Darwin College Research Report

DCRR-002

The Evolutionary Origins and Archaeology of Music

Iain Morley

October 2003
(electronic edition 12 January 2006)

Darwin College
Cambridge University
United Kingdom CB3 9EU
www.dar.cam.ac.uk/dcrr

ISSN 1749-9194

THE EVOLUTIONARY ORIGINS AND ARCHAEOLOGY OF MUSIC

or

An Investigation into the Prehistory of Human Musical Capacities and Behaviours, Using Archaeological, Anthropological, Cognitive and Behavioural Evidence

by

Iain Morley ***B.Sc. M.A. Ph.D.*****,**
Trinity Hall,
Cambridge

Originally submitted to the Faculty of Archaeology and Anthropology
in candidacy for admission to the degree of
Doctor of Philosophy of Cambridge University

October 2003

ABSTRACT

The Evolutionary Origins and Archaeology of Music: An Investigation into the Prehistory of Human Musical Capacities and Behaviours

The research presented in this dissertation examines the evidence for the emergence of the capacities underlying musical behaviours, their interrelationship, development and ultimate manifestation in the Palaeolithic. A multidisciplinary approach is taken, and the dissertation falls into four main sections.

Section I reviews and analyses ethnographic evidence regarding the use of music in four hunter-gatherer societies. It highlights a number of fundamental similarities between their musics, suggesting shared heritage or convergent development. In these traditions melody is predominantly vocal; instruments are primarily percussive, made of organic materials, and thus unlikely to leave an archaeological trace.

Section II comprises a comprehensive synthesis of the archaeological evidence for use of musical instruments in the Palaeolithic. The earliest currently known instruments date to c. 36,000 years ago; the evidence suggests that when modern humans arrived in Europe they were already carrying out developed instrumental behaviours. The diversity of evidence occurring after 30,000 years ago suggests that musical performance, or at least appreciation, was a group activity, rather than one involving a select few.

Section III examines the physiological and neurological apparatus involved in the production and processing of musical functions, and their evolution. From *Homo ergaster* there was co-evolution of various physiological and neurological capacities necessary to carry out vocalizations of increasing tonal range and duration. “Higher” analytical musical and linguistic functions emerged out of shared substrates concerned with vocal emotional expression and comprehension.

Section IV analyses the evolutionary bases for musical capacities and their integration. It shows that music, emotional expression and social behaviours, rhythmic and melodic abilities are fundamentally connected and share a heritage with linguistic prosody and corporeal control. Developed skills in these areas have potential selective benefits.

A synthesis of the findings from these various disciplines and sources of evidence concludes the thesis, proposing that musical capacities have their foundations in inter-specific tonal emotional vocal expression, and rhythmic-motor coordination of corporeal musculature in the execution of such vocalisations. These increased in complexity throughout the *Homo* lineage, and diverged from linguistic capacities with the development of lexicon and syntax; symbolic associations and diversity occurred with *Homo sapiens*, who were carrying out sophisticated instrumental musical behaviours upon their arrival in Europe.

Dedication

This thesis is dedicated to Lindsay MacLean Roffe, *in memoriam*, whose generosity has had, and will continue to have, a great influence on my life.

CONTENTS

INTRODUCTION

SECTION I
ETHNOGRAPHIC EVIDENCE FOR THE USE OF MUSIC

SECTION II
PALAEOLITHIC MUSIC ARCHAEOLOGY

SECTION III
EVOLUTION OF THE PHYSIOLOGICAL AND NEUROLOGICAL CAPACITIES FOR MUSIC

SECTION IV
EVOLUTIONARY RATIONALES FOR MUSIC

FIGURES

Acknowledgements

Over the course of this piece of research, I have come to owe a great deal, in many different ways, to a number of people. Foremost of these are Dr. Ian Cross, of the Faculty of Music and Wolfson College, and Dr. Preston Miracle, of the Department of Archaeology and St. John's College. Both have been model supervisors over the course of the last three years, very generous with their time and intellects, unfailingly helpful, friendly and stimulating, and inclusive in their academic pursuits. Their influences on the present work have been quite different, but entirely complementary; their influences on me in the form of the example they have set as professional academics have been formative. I could not have been better supervised.

Liaising between them, and coordinating in a way that is the marvel of the department, has been Jane Woods, instrumental in the institutional progress of this research, as well as being unstinting in her help and cheer. David Redhouse deserves thanks for his speedy and effective attention to computing issues. Sincere thanks are also due to Dr. Graeme Lawson and Dr. Chris Scarre, both of the MacDonald Institute for Archaeological Research, who kindly invited me to participate in their excellent conference on *Identifying intentionality in ancient use of acoustic space and structure* (27th-29th June 2003). I also owe both, independently, a debt of gratitude for other very valuable assistance provided at short notice, and for stimulating discussions.

Professor Paul Mellars, Department of Archaeology and Corpus Christi College, has been very generous in his support on many occasions, which has been sincerely appreciated. Timely interest and advice from Dr. Rob Foley, Department of Biological Anthropology and King's College, is also gratefully acknowledged. The input and supervision of Professor Stephen Mithen, University of Reading, in the earliest stages of my investigations into this area remains valuable.

Other important support in the early stages of the research was provided by Dr. Caroline Gale, without which the progress of this project would probably have been compromised. I owe appreciation to Dr. Paul Bahn, Professor Fred Coolidge, and Professor John Skoyles for their munificent provision of papers and correspondence at salient times.

I am extremely grateful to Trinity Hall for the Scholarship that I was awarded for this research, and especially to Dr. Christopher Padfield, Graduate Tutor, and Julie Powley, Graduate Officer, whose support with respect to funding, board and lodging has contributed significantly to the quality of my life, intellectual and social. On this count I am also indebted to the Master, Professor Peter Clarke, Fellows, Porters and many of my fellow members of Trinity Hall; my time thus far as a member of the college has been enormously rewarding and entertaining, and has provided more opportunities than any other time in my previous career. I shall always be proud to be a member of The Hall.

Obviously, a piece of work that constitutes the dominant component of ones' *raison d'etre* for three years will not be influenced purely by intellectual and financial input. There are many people who have directly contributed to making my time in Cambridge extremely enjoyable and valuable, in a wide variety of ways.

Much entertaining time has been spent with Laura Pugsley, Lisa Marlow, Masa Mlakar, Nicky Milner, William Fletcher, Helen Farr, Grahame Appleby, Jo Wilson, Krish Seetah, Kevin Lane and Alex Herrera, as well as Jessica Rippengal, Bryan Hanks, Dusan Boric and Emma Jenkins in the Grahame Clark Laboratory, and many other members of the Archaeology and Anthropology faculty who will know who they are. Of particular note are Preston and Lada Miracle, who have hosted several highly enjoyable parties with, on occasion, some significant forbearance. Sacha Jones, Eeva Vaanenen, Helen Bartlett, Rebecca Clamp, Sarah Cross and Gemma Fermer greatly enriched my time here. Likewise Gregory Davies, Tahl Holtzman, Jackson Armstrong, Richard Levett, Patrick Flynn and the other honourable brethren, plus Chris Handy, Dr. Simon MacLean and Dr. Rob Miller, none of whom underestimate the value of excellent conversation, beverages, and the appreciation of other fine things. My crewmates and other friends in Trinity Hall Boat Club have also been the source of many entertaining diversions.

In every sense, I would not have been here were it not for my parents, Michael and Mary. They remain an unfailing and generous source of inspiration, guidance (solicited and by example), fun, love and affection. Likewise, my sister, Ginette, and my brother, Alex, who remain not only great siblings, but are also great friends.

This research has been funded by a Scholarship from Trinity Hall, as noted, and in the latter two years by a Research Studentship from the Arts and Humanities Research Board; both contributions were essential. I am also grateful to the Wingate Foundation for the Scholarship they awarded me for this project; although I took it *in honorarium*, they have continued to be munificent and inclusive. Travel grants have been received from the Dorothy Garrod Trust, in the Department of Archaeology, and from Trinity Hall, for which I remain very grateful.

Retrospectively, I would also like to thank my examiners, Professor Paul Mellars and Professor Leslie Aiello, for their encouragement and support both at the time of my examination and since.

Declaration

This dissertation is the outcome of my own work, and includes nothing which was prepared in collaboration.

INTRODUCTION

“Since music is the only language with the contradictory attributes of being
at once intelligible and untranslatable, the musical creator is being comparable
to the gods, and music itself the supreme mystery of the science of man.”
(Levi-Strauss, 1970)

The state of the art and the aims of the present research

In the last ten years in particular, there has been a surge in interest and research into the evolution of the human brain and the development of human cognitive abilities. This research has spanned a number of disciplines, including Archaeology, Bioanthropology, Psychology, Philosophy and Linguistics, as well as generating several dedicated texts (e.g. Cartwright and Davies, 2000; Barrett *et al.*, 2001). Many areas of cognition have received attention in the context of this research, such as cultural and sexual behaviours (e.g. Dunbar *et al.*, 1999), intelligence and cognitive processing systems (e.g. Barkow *et al.*, 1992; Byrne 1995; Carter 1998; Deacon 1997; Dennett 1998), and the focus of a great deal of interest, the evolution of human language abilities (e.g. Aiello and Dunbar, 1993; Pinker, 1997; Jablonski and Aiello, 1998; Dunbar, 1998). Work in cognitive archaeology (e.g. Ingold & Gibson 1993; Mellars and Gibson 1996; Mithen 1996) has paralleled this research, drawing upon the fossil and archaeological record to make inferences about the development and nature of intelligence and cognition in human ancestors.

Within this body of research, however, one topic has been neglected by all but a few (e.g. Cross, 1999b, 2003b; Wallin *et al.*, 2000; D’Errico *et al.*, 2003): the evolution of musical abilities. It is only very recently that any unified approach to the issue has been attempted (Wallin, Merker and Brown, 2000), and then in an edited volume. Neuropsychological and developmental studies suggest that the human musical ability has a deep evolutionary history (e.g. Wallin 1991; Carter 1998) but this contrasts with the evidence (discovered to date) from the archaeological record (e.g. Scothern, 1992; Lawson *et al.*, 1998; Lawson and d’Errico, 2002), in which musical instruments and evidence for the possible use of acoustics do not appear until after 40,000 years ago. This is at least 80,000 years after the first anatomically modern humans appeared and a great deal longer after the emergence of various other important cognitive capabilities.

