal fold control, the nucleus ambiguus is also crucial for expiratory control, orofacial muscular control, and overall control of the laryngeal system (Jürgens 1998), and is directly adjacent to Broca’s area (Carlson 1994). This alone indicates how closely interrelated these functions are. As Davis et al. 1996 propose, ‘the PAG is a crucial brain site for mammalian voice production, not only in the production of emotional or involuntary sounds, but also as a generator of specific respiratory and laryngeal motor patterns essential for human speech and song’ (p. 34).

So it seems that these foundations for fine laryngeal and orofacial control necessary for vocalization have a long phylogenetic history, and that these foundations have since been built upon during the Homo lineage. Jürgens (1992) reports that in contrast to the primates studied, modern humans also have a direct connection to the nucleus

Hominid Physiological Evolution and the Emergence 121

ambiguus (site of the laryngeal motoneurons) from the primary motor cortex (responsible for planned motor control) (Jürgens 1976, 1992). This connection is not present in monkeys, and probably not in any non-human mammals (Jürgens 1992), and would seem to be what allows humans to carry out planned, structured vocal utterances. Such an ability is not present in other higher primates. As Lieberman (1994) reiterates, although apes can produce many vowel sounds and phonetic features of human speech, these are 'bound' into stereotyped calls based on affect, and they are limited in their production of novel voluntary motoric sequences. The evidence considered in the previous section concerning cervical and thoracic vertebral innervation suggested that controlled vocalizations of extended duration became possible between *Homo ergaster* and the common ancestor of *Homo neanderthalensis* and *Homo sapiens*. It seems reasonable to suggest that the primary connection between the motor cortex and the nucleus ambiguus for planning such utterances developed at the same time.

It is thus highly likely that the ability to perform voluntary, emotional vocal expression, on the basis of external stimuli and internal affective state, was present in all members of the *Homo* genus, and probably in all primates on the lineage between rhesus monkeys and humans. What separates us from other primates, however, is the vocal behaviour which involves voluntary control over the *structure* and *complexity* of vocal utterances, and the degree of the social functionality of these. Although monkeys have a separate pathway for controlling learned vocalizations, without the connection between the primary motor cortex and the laryngeal motoneurons in the nucleus ambiguus, other primates cannot show the human capacity for learning complex vocal patterns by imitation *and by invention*, and adapting them to novel situations.

Innate capacities for vocal and rhythmic complexity

So what vocal and perceptual abilities are inherent in humans at birth? 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 (1989a, 1989b 1992) showed that exactly this is 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 which 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 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.

It also seems that infants have an innate ability to accompany the emotive-tonal prosodic infant-directed speech (discussed in greater depth in the following section) 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 (Trevvarthen 1999). This can occur even in congenitally blind infants, who cannot possibly be imitating the movement of the parent.

As well as displaying sophisticated vocal-perceptual abilities soon after birth, human infants also undertake prolific vocal production at an early age. Elowson, Snowdon, and Lazaro-Perea (1998a) review and outline seven key features of human infant ‘babbling’ behaviour (a form of pre-linguistic vocal behaviour):

1. Babbling is universal and frequent irrespective of the infant’s cultural background.
2. Babbling is rhythmic and repetitive.
3. Babbling begins between six and ten months, peaking at seven months of age.
4. Babbling comprises a subset of the phonetic sounds found in adult speech.
5. Babbling has well-formed units with the consonant-vowel structure of adult speech.

Hominid Physiological Evolution and the Emergence 123

6. Babbling lacks apparent meaning with respect to how syllables are used by adults in language.
7. Babbling is instrumental in molding the caregiver-child bond in the infant's first year (p. 33, box 2).

Likewise, primate infants also display vocalization behaviours from an early age, and these may be valuable in informing us of the foundations for the way in which such behaviours develop, and the roots of their use. In those examples reported, primate infant vocalizations are dominated by social, interpersonal concerns. Following his study of trill vocalizations in white-faced capuchin monkeys *Cebus capucinus*, Gros-Louis (2002) observed that they seem to facilitate social interactions, and that such vocalizations were carried out to the greatest extent by immature and adult females. Infants trilled most when approaching other individuals, and those that did tended to interact with them in an affiliative way afterwards; Gros-Louis proposes that infants' trilling action may have an immediate effect on the listener's behaviour towards them, particularly with regard to such socially important activities as receiving grooming, touching, and inspecting food.

Similar findings were made by Elowson et al. (1998b), observing infant pygmy marmosets; response from a caregiving adult was far more likely to be given to an infant when vocalizing than when it was not. The only known study to date of non-human primate 'babbling' behaviour has identified considerable parallels with human infant babbling behaviour, on a number of counts (Elowson et al. 1998a).

In their long-term study of Amazonian pygmy marmosets *Cebuella pygmaea*, Elowson et al. (1998a) observed that '(1) both humans and pygmy marmosets have evolved a family-based social unit of cooperating individuals; (2) both humans and marmosets have an open, plastic system of communication where subtle changes of vocal production occur over the lifetime of the individual; and (3) both go through a period of babbling as part of their vocal development' (p. 36). In particular, 'the long, complex strings of vocalizations produced by the infant monkeys' (p. 32) have much in common with the babbling of human infants outlined above, such as 'universality, repetition, use of a subset of the adult vocal repertoire, recognizably adult-like vocal structure and lack of a clear vocal referent' (p. 31).

Future research will hopefully show whether this type of vocal development is exclusive to these particular primates, or whether

equivalent babbling behaviours are exhibited by other primate infants. It is conspicuous, however, that although repetitive, rhythmicity appears to be lacking from the vocalizations of the monkeys. What is especially interesting about this case is that the pygmy marmoset and other species in the *Callitrichidae* family are unique amongst primates in that they live in groups of extended family which are based on a sophisticated social system. These groups include 'in addition to infant twins, the two parents, older juvenile and subadult siblings, and often unrelated individuals . . . the analogy to human families is striking' (Elowson et al. 1998a, p. 32). In these groups, all individuals contribute to the care of the infants, in the form of carrying them, grooming them, and close physical contact in the form of huddling together at night and as protection against predators. It is possible that it is this form of social organization that leads to the early development and importance of such socially orientated vocal behaviour.

