

‘REVERSE-ENGINEERING’ THE HUMAN VOICE: EXAMINING THE ADAPTIVE PREREQUISITES FOR SONG AND LANGUAGE

Nicholas Bannan

International Centre for Research in Music Education, University of Reading, UK

ABSTRACT

Background. This research represents the modelling of vocal development and potential in three categories so as to devise means of examining how the specific attributes of human vocality evolved. Assuming singing to be universal and instinctive, the hypothesis is proposed that musical communication represented the bridge between animal communication and human language (sometimes referred to as ‘the continuity paradox’). Ontogeny is interpreted in the light of phylogeny, whereby it is proposed that language exploits neural structures adapted for communicative interaction whose purpose may have been different from speech and the effects of which were closer to what we would term Music.

Aims. A tripartite model of vocal potential is developed, drawing on: archaeological evidence of primate and hominid anatomy; the communication of modern primates; and the anatomical and neurological features of singing in living humans. This is related to a time-line for the evolution of the prerequisites for the eventual expression of which the modern vocal mechanism is capable.

Main Contribution. Referring to key evidence in a variety of disciplines, this study sets out to re-analyse the available data from the viewpoint of voice research informed by specifically musical properties of human vocality in addition to language capacity.

Implications. ‘Reverse-engineering’ human singing involves unravelling the various selective pressures which may have endowed advantages at different points in the 6 million year evolution of the human voice. The study illustrates how both adaptation and exaptation played a part in a complex process involving natural, sexual and group selection. Implications for further research and possible outcomes in music therapy and pedagogy are also suggested.

1. BACKGROUND: SONG AND SPECIES

This study adopts a framework for examining the phenomenon of human vocalization which places it within the category of animal communication. In this sense, it is proposed that song is, evolutionarily, the candidate vocal signature of *Homo Sapiens*. The models employed to substantiate this claim therefore differentiate between the mechanisms involved in speech and language. Evidence for how human vocal capacities could have evolved will be presented which illustrate the hypothesis that language developed out of anatomical and neural substrates already in existence, which were fashioned by the advantages to reproduction and survival conferred by a proto-musical form of communication.

1.1. Song and Language

Speech is predominantly a *serial* mode of communication, whereby individuals take turns to express and exchange representations which have the properties of recursion. This is captured in Grice’s formula (12) for the embedded seriality of referential speech (‘I believe that you believe that I believe *x*’). By contrast, Deacon proposes (6) that singing is ‘contagious’, allowing the sharing and exchange of emotions, and principally a *simultaneous* activity in humans. This is consistent with Merker’s view (24) that simultaneous chorusing may have been fundamental to the pre-linguistic nature of human communication.

1.2. Universality and Instinct

To argue that singing is therefore an adaptation shaped by selection requires that we consider this form of vocalization to be universal (2) and to be passed on in the genes as a potentially instinctive response: a behaviour which is exhibited in reaction to environmental cues through what Tinbergen (30) referred to as an *innate releasing mechanism*. Tinbergen poses four key questions in defining the instinctive nature of behaviour:

1. How has the capability evolved in the species?
2. How do individuals within the species develop the capability?
3. What happens in the nervous system when the capability is exercised?
4. Why is the capability exercised in a particular circumstance?

A full account of the nature of singing will address all four of these. This study will focus on evidence that illuminates the first two of Tinbergen’s concerns.

1.1. Ontogeny and Phylogeny

One aspect of how the development of a trait can be examined in a particular species is to trace the growth of the anatomical prerequisites for behaviour in the individual, from the embryonic stage to maturity (11). The nature and role of singing in child development provides insight into the capacity for vocalization in proto-human ancestors, as well as suggesting at least one mechanism whereby singing could have aided survival.

2. AIMS: THREE MODELS

The principal means of developing a robust theory for the evolution of human communication has been to establish complementary investigation (6, 17, 21, 22) in which comparisons made between the anatomical development of modern humans and other surviving primate species on the one hand parallel those between *Homo sapiens* and the remains of his ancestors. This triangulation of the engineering features of these distinct yet related communicative systems yields new questions as discoveries in one field have been interpreted in another. It is an asymmetric arrangement for obvious reasons: real-life communicative behaviour can only be observed in living organisms. But in experimental and observational work with primates (6, 24, 27), the agenda has been set for what archaeologists may find. This is especially the case where the engineering of the system has clearly undergone considerable changes whose morphology is traceable, and its developmental consequences observable in behaviour over the modern human life cycle.