The research presented in this dissertation aims to make a substantial contribution to the further development of cognitive archaeology and to our understanding of human musical ability. It examines the evidence for the emergence of the capacities underlying musical behaviours, their interrelationship, development and ultimate manifestation in the

Palaeolithic. The aim is to identify when musical capacities first developed, potential selective reasons for these developments within human ancestors, functional and cognitive links between the earliest language abilities and musical abilities, evolutionary rationales for human emotional response to music, and the possible nature of and reasons for the earliest use of music. To these ends, a multidisciplinary approach is taken, synthesising evidence from a broad diversity of fields, including palaeoanthropology, archaeology, ethnomusicology, neuroscience, developmental and social psychology and evolutionary biology. This research has four main strands, and the dissertation accordingly falls into four main sections.

The organisation of the dissertation

Section I reviews and analyses ethnographic evidence for the use of music in four hunter-gatherer societies: Plains Indians of America, Pygmies of the African jungle, Australian Aborigines of the Western Desert and Eskimos of the Arctic tundra. This section examines the nature of the music, its importance and role in society, and the range and nature of instruments used (especially of those unlikely to leave an archaeological trace). As well as identifying possible forms of instruments made from naturally occurring resources that may have been lost to the archaeological record, this section aims to identify potential uses of music amongst peoples who subsist by hunting and gathering.

Section II comprises a comprehensive synthesis of the archaeological evidence for use of musical instruments in the Palaeolithic. Previous investigations in this area have often dealt only with parts of the extant evidence; my investigation looks at forms of instruments and sound production, the spatial and temporal distribution of the evidence, and proposes models for the nature of musical practice in the Upper Palaeolithic.

Section III consists of an examination of the physiological apparatus involved in the production and processing of sound, and an analysis of our knowledge from palaeoanthropological evidence for the evolution of the neurology and physiology identified as dedicated to musical ability. This is followed by an examination of neuroscientific evidence for distinct regions of the brain which may be responsible for musical ability, and how they are interrelated. This allows conclusions to be drawn concerning when musical capabilities may have begun to emerge and evolve over time.

Section IV analyses the evolutionary bases for musical capacities and their integration. Issues addressed include the connections between music, emotion and social behaviours, the interrelationships between rhythmic and melodic abilities, evidence of innate capacities for music production and processing, and relationships between linguistic and musical abilities (especially in early stages of cognitive development). This is followed by an examination of roles that developed musical behaviours might have played in human evolution, and a consideration of the ways in which the development of musical behaviours fits with models of human cognitive evolution.

Each area of investigation undertaken in the present study can be shown to illuminate the others. A synthesis of the findings from these various disciplines and sources of evidence concludes the thesis, with the intention of addressing the aims stipulated above, and thus to contribute to an understanding of the origins and development of musical capacities in the human lineage.

A note on the definition of music

It is widely asserted in literature dealing with music psychology and anthropology that all cultures and societies have music (e.g. Clynes, 1982; Storr, 1992; Brown *et al*, 2000), but definitions of what constitutes music are few and far between. It is perhaps indicative of a sense that musical knowledge is somewhat intuitive that none of these authors consider it necessary to define the term. In studies of music cognition and psychology, music tends either to be dealt with holistically in terms of (Western) composed pieces or, in contrast, as components of music, such as discrete pitches or transitions; neither requires a definition of music to be specified. Developmental studies are also generally concerned with the emergence of *elements* of music perception and production, or of production and perception of Western music, again bypassing the necessity of defining the entity as a whole. Palaeolithic music archaeology has generally simply equated music with instrumentation (in contrast, d'Errico *et al.*, 2003, are noteworthy in eschewing this view).

Much of the problem stems from the difficulty of identifying universality in different cultures' conceptions of music. Indeed, the trend in ethnomusicology between the 1940s and 1970s was to frown upon attempts to identify musical universals, a pursuit seen as devaluing of the cultural diversity of behaviour. This view has gradually been replaced by an increasing interest in definitive features of music (Nettl, 2000). Bruno Nettl reports that "All societies have vocal music…. All societies have at least some music that conforms to a meter or contains a pulse…. All societies have some music that uses only three or four pitches, usually combining major seconds and minor thirds." (Nettl, 2000, p. 468). Rhythm may be produced by striking objects or the body (e.g. foot-stamping, clapping) or may be produced by rhythmic vocalisation. Melody may be produced instrumentally or with the voice. Musical behaviours, while seemingly universal, are very varied.

It should be evident from the above that a *conflict* of concepts of music between the disciplines considered in this dissertation is unlikely, as such concepts have remained unspecified. However, it is important that some parameters of the term be set, not least so that this research is not guilty of the same abstraction, or ethnocentrism, as has sometimes occurred in the past. This issue has been effectively addressed by Cross (2003b), who defines music as follows: "Music embodies, entrains and transposably intentionalises time in sound and action". This definition encompasses both the corporeal and auditory elements of musical performance and perception, and its potential to have multiple meanings – it means different things to different people in different contexts ("transposable intentionality"). It can also be applied to all experiences of music be they "live" and participatory, or recorded and solitary. It deliberately avoids cultural specificity, and whilst it does not stipulate some of the properties of music which may be universal (as will be explored in this thesis), the properties of music which it describes *are* universal. Whilst it is general enough to encompass any musical activity of any culture and era, it is precise enough such that any activity conforming to this definition would be considered musical. As such, it is this definition which is in mind when the terms "music" and "musical behaviours" are used in the present work. These, and some other terms that are used in the following chapters, are further elucidated below.

Terminology used

"Music" is used as a noun, to refer to the auditory outcome of "musical behaviours" or "musical activities", and is conceived of in terms of Cross' definition, above.

"Musical behaviours" is used to encompass vocal and instrumental melodic and percussive activities, including dance.

"Vocal" refers to the auditory output of the "vocal tract" (the combination of the larynx, upper respiratory tract and orofacial musculature). Hence also "vocalisation".

"Verbal" refers to vocalisations *with linguistic content*. Hence also "verbalisation".

The terms *Homo sapiens* and "anatomically modern human" are used interchangeably.

The terms *Homo neanderthalensis* and "Neanderthal" are used interchangeably.

The term "archaic *Homo sapiens*" is used to describe post-*Homo erectus* African and European hominids ancestral to *Homo sapiens* and/or *Homo neanderthalensis*, except where the example under discussion is identified specifically as *Homo heidelbergensis*.

SECTION I

ETHNOGRAPHIC EVIDENCE FOR THE USE OF MUSIC

"The eye of the flute is the doorway to thought"

(Fernando Librado, a.k.a. Kitsepawit of the Chumash, b.1804; ed. Hudson, 1977)

There are several principal objectives to this first section of the thesis. It aims to examine and illustrate the diversity of musical behaviours that exist, and demonstrate that our expectations about what constitutes these behaviours can be far removed from our own experience of music. In particular, it is important to overcome the perception that we would see evidence for instrumentation as soon as humans participated in musical behaviour. Coming from a western musical background, it is all too easy to equate music with instrumentation, after the western art-music (and popular music) tradition. This perception, consciously or otherwise, seems to have dominated archaeological speculation regarding the incidence of musical behaviours in Palaeolithic humans to date. As the studies below show, musical behaviour and the use of instrumentation are by no means interdependent.

Contrary, and complimentary, to the aim of illustrating the diversity of musical behaviours, is the aim of identifying shared features of musical behaviours across these cultures. Given the diversity of habitats and the temporal and spatial separation of the hunter-gatherer peoples considered, and that they are united only in their humanity and subsistence strategy, similarities in their musical behaviours could suggest information about the roots of these behaviours. Close parallels between the groups would imply either convergent development[1] of these behaviours or a shared cultural heritage. The former (convergence) would suggest that there are important evolutionary driving forces towards those common behaviours, either as a consequence of subsistence method or of human biology, these being the principal common factors between the groups. The latter situation (shared heritage) would indicate a very ancient tradition of musical behaviour, as these groups are probably separated by at least 50,000 years.

---

[1] Convergence, or *equifinality*, describes the situation whereby different species or genera (or, by extension, communities) develop the same physical or behavioural adaptation to a selective pressure under different circumstances.

A further aim of this section is to examine the nature and the uses of instruments, such as are used, how they are manufactured, and what resources are made use of in their construction. Are the resources used subject to complex modification to produce sound-making devices or are they used in their natural state? Are the instruments melodic, percussive or both? The answers to these questions could have important implications for the capability of prehistoric and pre-modern humans to undertake similar behaviours.

As a consequence of these analyses it should be possible to identify a diversity of types of instrumentation, including some which may have been available to be used by early humans, but which have been lost to the archaeological record. Through looking at the uses of music within societies which hunt and gather, it may also be possible to identify potential selective benefits to be had from the use of music within such a social organisation. These investigations form the content of Chapter 2. Before examining the ethnographic evidence itself, it is necessary to consider the issues and potential pitfalls, and strengths, of drawing upon ethnographic evidence in informing a picture of the past.

CHAPTER 1

Issues of Ethnography and Analogy

When examining ethnographic evidence as a source to aid in reconstructing scenarios of possible past human behaviour from the archaeological record, it is very important to be aware of a number of methodological and ideological issues that this approach raises.

The extent to which one can draw conclusions about past peoples from the archaeological record has, in itself, been a subject of great debate; the degree to which drawing analogies with current peoples can be useful in making these interpretations has been equally debated, particularly in the last fifty years. Whilst seeking to avoid presenting a complete summary of the debate (such is comprehensively performed by Wylie, 1985), the following section aims to highlight the salient issues which must be borne in mind when considering ethnographic evidence, and approaches that are (or are not) valid in applying it to gain insight into the past.