For Elowson et al. (1998a), the most interesting element of their findings is that it indicates that similar processes underlie (at least some) primate vocal learning to those that underlie human vocal learning, and they accordingly propose that pygmy marmoset babbling has a relevance to understanding the evolutionary processes of vocal development. In the context of the current research the most interesting implication is a phylogenetic one or an evolutionary one. If this is a form of early vocal behaviour and learning that is phylogenetically ancient and genuinely shared with other non-human primates, then it is interesting in that right alone, indicating that we might expect such behaviours to have been carried out by intervening ancestral hominid species. If, on the other hand, it is a type of behaviour that has emerged by convergence, it is equally interesting, as it could inform about the circumstances in which such behaviours are selectively useful to the individual and the group. In either case, these similarities indicate that lower primates have the cognitive capacities to develop and employ vocalizations in such a manner; that such vocalizations have an important social function at an early stage; and that there is no issue with the foundations of such behavioural capabilities having an ancient provenance. There is no reason for us to believe that any of our hominid ancestors could not have also exhibited such behaviours. Whether they would have encountered the selective pressure that shaped the ability to do so is a separate issue (see, for example, Morley 2003). If so, differences from species to

Hominid Physiological Evolution and the Emergence 125

species would presumably be represented in terms of the vocal range, control, and capabilities, which would ultimately bear upon the complexity of the utterances possible, and those most characteristically used.

Tonal and emotional content of vocalization and melody

In relation to the emotional state that they express, some vocal sounds and frequency changes are fundamental, invariant across cultures and even species (Scherer 1985, 1986; Morton 1977, 1994). Part of the reason for this is that facial expression has a fundamental influence on vocal quality, as orofacial musculature helps determine properties of vocalization such as frequency and vowel duration (Tartter 1980; Tartter and Braun 1994). Given the universality and innateness of certain fundamental facial expressions (e.g. Ekman and Friesen 1971; Ekman 1980; Carlson 1994), and the correspondence between these and characteristics of vocalizations, we can also expect characteristics of particular emotional vocalizations to be universal and innate too.

This correlation between facial expression and vocal quality also has an apparently ancient provenance, being shared by our nearest primate relatives, and similar correlations between vocal sound and emotional expression are also exhibited by several other species (Morton 1977; Falk 2004; Bermejo and Omedes 1999). Facial affect and vocal affect inform each other, and we employ and interpret them as signals inter-dependently in both production and perception (DeGelder and Vroomen 2000).

There is an additional factor that comes into play in the relationship between emotion, vocalization, and facial expression: the production of a particular facial expression whilst producing a particular vocalization results in some degree of feedback which actually affects emotional state (Levenson, Ekman, and Friesen 1990). Furthermore, there is a natural inclination to mimic such expressions (Kraut and Johnston 1979; Field et al. 1982) and consequently to feel such associated emotions. This physiological-emotional feedback may occur not only during production, but also during *perception* of such a stimulus (Wild, Erb, and Bartels 2001). This means that in producing, and even to an extent in perceiving, a particular sound, we generate some emotional response in ourselves due to the kinaesthetic feedback from the physiology required to produce that sound.

The universality of the vocal sounds and frequency changes that express particular emotions is especially evident in infant-directed (ID) speech, as the exaggeration of these elements of the vocalization is a characteristic feature of ID speech (e.g. Trainor et al. 2000; Fernald 1992b). This seems to be the case across cultures, and in both tonal and non-tonal languages (Greiser and Kuhl 1988; Fernald 1992b; Fernald 1993; Werker et al. 1994; Kitamura et al. 2002). ID vocalizations can tell us a great deal about the nature and role of the prosodic elements of speech and their relationship to melodic musical behaviour. It should be noted that the characteristic features of ID speech are also characteristics of the tonal (non-linguistic) elements of adult-directed (AD) speech, and they apparently share the same foundations and roles in emotional expression (Trainor et al. 2000).

Many of the properties of ID speech are shared with music. There are numerous parallels, in terms of: variable pitch contour; high rhythmicity; repetitive motifs; and the communication of affect, modulation of arousal, and eliciting of attention and affective response (e.g. Stern 1977; Fernald 1989a, 1989b, 1992b.; Fernald and Simon 1984; Papousek et al. 1991; Trainor et al. 1997; Trehub et al. 1993). Vocalizations of pre-school children in particular are often difficult to classify as either linguistic or musical (Mang 2000).

Dissanayake (2000) hypothesizes that this mother-infant behaviour emerged as a consequence of hominid infants being born progressively earlier as brain size increased, and that their thus being more helpless for longer after birth increased the need for emotional conjoinment at this early stage of life. This would, she suggests, have become particularly important at the time of archaic *Homo sapiens* (*Homo heidelbergensis*), which exhibited a large increase in brain size. 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 p. 401).

The benefits of such social-emotive interaction between individuals (emotional conjoinment and enculturation) is well argued and supported by much of the evidence cited above, and the connection between musical behaviours and affective vocalization is an important one (as evidenced by the neurological evidence considered previously). However, it may appear difficult to envisage hominids

Hominid Physiological Evolution and the Emergence 127

beginning to employ a behaviour which originally evolved as a ‘baby-talk’ in other situations in which they realized it ‘promoted affiliation and congruence in adult social life’ (Dissanayake, 2000 p. 401). As discussed previously, there are precedents for the importance of affective-communicative elements of such vocalizations in adult vocalizations, as well as in primates without language.

