

# Language out of Music: The Four Dimensions of Vocal Learning

---

*Nicholas Bannan*

Music, University of Western Australia

A growing consensus drawing on research in a wide variety of disciplines has, over the last fifteen years or so, argued the need to revisit Darwin's conjecture of 1871 that language may be descended from an existing, musical medium of communication that developed from animal calls. This paper seeks to examine, in an extension of Hockett's analysis of the design features required for linguistic communication, the nature of the acoustic information produced and perceived in human vocalisation, and to consider the anatomical and neural mechanisms on which these depend. An attempt is made to sketch an evolutionary chronology for key prerequisites of human orality. Cross-species comparisons are employed to illuminate the role of four acoustic variables (pitch, duration, amplitude and timbre), viewing the potential for human vocal productivity from the perspective of animal communication. Although humans are the only species to combine entrainment to pulse with attunement to precisely-tracked pitches, we also depend both for musical interaction and the production and perception of vowel sounds on precise and conscious control of the property of timbre. Drawing on, amongst others, Scherer's analyses of emotionally triggered sounds in a variety of species, and Fernald's presentation of the similarities of infant cries and adult production of infant-directed speech in a variety of cultures and languages, a case is made for the *instinctive* components of human communication being more music-like than language-like. In conclusion, historical and comparative data are employed to outline the adaptive and exaptive sequence by which human vocal communication evolved. The roles of selective pressures that conform to different adaptive models are compared—natural selection, sexual selection, group selection—leading to the proposal that all of these must have played their part at different stages in the process in a 'mosaic' model consistent with the development of other human traits.

## Introduction

Whereas conjecture regarding the origins of language in the writings of, amongst others, Vico (1976 [1744]), Rousseau (1966 [1761]) and their contemporaries had previously suggested a developmental musical bridge between animal communication and speech (Thomas 1995; Levman 2000), Darwin's position in *The Descent of Man* places this proposed model within the explanatory framework of the theory of evolution that he set out in *The Origin of Species* (1968 [1859]):

THE AUSTRALIAN JOURNAL OF ANTHROPOLOGY, 2008, 19:3, 272-293

I cannot doubt that language owes its origin to the imitation and modification of various natural sounds, the voices of other animals, and man's own instinctive cries, aided by signs and gestures. When we treat of sexual selection we shall see that primeval man, or rather some early progenitor of man, probably first used his voice in producing true musical cadences, that is in singing, as do some of the gibbon-apes of the present day; and we may conclude from a widely-spread analogy, that this power would have been especially exerted during the courtship of the sexes—would have expressed various emotions, such as love, jealousy, triumph—and would have served as a challenge to rivals. It is, therefore, probable that the imitation of musical cries by articulate sounds may have given rise to words expressive of various complex emotions. (Darwin 1871: 132-133)

Models of vocal communication and language have since arisen within the discipline of linguistics on lines established by Saussure (1959 [1916]), and the acquisition of speech analysed in terms of the features of syntax, grammar and lexical exchange that define a uniquely human achievement. By the end of the twentieth century, two positions had arisen with a polarity drawn according to principles consistent with the disciplines in which they were formed. On the one hand, following the theoretical approach to language and thought initially proposed by Chomsky (1957), researchers such as Pinker and Bickerton remained sceptical of language theories that admitted a role to musical communication. Pinker (1997) dismissed music as of no adaptive purpose; while Bickerton (1991) described the 'continuity paradox' whereby a language-using community cannot enter the mind of a species that can communicate without language. Thus it is seen as impossible to account for how a condition of vocal communication could have bridged the gap between animal calls, as suggested by Darwin, and modern human verbal ability.

A contrasting view has developed in the work of psychologists and anthropologists who have focused on the problem Bickerton outlined and sought solutions to it. Amongst these, Donald (1991) accorded a transitional role to mimesis that might allow the bootstrapping of linguistic development. Deacon (1997) perceived in the anatomy and neurology of the human voice a means of constructing intentionality and meaning on the foundations of instinctive responses that admit equally the development of song and speech. Wray (2002), a linguistics researcher with a background as a professional singer, proposed that language may have developed from holistic utterances that were subsequently atomised, as opposed to being built from arbitrary phonetic units combined together on the Chomskyan model. Mithen (2005) drew on these and others to suggest that language and song have developed independently from a single vocal behaviour in our ancestors that contained features of both, and painted the vivid if controversial picture of *The Singing Neanderthals*.

### Language from an animal communication perspective

In a paper that has proved influential in anthropology and musicology as well as in language studies, Hockett (1960) modelled the design features of vocal communication in a manner illustrating that, while many unrelated species possess communicative abilities similar to those of humans, our is the only species to synthesise them all into an integrated, variable behaviour capable of infinite yet meaningful productivity. Figure 1 sets out Hockett's model.

In the 1960 paper from which this is taken, Hockett additionally sketched a means of discriminating between human language and music, though the criteria defining his category 'music' include that this is, by implication, instrumental, and comprises 'only Western music since the time of Bach' (Hockett 1960: 94; see Figure 2). While Hockett's design features represent a useful basis for examining what happens when animals communicate vocally that provides an insight into the different contexts in which musical

**Figure 1. 13 Design-features of animal communication**  
Based on Charles F. Hockett *The Origin of Speech* 1960:89-96

1. **Vocal-auditory channel**  
(an acoustic system that leaves the limbs free for other purposes).
2. **Broadcast transmission/directional reception**  
(signal audible at distance/binaural direction-finding traces signal source).
3. **Rapid fading (transitoriness)**  
(leaves no trace: may explain fascination for echo/mimicry.  
Contrasts explicitly with communication by *olfactory signs*).
4. **Interchangeability**  
(user can both transmit and receive).
5. **Total feedback**  
(user 'hears what he says': internalisation leads to consciousness).
6. **Specialisation**  
(speech is a dedicated system, not a by-product of other behaviour).
7. **Semanticity**  
(capability of communication to be specific and interpretable).
8. **Arbitrariness**  
(<whale>=small word, big thing: <micro-organism>=*vice versa*).
9. **Discreteness**  
(significance of small variations in the code: e.g. *pin* not same as *bin*).
10. **Displacement**  
(the capacity to refer to absent objects, the past/ future, etc.).
11. **Productivity**  
(the capacity of the system to allow meaningful coinage).
12. **Traditional transmission**  
(capacity for the code to be used to develop/acquire the code).
13. **Duality of patterning**  
(economical generation: e.g. 'cat'/'tack'/'ct' share the same phonemes).

vocalisation occurs, his own application of it to music is limited and internally inconsistent.

Brown (2007) complements Hockett's approach by proposing design features for the way different species sing. For instance, we share with several avian species (such as parrots) the ability to control timbre and duration in producing the target acoustic objects of speech; and are able, like fireflies and certain crickets and frogs (Merker 2000), to entrain precisely to a common pulse. But we are the only species to combine these abilities so as to sing in unison, let alone to contribute differing but acoustically complementary pitches to collective performance through singing in harmony.

A full discussion of the implications of Hockett's claims for the differences between language and music is beyond the scope of this paper, but for present purposes the following can be stated:

- To the extent that musical behaviour is an evolved characteristic, vocal music, which is a function of anatomy, is far older than instrumental, which employs extra-somatic tools.

- If music is to be examined as a human universal, theorisation about its nature and structural properties cannot be limited to consideration of one culture or historical period.
- The infinite productivity of human language is arguably matched by our species' capacity for variety of song, in terms of mimicry, imitation, polyphony and styles of vocal production (Tran Quang and Bannan in press).
- Developments in the understanding of music cognition, both theoretical (Cox 2001) and clinical (Hodges 2000; Parsons 2001) illustrate that Displacement has a function in musical behaviour, and that Duality of Patterning also represents a feature of musical structure evident in both productivity and perception.

**Figure 2. Hockett's criteria for the application of his design features to music**  
(based on Hockett 1960: 94)

Instrumental music is as follows:

1. Auditory but NOT vocal.
2. It involves Broadcast Transmission and Directional Reception.
3. It involves Transitoriness.
4. It DOES NOT involve interchangeability.
5. It involves Total Feedback.
6. It involves Specialisation.
7. It does not (in general) involve Semanticity.
8. (the box for Arbitrariness is left blank).
9. It PARTLY involves Discreteness.
10. (the box for Displacement is left blank).
11. It involves Productivity.
12. It involved Traditional Transmission.
13. (the box for Duality of Patterning is left blank).