1.1. Problems with ethnographic analogy

Early uses of ethnographic examples to aid in explaining the archaeological record, in the late 19th and early 20th centuries, frequently made broad generalisations about the peoples concerned, both the contemporary society observed and the past one being interpreted. Certain parallels between one and the other were seen as indications of a wholesale equivalence between the two peoples concerned, to the extent that the existing ethnographic example was seen as a surviving analogue, or more particularly, *homologue*, of the Palaeolithic society in question.

This was largely due to a pervasive belief that the various so-called "savage" peoples of the world encountered to that date were genuinely surviving representatives of prehistoric societies. This belief fitted in well with classical evolutionist theory, as each of the ethnographic examples could be fitted into a hierarchical scheme whereby they had been displaced to marginalised regions of the world as a consequence of the emergence of better adapted (more "civilised") peoples. Possibly the best-known example of this view was expressed by Sollas in a series of talks delivered in 1906. He equated various Palaeolithic technologies from France with particular ethnographic examples; for example, he saw the Australian Aborigines as surviving Mousterians, African bushmen as Aurignacians, and native Americans and Eskimos as displaced Magdelenians (Wylie, 1985).

For many years this interpretation was accepted with very little question. Greater exploration of the archaeological record, however, increasingly demonstrated that the degree of parallelism between past and present hunters and gatherers was not as great as asserted in Sollas' model. Furthermore, more extensive study of existing hunter-gatherer populations soon showed that a great deal of cultural and adaptive variety may be found amongst peoples in similar environments and with similar technologies (Clark, 1953). Even peoples who had occupied the same geographical area continuously since prehistory did not necessarily display similarity of cultural expression.

As Sollas' interpretation and other analogues were shown to be deficient, there developed a considerable crisis of faith in the utility of ethnographic analogy in archaeological interpretation. The ethnocentricity of the existing models became increasingly obvious and a perception emerged that any interpretation of the archaeological record that made use of an analogy would be "constructed in the image of contemporary cultural forms or, in the image of the archaeologists' ethnocentric understanding of them" (Wylie, 1985: 68); i.e. the application of subjectively interpreted contemporary ethnographic examples to the already subjective interpretation of the archaeological record could only reduce, rather than increase, confidence in the interpretation by adding another layer of subjectivity. Thompson (1956), although acknowledging that the use of analogies could be helpful in providing a greater number of possible interpretations of the archaeological record, considers that there is no advantage in this; in fact, he asserts, it is positively disadvantageous, as it merely provides the archaeologist with a greater number of possible uncertain and unconfirmable subjective hypotheses, making the interpretation even less certain.

With the emergence of the New Archaeology, advocating scientific rigour and testable models in archaeological interpretation, there was no place for subjectivity and it was found that there were considerable difficulties in making analogical assertions that were subject to testing. The New Archaeologists (e.g. Binford, 1967 - and pre-empted by Strong, 1936, 1942, and Steward, 1942) suggested that ethnographic analogy should be used only to generate hypotheses which could then be tested using non-analogical methods. More extreme critics such as Freeman (1968) and Gould (1980) argue further, that the use of analogy inevitably distorts and adversely influences the perception of the archaeological evidence, so that certain interpretive possibilities are not considered because they do not have analogues today (Wylie, 1985: 64). Wobst (1978) also elucidates the potential limitations consequential of over-familiarity with the particular behavioural norms associated with the narrow sphere in which the ethnographic observer is located. These are particularly evident in situations dealing with external contact: in terms of interaction with other groups, the most stereotyped behaviours may in fact be exhibited the more *un*usual or unpredictable the situation being encountered, due to the need to exert predictability and control on the situation (Wobst, 1978). Wobst also cautions against a tendency to categorise different behaviours in wider circumstances as unusual deviations from the behavioural tendencies observed in normal contexts, rather than to treat them as novel behavioural tendencies in their own right. The apparent pitfalls and limitations of ethnographic observation were enough to lead Freeman (1968) to argue that theoretical frameworks to explain the archaeological record ought to be based only on comparison between those residues (Wylie, 1985: 88).

The fundamental issue, apart from the problem of subjectivity in the interpretation and application of the ethnography, is of the suitability of ethnographic sources as analogues for past populations. There is a great deal of variability in the archaeological record of hunter-gatherers in terms of technologies, ecological conditions, subsistence niches and, not least, physiology and neurology. This record, after all, covers most of the globe and many hundreds of thousands of years. However, there is considerably less variability amongst modern ethnographic examples of hunters and gatherers, and there is not necessarily an equivalent analogy to be made between aspects of past and present life-style. Not least of these aspects is habitat; most of the modern hunters and gatherers now occupy extreme and marginalised ecological niches, which in many cases were only occupied in relatively recent history or prehistory. There is thus no representation in the contemporary hunter-gatherer ethnography of many of the habitats occupied by prehistoric humans and hominids, and not necessarily any equivalence between the life-styles of prehistoric hunter-gatherers and those of modern surviving hunter-gatherer societies (Yellen, 1977). On the other hand, there is also a danger that certain features that modern hunter-gatherer societies share are taken for granted as applicable to all other examples of peoples subsisting in that manner (Yellen, 1977), when this may not have been the case.

In addressing these concerns, it must be said that, as for the issue of subjectivity in interpretation, this is a fact of our epistemology; our perception and experience of the world is intrinsically tied up, *in all circumstances*, with our previous knowledge and experience, to which we relate, in analogical ways, our current experiences, in order to understand them. It could thus be argued that a certain amount of subjectivity is impossible to avoid even if merely *describing* what one sees. But for an archaeologist to merely describe evidence without interpreting it is to leave a job half-done; the equivalent, to use an different analogy, of a medical doctor merely describing the symptoms he/she sees before him/her and administering no treatment. The question remains of what type of treatment to administer.

To examine a broad diversity of possible alternative explanations from a variety of cultural sources is to reduce the chance of a narrow and limited interpretation or explanation of evidence of behaviour constrained by the limits of personal experience. *Of course* one cannot be certain that such an interpretation will be correct, and *of course* it is subjective; these are risks we face every day in our understanding of any issue. And o*f course* it is a good thing to minimise that uncertainty. We can be thankful that there are sufficient archaeologists in the world to be able to venture their own interpretation that we need not be constrained by one opinion, but may be discursive and, if we are fortunate, reach a consensus.

It can be reasonably argued for archaeologists, as with the medical doctor, that it is their paramount responsibility to interpret the evidence they see before them, and not allow the patient to malinger untreated for fear of mis-applying their own expertise. So on what basis is it possible to make comparisons between ethnography and archaeology, and to draw analogy? What factors affecting a society should be considered to be definitive of its suitability as an analogical example to use? How far can certain parallels such as it *is* possible to detect be extended into other spheres of existence? Hawkes (1954) suggests that the more you move away from technological aspects of existence towards cultural factors, the greater the reduction in the reliability of any inferences made from modern example to past people. Gould (1980) advocates making interpretations of the archaeological record on the basis of natural laws or principles that will determine human behaviour (such as natural selection) and thus avoid uncertainty of analogues with present populations. He does not expound how such rules of behaviour should be identified without recourse to analogy though. Binford (1983) also advocates testable rule-based interpretations of evidence, but on the basis of ethnographically observed processes for the formation of such deposits. Wylie recommends building up a model of best fit from numerous ethnographic sources, while O'Connell (1995) argues that we can be much more likely to reach an understanding of the causes of archaeological evidence if we understand the behaviours, rather than just the processes of formation, which underlie that evidence.

1.2. Roles and application of ethnographic analogy

As O'Connell puts it (1995), archaeologists have only two sources of information to which to appeal in order to explain variability in past human behaviour: patterns in the nature and distribution of the products of past behaviour, and knowledge of human behaviour and its material consequences in the present day. Relying on only the former allows for little more than simple description of that evidence; use of the latter in addition provides a means for helping to interpret that evidence rather than just describe it.

In fact, to make any *interpretation* at all of the archaeological record requires analogy of one sort or another. To have any concept of what may constitute a reasonable cause for given evidence must depend upon drawing upon subjective experience, either of human behaviour in one's own culture or another. Chang (1967) expresses this issue, and goes on to assert that recourse to ethnographic analogy in interpretation does not merely allow conclusions to be made, but actually renders those conclusions far more likely to be true than they would be otherwise. As described above, this position has been countered to an extent by the fears of writers such as Freeman (1968) and Gould and Watson (1982), who draw attention to the danger of forcing archaeological evidence into frames of reference derived from modern populations, when past cultures may well have been unique.

The contention remains, however, that any interpretation of evidence is impossible without some level of subjective analogical reasoning; it could reasonably be argued that, given this fact of our epistemology, we would be most sensible to draw inspiration for our interpretations from sources with the most in common with the record we are attempting to interpret. The true value of this is in generating *most probable* explanations for a given pattern of evidence; as Yellen (1977) puts it, even "a probabilistic statement where probabilities are difficult or impossible to determine is better than no statement at all" (p.7). A difficulty is in determining which sources *do* have the most in common with the record we are interpreting, whilst avoiding circular reasoning. But there are ways of increasing the probability that a given interpretation is correct, and of avoiding the serious pitfalls expressed above associated with the wholesale application of analogy.

Firstly, it is not suggested that ethnographic analogy be used to the exclusion of other possible interpretations, but as an additional source of possible interpretations. In this way models of causal factors that may have otherwise been overlooked may be constructed. What is more, reference to a variety of examples in the ethnographic record should allow any interpretation, whether based on analogy or not, to be subject to testing on the basis of realism and practicality in a natural condition; a constraint that non-analogical interpretations have sometimes foregone in the past.

Rather than making indiscriminate analogues extended to all spheres of existence, good analogical reasoning makes clear the degree of difference or similarity between the example and the record being interpreted. The more diverse the examples examined with a view to making an analogical interpretation of a given record, the greater the opportunity to test or reinforce the credibility of the interpretations made and the suitability of the parallels that have been drawn. Through detailed study of the ways in which the cultural and material record may be formed the best ethnographic parallels for a given record can be chosen. Consequently, any aspects of analogy that are applied are well-considered and are not an indiscriminate application of parallels across every aspect of the record being interpreted.