Additionally, the fact that the vocal tract of *Homo heidelbergensis* seems to have been effectively fully modern (see above) would suggest that selective pressure for its use for wide-ranged tonal expression had come much earlier than Dissanayake proposes. This form of communication might gradually have evolved into music as Dissanayake suggests, or at least provided shared foundations, but it could also have been the basis for language amongst all of a population (cf. Aiello and Dunbar 1993; Scherer 1991; Brown 2000). It may be that the use of this mode of emotive utterance for communication between (pre-linguistic) infants and adults today is a vestige of this earliest form of social communication; after all, it is potentially still selectively useful, at least until a child gains full linguistic ability, and is still inherent in the tonal elements of *any* emotional vocal communication (Trainor et al. 2000). Thus an alternative, and perhaps more parsimonious, explanation to Dissanayake’s suggestion is that this (social-emotive vocalization) was a form of communication that came to be used *throughout* the social group at a much earlier time, without preference, both adult-adult and infant-adult, but is now perpetuated, in this predominantly non-lexical form, in adult-infant interactions, and in the prosodic content of adult speech. Furthermore, the shared prosodic pitch- and tempo-related properties of emotional vocalization (ID and AD) and music are not borrowed from one to the other, in either direction, but are, and always have been, a shared fundamental component of both.

Music elicits emotional responses in listeners for a variety of reasons. In eliciting an emotional reaction, music can generate genuine physiological responses such as changes in respiration, heart-rate, and temperature; and tingling (Krumhansl 1997; Sloboda 1991 1998; Panksepp 1995; Panksepp and Bernatsky 2002). Some emotional responses can be elicited, for example, as a consequence of learned association with particular circumstances; others as a consequence of direct resemblance to ecological phenomena to which we have instinctive or conditioned responses (Sloboda 2001; Sloboda and Juslin 2001; Davies 2001; Cross 2003). A third form of emotional

response occurs as a consequence of processing by, and stimulation of, auditory and kinaesthetic mechanisms associated with interpersonal interaction. We can react to and process music in many of the same ways as we do interaction with a person (Kivy 1989; Davies 2001; Lavy 2001); it shares many of the properties of such interaction, and is apparently processed by the same mechanisms (Watt and Ash 1998). Consequently, we can (although don't necessarily) emotionally empathize with music, if it elicits the same emotion as is being expressed. The context in which we experience music is also very important in determining the emotion, and intensity of emotion, experienced (Sloboda and Juslin 2001). Especially important in this respect is the social context—the extent to which the experience is shared with others.

Vocalization, corporeal expression, and rhythm

Gesture and speech are interdependent. Both are affected simultaneously in stutterers (Mayberry and Jaques 2000) and gestural and vocal behaviours are interrelated from the earliest babbling in infants (Masataka 2000; Mayberry and Jaques 2000; Locke et al. 1995). They can operate independently, but when the upper limbs are otherwise unoccupied they can be sequestered into speech-related gesture. It seems that there is a cognitive rhythmic motor co-ordinator which instigates such muscular sequences irrespective of the musculature that is used, and that the complex patterns of muscular movement of gesture (finger, hand, arm, shoulder, and joint musculature) and vocalization (orofacial, laryngeal, and respiratory musculature) are co-ordinated (Franz et al. 1992; cf. Alcock et al. 2000). The concordance between gesture and speech is instigated early in the vocalization process, by cyclical motor control, with gestural movements being associated with the nuclei of tone groups—prosodic rhythm—rather than the lexical elements of speech (McClave 1994; Mayberry and Jaques 2000). This is also evident in the gestures accompanying infant vocalizations. Greater lateralization of gesture occurs after babbling begins, implying left-hemisphere development at this time in the integration of the two actions (Locke et al. 1995). These earliest gestures are emotive and rhythmic rather than iconic (Trevarthen 1999) (corresponding with the above finding that gesture corresponds with prosodic rhythm rather than lexical content); and

Hominid Physiological Evolution and the Emergence 129

only start to be used iconically and in combination with words when lexical behaviour has started to develop (Messinger and Fogel 1998; Butcher and Goldin-Meadow 2000). In the case of speech, (as opposed to other, non-linguistic, vocalizations), linguistic meaning and narrative sentence structure are integrated into the gesture-speech system after the integration of gesture with prosody, but before their physical manifestation (Mayberry and Jaques 2000).

Affective content can apparently be interpreted equivalently in visual, auditory and kinaesthetic media, each of which can represent tension, release and particular emotions, underlining the cross-modality of such affective expression and interpretation (Krumhansl and Schenck 1997). As with facial expression, it seems that this can be a two-way process too, with feedback occurring to some extent between the modes of experience (Mitchell and Gallaher 2001). Consequently, our posture and movements can express a great deal about our emotional state, intentionally and unintentionally, and others' body-posture and movements thus provide important cues as to their emotional state. As well as being able to observe such cues, we can empathically experience something of their emotional state in mirroring them with our own bodies. Musicality and rhythmic movement involve deliberate control and sequencing of this system (Trevvarthen 1999). There seems 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, driven by the Intrinsic Motor Formation, which consists of the brain stem, basal ganglia and the limbic structures of the Emotional Motor System. The same systems that select and control movements also cause changes in the emotion-controlling elements of the limbic system. This can result in a self-directed feedback from movement into emotional state and, importantly, feedback and interaction between individuals, in terms of synchrony of movement and of emotional state (Trevvarthen 1999; McNeill 1995).