1.1. The Fossil Record

An approach to the evidence uncovered by archaeologists poses questions framed by research into the mechanics of the singing voice (3, 23). This exercise in ‘reverse-engineering’ (7) allows the breaking down of anatomical and neurological factors on which singing depends in a manner that can be set out chronologically. Thus the descent of the human species over the last six million years can be presented so as to permit comparisons between the proportions and inferred function of anatomical characteristics (see Fig. 1). While debate continues as new evidence comes to light, a consensus is beginning to emerge at least regarding the interpretation of these factors over the last 200,000 years, during which ‘modern’ human communication must finally have emerged in the form which is universal in our species today (13, 16, 17, 19, 20, 21, 22, 25, 29).

1.2. Animal Communication

The other reference-point for comparative analysis is provided by the acoustic, anatomical, behavioural and neurological characteristics of animal behaviour (27). The linguist Bickerton (1) put forward the ‘continuity paradox’ whereby it appears impossible to model how human speech could have emerged adaptively from any prior behaviour. The line of enquiry adopted in this study is to consider whether the continuity paradox could be bridged by musical communication (24).

2.3. The Anatomy of Human Song

Research into the anatomy of the vocal tract (21, 22) and its employment in harness with modern human respiration and posture to produce singing (3, 23) defines the taxonomy of features to be examined through making comparisons between present-day specimens of *Homo*, and related species such as the chimpanzee, with different stages in development, of which evidence has been preserved in the fossil record. The following features would represent prerequisites for human vocal communication:

- development of the auditory system on which human vocalisation depends;
- the evolution of upright posture, and its consequences for the nature of the human larynx and the capacities for resonance of the vocal tract;
- the development of voluntary breathing;
- the descent of the larynx;
- the neoteny of the adult skull-shape: the domed skull, ‘childish’, orthognathic lower jaw, and chin;
- the proportions of the naso-pharynx;
- the cerebral processing on which musical perception and productivity depend: brain size and the development of specific centres for the processing of speech and other functions related to singing;
- handedness and the integration of the senses;
- the development and consequences of omnivorous dentition;
- the development of the sinuses and Eustachian tubes.

A referenced commentary on these elements is provided in Fig 2 (4, 5, 8, 9, 10, 15, 18, 20, 21, 22, 28). Both in Fig. 2 and in the listing above, the ordering represents the best fit with the corresponding chronology adopted for the fossil evidence in Fig. 1.

3. MAIN CONTRIBUTION: SYNTHESIS

In representing the development of a progressively more elaborate anatomy shaped to the advantage of specialization in acoustic communication, both simultaneous and over distance, this evidence also reflects the ‘continuity paradox’ in tracing the capacity for sound-making of successive species from the purely instinctive to the state which we consider today as symbolic and representational. Deacon (6) alerts us to the need to recognize the continuing role of emotional motivation in modern human communication, however rational and symbolic, not least because sound-making reflexes are first displayed even before birth. All human language develops on a scaffolding of such instinctive responses through which the brain organizes itself during infancy in a process involving both perception and productivity. Davis et al (5) argue that infant cries depend on very old neural pathways that developed early in mammalian evolution long prior to the chimpanzee-hominid branch.

Falk’s response to the evidence is to suggest that Australopithecines had ‘a rich repertoire of calls employed in social contexts and used to express emotions’ (8, p. 211). Judging by brain growth and proportion (15), subsequent species in the hominid line have clearly possessed increasing

intelligence suggesting both more complex social interaction, and specialization for a more sophisticated response to the challenge of the environment represented by tool-making and the use of fire. Both of the latter begin to appear as evidence supplementary to biological material. In turn, dependence on artifacts is consistent with the emergence of culture: genes alone do not confer sufficiency for survival, which is supplemented by learning. The mechanism for information exchange which allows learning could have been gestural, and certainly became language: but, with more limited acoustic means closer to the vocal capacities of chimpanzees or, more appropriately, gibbons (24), early hominids could have expressed themselves through organized sound closer in its acoustic properties to the shapes and structures of music than to semantic speech. This claim – for a musical basis for linguistic evolution – is at least consistent with the anatomical data. However, such speculation can only inform our examination of the available evidence if it is related to models by which adaptation is believed to occur.