Thus, in contrast to the accusation of shoehorning a given set of evidence into a model wholly based on a single example that is unlikely to be representative of the culture that created the record, judicious use of ethnographic examples can allow a model to be built up selectively from a number of analogical sources, thereby providing a well-grounded unique model of causal factors that may be unlike any single modern analogue.

According to O'Connell (1995), though, in the past there has been a limiting factor in constructing models of past behavioural tendencies on the basis of ethnographic analogy because, he says, ethnoarchaeologists have not concerned themselves with explaining the behaviours that underlie the deposition of material remains, but merely with describing those material consequences (O'Connell, 1995: 206). This latter is, of course, exactly the approach that was advocated by protesters against analogy, such as Freeman (1968).

O'Connell suggests that, because modern hunter-gatherers are neither ecologically or physiologically equivalent to past hunter-gatherers, being themselves the product of evolutionary history, we cannot expect to be able to draw direct parallels between them; consequently, models based on modern examples are unable to account for behaviours which may be present in the archaeological record but which are not represented in the modern ethnographic record, because they model the relationships between a given behaviour and its consequences in the record, without attempting to model the reasons for the behaviours themselves. Thus they can only ever be applied in circumstances where a given archaeological product is present that has been observed ethnographically, and cannot be extended beyond that into new, unfamiliar, behavioural patterns (O'Connell, 1995: 217), which should, in fact, be of primary interest for the archaeologist.

O'Connell (1995) thus recommends that ethnoarchaeological interpretation be linked with Neo-Darwinian behavioural ecology to provide a framework for devising models and explanations for past behaviours, that may be tested by reference to the principle of the behavioural theory and with the archaeological record. These models for behaviours should be made on the basis of potential benefit and cost factors, informed and subsequently tested by evidence from modern human populations and other organisms.

Through examining the role of music and music-related activities in existing hunter-gatherer societies, it should be possible to at least draw informed conclusions regarding the selective and practical viability of proposed models of musical behaviours amongst past peoples subsisting in this manner.

CHAPTER 2

Musical Behaviours and Instruments of Hunter-Gatherer Societies

As stated at the start of this section, this chapter examines anthropological and ethnomusicological sources for evidence of the nature of the use of music and instrumentation in hunter-gatherer societies today, and in recent history. As a consequence of this analysis it should be possible to identify types of instrumentation from outside the repertoire of the western art-music tradition which may have been used by early humans, but which have been lost to the archaeological record. Through looking at the uses of music within societies which hunt and gather, it may also be possible to identify potential selective benefits to be had from the use of music within such a social organisation.

Despite a particular interest in the nature and uses of music in “traditional” and hunter-gatherer societies in the late nineteenth and early twentieth century, the majority of the anthropological literature of the last forty years deals extremely cursorily with the topic of music, if at all. This is probably as a consequence of the creation of the separate discipline of *ethnomusicology* from within anthropology. Ethnomusicology has subsequently often focused more on analysis of the music itself and the methodology of its study than the traditional uses of the music and its instrumentation.

The early anthropological sources, although at least taking an interest in its nature and importance, often deal with music in a rather anthropocentric way, with many references to “savages” and “primitives”, up until at least the nineteen-fifties. More recent authors who do address the topic tend to concentrate on the uses of music for political communication and as an indication of social change today, rather than on the traditional uses of music within the cultures (e.g. Barac, 1999; Broughton *et al.*, 1994). Even the recent “Cambridge Encyclopedia of Hunters and Gatherers” (1999), despite including excellent sections dealing with the language, traditions, lifestyles and habitats of most of the world’s hunters and gatherers, dedicates only six of its 500 pages specifically to music in those societies, four and a half of those concerned with modern use of music (Barac, 1999). Many other volumes pertaining to hunter-gatherers contain no mention of music, instruments or singing at all (e.g. Bettinger, 1991; Schrive, 1984, to name just two).

Fortunately, the work of ethnomusicologists such as John Blacking, Bruno Nettl, David McAllester and David Locke does make detailed study of the instrumentation and purposes of the music of the cultures they studied. Considering that their studies suggest that music impinges on a great variety of aspects of life in these cultures, it is all the more surprising that there is such a dearth of reporting and discussion elsewhere.

Bearing in mind the issues, limitations (and strengths) associated with making analogies between past societies and the present ethnographic record, as discussed in the previous chapter, the aims of the present chapter are quite specific. It does not, at this stage, seek to explain Palaeolithic musical evidence by drawing direct parallels with the present or recent history. Instead, it seeks to explore the diversity of musical behaviours and instrumentation evident to allow for a broader conception of how such behaviours may have been manifested, and why.

This chapter describes the types of instruments used traditionally by four groups of modern hunter-gatherers from around the world: the Native Americans of the plains (Blackfoot and Sioux), the Aka and Mbuti African Pygmies, the Pintupi-speaking Australian Aborigines, and the Yupik Eskimos of Southwest Alaska. These groups occupy four very different types of environments: rolling temperate grasslands, wet rainforest, arid desert, and Arctic tundra respectively; also, with the exception of Asia, between them there is a representative from each of the continents still inhabited by hunter-gatherers. In each case, the ways in which music is used within those cultures, the use of instruments, and their materials of manufacture are examined.

Although modern hunter-gatherers are not to be considered to be a direct analogy for Palaeolithic humans, their subsistence methods may be similar. These define their lifestyles to a large degree, so the contexts within which musical behaviours are undertaken may also be similar. Furthermore, the tools and raw materials which they have available to them for the creation of instruments often resemble those available to early humans. For example, before contact with westerners, the Plains Indians (Taylor, 1991), Pygmies (Ichikawa, 1999) and Australian Aborigines (Morton, 1999) made no use of metal in the manufacture of their tools and artefacts, using only wood, stone and occasionally bone.

2.1. Native Americans of the plains (Blackfoot and Sioux)

The first group of hunter-gatherers to be looked at are the Native Americans of the plains of central North America. Although there were at least 32 tribes occupying this region in the year 1800, some of the best documentation relates to the Sioux and the Blackfoot tribes. These tribes both lived in areas of relatively high humidity, producing rolling grasslands as their major habitat, the Blackfoot in the northern plains, and the Sioux in the east (Taylor, 1991). This plains environment may in fact be more reminiscent of the habitat occupied by ancestral hunter-gatherers in Africa than those habitats that are home to hunter-gatherers in Africa today. This is because hunter-gatherers in Africa today occupy areas either of very low rainfall (e.g. the !Kung San of the Kalahari desert) or of very high rainfall (e.g. the rainforest-dwelling Pygmies), with all of the temperate grassland now occupied by settled agriculturalists (Foley, 1992).

The Blackfoot were traditionally nomadic hunters of antelope and bison (buffalo) and, until the introduction of horses by Europeans in the eighteenth century, hunted and travelled on foot. Interestingly, they used a hunting method which was also used by Middle and Upper Palaeolithic hunters, both Neanderthal and modern humans (Chase, 1989). The men would drive a herd of the animals into a v-shaped drive leading over a cliff edge or ditch until the animals fell to their deaths, ready to be collected and processed (Taylor, 1991). Particularly relevant is the use of song in this procedure amongst the Blackfoot. The herd was initially enticed towards the drive area by a young man singing a spiritually potent song in the manner of a bleating calf (Kehoe, 1999).

All able-bodied members of the group would help process such a kill. Blackfoot women also harvested plant foods from the surrounding environment (and some limited cultivation), such as berries, bulbs and turnips (Kehoe, 1999). They followed the buffalo herds' annual movements, living in bands of 10-20 tipis, with around 8 persons in each, moving to grasslands in the spring, meeting with other groups in the summer, and then to sheltered river-valleys in the autumn (Epp, 1988).

The summer meeting between groups involved the resolution of disputes, policies and trade, as well as the performance of the "Sun Dance", to bring prosperity and health. Groups had leaders, and certain families were seen as privileged, such that they did not have to participate in daily menial work, but were instead concerned with leadership activities. However, leadership was subject to the views of the members of the group (including children), and individual autonomy was also highly valued (Kehoe, 1999). Although the nomadic nature of their existence is curtailed by the reservation system today, many of the other activities of the society, including music, are still practised (Nettl, 1992).

The music of the Plains Indians resembles that of the majority of the rest of the Native Americans, in that it is almost always *monophonic* (contains only one melody proceeding at any one time), and the melody is nearly always vocal rather than instrumental (Nettl, 1992; McAllester, 1996). Such instrumentation as there is consists predominantly of percussion in the form of drums or rattles, which are used to accompany the vocal melody. This limited variety of types of instruments is in fact typical of the whole of the North American continent, where the selection of musical instruments used by the native inhabitants is surprisingly small, considering the widespread use of music and variety of peoples.

Although almost all conform to basic types of idiophone (instruments whose bodies vibrate to produce their sound), the instruments within this category are very varied. There are rattles made from dried gourds (which rattle due to the dry seeds or stones inside), tree bark and spiders' nests, as well as deer hooves and turtle shells, for example (McAllester, 1996). The deerhoof rattles consist of around twenty doe hooves suspended from a stick. Drums tend to be either frame-drums or barrel drums, made from wood and skin, some consisting only of a piece of rawhide suspended from a stake (Nettl, 1992). Sometimes simply a plank of wood is beaten (Nettl, 1956) and idiophones made from a notched stick are also used (Nettl, 1992). Much less common, but also used, are some aerophones such as bullroarers and whistles/flutes. The latter often have no finger holes, so produce a single tone, and are made of wood or bird-bone; others have up to six finger holes (Nettl, 1956). It is noteworthy that all of the instruments used are manufactured from organic matter, occurring naturally in the Plains Indians' environment, and that many require little, if any, modification before use. Figure 2.1 illustrates some examples of Native American instruments, from a selection of groups.