Hockett's model was intended to account for language as a set of distinct but overlapping traits with an evolutionary history shared with ancestor species (1960: 93). He did not set out to consider, as Darwin had proposed, that song may have played a part in this process. An account of the evolution of human song intended to complement Hockett's model will need to recognise the different conditions in which song is practised: for instance, as simultaneous chorus, or wordless emotional expression; as well as how these may have conferred survival value in specific social/environmental contexts (Foley in press; Gamble in press). Essential to an account of adaptation on Darwinian lines is a theoretical framework for the conditions that may have shaped it. Tinbergen's (1951) formula for the study of instinct provides a guide:

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?

Tinbergen argued that instinctive behaviours in an organism respond to an *Innate Releasing Mechanism* that governs when and under what conditions (e.g. temperature; acoustic stimulus; hours of daylight; humidity) they may be expressed. In reviewing the positions adopted by Pinker and Wray, and considering whether in either phylogeny or

ontogeny an innate releasing mechanism might have played a role in the acquisition of speech and song, it would seem logical to place the nature of singing, or at least song-like behaviours, as closer than speech to the kind of instinctive behaviour Tinbergen's model requires us to explain. On this basis, the capacity for song in our early hominin ancestors conferred survival advantages for, say, reproductive competition and/or collective sonic defence (or both, see Merker 2000) well before anything that could be classified as language evolved. Could it then be that it is music that has been carried in our genes since the ancestral stage Darwin proposes, and that language has developed on these foundations?

### **Musical behaviour: parameters of musical response, control and learning**

Music, like language, varies almost infinitely as both process and product across the widespread and historically independent cultures in which it is universally practised. More, perhaps, than language, music tends to be recalled and valued as product: a favourite artwork (song, dance, symphony); an artist whose CD one covets; a tune one learned as a child; a mnemonic that recalls information. Moreover, while all normal humans acquire language, the awkward feature presents itself that, at least in certain cultures, individuals shy away from active participation and may not even see themselves as 'musical' at all. Indeed, this apparent imbalance in human musical productivity has been taken as evidence that music may not be adaptive, since 'unmusical' people should not emerge from a genetic predisposition for musicality (see Pinker 1997: 529, 535). But it is clear both that judgements of musical ability are culturally constructed (Blacking 1987), and that attainment of musical productivity varies enormously in relation to motivation and opportunity (Welch 2001).

In order to dissociate the universal, evolved capacity for music from valuation of, or confusion with, the varied products to which this ability has given rise, it may help to coin the term *musical behaviour* to designate those features of perception, production and anatomical response that manifest themselves under conditions appropriate to their release. The use of this term allows us to focus on process, and the evolved capacity for it, rather than product. On this basis 'shower singers' who are musically active only in private are as much behaving musically on these occasions as extroverts chanting on the football terrace or infants moving in time to a mother's song, let alone expert performers employing the properties of designed tools such as musical instruments. Musical behaviour embraces activities that are consistent with a set of evolved characteristics employed to make music even where no explicitly musical product is intended, such as in cooperative rowing or other rhythmically-coordinated work, infant-directed speech, acoustic response to the environment, mnemonic chanting, and a wide variety of further instances that are presented in the following figure (Figure 3).

The evidence provided by these widely varying practices supports the view that humans carry the capacity for musical behaviour in their genes, and this helps to explain the universal practice of musical activities in every recorded human society (Blacking 1987; Brown 1991). Archaeologists have begun to recognise that their discipline has tended to ignore this essential feature of human social and cultural practice, and during the last twenty years the discipline of archaeomusicology (Hickmann 2006) has played a part in defining the wider concerns of evolutionary musicology (or what Wallin 1991, termed *biomusicology*) in work that has analysed the environmental and acoustic features that may have stimulated musical responses, as well as evidence that prehistoric art may have been associated with ritual involving music and dance (Schneider 1986; Garfinkel

1998, 2001, 2003; Watson and Keating 1999; Mithen 2005; Cross and Watson 2006; Gamble in press;). While attention has been paid to the presence in the fossil and artefact record of claimed musical instruments that have survived (Stockmann 1986; Kunej and Turk

### Figure 3. When and why humans sing

This taxonomy, the ordering of which reflects the evolutionary/developmental characteristics of these activities, exemplifies some of the motivations for song:

- for personal amusement (Brailoui 1984 [1949]: 56; Wishart 1979, 1996; Nattiez 1990, 1999, 2000; Trevarthen 1994; Cook 2000);
- in children's social development of mediation and group play (Opie and Opie 1985; Blacking 1987; Cook 2000);
- in relating, via mimicry, to the environment (Nattiez 1990; Ridington 1991; Watson and Keating 1999);
- as a means of recalling information; (Ellis 1980, 2001; Moyle 1980; Sacks 1985: 7-21; Chatwin 1987);
- as an efficient carrier across distance or in competition with other signals (Classe 1957; Sundberg 1999);
- as display or entertainment (Hines 1983; Page 1987; Miller 2000; Potter 2000; Toop 2000);
- in communication with animals (Wallin 1991);
- in soothing infants and teaching children (Woodward 1992; Trevarthen 1994; Nelson 1997);
- to introduce states of mind appropriate to specific activities (Day 1992; Oubré 1997; Tolbert 2000);
- to heighten speech in ritual (Hollenweger n.d.; Schleifer 2001);
- to create a medium of communication appropriate to an intended audience (Potter 2000; Toop 2000);
- as an accompaniment to work (Harker 1985; Carlton 1996; Garfinkel 1998, 2001, 2003);
- as a cohesive force in social and political organisation (Thrills 1998; Merker 2000; Baily 2004; Howard 2004) and
- in response to the specific generative potential of musical structures (Nattiez 1990; Cox 2000).

2000) whether or not they represent objects that were intentionally constructed for musical performance (Morley 2006), we have to accept that the principal candidates for earliest known instrument have not fossilised: the soft tissue of the human respiratory tract that may, over a period of some two million or more years, have permitted progressively more elaborate musical vocalisation; and organic matter employed in the musical practices of surviving hunter-gatherer societies such as bamboo, grass, seaweed, moth chrysalises and gourds that form the material from which instruments can be created with relatively little technological ingenuity (Espí-Sanchis and Bannan in press).<sup>1</sup> Further, it would seem likely that these kinds of musical *objets trouvés* extended existing musical functions (vocalisation, clapping, stamping) modelled on imaginative exploitation of the resonances and capacities for sound production of the human body itself. With this in mind, it would seem fruitful to commence an examination of human vocal ability with an analysis of the

capacities for vocal imitation, perception and learning to which our evolved anatomy has given rise.

The complex anatomy of the human vocal instrument allows control over the following four 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)

Bearing in mind the frameworks for studying such phenomena provided by Hockett and Tinbergen, we could approach accounting for human vocalisation from the perspective of animal communication (c.f. also Hauser 1997; Fitch 2000). But, while humans are the only species to have acquired productive language competency (Premack 1986), we are by no means the only species to sing, or to employ vocal learning in doing so (Nottebohm 1975). For instance, the same apparent capacity for vocal control outlined above (over the four parameters: fundamental frequency; timbre; duration and rhythm of component notes; and volume), in a similar overall range to that of the human voice, governs the varied communication of the Australian raven *Corvus coronoides* (Rowley 1973). While Rowley discusses the ethology of 'raven language', there are clearly differences both in kind and degree between what humans and ravens achieve, and we can assume that Rowley employs this term metaphorically. Nevertheless, the comparison brings home how important it is that we account for perception and productivity that employs all four of these acoustic parameters, and that we consider how abilities of these kinds may have developed in each species as well as how they are acquired by individuals.

Rowley (1973) sets out the eighteen or so distinct behaviours that give rise to raven calls, such as territorial protection and pair-bonding, but is less clear on the precise acoustic properties that distinguish one from another. A study that analysed the Australian raven calls as a combinatorial system that employs variation in pitch, timbre (including distinctions that map eerily onto human-like vowel sounds), duration and volume might well throw light on how the 'language' is acquired, and what constraints operate on characteristics associated with individuals: the relationship between phylogeny and ontogeny. This is the same agenda that requires careful examination in the case of human pre-verbal vocalisation.