Production of complex vocalization relies on priming of the whole motor system (Blount 1994). Particular non-lexical vocal utterances (and non-linguistic content of speech) are accompanied by quite specific involuntary body-language (Kappas et al. 1991; Wagner 1989; Lavy 2001); they share an affective origin in a system in which vocalization and corporeal expression, or to put it another way, vocal and corporeal gesture, are intimately linked (Blount 1994; cf. Clynes' (1977) 'sentic modulation' and Donald's (1991)

‘mimesis’). Note that this is at least equally applicable to body movements accompanying musical vocalization, and vocalization accompanying dance. There are not only shared neurological foundations between the ability to execute vocal and manual gestural sequences, but also a link between vocal and manual *rhythmic* capability. Humans are the only primates to carry out vocalizations with rhythmic content (Geissman 2000), and the capacity to perform rhythms, both manually and verbally, forms an important component of oral/praxic ability—detriments to one result in detriment to the other. This integration occurs whether utterances are linguistic or not (Alcock et al. 2000).

Whilst rhythm and melody are apparently produced and processed in neurologically specialized and distinct areas of the brain which are somewhat independent of each other (as illustrated by the neurological evidence discussed previously), there is also clearly important integration of these systems, with rhythmic muscular movements being co-ordinated with prosodic elements of vocalization in their production. Specifically, they are interdependent in the planning and execution of sequences of muscular movement associated with instigation of vocalization, rhythmic physiological movement, and expression of emotional state in these media. The production and perception of tonal *content* in vocalizations does not appear to require any input from rhythm-controlling systems in the left hemisphere (see above) but the planning and execution of sequences of such vocalizations does.

The role of this system in evolution

There is evidence that group size increased over the course of hominid evolution, and it has been suggested that complex vocalization and language became increasingly important as a consequence (Aiello and Dunbar 1993; Dunbar 1998; Kudo and Dunbar 2001). The role of complex vocalization of affect, and sensitivity to the prosodic content of vocalizations, has already been elaborated; there is much evidence to suggest that full language also remains important in this role (Dunbar et al. 1997; Emler 1992).

Melody and linguistic speech share common features in the form of intonational phrasing and combinatorial syntax, and share a ‘phonological level’ of meaning, an ‘acoustic mode’ involving emotive

Hominid Physiological Evolution and the Emergence 131

meaning and interpretation (Brown 2000). Whilst music often does use the full range of the vocal tract, full language does not need to in order to communicate effectively, as linguistic structures disambiguate meaning (e.g. in whispering; or in written text). This would not have been the case for a pre-syntactic, pre-lexical proto-language.

The fossil and neurological evidence discussed in the preceding sections suggests that the ability to produce short, tonally varied utterances emerged before the ability to produce extended sequences of vocalizations. It seems likely that discrete tonally-contoured units expressive of affective state would have become combined into progressively larger globally-contoured units. What must have emerged in the course of the evolution of pitched-contoured vocalizations is an *increased range and control* of pitch contour, allowing greater vocal versatility, expressiveness, and thus efficiency, in proto-linguistic vocal affective communication. This was coupled with the emergence of greater control over the duration of such utterances, and the ability to plan them. This need not have emerged as a new system of vocalization initially, however, instead building upon the limited pitch control already used for emotive-tonal-social vocalization by higher primates. The selective advantages associated with the possession of such capabilities resulted in the continual evolution of the vocal tract, and control over it, through the lowering of the larynx and increased innervation of the associated laryngeal and upper vocal tract musculature.

The lowering of the larynx allowed greater versatility of vocal tone to be produced (and lower formant frequencies) and also allowed greater control and movement of the tongue. It is suggested (Fitch 2000a) that the ability to *temporarily* lower the larynx in order to produce more complex vocalizations may have led to a preliminary permanent lowering of the larynx. This would subsequently have continued to lower, with concomitant increasing control over the sound-producing capabilities of the vocal tract and orofacial articulators.

Whilst extremely important in the type of information they express, non-verbal vocal utterances, in meaning, are limited to acting as expressions of personal state and relationships between the self and other. In terms of the emergence of language from such vocalizations, however, it seems likely that such utterances formed the first *vocalized* form of *conceptualizations* relating the self to object/other, and that this ability to vocalize such concepts was subsequently built upon by

more complex vocal and gestural sequences (see Armstrong et al. 1994). In fact, it is at that stage that the syntactic potential of the coupling of vocal and *iconic* gestural communication may have become fundamentally important, bridging the gap between non-lexical vocalizations of affect, and multi-part vocalizations with syntactic structure and iconic associations (Armstrong et al. 1994).

The socially important, emotionally communicative elements of such vocalization would likely have remained the dominant element initially, with iconic and then abstract lexical associations gradually increasing in importance. It is the latter—iconic and abstract lexical content—which relies on symbolic and analogical capacity, and which probably could not have emerged prior to *Homo sapiens*; the former do not rely on symbolic capacity, and, as illustrated, would have been selectively important at a much earlier time. Even with the emergence of full lexical and syntactic language, the social-affective communicative foundations of prosodic contour remained a fundamental element of vocal communication. It is these which formed the foundations of musical behaviours.

In summary, there is a large amount of diverse evidence which suggests that musical behaviours have as their foundation the vocalizations and gestures associated with the expression of emotion and maintenance of social bonds. Accounted for is the production and perception of such prosodically-contoured vocalizations, and the interdependence with them of corporeal and rhythmic expression of the same. Music is built on fundamentally important social and physiological mechanisms and, at an essential level, is processed as such. It uses capacities crucial in situations of social complexity: the vocal, facial, and interactive foundations of these capabilities are evident in other higher primates, and such capacities would have become increasingly important and sophisticated as group size and complexity of social relations increased. Vocal emotional expression, interaction, sensitivity to others' emotional state, and empathic ability would have been selectively important abilities; those individuals in which such capabilities were more developed would have been selectively fitter. Music draws upon these abilities as it uses emotional (vocal and physiological) stimuli as its foundation. Fundamentally integrated into the planning and control of complex sequences of vocalizations, and related to the prosodic rhythm inherent in such sequences, is rhythmic motor co-ordination. The motor system is primed in the instigation of such vocal behaviours, and corporeal

Hominid Physiological Evolution and the Emergence 133

gesture is consequently incorporated into the execution of the vocal behaviour.