It may seem that the music is rather simplistic, but amongst the Plains Indians the value of the music is not measured in terms of its complexity. Instead, it is its ability to integrate ceremonial and social events, to integrate society in general and represent it to outsiders, and to evoke supernatural influence that is important (Nettl, 1992). The Plains Indians traditionally believed that music came to people through supernatural input in dreams, so little credit for agency in composition (in the contemporary western sense) was given to individuals either (Nettl, 1956), although considerable credit could be given for being the recipient of such supernatural input. The use of music to evoke supernatural power is particularly prevalent amongst the Blackfoot Indians, who have specific songs for each act in a ceremony. For example, in a ceremony to influence the weather, a bundle of objects is opened and the correct song must be sung for each, in the right order, to "activate" each object. Some of these ceremonial bundles can contain over 160 items (Nettl, 1992).

Such religious activities are very frequently accompanied by dancing as well as the music, and this is the most common use of music amongst the plains Indians. The second most common use is to accompany *social* dancing. War dances and puberty rites are other important social and ceremonial occasions accompanied by music and dance, and the

Figure 2.1: A selection of Native American Musical Instruments (*From Taylor, 1995*)

(a) Rattle made of cocoons with small pebbles inserted	(d) Pair of split-stick rattles
(b) Wooden rasp and stick (scraped idiophone)	(e) Elder-wood end-blown flute
(c) Cocoons strung on a cord, worn as leg rattles	(f) Cane whistle
	(g) Wooden rattle
	(h) Bone whistle
	(i) Gourd rattle

aforementioned deer-hoof rattles accompany the latter in particular (Nettl, 1956). In all these instances, the men perform the majority of the dancing and singing. In the religious activities involving the opening of bundles of items, the performance is limited to the medicine man performing the ritual, but in the more social activities, the performance involves many members of the group. This includes the women, who join in certain of the men's songs, and walk around the periphery of the dancing area, rather than dancing themselves (Nettl, 1992).

Both the vocal and the rhythmic elements of the Blackfoot and Sioux music consist of a great deal of repetition, with very subtle variations on a theme (McAllester, 1996). The vocal technique used in the Sioux "Grass Dance" (a war dance) is of particular interest. This uses "vocables" (non-lexical, meaningless syllables) with high emotive input rather than translatable words as the basis of the song (McAllester, 1996). This is, in fact, a common feature of virtually all Plains Indian songs, with any words occupying only a tiny portion of the melody (Nettl, 1992). For example, the aforementioned Blackfoot Sun Dance ceremony contains only the words "Sun says to sing", and a medicine bundle ceremony contains only the words "It is spring, let others see you"; all the rest of the melody, although verbal, consists of vocables (Nettl, 1989).

Although this vocable music is clearly used in close association with symbolic activities, and a particular song can relate very specifically to a particular activity, the music itself is said to have no symbolic content. According to Nettl: "native informants are able to say almost nothing on the symbolic aspect of their [non-lexical] music" (Nettl, 1956, p. 25). It would seem that the main purpose of this type of vocalisation is to contribute to the emotional responses evoked by the music (McAllester, 1996), and has no (conscious) symbolism behind it. This is in contrast to lyrical music, or to instrumental pieces which deliberately attempt to evoke the various traits of their subjects, be they *Mars, the war-bringer*, *Winter*, or the sounds of the Australian bush, as is the case with the Pintupi described below. The only exception to this amongst the Plains Indians would appear to be the imitation of a bleating calf used in the hunting tactic described at the beginning of this section.

2.2. African Pygmies of the equatorial forest (Aka and Mbuti)

The second group of hunter-gatherers to be examined are the Pygmies of the Ituri forest of equatorial Africa, in particular the Aka (or BaAka) and Mbuti (or BaMbuti) communities. These two groups are particularly worthy of note as they still receive the majority of their subsistence from wild foods, which they forage and hunt. This is in contrast to the majority of the other Pygmy groups (such as the Efe and Baka) who now tend to live near to villages of farmers, and provide services in exchange for a diet of more cultivated food (Hitchcock, 1999).

Both groups live in dense humid forest near the equator, in central Africa. Their principle habitation is family domes made from bent branches and thatched with leaves, organised into communities of 30-100. These communities are nomadic, and tend to move on every month or two, following their quarry. The communities are egalitarian and co-operative, with no formalised hierarchy, although individuals may be acknowledged as experts in particular skills. Ownership of personal possessions is minimal, largely as a product of the nomadic lifestyle (Turino, 1992), and they do not make any use of metal or earthenware in manufacturing their tools and containers (Ichikawa, 1999), which tend to be made of wood or bark (Bahuchet, 1999).

Hunting is carried out using nets, spears and poisoned darts (Bahuchet, 1999; Ichikawa, 1999). Amongst the Aka, subsistence from hunting accounts for about two-thirds of their diet. The other third is made up of foraged plants, fungi and animals, such as wild yam, leaves, nuts, snails, tortoises, weevils, beetles and caterpillars. This activity is predominantly carried out by women, whereas the men collect honey and do most of the hunting of large game. The game is mainly elephants, antelope, river hogs, gorillas, monkeys and chimpanzees. Monkeys are hunted with bows and poisoned darts; large game is hunted with broad-bladed spears. During the dry season, when many Aka communities band together, communal hunting often occurs, with all the men, women and children getting together to “beat” duiker into rings of nets (Bahuchet, 1999). This activity is also carried out by the Mbuti, although hunting among the Mbuti accounts for less of their subsistence than among the Aka. The Mbuti also rely on the foraging of wild vegetables, honey, nuts and invertebrates as a portion of their diets, although in recent years more of their diet has been provided by nearby farmers (Ichikawa, 1999).

The music of the Aka and Mbuti is, like most aspects of the culture, a communal activity, without specialist musicians. Like that of the Plains Indians, it is predominantly vocal, with little instrumentation; this is one of several features which the Pygmy and Plains Indian music share. Aka and Mbuti music is also considered to be of supernatural origin, told to individuals by their ancestors in their sleep, and is used as a way of communicating with the spirits of their surrounding environment (Ichikawa, 1999). They believe the way to communicate with the divine is through sound alone; as a result song texts are minimal, often consisting of only one line such as “the forest is good” among the Aka (Turino, 1992) or “we are the children of the forest” among the Mbuti (Turnbull, 1962). The rest of the vocal element of the music consists, as with Plains Indians, of *vocables* (non-lexical vocal syllables) (Locke, 1996).

Some music is performed individually, such as lullabies, but communal singing for ceremonial activities is most important (Turino, 1992). Such music is often related to specific subsistence activities, with particular songs being performed for each activity. Thus, there are songs for net hunting, elephant hunting, and honey collecting. For example, the *mabo* song and dance of the Aka is a net hunting dance, and is used for the ritual purpose of the administration of hunting medicine. It is also used as a method for instruction in dance, including between neighbouring settlements (Kisliuk, 1991, cited in Locke, 1996). In this respect, it acts both as a philological aid and as a diplomatic mechanism. The *mobandi* ritual dance is related to honey collecting, and is unusual in that it is the only one that is seasonal rather than being carried out before or after the hunt. It is performed after the rains when the *mbaso* tree flowers, and is a purification ritual. The Aka believe that any misfortune to occur in the hunting and gathering is the result of human misconduct, which jeopardises the help of the spirits in the activity. During the *mobandi* all the members of the group gently hit themselves with branches to transfer any evil forces out of themselves and onto the branch (Bahuchet, 1999), avoiding this risk.

Songs also mark formal “rites of passage” ceremonies such as circumcision, girls’ puberty, marriages and funerals (Ichikawa, 1999). On these occasions they are usually performed by women, unlike the hunting-related songs, which are performed by men (Turino, 1992). It is only in these ceremonies that gender or age distinctions are made in the performance of music; in other performances, anyone may participate who wishes to (Turino, 1992).

Instrumentation consists mainly of percussive accompaniment to the vocal tunes using claps, drumming on skins which cover the end of cone-shaped logs and rapping sticks on the drum body (Locke, 1996). There are also a few rattles, as well as some end-blown flute pipes made from cane (Turino, 1992). Like all of the Pygmies’ traditional artefacts, these instruments are made from organic materials occurring in their environment. In these respects, the instrumentation is similar to that of the Plains Indians described above. However, despite its similarities, the music of the Pygmies differs from that of the Plains Indians in that it is *polyphonic*; a performance consists of several different melodies sung simultaneously by different groups, which “interlock” with each other to form a multi-layered piece of music (Turino, 1992).

The majority of the music itself appears to lack any direct symbolic content (in terms of lexical meaning or mimicry), although it is clearly associated with activities which do have symbolic content and associations. The melodies do not attempt to emulate or evoke any other thing, but follow particular structures for different activities and dances. The exception to this is the *molimo* music, which is performed specifically to wake the forest if hunting is bad. This ceremony is performed at night and can last several months, and is rooted in the belief that if hunting is poor, it is because the forest is sleeping. It uses the molimo trumpet, a single end-blown tube, which is supposed to mimic the sounds of the forest and answer the men's singing (Turino, 1992). As such, the molimo sound does have a direct symbolic association.

Fundamentally, musical and dancing activity forms the focus of group social activity on a large scale, and is the main occasion where social relationships manifest themselves (Bundo, 2002). It is notable that these roles would in the past have been fulfilled by both hunting and musical activities, but that in recent years as the importance of large-group hunting activities has declined the role is most commonly fulfilled by musical activities.

2.3. Australian Aborigines of the Western Desert (Pintupi)

The Aboriginal population of Australia falls into two main language groups, with peoples being divided more by geographical features than by "tribal" groupings (Peterson, 1999). Communities exist within these categories, but widespread systems of exchange across the continent, of goods, ceremonies, social practices and some intermarriage, have led to an indigenous way of life which is fairly homogeneous throughout the continent. Archaeological evidence suggests that systems of long-distance exchange were in place at least 30,000 years ago (Mulvaney, 1976); amongst historical populations at least, these exchanges were as much for the benefit of social relationships as for practical purposes (Peterson, 1999).