An extension of Hockett's (1960) framework thus helps illustrate the flexibility and plasticity of human vocal production. Like ravens, we can announce our presence, doing so in terms that can manipulate responses. Like gibbons, we can express sexually dimorphic contributions to cooperative activity. Like mobbing lorikeets or chimpanzees, but with infinitely greater precision, we can produce summative vocal displays that create a product with increased volume: but humans are unique in having developed the ability to incorporate the additional richness of precisely tuned harmonic reinforcement. This special adaptation represents a 'three-dimensional' response to pitch relationships for which we may not be the only species to have acquired a perceptual grasp (Wright *et al.* 2000), but which we are, significantly, the only species able to exploit both within gender groups, across the summation of female and male cooperation, and in joint activities involving infants and adults. Indeed, the human life-cycle is one which demands plasticity of timbral and pitched production as voices alter in range in relation to maturation and other biological features: adolescent voice change in males (Cooksey 1997); variation in range and timbre due to menarche, pregnancy and menopause in women (Abitbol *et al.* 1989; Lã

and Davidson 2005). This aspect of human vocal capacity is so important that it needs stressing further: while several species of bird and cetacean sing expressively, and some have extraordinary capacities for the imitation of other species and sound sources, human beings are the only species to have evolved the ability to sing in unison in both the dimensions of rhythmic co-ordination and precise pitch attunement. The further widespread ('almost universal', Brown 1991) feature of singing different notes that aggregate to make chords is equally something that only the later stages of the hominid line appears to have attained. Not only is this as astonishing an achievement as language itself: one explanation for the evolution of this ability is that it may have been a necessary condition for the development of the anatomical, aural and neural substrates on which language depends.

What this enriched ability for a precise relationship between pitch and timbre in human vocalisation represents is a distinctive capacity for production that has exploited the existing potential of our primate, indeed mammalian, aural perception:

Our ears have developed in such a way as to be able to analyze the pattern formed by the harmonic content of sounds in all its amazing intricacy. And this ability is not unique to humans. All animals have a capacity to recognize the special harmonic or spectral characteristics of the natural sounds of their environment. It enables animals to recognise their own kind and their enemies and provides other information they need for survival. A most interesting facet of this concerns how humans use spectral content not only to tell the difference between musical instruments or different people's voices but also as a source for deriving pitch from any complex periodic sounds. (Walker 1990: 30)

On the basis that form precedes function, we need to investigate how this special adaptation may have emerged, and the role it plays in individual vocal learning.

### **The development of vocal communication**

The human infant is born with the anatomical capacity to vocalise efficiently: the individual's unique sound signature (voice) commences development at birth, primed by aural experience in the womb (Shetler 1990; Woodward 1992; Lecanuet 1996) and expressed initially in crying. Interaction with others develops quickly (Trevvarthen 1994); entrainment to precise simultaneity with others follows; both are independent of language, and may remain so in cases of impaired language development (El Mogharbel *et al.* 2003). Locke (2006) theorises that infant instinctive vocalisation may provide cues to which parents respond in their quality of care-giving—what he terms *parental selection of vocal behaviour*. On this basis, the earliest sounds uttered without specific intention by human infants may play a part in their capacity to survive. Soltis (2004) argues similarly that the crying of the neonate represents an 'honest signal' that has evolved in tandem with the response to it of the care-giver, and suggests that it represents the basis for development whereby the infant subsequently crosses the threshold between instinctive reaction and intentionality in a manner consistent with Deacon's account (1997) of the myelination of vocal behaviour in response to environmental stimulus. The following diagram (Figure 4) sets out to illustrate the ways in which voice development proceeds as expressive and intentional vocal competences build on such early, instinctive behaviours. Infants vocalise when alone, in play and experimentation, but also with increased attentiveness and arousal in response to carers (Papousek 1992; Trehub 2001).

In thus modelling the acquisition of vocal competence in human infants, as their consciousness of the medium they exploit in communication develops in relation to their consciousness of others and the surrounding environment, giving rise to communication



that crosses the frontier from instinctive response to intentionality, a means of distinguishing between song and speech becomes apparent. In particular, while language is fundamentally *serial* (Grice 1969), like the 'send-receive' function of a two-way radio, singing is characteristically *simultaneous* (and 'contagious': Deacon 1997; Brown 2007). The latter would be consistent with the view that the roots of some modern human language abilities may have emerged from synchronous chorusing (Merker 2000).

Through looking at song in addition to language, we can place the development of human vocalisation in a broader context that conforms well to the demands of Tinbergen's framework. 'Reverse-engineering' the full range of human vocal activity cross-culturally reveals properties of song in comparison with speech that could be taken to confer survival advantages. These represent candidate adaptations on natural selection lines where specific anatomical features have been exploited:

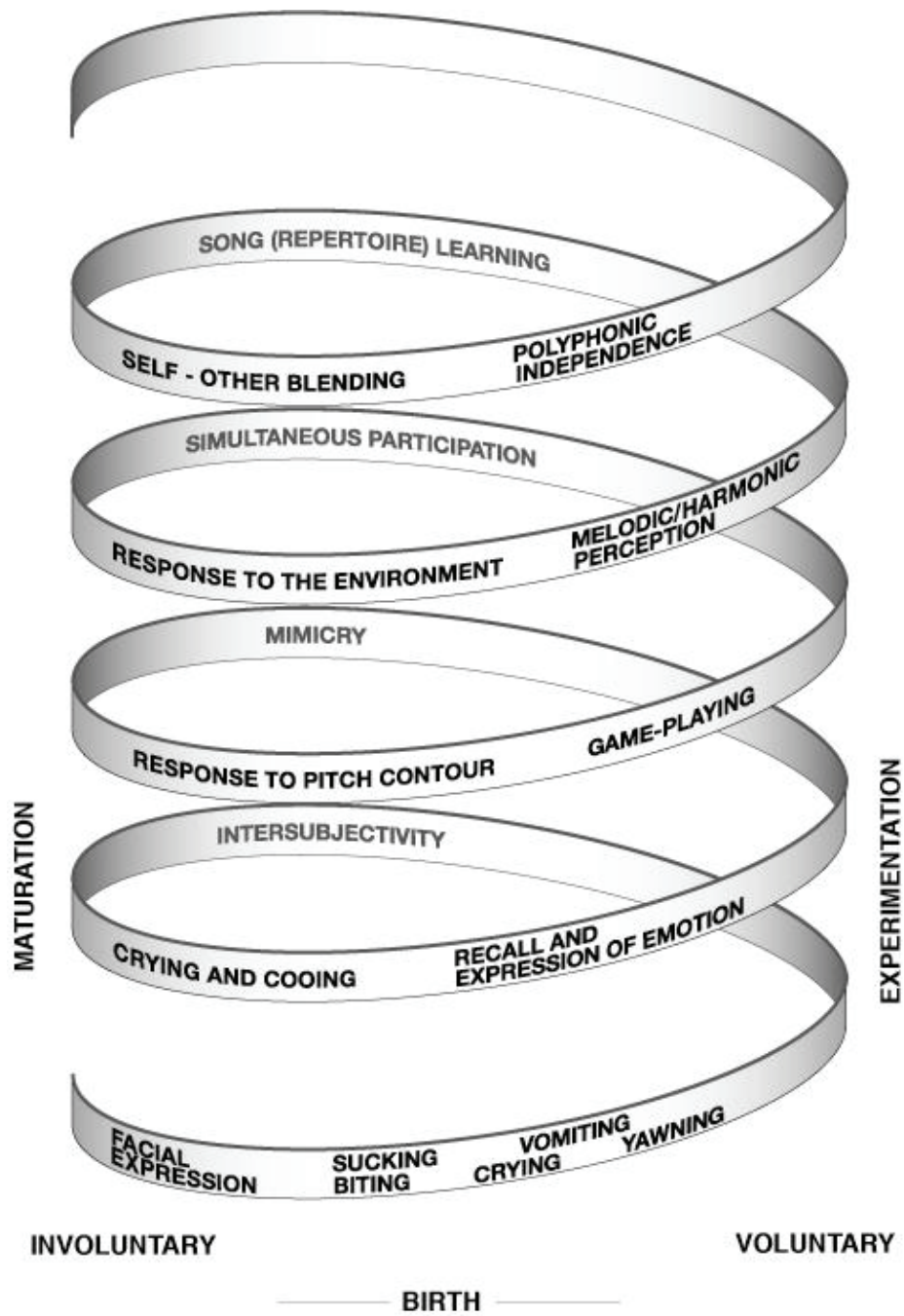
- (i) cost-benefit: expenditure of energy is minimised where respiration can be slowed without harm to the organism in support of variable and resonantly amplified utterance (this may have a role in empathetic vocal affect, such as songs for calming both listeners and/or the singer himself—singing can evoke a sense of well-being);
- (ii) a wide variety of responses to stimuli is observed, both emotionally-motivated and under voluntary control; and
- (iii) simultaneous vocalisation can prove effective as aggressive display, but also has the capacity for eliciting a variety of shared emotions.