REFERENCES

- Aiello, L. C. and R. I. M. Dunbar (1993). 'Neocortex Size, Group Size, and the Evolution of Language', *Current Anthropology*, 34: 184–93.
- Aitken, P. G. (1981). 'Cortical Control of Conditioned and Spontaneous Vocal Behavior in Rhesus Monkeys', *Brain and Language*, 13: 171–84.
- Alcock, K. J., R.E. Passingham, K. Watkins et al. (2000). 'Pitch and Timing Abilities in Inherited Speech and Language Impairment', *Brain and Language*, 75: 34–46.
- Arensberg, P., L. A. Schepartz, A. M. Tillier et al. (1990). 'A Reappraisal of the Anatomical Basis for Speech in Middle Palaeolithic Hominids', *American Journal Of Physical Anthropology*, 83: 137–46.
- Armstrong, D. F., W. C. Stokoe, S. E. Wilcox SE (1994). 'Signs of the Origins of Syntax', *Current Anthropology*, 35: 349–68.
- Benson, D. F. (1985). 'Language in the Left Hemisphere'. In: D. F. Benson and E. Zaidel (eds.) *The Dual Brain*, pp. 193–203. New York: Guildford Press.
- Bermejo, M. and A. Omedes (1999). 'Preliminary Vocal Repertoire and Vocal Communication of Wild Bonobos (*Pan paniscus*) at Lilungo (Democratic Republic of Congo)', *Folio Primatologica*, 70: 328–57.
- Blount, B. G. (1994). 'Comment on Armstrong, Stokoe and Wilcox', *Current Anthropology*, 35: 358–9.
- Bogen, J. E. (1985). 'The Dual Brain: Some Historical and Methodological Aspects'. In: D. F. Benson and E. Zaidel (eds.) *The Dual Brain*, pp. 27–43. New York: Guildford Press.
- Borchgrevink, H. M. (1980). 'Cerebral Lateralization of Speech and Singing after Intracarotid Amytal Injection'. In: M. Taylor Sarno and O. Hooks (eds.) *Aphasia: Assessment and Treatment*, pp. 186–91. Stockholm: Almqvist & Wiksell.
- (1982). 'Prosody and Musical Rhythm Are Controlled by the Speech Hemisphere'. In: M. Clynes (ed.) *Music, Mind and Brain*, pp. 151–7. New York: Plenum Press.
- (1991). 'Prosody, Musical Rhythm, Tone Pitch and Response Initiation during Amytal Hemisphere Anaesthesia'. In: J. Sundberg, L. Nord and R. Carlson (eds.) *Music, Language, Speech and Brain*, pp. 327–43. Basingstoke: MacMillan.

- Borg, E. and S. A. Counter (1989). 'The Middle-Ear Muscles', *Scientific American*, 261 2: 63–8.
- Brown, S. (2000). 'The 'Musilanguage' Model of Music Evolution'. In: N. L. Wallin, B. Merker, and S. Brown (eds.) *The Origins of Music*, pp. 271–300. Cambridge, MA: MIT Press.
- Buchanan, T. W., K. Lutz, S. Mirzazade et al. (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: J. Wind, A. Jonker, R. Allot et al. (eds.) *Studies In Language Origins*, iii, pp. 1–19. Amsterdam: Benjamins.
- Butcher, C. and S. Goldin-Meadow (2000). 'Gesture and the Transition from One- to Two-Word Speech: When Hand and Mouth Come Together'. In: D. McNeill (ed.) *Language and Gesture*, pp. 235–57. New York: Cambridge University Press.
- Calvin, W. H. (1996). *How Brains Think*. London: Weidenfeld and Nicholson.
- Carlson, N. R. (1994). *The Physiology of Behavior*. Boston, MA: Allyn and Bacon.
- Clynes, M. (1977). *Sentics: The Touch of Emotions*. New York: Souvenir Press.
- Cross, I. (2003). 'Music and Emotion'. [Online] Available: <<http://www.mus.cam.ac.uk/~ic108/IBMandS/emotion/musicemotion.html>> accessed November 2011.
- Davies, S. (2001). 'Philosophical Perspectives on Music's Expressiveness'. In: P. N. Juslin and J. A. Sloboda (eds.) *Music and Emotion: Theory and Research*, p. 23–44. Oxford: Oxford University Press.
- Davis, P. J., S. P. Zhang, A. Winkworth et al. (1996). 'Neural Control of Vocalization: Respiratory and Emotional Influences', *Journal of Voice*, 10: 23–38.
- DeGelder, B. and J. Vroomen (2000). 'The Perception of Emotions by Ear and by Eye', *Cognition and Emotion*, 14: 289–311.
- Dissanayake, E. (2000). 'Antecedents of the Modern Arts in Early Mother-Infant Interaction'. In: N. L. Wallin, B. Merker, and S. Brown (eds.) *The Origins of Music*, pp. 389–410. Cambridge, MA: MIT Press.
- Donald, M. (1991). *Origins of the Modern Mind: Three Stages in the Evolution of Culture and Cognition*. Cambridge, MA: Harvard University Press.
- 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: J. R. Hurford, M. Studdert-Kennedy, and C. Knight (eds.) *Approaches to the Evolution of Language*, pp. 92–110. Cambridge: Cambridge University Press.