Of particular note amongst the Aboriginal groups are the Pintupi-speaking people of the Western desert, because they were the last Aboriginal population to be in any way incorporated into the settler society, during the 1960s (Myers, 1999). Consequently, they were (until then) less severely influenced by Western lifestyle than many of the other groups, and recent records exist of their traditional way of life. During the 1970s they started to create a few autonomous satellite communities around Papunya, and by the early 1980s, most had returned to their traditional territories of the Gibson desert (Myers, 1999). The traditions outlined below are related in the past tense, as they describe the features of the Pintupi before their re-settlement. However, since members of the culture have begun to return to their traditional territories today, hopefully they will perpetuate many of these cultural elements and subsistence methods, as have other Aborigine groups in recent years.

The Gibson desert is an area of sandy dunes, plains and hills, with sparse vegetation. Rainfall is minimal, mostly falling only during January and February. The main fauna are lizards, feral cats, kangaroos, wallabies and emu. These animals were traditionally hunted by the men using stone axes, spears and spear-throwers. Women collected seeds, fruit, vegetables, honey-ants and grubs using digging-sticks and containers made from tree-bark. These foods gathered by the women constituted the majority of the Pintupi diet (60-80%) for most of the year (Gould, 1969); the majority of protein was provided by lizards, caught by both men and women (Myers, 1999). Hunting generally occurred alone or in small groups, and large game was shared amongst all the families of a residential group (10-30 members) (Myers, 1999). As with the Plains Indians and the Pygmies, the dry season (dri*est* season, in the case of the Pintupi) was the time when many separate communities of the same group would gather together for social and ceremonial activities; amongst the Pintupi, up to ten such groups would gather in these instances. Traditional shelters for the family groups were made of brush and branches, and would house a nuclear family. Groups of unmarried men still going through initiation and widowed women sometimes had their own habitations (Myers, 1999).

Ownership of the land was seen as being dictated by the content of songs, stories and rituals pertaining to the Dreaming (see below), with each group being guardians of a particular set of this information (Breen, 1994). This “ownership”, however, did not relate strongly to the actual group boundaries, which were quite flexible (Myers, 1999). Social organisation was largely non-hierarchical, without designated leaders, but additional respect was given on the basis of age or the successful performance of rituals. Gender relations were predominantly egalitarian (Myers, 1999).

Like that of the Arrernte (Morton, 1999), Cape York (Martin, 1999), and Ngarrindjeri (Tonkinson, 1999) Aborigines, music amongst the Pintupi (Myers, 1999) was tied up closely with story and ritual describing their hereditary land. These combined renditions acted out events from the *Tjukurrpa*, or “The Dreaming”. During Tjukurrpa, ancestor spirits travelled the land leaving their marks on the landscape and the essence of people and animals on the land. Pintupi songs incorporate these mythologies into the performances, and describe the world around them (Myers, 1999). The precise preservation of these songs was seen by them as fundamental to their continued survival and the constant renewal of nature, so was considered extremely powerful and valuable. As they had no written records and notation, this oral tradition was the only way in which this information could be passed on and preserved, and the musical form had important mnemonic benefits to this end (Jones, 1983).

Unlike the music of the Plains Indians and the Pygmies described so far, the music of the Pintupi was lyrically dominated; in fact, the perpetuation and communication of lyrical information was central to the purpose of much of the singing. However, the music does resemble that of the Plains Indians and Pygmies (and other Aboriginal groups) in that it was predominantly vocal, and featured little instrumentation (Breen, 1994). In fact, much of the music relied on percussion produced without instruments at all, such as clapping hands, slapping the body and stamping feet on the ground. When instrumentation was used, it was the same as that used by the majority of other Aboriginal groups (Nettl, 1992). Percussion instrumentation consisted traditionally of two eucalyptus sticks bashed together (Nettl, 1992) called *bilma* (Breen, 1994).

The only non-percussive instrument used was the *didjeridu*, a long pipe made from the limb of a eucalyptus tree, hollowed out by termites (Breen, 1994). This is played like a trumpet, creating a low single-tone droning sound, the "colour" of which can be varied by changing the shape of the mouth (Nettl, 1992). This originated amongst the Aboriginals of the North of the continent, but is now used universally. Originally, it was only played by men specially selected by elder members of the group, and to learn it, players were encouraged to listen to all of the sounds of the bush, in order to be able to imitate the sounds of nature (Breen, 1994).

The music of the Pintupi had, as other Aboriginal music still does, explicit symbolic associations. This symbolism was seen as very important, describing the world and history of the people, and its preservation was seen as fundamental to the well-being of the group. The sound of the only non-percussive instrumentation that there is, the didjeridu, also had symbolic meaning; players aimed to mimic the sounds of the surrounding environment, adding to the symbolism of the song it accompanied (Breen, 1994).

2.4. The Eskimo of South-West Alaska (Yupik) and Canada (Inuit)

The preceding sections have considered hunter-gatherers of the temperate plains, equatorial jungle and the desert; this section looks at hunter-gatherers living in sub-arctic tundra. This environment is a demanding one in which to subsist, with a limited range of resources and difficult environmental conditions.

The peoples inhabiting the area to the south of the Bering Strait and Norton Sound in south-west Alaska are known collectively as the Yupik. The vast majority of the research specifically regarding the music of the Yupik has been carried out by Johnston, who provides the only comprehensive description of their musical traditions (Johnston, 1999). As such, much of the following section draws upon this work. The climate in this area is milder than that further north, where the Inupiaq Eskimo live, and consists mainly of low marshy tracts close to sea-level. This area is largely treeless, although the Eskimo do move upstream to penetrate the forested inland areas too (Nelson, 1899/1983). Although there are no fully-formed trees in this area, there is a variety of other vegetation. This particularly grows in sheltered areas on hillsides and along the courses of river valleys, and includes willows and alders, as well as the grasses, sphagnum and flowering plants that occur on the rest of the landscape (Nelson, 1899/1983).

The population of Yupik numbers around 17,000 today, and they subsist in the traditional manner on sea-mammal, salmon-fishing and reindeer, as well as wild berries. These are gathered, mainly by women and older children, and are stored until needed in winter in underground chambers (Johnston, 1989). This population exists in groups of up to 100 people, who associate themselves with particular tracts of land. This occupancy is not normally considered as ownership of the land or its resources, but merely as the community being the “traditional occupants” of that particular tract (Riches, 1995). The social groups are best described as having a structure that is flexible, egalitarian and individualistic (Gardner, 1991), with little ranking within the group, or between groups. Leadership tends to be attributed on the basis of exhibiting exemplary behaviour in some aspect of life, and the security of the leaders’ position rests purely with the fact that others have *chosen* to follow them (Riches, 1995). The provision of food is a reciprocal system, with all members of a “task-group” (up to 30 people) contributing to the total, and those contributing the most receiving some prestige as a consequence (Riches, 1984, quoted in Riches, 1995).

The music of the Yupik is predominantly vocal, with only two different instruments used, if any. All melody is carried by the vocal parts, whilst the instruments provide percussion. The most common instrument is a simple frame drum, made from a round wooden frame covered with a sea-mammal membrane; each village tends to have a specialist maker of traditional drums. The other instrument sometimes used is a rattle; this may be made, for example, from a hoop strung with puffin-bills which rattle against each other when shaken (Johnston, 1989).

The Yupik themselves classify their songs and musical styles into thirteen categories; these are described in some detail by Johnston (1989), and his findings are summarised below. The first six of these are for adults: dance songs, shamans’ songs, hunting songs, teasing songs, travelling songs and berry-picking songs. The remaining seven are for children (either to be performed by children, or by adults for children): story songs, juggling-game songs, jump-rope game songs, ghost-game songs, bird identification songs, fish identification songs and *inqum* "cooing" songs (Johnston, 1989).

Dance songs tend to feature drumming. This is performed by four to eight people simultaneously who drum in synchrony, whilst chanting the words of the song. Drummers do not participate in dancing, and dancers do not sing either, although the audience may sing along. The drummer-singers are almost always male, although there is no formal rule against women drumming (Johnston, 1989). There are three main areas of subject matter for Yupik dance songs: the relating of past adventures of cultural/community heroes, description of comic incidents and portrayal of animal or bird-life. Since the songs often form a repository of knowledge of past adventures and lore for the community, the singers are usually those who have most of interest to relate. The dance motions reflect this story-telling nature of the song, and mimic activities of daily life. Men's dance actions tend to resemble activities such as sledding and harpooning whilst the women's dance motions mimic activities such as carcass cutting and feather plucking.

The melody of the songs tends to consist only of two or three notes, with the note changing with each syllable of the song. As well as the narrative words there is a lot of use made of *vocable* non-linguistic syllables. Many of these communal songs place an emphasis on comic occurrences and events, and can have a strong bonding and socially lubricating effect on the community. As Johnston (1989) points out, this can be very important in communities of people confined in small quarters in some of the harshest environmental conditions in the world.

Another way in which song is used as a social lubricant is in the so-called teasing songs. These are used as a form of legal mechanism and as a way of avoiding direct conflict between two protagonists. The offended individual would compose a witty, chiding, song aimed at the offender, which they would perform in front of the community. This would embarrass the latter, and hopefully also act as a deterrent for future repetitions of the offending action (Johnston, 1989).

Music and song are not only used as aids to social interactions, but also to influence the world around the Yupik. The shaman of the Yupik were traditionally the intermediaries between the hunters and the gods of hunting, who control the success of the hunt. It was thought that the shaman's soul could leave his body and travel to the moon, where the hunting gods resided, in order to commune with them; this process was aided by the shaman's drumming as he sang. These songs are still often used at important festivals, and composed for the event itself (Johnston, 1989). The hunters themselves also use music to influence the unpredictable environment in which they have to operate, using song to affect the hunt. The hunting songs are performed during the hunt itself and are thought to have the power to stop wounded game from escaping.

Songs may also be used to hold important knowledge, whilst being a source of entertainment; this is the case with the travelling songs and the berry-picking songs. Yupik frequently travel long distances for subsistence purposes and for the maintenance of alliances between different communities. Whilst they travel, they sing songs which often relate adventure narratives; the importance and significance of the Yupiks' awareness of their environment is reflected in that these songs frequently include highly detailed descriptions of prevailing weather conditions as the journey is described.