These features of vocal interaction correlate strongly, in both their purpose and the mechanisms by which they operate, with emotional exchange: what, based on Deacon (1997: 59), could be termed *recruitment arousal* (see also pp. 57-58 for the background to this in 'contagious' vocal behaviour). Unlike language, which is a superior medium for exchanging propositional logic, the reliability of a song-like system of vocal interaction relates almost exclusively to features of the manipulation and mediation of emotional responses. The instinctive basis of a system of this kind is discernible in the early acquisition of manipulative crying in human infants between birth and three months (Soltis 2004; see also Trevarthen 1994), a behaviour entirely absent in the infants of our nearest genetic relative, the chimpanzee. In considering the role that musical vocalisation may have played in the social life of our ancestors, we can model whether developments *within* the set of skills that broadly comprise the capacity to sing may represent the bridge between animal calls and the emergence of the potential for language. We need also to evaluate how specific selection pressures may have shaped the anatomical and neural adaptations on which a given performance advantage or perceptual sensitivity depends. We will return to this theme once we have examined the mechanisms by which some key aspects of vocal learning are acquired.

### **Intentionality, honesty, reliability and cost in human vocal communication**

Vocal learning, then, involves mimicry and rehearsal that develops control of the four parameters of production (fundamental frequency, duration, volume, timbre), and this emerges from creative play behaviour as well as intersubjective response including simultaneous participation. Communication becomes possible as instinctive acoustic production is harnessed intentionally. A key feature of this process is the achievement of intentional *monotony*, whereby fluctuation in fundamental frequency, that is in neonates solely subject to emotional motivation or state of health (Soltis 2004), is brought under control.

Figure 4. The acquisition of vocality in infancy



The acquisition of fine control of fundamental frequency confers benefits to the development of both language and music, through divergent exploitation of its potential.

However the productive and perceptual abilities that permit these behaviours originally evolved, they have been harnessed in quite different ways in the two media: (1) in language perception and production, it is the variation of formants in relation to fundamental frequency that provides vowels, the phonetic engines of efficient vocal continuity and significant carriers of distinct meaning; (2) in music, control of fundamental frequency ( $F_0$ ) allows the establishment of *tonality*, the means by which pitch relationships operate on short-term memory to achieve musical coherence in relation to the passing of time.

The link between these otherwise distinct functions in language and music arises due to the evolved perception of categorical boundaries between consonance and dissonance that are themselves consequences of the neuro-anatomy of aural discrimination. Pitches that are related by whole-number ratios have greater coherence than those that we perceive as inharmonic because their frequency ratios are more complex. The Harmonic Series (Figure 5) that governs such intervallic properties thus gives rise to a capacity for categorical *timbre* perception where formants approximate to its member ratios in relation to a perceived or recalled  $F_0$ . Human employment of these features in the 'total feedback' mode of Hockett's model of the relationship between production and perception suggests a greater reliability of discrimination of timbral relationships in comparison with categorical perception of  $F_0$  pitch contour; precisely the feature exploited by language, which requires perceptual and productive distinction between vowel sounds but, with the significant exception of tonal languages, does not exploit precise frequency differences (intervals) in the way that music does.

#### Figure 5. The Harmonic Series

When a string or a column of air vibrates, it sets up harmonics that enrich the character and resonance of its fundamental frequency (what we normally perceive as its pitch). These represent whole-number divisions of the length of the string or tube, resulting in whole-number multiplications of the fundamental frequency. Thus, the 'open' A string of a violin at 440 cycles per second contains, as part of its sonic character, further activation at 880 Hz (the Octave), 1320 Hz (the Twelfth), 1760 (the Fifteenth), and so on: these can all be isolated as distinct pitches—on the violin by lightly touching the node of the string at the precise point that suppresses other vibrations; on wind instruments by varying the pressure and direction of the air supply. Practised singers in countries as widely separated as Southern Africa and Mongolia can achieve the same effect in a style known as Overtone Singing.

Meanwhile, the aggregate sound of performing at the same time notes related to one another in this way results in a chord: e.g.

A (440 Hz) + E (1320 Hz/ $3^{\text{rd}}$  Harmonic) + C# (2200 Hz/ $5^{\text{th}}$  Harmonic) produces the euphonious chord known as A Major.

Differences between spoken vowels depend on similar features of resonance, suggesting that vowels employed in speech may represent an adaptation to language of harmonically differentiated properties of high-energy singing. Evidence for this would include: that resonant singing, especially that which depends on blend between several singers performing simultaneously, requires attunement to collective timbre as well as relative pitch; that infant-directed speech tends towards greater clarity of vowel production as well as heightened pitch contour.

This capacity for the discrimination of categorical boundaries between vowels in speech not only approximates to the frequencies of Harmonics over fundamental frequency: it also permits the calibration of adjustments that render meaningful the speech of people whose accent, especially their production of vowel sounds, differs from our own. Figure 6 sets out a variety of English words whose meanings can only be discriminated by clear perception of their vowel content.

<Boot> has a vowel with little harmonic activation: that is the condition that allows us to recognise the word when it is uttered. Successively, each word in the left-hand column introduces harmonic content of higher frequency. We can both hear this in the result and, as a practical rehearsal in voice acoustics, *sense* the process in the different feeling associated with the production of each vowel in the vocal mechanism as we pronounce the words. The words in the central and right-hand columns indicate further variants to these sounds achieved by other means.

Acoustic analysis of the universal practice of Infant Directed Speech (Fernald 1992; Trevarthen 1999; Trehub 2001) illustrates that vowel sounds that map the varied formants of adult speakers approach more closely precise focal harmonics. This is as much a feature of the perceived ‘musicality’ of Infant Directed Speech as raised pitch levels, expressive contour, continuity of phonation and slowed tempo.

Discrimination of vowels is a reliable system on which perception of meaning crucially depends: I need to know whether you are talking about <bat>, <bet> or <bit>. Human response to the *gestalt* relationship between vowel qualities permits us to modify production so as, for instance, to deceive through imitating voices not our own, or through communicating an impression that is intended to convey or conceal our emotional or physical state. Similarly, we re-calibrate our perceptual mechanism in discerning whether a speaker has learned language in a manner that employs the distinction between vowels in the

**Figure 6. The Main Vowels in English**

Illustrated by words and names beginning with  
a **B** sound and ending with a **T** sound

Boot (English)	[Boo-wot (UK ‘geordie’)]
Boat	
Bought	
Bott	
Bart	Burt
Bat	Beart
Bet	But
Bate	
Bit	
Beet	
Boot (Scottish <Büt>)	Bute

same way that we do ourselves. So, I can tell when I listen to a South African speaker of English that, when he says, <bit>, he means what I call <bat> and not what I would myself pronounce as <bit>; while, as a UK-educated English speaker, if listening to an Australian I make a corresponding adjustment to the pronunciation <bet>.<sup>2</sup> Speakers can also manipulate these techniques to imitate accents other than their own. Our ability to make such adjustments in order to discern meaning where a Babel-like chaos would otherwise descend depends on a sophisticated acquired response to the properties of the harmonic series that is built on the innate properties of our evolved auditory mechanism (Figure 7).

### Figure 7. Vowels as Harmonics: music fossilised in speech?

If one performs the following set of vowel sounds as a sung glide from one to the next on a controlled monotone, one can learn to differentiate each specific sound in terms of a distinct higher pitch (Harmonic) produced simultaneously with the maintained pitch produced by the vocal folds. This is the basis of tonality in music, as well as of vowel discrimination in language.

oo – oa – or – o – ar – a – e – é – i – ee – ü

The foundation for discriminating differences between the formants that express each distinct vowel sound is detection of the whole-number ratio between  $F_0$  and the corresponding harmonic. Each vowel thus represents the focal resonance providing the cue for discrimination of the categorical boundaries between any harmonic and its neighbours.