1.1. Interpreting the Evidence

‘Reverse-engineering’ modern human vocalisation cross-culturally reveals properties of song in comparison with speech which could be taken to confer survival advantages:

- (i) cost-benefit: expenditure of energy is minimised where respiration can be slowed without harm to the organism in support of variable and resonantly amplified utterance. This may have a role in empathetic vocal affect, such as songs for calming both listeners and/or the singer himself;
- (ii) a wide variety of responses to stimuli is observed, both emotionally-motivated and under voluntary control.

The complex anatomy of the human vocal instrument allows control over the following parameters of potentially meaningful communication:

- pitch;
- volume (both *capacity for* and *control of*);
- duration (both maximisation of efficient employment of a single breath; and capacity to match durational units to entrained tempi);
- timbre (capacity to vary formants over a sustained or repeated fundamental; ability to employ this feature to disguise or mimic the vocal signals or emotions of others).

It is no coincidence that these same features comprise what we think of as the elements of music.

1.2. Relating Evidence to Function

If these are the evolved functions to which we may attribute modern vocal productivity – both speech and song, whereby the former has exploited the existing mechanisms of the latter – then the search for the appearance of anatomical features on which each depends will also require our addressing the implication of Tinbergen’s first question: what selective advantage can such properties have conferred such that they shaped the product of genetic transmission?

In this respect, voluntary breathing links closely to efficiency of vocalization in marked contrast to the pant-hoot of the chimpanzee. The ‘two-tube’ vocal tract (21, 22) has conferred a capacity for resonance which aids communication at a distance, but also has a more radical outcome: the capacity to out-compete rivals in certain parts of the frequency spectrum – a form of ‘acoustic jamming’ (14, p. 127). Two properties of human vocalization descend from this: the capacity to produce and perceive as distinct vowel sounds variants of the timbral character of pitched sounds; and the consequences of sexual dimorphism whereby adult humans have voices of differing range. The first of these, vowels, are the foundation of all spoken language. The second may have played a role in the means by which humans achieved both reproductive success and achieved specific attributes of child care from which learning and thereby culture developed.

1.3. Models for Vocal Evolution

If *Homo* is the musical primate, then we need to consider the mechanism whereby musical instincts entered the human genome. An argument for this based on natural selection would propose that humans can use their voices to locate food (perhaps by mimicking animals) and avoid or scare away predators. This weapon/armour model seems insufficient, not least because human music-making is so clearly a social behaviour. In this respect, it might be claimed that vocalization has emerged as a characteristic of sexual selection, as with many species of bird, ensuring that fit males attract suitable females. But this also fails to take into account the importance of singing in childhood as well as in other related contexts – as a mnemonic accompanying individual tasks; or as mechanism for large-group co-ordination.

Merker (24) argues persuasively for a group-selection model that nevertheless regulates reproductive success, involving male co-operation in attracting unrelated migrant females. While this might explain the differing range of adult male and female voices, it leaves unexplained the phenomenon whereby, at least in some cultures, differing voices can be combined within family and other groupings (infant plus female; infant plus male; male plus female; the full chorus).

An integrated model of vocal evolution is required which represents features of natural, sexual and group selection in relation to the chronology of the available fossil evidence, and which embraces both adaptive and exaptive processes. In particular, the re-employment of an existing anatomical structure for a different purposes will represent a mental adaptation.

4. IMPLICATIONS: SONG AND CULTURE

The musical characteristics of human communication have been overlooked, notably by Pinker, who claims that music is not an adaptation and that it is, neurologically, 'cheese cake for the ears' (26, p. 538). This is neither borne out by analysis of the archaeological and anthropological data, nor cognizant of the multiplicity of uses to which human beings put music today. In terms of modern human behaviour, the neurological implications of viewing musical processing as foundational to language are consistent with brain processing revealed by aphasias (6) and the practice of music therapy. Similarly, the potential of vocalization for preserving emotional motivation and its role in memory can be addressed both in the treatment of the victims of strokes and Alzheimer's, and in continuing to care for the young of our species through a medium to which they are genetically fitted to respond.