Hominid Physiological Evolution and the Emergence 135

- N. Duncan, and A. Marriott (1997). 'Human Conversational Behaviour', *Human Nature*, 8: 231–46.
- Ekman, P. (1980). *The Face of Man: Expressions of Universal Emotions in a New Guinea Village*. New York: Garland SPTM Press.
- and W. V. Friesen (1971). 'Constants across Cultures in the Face and Emotion', *Journal of Personality and Social Psychology*, 17: 124–9.
- Elowson, A. M., C. T. Snowdon, and C. Lazaro-Perea (1998a). "'Babbling" and Social Context in Infant Monkeys: Parallels to Human Infants', *Trends in Cognitive Sciences*, 2: 31–7.
- (1998b). 'Infant "Babbling" in a Non-Human Primate: Complex Vocal Sequences with Repeated Call-Types', *Behaviour*, 135: 643–64.
- 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: Henry Holt.
- (2004). 'Prelinguistic Evolution in Early Hominins: Whence Motherese?' *Behavioural and Brain Sciences*, 27: 491–503.
- Fernald, A. (1989a). 'Emotion and Meaning in Mothers' Speech to Infants'. Paper presented to *The Society for Research in Child Development*, Kansas City, April 1989.
- (1989b). 'Intonation and Communicative Intent in Mothers' Speech to Infants: Is Melody the Message?' *Child Development*, 60: 1497–510.
- (1992). 'Meaningful Melodies in Mothers' Speech to Infants'. In: H. Papousek, U. Jürgens, and M. Papousek (eds.) *Nonverbal Vocal Communication: Comparative and Developmental Approaches*, pp. 262–82. Paris and Cambridge: Editions de la Maison des Sciences de l'Homme/Cambridge University Press.
- (1992b). 'Human Maternal Vocalizations to Infants as Biologically Relevant Signals: An Evolutionary Perspective'. In: J. Barkow, L. Cosmides, and J. Tooby (eds.) *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*, pp. 391–428. Oxford: Oxford University Press.
- (1993). 'Approval and Disapproval—Infant Responsiveness to Vocal Affect in Familiar and Unfamiliar Languages', *Child Development*, 64: 657–74.
- and T. Simon (1984). 'Expanded Intonation Contour in Mothers' Speech to Newborns', *Developmental Psychology*, 20: 104–13.
- Field, T., R. Woodson, R. Greenberg et al. (1982). 'Discrimination and Imitation of Facial Expression in Neonates', *Science*, 218: 179–81.
- Fitch, W. T. (2000a). 'The Phonetic Potential of Nonhuman Vocal Tracts: Comparative Cineradiographic Observations of Vocalising Animals', *Phonetica*, 57: 205–18.
- Franz, E. A., H. N. Zelaznik, and A. Smith (1992). 'Evidence of Common Timing Processes in the Control of Manual, Orofacial, and Speech Movements', *Journal of Motor Behavior*, 24: 281–7.

- Frayer, D. W. and C. Nicolay (2000). 'Fossil Evidence for the Origin of Speech Sounds'. In: N. L. Wallin, B. Merker, and S. Brown (eds.) *The Origins of Music*, pp. 217–34. Cambridge, MA: MIT Press.
- Fujii, T., R. Fukatsu, S. Watabe et al. (1990). 'Auditory Sound Agnosia without Aphasia Following Right Temporal Lobe Lesion', *Cortex*, 26: 263–8.
- Geissman, T. (2000). 'Gibbon Songs and Human Music in an Evolutionary Perspective'. In: N. L. Wallin, B. Merker, and S. Brown (eds.) *The Origins of Music*, pp. 102–23. Cambridge, MA: MIT Press.
- Geschwind, N., F. Quadfasel, and J. Segarra (1965). 'Isolation of the Speech Area', *Neuropsychologia*, 6: 327–40.
- Greiser, D. and P. Kuhl (1988). 'Maternal Speech to Infants in a Tonal Language: Support for Universal Prosodic Features in Motherese', *Developmental Psychology*, 24: 14–20.
- Gros-Louis, J. (2002). 'Contexts and Behavioural Correlates of Trill Vocalizations in Wild White-Faced Capuchin Monkeys (*Cebus capucinus*)', *American Journal of Primatology*, 57: 189–202.
- 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.
- (1983). 'Human Palaeontological Evidence Relevant to Language Behavior'. *Human Neurobiology*, 2: 105–14.
- Johanson, D. and B. Edgar (1996). *From Lucy to Language*. London: Weidenfeld and Nicolson.
- Jürgens, U. (1976). 'Reinforcing Concomitants of Electrically Elicited Vocalizations', *Experimental Brain Research*, 26, 2: 203–14.
- (1992). 'On the Neurobiology of Vocal Communication'. In: H. Papoušek, U. Jürgens and M. Papoušek (eds.) *Nonverbal Vocal Communication: Comparative and Developmental Approaches*, pp. 31–4. Paris and Cambridge: Editions de la Maison des Sciences de l'Homme/Cambridge University Press.
- (1998). 'Neuronal Control of Mammalian Vocalization, with Special Reference to the Squirrel Monkey', *Naturwissenschaften*, 85: 376–88.
- and D. Von Cramon (1982). 'On the Role of the Anterior Cingulate Cortex in Phonation—A Case Report', *Brain and Language*, 15: 234–48.
- and P. Zwirner (1986). 'The Role of the Periaqueductal Grey in Limbic and Neocortical Vocal Fold Control', *Neuroreport*, 7: 2921–3.
- Kappas, A., U. Hess, and K. R. Scherer (1991). 'Voice and Emotion'. In: R. S. Feldman and B. Rimé (eds.) *Fundamentals of Nonverbal Behavior: Studies in Emotion and Social Interaction*, pp. 200–38. Cambridge: Cambridge University Press.