That this is a *musical* phenomenon is captured in the means by which it is exploited in the range of responses to the properties of harmonics evident in Xhosa musical practice (see below).

Evidence for this transference to language of the musical properties of vocalisation occurs in the nonsense-phrases and onomatopoeia of children's speech and verbal games, especially in mnemonic pairs and sequences. No horse goes 'clap clip': in such expressions across various languages, the vowel with the 'high' harmonic always precedes the lower, as in a musical cadence. The following are examples of this: *ding dong, tick tock, clip clop, spicks and specks, riff-raff, pring prang*.

The latter is a category of children's singing game from Javanese Sunda (Romet 1992). 'Riff-raff' was first recorded in the Glossary of Anglo-Indian (1886) edited by Yule and Burnell, titled *Hobson-Jobson*.<sup>3</sup> Such acoustic pairings would seem to be widespread, and have their origins in the relationship between production and perception. Words with entirely unrelated meanings (or no meaning at all) are linked in the mind as a consequence of their articulatory demands. Proximal and dependent features of neurology give rise to a range of verbal features such as puns, malapropisms, spoonerisms and 'Freudian slips'. Such phenomena represent the revenge within the human mind of the acoustic, right-brain aspects of processing on the ordered generation of unambiguous meaning of the left. Play (as well as mistakes) associated with such productivity is arguably as characteristic of human vocalisation as are the precise utterances dissected by Chomsky.

### Transferability of parameters between language and music

The consequences of this capacity for sonic patterning and its perception can be detected in the overlap between aspects of musical and linguistic production in a variety of cultures. For instance, Dargie's (1988) comprehensive analysis of the employment of pitch systems in Xhosa vocalisation and instrumental practice illustrates how sensitivity to the properties of the Harmonic Series underpins a wide variety of behaviours in this culture. Although Dargie was unable to discern that his Xhosa informants possessed an explicit theory of music that could explain these phenomena, it is clear that implicit familiarity with the categorical boundaries of Harmonic Series relationships informed musical productivity. The activities presented in Figure 8 all represent responses in different media—language, various types of song, various techniques of instrumental performance, children's play—to the properties of the Harmonic Series. Dargie's account of these practices shows that

experience of one reinforces another in a network of socially and ritually determined skill acquisition. Similar relationships between vocalisation and instrumental performance exist in Australian aborigine music (Tarnopolsky *et al.* 2006), where, again, the didjeridu exploits the properties of the Harmonic Series; in the practice of vocalising *tabla* rhythmic patterns in Indian music; in the scat vocalisation of Louis Armstrong and Ella Fitzgerald, where the playing characteristics of instruments are often imitated vocally; and in varying examples of procedures for the combination of voices and instruments in every culture. Indeed, musical performance and pedagogy rely so commonly on the phenomenon of vocal-instrumental imitation that this, too, could be claimed to be universal. Similarly, the imitative capacities that link language and song to accurate and expressive animal mimicry have been studied in cultures as widely separated as New Guinea (Feld 1990), Central Africa (Lewis 2002) and Amazonia (Demolin 2006).

The social interaction derived from the learning of such practices within a culture such that they can be meaningfully imitated provides evidence of theory of mind. The implicit 'theories of harmony' that underpin these behaviours in Xhosa culture could well have arisen early in human evolution. Harnessed to a 'theory of mind', a theory of harmony acts as a bridge between musical vocalisation and the properties of categorical discrimination that are demanded by unambiguous language. The theory of mind that permits unison- and harmony-singing comprises a unique cognitive adaptation that could be seen as the candidate human equivalent to the capacity of birds instinctively to fly in flocks and fish to shoal for their collective protection. Precisely what shaped this behaviour, and when in our ancestral past, is a matter for speculation: but the consequences for the evolution of human culture are widespread, various and of great significance, and the phenomenon cries out for further investigation.

**Figure 8. Transferable features of an implicit theory of harmony**

Xhosa vocal, linguistic and instrumental practice (Dargie 1988)

vowels differentiated in speech	jew's harps, various varieties
parallel harmony in group singing	percussive bow playing
	continuous bow-playing
sub-harmonic singing	calabash resonance differentiation
employment of an impaled beetle as an oscillator to perform melodies through vocal filtering	

In historically recorded time, the harnessing of theory of mind to theory of harmony has given rise to profuse and varied elaboration within the medium of music, embracing phenomena such as the singing of chords and polyphony in which different performers contribute pitches derived from an agreed fundamental; singing in octave unison that exploits the complementary ranges of men and women, or men and children; and a range of musical structures and learning strategies that are associated with the practice of group singing on these lines. Theory of mind is evident in the way in which such practices are linked to rituals that mediate social relations and cope with death (Tolbert 2000). Just as Dargie found correspondences between vocal and instrumental practice amongst the Xhosa subjects he studied, one can suggest a vocal origin or foundation to a wide variety of

musical phenomena. On this basis, the Western orchestra is an extrasomatic choir, while performance on the Geissenklösterle flutes by their original makers some 35,000 years ago (Conard *et al.* 2004) can be seen as symbolic in its translation to a new pitch range of the capacity for efficient exploitation of exhalation, and of melodic relationships that could previously have been vocally performed only in a 'normal', or less costly, vocal range. Flute sounds extend the vocal phenotype to a degree consistent with perceived magical properties: they are both vocal, in that they can do what voices do in employing breath to shape sounds, and non-vocal, in that they achieve this in a musical range that voices do not normally inhabit. Nor should it escape notice that they could be constructed from a material, the wing bone of a water fowl, that was once itself part of an organism that was vocally alive and, at least on the evidence of the somewhat more recent burial of a child on a swan's wing at Vedbaek (Mithen 2003: 181), may have been associated with the transition to the spirit world.

### The selective mechanisms for voice acquisition

In conclusion, we can attempt to place the survey of vocal practices and their developmental features in the context of evolutionary theory. The selective pressures for the kinds of vocal advantages that represent the development of pre-human oral communication would have represented armaments in the war of the frequency spectrum. It seems likely that the initial acoustic adaptations that shaped a distinctly separate path leading eventually to our species were naturally-selective: an increased capacity to make more noise, or to do so over a longer duration, or in harness with co-participants, so as to evade predators and protect kin. The position outlined in this article focuses on the need to consider the properties of song, in addition to language, in any robust account of the development of human vocal communication. Figure 9 identifies three adaptive transitions that must have occurred between our last common ancestor with the chimpanzee and our modern human abilities.

#### Figure 9. The advantages of human-style vocalisation

Three critical adaptations for song (in which phylogeny is recapitulated in ontogeny)

1. Voluntary breathing  
(from pant-hoot to controlled exhalation).
2. Monotony  
(vocal control of  $F_0$  transcends emotional motivation).
3. Tonality  
( $F_0$  encoded in short-term memory permits structurally significant variation as well as summative group participation).

Future research needs to focus on the developmental and neurological features of voice acquisition, discriminating clearly between those properties that can be viewed as specifically linguistic and those that are, in a wider sense, musical. Vocal musicality needs to be studied throughout the human life-cycle, embracing the adaptive aspects of musical communication that feature in the music therapy and clinical literature (Morgan and Tilluckdharry 1982; Sacks 1985: 7-21; Aldridge 1996; El Mogharbel *et al.* 2003; Bannan and Montgomery-Smith 2008), especially where some humans who are unable to speak prove able to sing or whistle. Attention needs also to be paid to the features of musical

behaviour that are retained in the language production of modern humans, extending the existing work of researchers such as Tsunoda (1985), Fernald (1992), Scherer (1992), Deutsch (1999) and Welch (2000, 2001). Fernald's (1992) presentation of the acoustic similarities of infant cries and adult production of infant-directed speech across a variety of cultures and languages conveys a sense that the instinctive elements of vocal communication remain universal, and continue to play an instrumental part in language utterance. More radically still, Scherer's (1992) analysis of emotionally triggered sounds in a variety of species including humans reminds us that the animal call remains a potent underlying influence upon our productive and perceptive mechanisms. Welch (2000, 2001) has demonstrated the extent to which young singers retain an ability to vocalise musically until this becomes inhibited by the demands of language. Amongst others, he has illustrated that the condition referred to as 'tone-deafness' is extremely rare, and that people in Western-style societies who believe they cannot sing have frequently been inhibited from doing so in early life through embarrassing or negative social judgement. Deutsch's (1999) work illustrates the musical foundations of the way our auditory perception primes for language acquisition, tuning the way we hear ourselves and others in the processes that relate auditory memory to self-awareness; while Tsunoda's (1985) presentation of the way rhythmic separation in consonantless vowel strings endow meaning in Japanese suggests a lateral predisposition for such acoustic organisation in a manner that complements the tongue-twisters of consonant-rich languages such as English and Xhosa. The examples in Figure 10 convey a sense of the extent to which spoken language has song-like origins.