4. REFERENCES

1. Bickerton, D. (1991) *Language and Species*. Chicago: University of Chicago Press
2. Brown, D. E. (1991) *Human Universals*. New York: McGraw-Hill
3. Bunch, M. (1995) *The Dynamics of the Singing Voice*. New York: Springer-Verlag
4. Cartmill, M. (1992) Non-human primates, in S. Jones, R. Martin and D. Pilbeam (Ed.s) *The Cambridge Encyclopedia of Human Evolution* (pp. 24-32). Cambridge University Press
5. Davis, P., Zhang, S., Winkworth, A., and Bandler, R. (1996) Neural control of vocalisation: respiratory and emotional influences, in *Journal of Voice* Vol. 10 No. 1, pp. 23-38. Philadelphia: Lippincott-Raven
6. Deacon, T. (1997) *The Symbolic Species* London: Allen Lane
7. Dennett, D. (1995) *Darwin's Dangerous Idea* London: Allen Lane
8. Falk, D. (2000) Hominid brain evolution and the origins of music, in N. Wallin, B. Merker and S. Brown, (Ed.s), *The Origins of Music* (pp. 197-216). Cambridge, MA: MIT Press
9. Franciscus, R. G. and Trinkhaus, E. (1988) Nasal morphology and the emergence of *Homo erectus*. *American Journal of Physical Anthropology* 75, pp. 517-28
10. Frayer D. W. and Nicolay, C. (2000) Fossil evidence for the origin of speech sounds, in N. Wallin, B. Merker and S. Brown, (Ed.s), *The Origins of Music* (pp. 217-34). Cambridge, MA: MIT Press
11. Gould, S. J. (1982) (1977) *Ontogeny and Phylogeny* Cambridge, Mass: Harvard University Press
12. Grice, H. P. (1969) Utterers' meaning and intentions. *Philosophical Review* 78, pp. 147-77
13. Haile-Selassie, Y. (2001) Late Miocene Hominids from the Middle Awash, Ethiopia. *Nature*, 412, pp. 178-81
14. Hauser, M. D. (1997) *The Evolution of Communication* Cambridge, MA: MIT Press
15. Jerison, H. (2000) Paleoneurology and the Biology of Music, in N. Wallin, B. Merker and S. Brown, (Ed.s), *The Origins of Music* (pp. 177-96). Cambridge, MA: MIT Press
16. Johanson, D. C. and Edgar, B. E. (1996) *From Lucy to Language*. New York: Simon & Schuster
17. Jones, S., Martin, R. and Pilbeam, D. (1992) *The Cambridge Encyclopedia of Human Evolution*. Cambridge: Cambridge University Press
18. Kupper-Smith, R. B. (1996) *Eustachian tube function and dysfunction*. The Bobby R. Alford Department of Otorhinolaryngology and Communicative Sciences, Baylor College of Medicine, Houston, Texas. <http://www.bcm.tmc.edu/oto/grand/71196.html>
19. Lemonick, M. D. and Dorfman, A. (2001) One giant step for mankind. *Time*, vol. 158, no. 4, pp. 54-61
20. Lieberman, D. E. (1998) Sphenoid shortening and the evolution of modern human cranial shape. *Nature*, vol. 393, pp. 158-62
21. Lieberman, P. (1975) *On the origins of language*. New York: Macmillan
22. Lieberman, P. and Crelin, E. S. (1971) On the speech of Neanderthal man. *Linguist. Inquiry*, 2, pp. 203-22
23. McKinney, J. (1994) *The diagnosis & correction of vocal faults: a manual for teachers of singing and for choir trainers*. Expanded reprint of 1st edition (Nashville: Broadman Press, 1982): Genevox.
24. Merker, B. (2000a) Synchronous chorusing and human origins, in N. Wallin, B. Merker and S. Brown, (Ed.s), *The Origins of Music* (pp. 315-327). Cambridge, MA: MIT Press
25. Mithen, S (1996) *The prehistory of the mind* London: Thames and Hudson
26. Pinker, S (1997) *How the Mind Works*. New York: W. W. Norton
27. Premack, D. (1986) "Gavagai!" or the future history of the animal language controversy. Cambridge, MA: MIT Press

28. Reeser, H. (2001) Selective brain cooling in early hominids: phylogenetic and evolutionary implications. Unpublished paper, University of Florida. reeser@flmnh.ufl.edu
29. Senut, B., Pickford, M., Gommery, D., Mein, P., Cheboi, K. and Coppens, Y. (2001) First hominid from the Miocene (Lukeino Formation, Kenya). *Comptes Rendus de l'Académie de Sciences*, vol. 332, pp. 137-44
30. Tinbergen, N. (1951) *The study of instinct* Oxford: Clarendon Press
31. Walker, A. (1996) *The wisdom of bones: In search of Human Origins*. New York: Alfred Knopf