Hominid Physiological Evolution and the Emergence 137

- Karow, C. M., T. P. Marquardt and R. C. Marshall (2001). 'Affective Processing in Left and Right Hemisphere Brain-Damaged Subjects with and without Subcortical Involvement', *Aphasiology*, 15: 715–29.
- Kirzinger, A. and U. Jürgens (1982) 'Cortical Lesion Effects and Vocalization in the Squirrel Monkey', *Brain Research*, 233: 299–315.
- Kitamura, C., C. Thanavisthuth, D. Burnham et al. (2002). 'Universality and Specificity in Infant-Directed Speech: Pitch Modifications as a Function of Infant Age and Sex in a Tonal and Non-Tonal Language', *Infant Behaviour and Development*, 24: 372–92.
- Kivy, P. (1989). *Sound Sentiment*. Philadelphia, PA: Temple University Press.
- Kochetkova, V. I. (1978). *Paleoneurology*. Washington, DC: Winston & Sons.
- Kraut, R. E. and R. Johnston (1979). 'Social and Emotional Messages of Smiling: An Ethological Approach', *Journal of Personality and Social Psychology*, 37: 1539–53.
- Krumhansl, C. L. (1997). 'An Exploratory Study of Musical Emotions and Psychophysiology', *Canadian Journal of Experimental Psychology*, 51: 336–52.
- and D. L. Schenck (1997). 'Can Dance Reflect the Structural and Expressive Qualities of Music? A Perceptual Experiment on Balanchine's Choreography of Mozart's *Divertimento* No. 15', *Musicae Scientiae*, 1: 63–85.
- Kudo, H. and R. I. M. Dunbar (2001). 'Neocortex Size and Social Network in Primates', *Animal Behaviour*, 62: 711–22.
- Kuhl, P. (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: 20–7.
- R. C. Heimbuch, and 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.
- (1982). 'The Basicranium of Plio-Pleistocene Hominids as an Indicator of their Upper Respiratory Systems', *American Journal of Physical Anthropology*, 59: 323–44.
- and 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.
- Lavy, M. M. (2001). 'Emotion and the Experience of Listening to Music: A Framework for Empirical Research', PhD thesis, University of Cambridge.
- Leakey, R. (1994). *The Origin of Humankind*. London: Weidenfeld and Nicolson.
- Levenson, R. W., P. Ekman, and W. V. Friesen (1990). 'Voluntary Facial Action Generates Emotion-Specific Autonomic Nervous System Activity', *Psychophysiology*, 27: 363–84.

- Lieberman, A. M., F. S. Cooper, D. P. Shankweiler et al. (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.
- (1989). 'The Origins of Some Aspects of Human Language and Cognition'. In: P. Mellars and C. Stringer (eds.) *The Human Revolution: Behavioural and Biological Perspectives in the Origins of Modern Humans*, pp. 391–414. Edinburgh: Edinburgh University Press.
- (1994). 'Comment on Armstrong, Stokoe and Wilson', *Current Anthropology*, 35: 362–3.
- Locke, J. L., K. E. Bekken, L. McMinn-Larson et al. (1995). 'Emergent Control of Manual and Vocal Motor Activity in Relation to the Development of Speech', *Brain and Language*, 51: 498–508.
- MacLarnon, A. M. and G. P. Hewitt (1999). 'The Evolution of Human Speech: The Role of Enhanced Breathing Control', *American Journal of Physical Anthropology*, 109: 341–63.
- Magriples, U. and J. T. Laitman (1987). 'Developmental Change in the Position of the Fetal Human Larynx', *American Journal of Physical Anthropology*, 72: 463–72.
- Mang, E. (2000). 'Intermediate Vocalizations: An Investigation of the Boundary between Speech and Songs in Young Children's Vocalizations', *Bulletin of the Council for Research in Music Education*, 147: 116–21.
- Marin, O. S. M. and D. W. Perry (1999). 'Neurological Aspects of Music Perception and Performance'. In: D. Deutsch (ed.) *The Psychology of Music*, 2nd ed, pp. 653–724. New York: Academic Press.
- Masataka, N. (2000). 'The Role of Modality and Input in the Earliest Stages of Language Acquisition: Studies of Japanese Sign Language'. In: C. Chamberlain, J. Morford and R. I. Mayberry (eds.) *Language Acquisition by Eye*, pp. 3–24. Mahwah, NJ: Erlbaum.
- Mayberry, R. I. and J. Jaques (2000). 'Gesture Production during Stuttered Speech: Insights into the Nature of Gesture-Speech Integration'. In: D. McNeill (ed.) *Language and Gesture*, pp. 199–214. Cambridge: Cambridge University Press.
- Mazzucchi, A., C. Marchini, R. Budai et al. (1982). 'A Case of Receptive Amusia with Prominent Timbre Perception Defect', *Journal of Neurology, Neurosurgery, and Psychiatry*, 45: 644–7.
- McClave, E. (1994). 'Gestural Beats: The Rhythm Hypothesis', *Journal of Psycholinguistic Research*, 23: 45–66.
- McNeill, W. H. (1995). *Keeping Together in Time*. Cambridge, MA: Harvard University Press.
- Messinger, D. S. and A. Fogel (1998). 'Give and Take: The Development of Conventional Infant Gestures', *Merrill-Palmer Quarterly: Journal of Developmental Psychology*, 44: 566–90.