**Figure 10. Japanese phrases that employ only vowels, duration and prosody (after Tsunoda 1985)**

*ue o ui, o ooi, ai o ou, ai ue o*

[worried about hunger, concealing old age, he seeks love, a love-hungry man]

*u uu*

[a cormorant is hungry]

*ii i, ii i, ii ii*

[speaks of good rice and a good stomach]

*oo, oo o oo*

[oh let's chase the king]

*ooo, oooo, oo ooo*

[the courageous king conceals his tail when he goes out]



## Conclusions

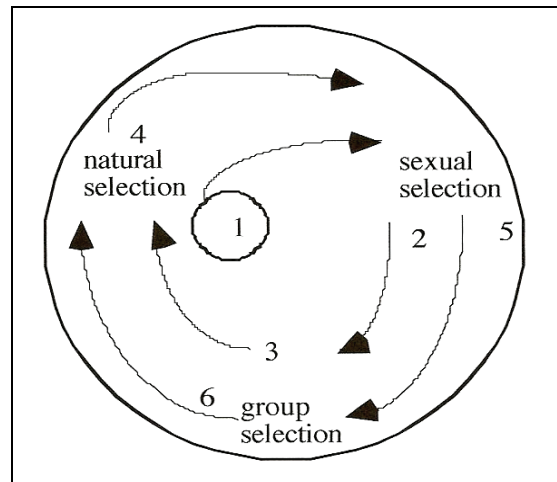
The variety of musical behaviours that have been reviewed in both musical and language vocalisation would suggest a mosaic pattern of vocal evolution over a considerable period of evolutionary time (Foley in press). Some behaviours suggest a response to the need for calming and introspection; others for alertness and arousal. Some are associated with privacy, others with public display; some with adulthood; others with childhood; others with relations across generations. It is difficult to imagine that a single macro-mutation could have given rise to this array of inconsistent behaviours: rather, human musical ability would seem to have arisen as a suite of characteristics shaped by the varied and complex social and environmental challenges faced by our species over the extensive period of its gradual evolution. The means by which each element in this suite of behaviours may have developed is suggested in Figure 11, distinguishing between various selective mechanisms.

### Figure 11 Models for Vocal Evolution

- Natural Selection (the anatomical arms race).
- Sexual Selection (mate choice and enhanced genetic inheritance).
- Group Selection (evolved behavioural characteristics including synchrony and altruism).
- A synthesis dependent on all of these over evolutionary time (exaptations emerging as adaptations).

A mosaic pattern on the lines depicted in Figure 12, running through successive iterations, would admit musically-organised human vocalisation as a set of candidate behaviours able to act as a developmental bridge between animal calls and modern human language. The possibility that song could have provided a system of both affective and referential communication able to sustain the cognitive and emotional exchanges of an ancestor species such as *Homo Erectus*, let alone Mithen's Singing Neanderthals, is consistent with a tantalising suggestion proposed by Dawkins (2004: 126). In establishing his theory that we should look to the capacities of common ancestors for developmental features, rather than exclusively making direct comparisons with other surviving species

**Figure 12. The integration of selective models for vocal development**



today, Dawkins speculates on whether modern humans may have more in common with their distant common ancestor the gibbons (18 million years ago), whose vocal behaviour and its social purpose may prefigure certain human traits, than with more recent, African common ancestors such as with the chimpanzee (*ca* 8 mya), in whose daughter species the capacity for vocalisation and patterns of social interaction may have diverged from this older common line.

Although it is unarguable that modern human vocalicity in both speech and song depends principally upon the evolved powers and architecture of the brain, it should remain on the agenda of research in neurology, zoology, language studies and anthropology to consider whether the critical neural adaptations on which these communicative behaviours depend were not themselves shaped by a response to existing anatomical advantages that would be consistent with a form of musical proto-language on lines first argued by Darwin.

### Acknowledgements

I wish to record my gratitude to the three anonymous referees whose just but constructive criticism of my original submission allowed me to review its argument and shape it for the readership of this publication.

### Notes

1. Dargie (1988) provides the example in Xhosa children's musical play of live beetles impaled on a stick that is placed against the jaw to act as an oscillator whose vibrations are amplified and harmonically varied by movements of the throat and lips. Similarly, the Australian didgeridu represents an instrument that requires human ingenuity to play, but is actually constructed by the action of termites.
2. As a native English speaker whose language was acquired through a childhood in Scotland, Wales and Kent, this is how I 'hear' these relationships. Since moving to Australia, I recognise that the aural 'point of view' of native Australian English speakers is quite different: this was referred to by one of the reviewers of the original submission of this paper. Far from invalidating the point I am making, this divergence of perception supports it: each of us depends on this set of harmonic-series approximations to locate in the relative vowel production of others the characteristic differences between the way they speak and the way we normally do ourselves. This feature of human auditory perception and the link between linguistic and musical properties has been thoroughly explored in the work of Diana Deutsch (see Deutsch 1999; Ragozzine and Deutsch 1994).
3. 'Hobson-Jobson', the term for Anglo-Indian patois, is itself reputed to be an anglicised version of the Islamic oath 'Ya Hasan! Ya Hosain!' (McCrum, Cran and MacNeil 1986).

### References

- Abitbol J., J. de Brux, G. Millot, M-F Masson, O. Languille Mimoun, H. Pau and B. Abitol 1989. Does a hormonal vocal cord cycle exist in women? Study of vocal premenstrual syndrome in voice performers by videostroboscopy-glottography and cytology on 38 women. *Journal of Voice* 3(2): 157-162.
- Aldridge, D. 1996. *Music Therapy Research and Practice in Medicine: From Out of the Silence*. London: Jessica Kingsley.
- Baily, J. 2004. Music censorship in Afghanistan before and after the Taliban. In M. Korpe (ed.) *Shoot the Singer! Music Censorship Today*, pp.19-28. London: Zed Books.
- Bannan, N. and C. Montgomery-Smith 2008. 'Singing for the brain': reflections on the human capacity for music arising from a pilot study of group singing with Alzheimer's patients and carers. *Journal of the Royal Society for the Promotion of Health* 128(2): 73-78.
- Bickerton, D. 1991. *Language and Species*. Chicago, IL: University of Chicago Press.