Environmental information is also contained within the berry-picking songs performed by the women and children as they travel to pick the cloudberries and blueberries for storage for the winter. These songs are often concerned with the tundra and its features, particularly descriptive names of features which aid in their location. In this respect, these songs closely parallel the content and use of the songs of the Pintupi (and other) Australian Aborigines (see above), specifically used as a mnemonic aid describing the surrounding environment.

The Yupik songs performed for or by children also seem to have educational purposes: as an aid to developing co-ordination skills (the juggling-game, string-figure-game and the jump-rope songs), as an educational medium concerning the natural world (bird-identification and fish-identification songs), encouraging creativity and interaction (the ghost-game songs) and as examples of triumph over adversity (the story songs).

The jump-rope songs take a question-answer form, with the rope-holders asking a question and the rope-jumper answering whilst avoiding getting caught in the skipping-rope. The juggling-game songs are performed both by adults and children, and are reported in communities throughout Greenland, Canada, Alaska and Siberia; it is thus likely that they are an extremely old form of song. The performer sings the song, which tends to have satirical or rather indelicate subject matter, whilst juggling stones in one hand. The string-figure game is a variation of cats-cradle, with a loop of string held between the hands and feet or forehead of the performer. As they perform the song, the individual winds extra twists and loops into the string in time with the rhythm and syllables of the song. This game develops dexterity and spatial-directional skills whilst the juggling game develops hand-eye co-ordination skills (Johnston, 1989).

The bird and fish identification songs teach important information about the appearance and behaviour of birds, small animals and fish. They are sung by elders to children and contain information about markings, nesting, courting, migrations and the way these relate to the environment and the weather. They also aid in the hunting and catching of these creatures for subsistence, which children are encouraged to do at a young age. The Yupik also aim in the context of these songs to mimic the sounds and behaviours of the animals they find in their surroundings. A seal hunter may, for example, entice curious seals by accurately mimicking seal noises. Fish that are caught are often sung to in the hope that they will enjoy being caught and return the next year in their re-incarnated form to be caught again (Johnston, 1989).

Like the jump-rope songs, the ghost-game songs are bipartate, consisting of question and answer between the participants; however, in this instance they are improvised replies to the question (usually concerning the whereabouts of an undesirable spirit). The children are thus required to be spontaneously creative; any participant who fails to come up with an answer becomes the ghost and has to chase and catch the other children (Johnston, 1989). It is thus a game of physical activity too.

It can be seen that there is a very great variety of song and music amongst the Yupik of the Sub-Arctic tundra, and that they also serve a variety of purposes, on an individual and community level. Music and song, for the Yupik, can be informative and educational, a social lubricant, an aid to hunting, a way of learning physical and mental skills, or purely a pass-time.

Very different, but worthy of note are the throat-games of the Inuit, known as Katajjait. These games are found in eastern and central Canada, northern Quebec and South Baffin Land, performed by Igloolik, Caribou and Nelsilik Inuit (Nattiez, 1983; Baghemil, 1988). They are generally played by two women who face each other at close range (Baghemil, 1988), sometimes even holding each other's shoulders (Nattiez, 1983). Both participants produce sequences of short repeated vocalisations. These are strung together into longer phrases and syncopated into complex sound structures which are a mixture of droning and complex gutteral sounds (Baghemil, 1988). The game is a sort of endurance test, in that it ends when one participant runs out of energy, inspiration, or starts to laugh (Baghemil, 1988). The game is performed playfully, but the performance is valued on the basis of endurance, virtuosity and aesthetics of sound – participants have to win with "beautiful sounds" (Nattiez, 1983).

The sounds themselves take the form of morphemes and vocables (Beaudry, 1978), and form repeated motifs "…made up of a morpheme, a rhythm, an intonation contour and a pattern of voiced and voiceless, inspirated and expirated sounds" (Nattiez, 1983). There tend to be no more than three or four tone levels used for the utterances (Baghemil, 1988). This small number of elements is constantly re-combined in novel ways during the course of the competition. The sounds produced may have subject matter in the sense of using real words, which may form a narrative, but most often there is no narrative at all. More frequently the sounds imitate animal noises such as goose-cries, ducks, walruses, panting dogs or mosquitos. Alternatively they may just be abstract sounds made for aesthetic effect (Baghemil, 1988). Baghemil (1988) reports that the Inuit themselves consider the Katajjait to be the language spoken by the Tunnituaruit, or "flying heads", mythical creatures that are half-woman, half-bird. The dead are also thought to communicate with this language, especially as manifested as the Aurora Borealis. Nattiez (1983), however, reports contrarily that the superstitious/spiritual language element is only rarely linked to the games and that there does not seem to be any consistent "deep meaning" or symbolism to the games; they are usually performed for their own sake as entertainment. This is also consistent with what Beaudry (1978) was told by the performers she observed. The degree of symbolism and perceived content may be something that has varied historically. According to Nattiez (1983), the games used to be performed by up to four people at once, at any time of day, month or year. The performers, although usually women, could also be men or boys. The games seemed to be multifunctional, and could be performed as entertainment, celebration or just to keep a baby quiet, for example (Nattiez, 1983).

These throat games are usually studied as a form of music, but Baghemil (1988) argues that they should be better analysed as a form of language. They seem in some ways to straddle both music and language in the vocalisations they use. Baghemil (ibid.) observes that the voicing and airstream mechanisms used in the games have more in common with several world languages, and only rarely occur in song. Interestingly, it also seems to be a system independent of the Inuktitut language – the features of vocalisation that occur in the games have not been transferred across from the language as they would be in traditional song (Baghemil, 1988).

2.5. Chapter summary and conclusions

2.5.1. The uses and nature of music

It can be seen from the preceding sections that these four societies of hunter-gatherers (Plains Indians, Pygmies, Pintupi Aborigines and Alaskan Eskimos) have a number of features in common, despite being separated across three continents, and occupying four very different environments:

All four groups meet up with fellow communities during their most difficult subsistence season, during which time there is increased performance of ceremonial and communal music and dance. In all four cultures, the ceremonial and social use of music is very important, is communal and is almost always accompanied by rhythmic dancing. In all four cultures music is also performed purely as a communal activity for pleasure.

All four have music which is predominantly vocal, and accompanied mainly by percussion instruments. The use of melodic instruments is minimal and, when used, invariably consists of end-blown pipes. These are usually single-toned instruments. All the instruments, whether percussive or melodic, are made from naturally occurring organic materials.

Amongst the Blackfoot and Sioux Plains Indians and the Aka and Mbuti Pygmies, song has minimal lyrical content, vocalisations instead consisting of *vocables*, which are emotive sounds with no obvious symbolism. In these two cultures, music accompanies rites of passage and rituals relating to subsistence and hunting. The Pintupi are a contrast in this respect, in that their music is almost exclusively lyrical, and is a repository for community knowledge and mythologies. The Inuit and Yupik music contains both types of content. Much consists of non-lyrical vocables and animal sounds, or sounds made just for aesthetic purposes. The greater proportion relates stories, descriptions of events, environments, journeys and subsistence sources, so constitute an important repository of knowledge for both adults and children. Some songs and music are inextricably related to games designed to develop important skills of subsistence and coordination. It is worth noting here that with the exception of the Yupik, in all the examples examined above the literature does not mention the role and nature of music as related to children. Whether this is a consequence of an absence of such musics in these other cultures, or whether it is a consequence of a lacuna in the research is not clear.

All four peoples believe themselves to have come from the land, to be akin with the other fauna of their environment, and use music to try to influence the world around them. Music and dance can have important uses in engendering group cohesion, altering mood, as an aid to the teaching of dance, and can facilitate group interactions and communality, within and between groups. In the majority of these instances the music itself has no inherent symbolism, but it can be used to accompany symbolic activities. As with the Yupik, amongst the Pintupi, song is a repository for knowledge of tribal history and legend, and is a mnemonic aid. The mnemonic benefits of a musical structure to knowledge are also mentioned by John Sloboda:

> "Songs and rhythmically organised poems and sayings form the major repository of knowledge in non-literate cultures. This seems to be because such organised sequences are much easier to remember than the type of prose which literate societies use in books." (Sloboda, 1985, p.267).

2.5.2. Methods and materials of construction of instruments: implications for the archaeological record

There appears to have been in the past a general conception when considering Palaeolithic music that it is unlikely to have existed before the advent of melodic instruments. This is probably based on a false premise which is quite ethnocentrically western, and due to the structure of our own music and orchestras.

The music of the hunter-gatherers described in this chapter is largely non-instrumental, with melody being carried by vocalisation alone and much of the percussion being provided by clapping, slapping and foot-stamping. Instrumentation, when it is used, is almost exclusively percussive, and is constructed from readily occurring natural resources with relatively little modification. Most use skins, wood, cane, or other vegetable matter such as gourds. Wood is used more readily than bone for scrapers, flutes, drums and drumsticks; the use of bone is rare among the Plains Indians and there is no mention of its use at all amongst the Pintupi and Pygmies. Although ivory is used by the Yupik and Inuit, there it does not tend to be used for instrumentation at all. Apart from bone, none of these instruments would leave an archaeological trace, as vegetable or animal matter degrades rapidly under most natural conditions. Had they been used by early humans, there would be no evidence of it in the archaeological record.

As for the capability of earlier hominids to produce such artefacts, the Schöningen spears (dated to 420-380 Kya) provide ample evidence of the ability of *Homo heidelbergensis* to work wood very finely (Dennel, 1997). The ability to cut skins would have been a necessity from the earliest scavenging or hunting, for the processing of carcasses. Furthermore, many of the instruments described in this chapter require very little modification before use anyway. A naturally dried gourd will rattle as the seeds inside move freely, cane and bird bones form hollow pipes in their natural state, and a tree trunk can be hollowed by termites, as is the case with the didjeridu. Much of the traditional percussion used by the hunter-gatherers described consists simply of two pieces of wood, or a piece of wood and a tree-stump being struck together. It is clear that the earliest archaeological evidence for musical instruments should not be assumed to represent the first incidence of musical behaviours in humans.