Hominid Physiological Evolution and the Emergence 139

- Mitchell, R. W. and M. C. Gallaher (2001). 'Embodying Music: Matching Music and Dance in Memory', *Music Perception*, 19: 65–85.
- Morley, I. (2002). 'Evolution of the Physiological and Neurological Capacities for Music', *Cambridge Archaeological Journal*, 12: 195–216.
- (2003). 'The Evolutionary Origins and Archaeology of Music: An Investigation into the Prehistory of Human Musical Capacities and Behaviours', PhD thesis, Cambridge University.
- Morton, E. Ss (1977). 'On the Occurrence and Significance of Motivation-Structural Rules in Some Bird and Mammal Sounds', *American Naturalist*, 111: 855–69.
- (1994). 'Sound Symbolism and Its Role in Non-Human Vertebrate Communication'. In: L. Hinton, J. Nichols and J. J. Ohala (eds.) *Sound Symbolism*, pp. 348–65. Cambridge: Cambridge University Press.
- Noffsinger, D. (1985). 'Dichotic-Listening Techniques in the Study of Hemispheric Asymmetries'. In: D. F. Benson and E. Zaidel (eds.) *The Dual Brain*, pp. 127–41. New York: Guildford Press.
- Panksepp, J. (1995). 'The Emotional Sources of 'Chills' Induced by Music', *Music Perception*, 13: 171–208.
- and G. Bernatsky G (2002). 'Emotional Sounds and the Brain: The Neuro-Affective Foundations of Musical Appreciation', *Behavioural Processes*, 60: 133–55.
- Papousek, M., H. Papousek and D. Symmes (1991). 'The Meanings of Melodies in Motherese in Tone and Stress Languages', *Infant Behavior and Development*, 14: 415–40.
- Patel, A. D., I. Peretz, M. Tramo et al. (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.
- R. Kolinsky, M. Tramo et al. (1994). 'Functional Dissociations Following Bilateral Lesions of Auditory-Cortex', *Brain*, 117: 1283–301.
- Peterson, S. E., P. T. Fox, M. I. Posner et al. (1988). 'Positron Emission Tomographic Studies of the Cortical Anatomy of Single-Word Processing', *Nature*, 331: 585–9.
- Scherer KR (1985). Vocal affect signalling: a comparative approach. *Advances in the Study of Behaviour*, 15, 189–244.
- (1986). 'Vocal Affect Expression: A Review and Model for Future Research', *Psychological Bulletin*, 99: 143–65.
- (1991). 'Emotion Expression in Speech and Music'. In: J. Sundberg, L. Nord and R. Carlson (eds.) *Music, Language, Speech and Brain*, p. 146–56. Basingstoke: MacMillan.

- Slater, P. J. B. (2000). 'Birdsong Repertoires: Their Origin and Use'. In: N. L. Wallin, B. Merker, and S. Brown (eds.) *The Origins of Music*, pp. 49–63. Cambridge, MA: MIT Press.
- Sloboda, J. A. (1991). 'Music Structure and Emotional Response: Some Empirical Findings', *Psychology Of Music*, 19: 110–20.
- (1998). 'Does Music Mean Anything?' *Musicae Scientiae*, 2: 21–31.
- (2001). 'The Psychology of Music: Affect'. In: L. Macy (ed.) *The New Grove Dictionary of Music Online*. Available: <<http://www.grovemusic.com>> accessed July 2003.
- & P. N. Juslin (2001). 'Psychological Perspectives on Music and Emotion'. In: P. N. Juslin and J. A. Sloboda (eds.) *Music and Emotion: Theory and Research*, pp. 71–104. Oxford: Oxford University Press.
- Snow, D. (2000). 'The Emotional Basis of Linguistic and Nonlinguistic Intonation: Implications for Hemispheric Specialization'. *Developmental Neuropsychology*, 17: 1–28.
- Spoor, F. (1996). 'The Ancestral Morphology of the Hominid Bony Labyrinth: The Evidence from *Dryopithecus*', *American Journal of Physical Anthropology*, 22 (Supplement): 219.
- B. Wood, and F. Zonneveld (1994). 'Implications of Hominid Labyrinthine Morphology for Evolution of Human Bipedal Locomotion', *Nature*, 369: 645–8.
- Spoor, F. and F. Zonneveld F (1998). 'Comparative Review of the Human Bony Labyrinth', *Yearbook of Physical Anthropology*, 41: 211–51.
- Stern, D. N. (1977). *The First Relationship: Infant and Mother*. Cambridge, MA: Harvard University Press.
- Tartter, V. C. (1980). 'Happy Talk: Perceptual and Acoustic Effects of Smiling on Speech', *Perception and Psychophysics*, 27: 24–7.
- and D. Braun (1994). 'Hearing Smiles and Frowns in Normal and Whisper Registers', *Journal of the Acoustical Society of America*, 96: 2101–7.
- Trainor, L. J., E. D. Clark, A. Huntley et al. (1997). 'The Acoustic Basis of Infant Preferences for Infant-Directed Singing', *Infant Behavior and Development*, 60: 383–96.
- C. Austin, and R. N. Desjardins (2000). 'Is Infant-Directed Speech Prosody a Result of the Vocal Expression of Emotion?' *Psychological Science*, 11: 188–95.
- Trehub, S. E., L. J. Trainor, and A. M. Unyk (1993). 'Music and Speech Processing in the First Year of Life', *Advances in Child Development and Behaviour*, 24: 1–35.
- Trevarthen, C. (1999). 'Musicality and the Intrinsic Motive Pulse: Evidence from Human Psychobiology and Infant Communication', *Musicae Scientiae*, Special Issue: 155–215.

Hominid Physiological Evolution and the Emergence 141

- Vanderhorst, V. G. J. M., E. Teresawa, and H. J. Ralston (2001). 'Mono-synaptic Projections from the Nucleus Retroambiguus Region to Laryngeal Motoneurons in the Rhesus Monkey', *Neuroscience*, 107: 117–25.
- Wagner, H. (1989). 'The Physiological Differentiation of Emotions'. In: H. Wagner and A. Manstead (eds.) *Handbook of Social Psychophysiology*, pp. 77–89. Chichester: Wiley.
- Watt, R. J. and R. L. Ash (1998). 'A Psychological Investigation of Meaning in Music', *Musicae Scientiae*, 2: 33–54.
- Werker, J. F., J. E. Pegg and P. J. McLeod (1994). 'A Cross-Language Investigation of Infant Preference for Infant-Directed Communication', *Infant Behaviour and Development*, 17: 323–33.
- Wild, B., M. Erb, and M. Bartels (2001). 'Are Emotions Contagious? Evoked Emotions while Viewing Emotionally Expressive Faces: Quality, Quantity, Time Course and Gender Differences', *Psychiatry Research*, 102: 109–24.
- Wind, J. (1990). 'The Evolutionary History of the Human Speech Organs'. In: J. Wind, E. G. Pulleybank, E. de Grolier et al. (eds.) *Studies In Language Origins*, i, pp. 173–97. Amsterdam: Benjamins.