- Blacking, J. 1987. *A Commonsense View of all Music: Reflections on Percy Grainger's Contribution to Ethnomusicology and Music Education*. Cambridge: Cambridge University Press.
- Brailoui, C. 1984 [1949]. *Problems of Ethnomusicology*. A. L. Lloyd (ed. and trans.). Cambridge: Cambridge University Press.
- Brown, D. E. 1991. *Human Universals*. New York, NY: McGraw-Hill.
- Brown, S. 2007. Contagious heterophony: a new theory about the origins of music. *Musica Scientiae* 11(1): 3-26.
- Carlton, R. 1996. Man as a musical animal. *International Review of the Aesthetics and Sociology of Music (IRASM)* 27(1): 79-86.
- Chatwin, B. 1987. *The Songlines*. London: Jonathan Cape.
- Chomsky, N. 1957. *Syntactic Structures*. The Hague: Mouton.
- Classe, A. 1957. Phonetics of the Silbo Gomero. *Archivum Linguisticum* ix: 43-61.
- Conard, N., M. Malina, S. Münzel and F. Seeberger 2004. Eine Mammutfelßenbeinflöte aus dem Jungpaläolithikum auf der Schwäbischen Alb. *Archäologisches Korrespondenzblatt* 34(4): 447-462.
- Cook, G. 2000. *Language Play, Language Learning*. Oxford: Oxford University Press.
- Cooksey, J. 1997. Voice transformation in male adolescents. In L. Thurman and G. Welch (eds) *Bodymind and Voice: Foundations of Voice Education*, pp. 405-515. Minneapolis, MN: The VoiceCare Network.
- Cox, A. 2001. The mimetic hypothesis and embodied musical meaning. *Musica Scientiae* 5(2): 195-212.
- Cross, I. and A. Watson 2006. Acoustics and the human experience of socially organised sound. In C. Scarre and G. Lawson (eds), *Acoustics, Space and Intentionality: Identifying Intentionality in the Ancient Use of Acoustic Spaces and Structures*, pp.107-116. Cambridge: McDonald Institute for Archaeological Research.
- Dargie, D. 1988. *Xhosa Music: Its Techniques and Instruments, with a Collection of Songs*. Cape Town: David Phillip.
- Darwin, C. 1871. *The Descent of Man, and Selection in Relation to Sex*. London: John Murray.
- Darwin, C. 1859/1968. *The Origin of Species*. London: Penguin.
- Dawkins, R. 2004. *The Ancestor's Tale*. London: Weidenfeld and Nicholson.
- Day, T. 1992. *Why Catholics Can't Sing: The Culture of Catholicism and the Triumph of Bad Taste*. New York: Crossroad.
- Deacon, T. 1997. *The Symbolic Species*. London: Allen Lane.
- Demolin, D. 2006. The evolution of language and music in a comparative perspective. Paper presented at the *Cradle of Language* conference, University of Stellenbosch, South Africa, 6-10 November 2006. Abstract available at: URL <http://www.clc.sun.ac.za/Programme.htm>
- Deutsch, D. 1999. *The Psychology of Music*. London: Academic Press.
- Donald, M. 1991. *Origins of the Modern Mind: Three Stages in the Evolution of Culture and Cognition*. Cambridge, MA: Harvard University Press.
- Ellis, C. 1980. Australia, §II, 3: Southern Aboriginal music. In S. Sadie (ed.) *The New Grove Dictionary of Music and Musicians*, Vol. 1, pp. 722-8. London: Macmillan.
- Ellis, C. 2001. Song for ages and stages, in *Australia 2, Central Aboriginal music ii*. L. Macy (ed.) *Grove Music Online* URL <http://www.grovemusic.com>
- El Mogharbel, C., I. Laufs, M. Wenglorz and W. Deutsch 2003. The sounds of songs without words. In R. Kopiez, A. Lehmann, I. Wolther and C. Wolf (eds) *Proceedings of the Fifth Triennial Conference of the European Society for the Cognitive Sciences of Music*. Hanover: Institute for Research in Music Education Monograph No. 6. CD-ROM.
- Espi-Sanchis, P. and N. Bannan in press. Found objects in the musical practices of hunter-gatherers: implications for the evolution of instrumental music. In N. Bannan and S. Mithen (eds), *Music, Language and Human Evolution*. Oxford: Oxford University Press.
- Feld, S. 1990. *Sound and Sentiment: Birds, Weeping, Poetics and Song in Kaluli Expression*. Philadelphia, PA: University of Pennsylvania Press.
- Fernald, A. 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: Éditions de la Maison des Sciences de l'Homme/Cambridge University Press.

- Fitch, W. T. 2000. The evolution of speech: a comparative review. *Trends in Cognitive Sciences* 4(7): 258-267.
- Foley, R. in press. Music and mosaics: the evolution of human abilities. In N. Bannan and S. Mithen (eds), *Music, Language and Human Evolution*. Oxford: Oxford University Press.
- Gamble, C. in press. When the words dry up: music and material metaphors half a million years ago. In N. Bannan and S. Mithen (eds), *Music, Language and Human Evolution*. Oxford: Oxford University Press.
- Garfinkel, Y. 1998. Dancing and the beginning of art scenes in the early village communities of the Near East and Southeast Europe. *Cambridge Archaeological Journal* 8(2): 207-237.
- Garfinkel, Y. 2001. Dancing or fighting? A recently discovered predynastic scene from Abydos, Egypt. *Cambridge Archaeological Journal* 11(2): 241-254.
- Garfinkel, Y. 2003. *Dancing at the Dawn of Agriculture*. Austin, TX: University of Texas Press.
- Grice, H. P. 1969. Utterers' meaning and intentions. *Philosophical Review* 8(2): 147-177.
- Harker, D. 1985. *Fakesong: The Manufacture of British 'Folksong' 1700 to the Present Day*. Milton Keynes: Open University Press.
- Hauser, M. D. 1997. *The Evolution of Communication*. Cambridge, MA: MIT Press.
- Hickmann, E. 2006. Archaeomusicology. In L. Macy (ed) *Grove Music Online*. URL <http://www.grovemusic.com>
- Hines, J. 1983. *Great Singers on Great Singing*. London: Gollancz.
- Hockett, C. F. 1960. The origin of speech. *Scientific American* 203: 89-96.
- Hodges, D. 2000. Why are we musical? Support for an evolutionary theory of human musicality. In C. Woods, G. Luck, R. Brochard, F. Seddon and J. A. Sloboda (eds) *Proceedings of the Sixth International Conference on Musical Perception and Cognition*. CD-ROM, Department of Psychology, University of Keele, UK.
- Hollenweger, W. J. n.d. Singing in tongues. In L. Macy (ed.) *Grove Music Online*. URL <http://www.oxfordmusiconline.com>
- Howard, D. 2004. Measuring the tuning accuracy of thousands singing in unison: an English premier football league table of fans' singing tunefulness. *Logopedics Phoniatics Vocology* 29(2): 77-83.
- Kunej, D. and I. Turk 2000. New perspectives on the beginnings of music: archaeological and musical analysis of a Middle Paleolithic bone 'flute'. In N. Wallin, B. Merker and S. Brown (eds) *The Origins of Music*, pp.235-68. Cambridge, MA: MIT Press.
- Lã, F. and J. W. Davidson 2005. Investigating the relationship between sexual hormones and female Western classical singing. *Research Studies in Music Education* 24: 75-87.
- Lecanuet, J.-P. 1996. Prenatal auditory experience. In I. Deliège and J. Sloboda (eds) *Musical Beginnings: Origins and Development of Musical Competence*, pp.3-34. Oxford: Oxford University Press.
- Levman, B. 2000. Western theories of music origin, historical and modern. *Music Scientiae* IV(2): 185-212.
- Lewis, J. 2002. Forest Hunter-Gatherers and their World: A Study of the Mbendjele Yaka Pygmies of Congo-Brazzaville and their Secular and Religious Activities and Representations. PhD Thesis, Dept. of Social Anthropology, London School of Economics.
- Locke, J. L. 2006. Parental selection of vocal behaviour. *Human Nature* 17(2): 155-168.
- McCrum, R., W. Cran and R. MacNeil 1986. *The Story of English*. London: Faber and Faber.
- Merker, B. 2000. Synchronous chorusing and human origins. In N. Wallin, B. Merker and S. Brown (eds) *The Origins of Music*, pp.315-327. Cambridge, MA: MIT Press.
- Miller, G. 2000. *The Mating Mind: How Sexual Choice Shaped the Evolution of Human Nature*. New York: Doubleday.
- Mithen, S. 2003. *After the Ice: A Global Human History 2,000-5,000 BC*. London: Weidenfeld and Nicholson.
- Mithen, S. 2005. *The Singing Neanderthals*. London: Weidenfeld and Nicholson.
- Morgan, O. S. and R. Tilluckdharry 1982. Presentation of singing function in severe aphasia. *West Indian Medical Journal* 31: 159-61.
- Morley, I. 2006. Mousterian musicianship? The case of the Dive Babe I bone. *Oxford Journal of Archaeology* 25(4): 317-333.
- Moyle, A. M. 1980. Australia, §II, 2: Northern Aboriginal music. In S. Sadie (ed.) *The New Grove Dictionary of Music and Musicians*, vol. 1, pp.713-721. London: Macmillan.

- Nattiez, J-J. 1990. *Music and Discourse: Towards a Semiology of Music*. Trans. C. Abbate. Princeton, NJ: Princeton University Press.
- Nattiez, J-J. 1999. Inuit throat games. *Ethnomusicology* 43(3): 399-418.
- Nattiez, J-J. 2000. Linguistic models and the analysis of musical structures. *Rivista Italiana di Musicologia* 35(1-2): 379-410. Trans. J. Dunsby.
- Nelson, J. 1997. Lullabies as human adaptation: a cross-cultural analysis of children's bedtime songs. In D. Schneck and J. Schneck (eds) *Music in Human Adaptation*, pp.61-78. Blacksburg, VA: Virginia Polytechnic Institute and State University Press.
- Nottebohm, F. 1975. A zoologist's view of some language phenomena, with particular emphasis on vocal learning. In E. H. Lenneberg and E. Lenneberg (eds) *Foundations of Language Development*, pp.61-103. New York, NY: Academic Press.
- Opie, I. and P. Opie 1985. *The Singing Game*. Oxford: Oxford University Press.
- Oubré, A. 1997. *Instinct and Revelation: Reflections on the Origins of Numinous Perception*. Newark, NJ: Gordon and Breach.
- Page, C. 1987. *Voices and Instruments of the Middle Ages: Instrumental Practice and Songs in France 1100-1300*. London: J. M. Dent.
- Papousek, M. 1992. Early ontogeny of vocal communication in parent-infant interactions. In H. Papousek, U. Jürgens and M. Papousek (eds) *Nonverbal Vocal Communication: Comparative and Developmental Approaches*, pp.230-261. Paris and Cambridge: Éditions de la Maison des Sciences de l'Homme/Cambridge University Press.
- Parsons, L. 2001. Exploring the functional neuroanatomy of music performance, perception, and comprehension. *Annals of the New York Academy of Sciences* 930: 211-231.
- Pinker, S. 1997. *How the Mind Works*. New York, NY: W. W. Norton.
- Potter, J. 2000. *The Cambridge Companion to Singing*. Cambridge: Cambridge University Press.
- Premack, D. 1986. 'Gavagai!' or the Future History of the Animal Language Controversy. Cambridge, MA: MIT Press.
- Ragozzine, F. and D. Deutsch 1994. A regional difference in perception of the Tritone Paradox within the United States. *Music Perception* 12(2): 213-225.
- Ridington, R. 1991. Northern hunters. In A. M. Josephy (ed.) *America in 1492: The World of the Indian Peoples Before the Arrival of Columbus*, pp.21-47. New York, NY: Vintage Books.
- Romet, C. 1992. Song acquisition in culture: A West Javanese study in children's song development. In H. Lees (ed.) *Music Education: Sharing Musics of the World. Proceedings of the 20th World Conference of the International Society for Music Education, Seoul, Korea*, pp.164-173. Christchurch, NZ: ISME/University of Canterbury.
- Rousseau, J-J. 1966 [1761]. Essay on the origin of languages. In J. H. Moran and A. Gode (eds and trans.) *On the Origin of Language*, pp.3-73. New York, NY: Frederick Ungar.
- Rowley, I. 1973. The comparative ecology of Australian corvids II: social organisation and behaviour. *CSIRO Wildlife Research* 18(1): 25-65.
- Sacks, O. 1985. *The Man who Mistook his Wife for a Hat*. London: Duckworth.
- Saussure, F. de 1959 [1916]. *Course in General Linguistics*. New York, NY: McGraw-Hill.
- Scherer, K. R. 1992. Vocal affect expression as symptom, symbol and appeal. In H. Papousek, U. Jürgens and M. Papousek (eds) *Nonverbal Vocal Communication: Comparative and Developmental Approaches*, pp.43-60. Paris and Cambridge: Editions de la Maison des Sciences de l'Homme/Cambridge University Press.
- Schleifer, E. 2001. Biblical cantillation. Jewish music, §III, 2: synagogue music and its development (ii). In L. Macy (ed.) *Grove Music Online* URL <http://www.oxfordmusiconline.com>
- Schneider, A. 1986. ΑΡΧΑΙΟΛΟΓΕΩ. Some comments on methods and sources in music archaeology. In C. S. Lund (ed.) *Second Conference of the ICTM Study Group on Music Archaeology*, pp.195-224. Stockholm: Royal Swedish Academy of Music.
- Shetler, D. J. 1990. The inquiry into prenatal musical experience. In F. R. Wilson and F. L. Roehamann (eds) *Music and Child Development*, pp.44-62. St. Louis, MO: MMB Music.
- Soltis, J. 2004. The signal functions of early infant crying. *Behaviour and Brain Sciences* 27(4): 443-490.
- Stockmann, D. 1986. On the early history of drums and drumming in Europe and the Mediterranean. In C. S. Lund (ed.) *Second Conference of the ICTM Study Group on Music Archaeology*, pp.11-28. Stockholm: Royal Swedish Academy of Music.
- Sundberg, J. 1999. The perception of singing. In D. Deutsch (ed.) *The Psychology of Music*, pp.171-214. London: Academic Press.

- Tarnopolsky, A., N. Fletcher, L. Hollenberg, B. Lange, J. Smith and J. Wolfe 2006. Vocal tract resonances and the sound of the Australian didjeridu (yidaki) I. Experiment. *Journal of the Acoustic Society of America* 119(2): 1194-1204.
- Thrills, A. 1998. *You're Not Singing Anymore*. London: Ebury Press.
- Thomas, D. A. 1995. *Music and the Origins of Language: Theories from the French Enlightenment*. Cambridge: Cambridge University Press.
- Tolbert, E. 2000. Music and meaning: an evolutionary story. In C. Woods, G. Luck, R. Brochard, F. Seddon and J. A. Sloboda (eds) *Proceedings of the Sixth International Conference on Musical Perception and Cognition*. CD-ROM, Department of Psychology, University of Keele, UK.
- Tinbergen, N. 1951. *The Study of Instinct*. Oxford: Clarendon Press.
- Toop, D. 2000. The evolving language of rap. In J. Potter (ed.) *The Cambridge Companion to Singing*, pp.42-52. Cambridge: Cambridge University Press.
- Tran Quang, H. and N. Bannan in press. Vocal traditions of the world: towards an evolutionary account of voice production in music. In N. Bannan and S. Mithen (eds) *Music, Language and Human Evolution*. Oxford: Oxford University Press.
- Trehub, S. E. 2001. Musical predispositions in infancy. In R. J. Zatorre and I. Peretz (eds) *The Biological Foundations of Music*, pp.11-16. Annals of the New York Academy of Sciences, Vol. 930.
- Trevarthen, C. 1994. Infant semiosis. In W. Nöth (ed.) *Origins of Semiosis*, pp.219-252. Berlin: Mouton de Gruyter.
- Trevarthen, C. 1999. Musicality and the intrinsic motor pulse: evidence from human psychobiology and infant communication. *Musica Scientiae* Special Issue: Rhythm, Musical Narrative and Origins of Human Communication, pp.155-211. Brussels: ESCOM
- Tsunoda, T. 1985. *The Japanese Brain*. Tokyo: Taishukan.
- Vico, G. 1976 [1744]. *The New Science of Giambattista Vico*. Trans. T. G. Bergin and M. H. Fisch. New York, NY: Cornell University Press.
- Walker, R. 1990. *Musical Beliefs: Psychoacoustic, Mythical and Educational Perspectives*. New York, NY: Teachers College Press.
- Wallin, N. 1991. *Biomusicology: Neurophysiological, Neuropsychological and Evolutionary Perspectives on the Origin and Purposes of Music*. Stuyvesant, NY: Pendragon Press.
- Watson, A. and D. Keating 1999. Architecture and sound: an acoustic analysis of megalithic monuments in prehistoric Britain. *Antiquity* 73: 325-336.
- Welch, G. 2000. The genesis of singing behaviour. In C. Woods, G. Luck, R. Brochard, F. Seddon and J. A. Sloboda (eds), *Proceedings of the Sixth International Conference on Musical Perception and Cognition*. CD-ROM, Department of Psychology, University of Keele, UK.
- Welch, G. 2001. *The Misunderstanding of Music*. London: Institute of Education, University of London.
- Wishart, T. 1979. *The Book of Lost Voices*. York: University of York.
- Wishart, T. 1996. *On Sonic Art*. Amsterdam: Harwood Academic Press.
- Woodward, S. 1992. The Transmission of Music into the Human Uterus and the Response to Music of the Human Fetus and Neonate. PhD Thesis, University of Cape Town, South Africa.
- Wray, A. 2002. *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.
- Wright, A. A., J. J. Rivera, S. H. Hulse, M. Shyan and J. J. Neiworth 2000. Music perception and octave generalization in rhesus monkeys. *Journal of Experimental Psychology: General* 129(3): 291-307.
- Yule, H. and A. C. Burnell 1979 [1868]. *Hobson-Jobson: A Glossary of Anglo-Indian Words and Phrases, and of Kindred Terms, Etymological, Historical, Geographical and Discursive*. New Delhi: Munshiram Manharlal.