The preceding evidence illustrates that there should be no expectation that we would see evidence for instrumentation as soon as humans participated in musical behaviour, or that the first incidence of instrumentation indicates the first incidence of musical behaviour. Secondly, it also highlights fundamental similarities in the nature and roles of music between these diverse groups of hunter-gatherers; this could imply either convergent development of these behaviours in the different environments in which we find these peoples today, or a shared cultural heritage (or both). The former (convergence) would suggest that there are important evolutionary driving forces towards those common behaviours, either as a consequence of subsistence method or of human biology, these being the principal common factors between the groups. The latter situation (shared heritage) would indicate a very ancient tradition of musical behaviour, as these groups are probably separated by at least 50,000 years.

Given these considerations, the following chapter explores the nature of the archaeological evidence for musical behaviours that we find in Europe in the Upper Palaeolithic, and what it can tell us about the nature of musical behaviours at this time, when they were clearly developed.

SECTION II

PALAEOLITHIC MUSIC ARCHAEOLOGY

This section examines the archaeological evidence for the use of instruments and music in the Middle and Upper Palaeolithic, and the possible nature of music at this time. Whilst a major objective of this thesis is to address the questions of the origins of the *capacities* for musical behaviours in the human lineage, its overriding theme is the development of a picture of the emergence of the practise of musical behaviours in the Palaeolithic. In fact, past considerations of the prehistory of music have frequently considered *only* the direct archaeological evidence from the Upper Palaeolithic, and treated this as synonymous with the origin of music. As will be illustrated by the evidence considered in the latter sections of this thesis, this is to neglect a great deal of evidence of other forms from previous stages in human and hominid evolution. However, this is not to say that the significance of the archaeological evidence of the Upper Palaeolithic has been overstated; its considerable value lies in that it is the first concrete evidence of recognisable instrumental behaviours in modern humans, and its potential to demonstrate the importance of musical behaviours in human societies. What that evidence can tell us about the forms that it might have taken and roles that it might have fulfilled will be fundamental in building up a picture of music's place in the emergence of recognisable human behaviours.

CHAPTER 3

Archaeological Evidence for the Use of Music in the Palaeolithic

This chapter consists of a critical analysis of the archaeological evidence for the use of musical instrumentation in the Middle- and Upper Palaeolithic (c.200,000-12,000 years ago). Research in this area has hitherto been disparate, and many of the finds have been open to debate as regards their anthropic origin or use as sound-producers. The chapter examines the possible forms of instrumentation and sound production represented in the archaeological record of the Palaeolithic, including the use of acoustic features of the environment such as caves and other lithic objects, as well as examining the spatial and temporal distribution of the evidence. Overall, the aims are to illustrate the diversity of evidence available, to appraise it critically, and to model the nature and possible purposes of music as practised in the Upper Palaeolithic. This analysis is complemented by a comprehensive database of all reputed musical finds from the Middle- and Upper Palaeolithic, their locations, dates and status as possible sound-producers (located in Tables 1 and 2, Appendix).

Palaeolithic objects which could be musical instruments fall into five main types. The most prolific are those that are thought to be flutes or pipes. Secondly, there are pierced phalanges, often interpreted as "phalangeal whistles". A third type of aerophone are objects which have been interpreted as bullroarers (Dams, 1985; Scothern, 1992). In addition, there are some sections of bone which have been notched with parallel grooves, which may be rasps (Huyge, 1990). Finally, some authors have suggested that caves themselves, and features of caves, were used as sounding devices (Glory, 1964, 1965; Dams, 1984, 1985; Reznikoff and Dauvois, 1988; Dauvois, 1989). With the exception of the latter case, all of the known or reputed examples of sound-producers from the Palaeolithic are made from bone. This is in contrast to ethnographic examples of hunter-gatherer cases in recent history (examined in Chapter 2), where the use of bone for the manufacture of sound-producers is relatively rare, with naturally occurring plant-based (or other organic) products being by far the dominant material use. It is worth noting at this early stage of the discussion that the apparent exclusivity of bone as a raw material for the manufacture of musical instruments in the Palaeolithic is likely to be the result of differential preservation, rather than the exclusive choice of that material on the part of those peoples carrying out the manufacture.

3.1. Bullroarers ("free aerophones")

Although less prolific than items considered to be flutes or pipes, dealt with below, another sound-producer of which there may be evidence in the Palaeolithic is the bullroarer. This instrument consists of a flat perforated piece of wood or bone on the end of a cord, which creates a whirring sound when spun in a circular motion, and is, or has been, used in a great diversity of global cultures, in both spiritual and functional contexts. For example, amongst some Australian Aborigines it is used to imitate the spirits occurring in the natural sounds of nature, and Malayans use it to scare animals away from plantations (Scothern, 1992). They are also used by many North American groups, the Mauri of New Zealand, in New Guinea, and amongst the Sami of Scandinavia.

Perforated pieces of bone have been found in Palaeolithic contexts associated with Aurignacian (c.38-28 Kya) through to Gravettian technological complexes (c.29-22 Kya, Mellars, 1994). Various possible functions for these have been proposed on the assumption that the perforations were man-made, including as fishing weights (Zervos, 1978) and as possible bullroarers (Scothern, 1992). Some of the most important work regarding the anthropic origin of pierced Palaeolithic bones, and thus "authenticity" of Palaeolithic musical instruments, has been carried out by d'Errico and Villa (1997) and d'Errico *et al.* (1998). They compared natural assemblages of faunal bones found at hyena dens and caves with those found in a variety of Palaeolithic contexts which have produced items variously claimed as worked bone, pendants or pipes. They also examined the relative quantities of different bones and parts of bone which survived in each case, and they looked closely at the effects on bones of carnivore activity, in terms of both gnawing action and digestive action. This included the production of lines and "incisions", smoothed "polished" surfaces and holes in bones.

As a result of these comparisons they convincingly argue that a large number of Palaeolithic "pendants" are in fact the product of hyena activity, not human behaviour; this includes objects such as those from Kulna, Bois Roche, Pech de l'Azé and Bocksteinschmeide. Although none of these pendant-like pieces had been claimed to be musical instruments, several similar ones from Aurignacian to Gravettian contexts have been suggested as possible bullroarers (Scothern 1992); as many small pendant-like pieces of bone with a single perforation seem to be the product of carnivore gnawing and digestion, until the reputed bullroarer artefacts have been re-analysed according to d'Errico and Villa's criteria, it will be impossible to make any assertions about their anthropic origin, let alone their function. If human agency in the creation of any of these objects can be determined with any confidence in the future, it may then be possible to carry out use-wear analysis of the area around the perforation, to determine the function of the objects.

However, there are examples of reputed bullroarers of which there can be no doubt of their human origin. A particularly spectacular example is an artefact from Magdalenian layers at La Roche de Birol, in the Dordogne (see Figure 3.1). This oval carved piece of reindeer antler is approximately 18cm long and 4cm wide at its broadest point. It is incised all over with long lines running along its length, and shorter lines perpendicular to these. It is pierced with a hole at one end, through which, it is presumed, would have been threaded a cord, and it is covered with red ochre. It is possible, of course, that this artefact might have served some other purpose than being a bullroarer; it may have been a very large pendant, for example. The size of the artefact and the fact that it is covered with red ochre would mitigate against it being used as a fishing weight, however. It could certainly have been extremely effective as a sound-producer, if that was its purpose. This was amply illustrated by Dauvois (1989) who produced a replica of the artefact out of reindeer antler using a silex tool, and found that it hummed (*vrombit*) at frequencies between 130 and 174Hz. As a comparison, a known bullroarer from New Guinea produced frequencies between 90 and 125 Hz (Dauvois, 1989); this was a particularly large artefact, at 370mm being more than twice the length of the La Roche example, and produced a deeper tone as a consequence. The smaller the artefact, the higher the frequency (and tone) of the sound produced by it. According to Dauvois, the La Roche example produces a particularly impressive sound when twirled in a cave, and even more so when it is played in the total darkness inside the cave, "vraiment un *son volant*" (Dauvois, 1989, p. 10).

In total, Dauvois describes or illustrates seven likely bullroarers associated with Solutrean and Magdalenian contexts. Figure 3.2(1) illustrates again the Upper Magdalenian example from La Roche, Lalinde (Dordogne). Fig. 3.2(2) represents an engraved example, 90mm long, from Upper Solutrean levels at Lespugue, Haute-Garonne. Fig. 3.2(3) is from Laugerie Basse, Les-Eyzies-De-Tayac, Dordogne, and is 107mm long, and Fig. 3.2(4) represents a 190mm long Solutrean example from Badegoule, Bersac, Dordogne. It is perhaps worth noting that both the Haute-Garonne area and Laugerie Basse/Les Eyzies have produced several examples of other sound producers (see Table 1, Appendix, 117 UP, and Table 2, Appendix, W030 G/P, W038-039 Ma, W051-056 Ma, W060 Ma and W083 UP; see section 1.3, below, for an explanation of these tables).

In addition there are three further examples from Upper and Final Magdalenian layers on bone described by Dauvois (1989). These are particularly interesting in that they are all engraved with images of bovids. An example from abri Fontalès, Tarn-et-Garonne is engraved on one surface, and examples from abri Morin (Pessac-sur-Dordogne, Gironde) and Lortet (Tarn-et-Garonne) are engraved on both sides with bovid images. Dauvois does, however, caution against the temptation to equate the sound of a *rhombe* with the mooing of a cow, lest we impose our own single interpretation on the evidence and fail to consider other possibilities (p.11). It must be said that this is a particularly great temptation given the English name for the items in question.

Again, it is possible that future use-wear analysis around the perforation may help to identify the way in which such items were used, and remove ambiguity from their interpretation. In addition to Dauvois' work (1989), other experimental reconstructions (Alebo, 1986) have tested the sound-making capabilities of a reputed bullroarer, as well as the process of manufacture of tendon string for the purpose, but there seems to have been no use-wear analysis carried out. Future experimental re-construction and use of bone bullroarers may allow patterns of abrasion around the perforation to be identified, caused

Figure 3.1: Magdalenian "bullroarer", covered with red ochre and incised with linear motifs, from La Roche, Lalinde, Dordogne. 180mm long, 40mm wide at widest point (*from Bahn, 1997, p. 85*).

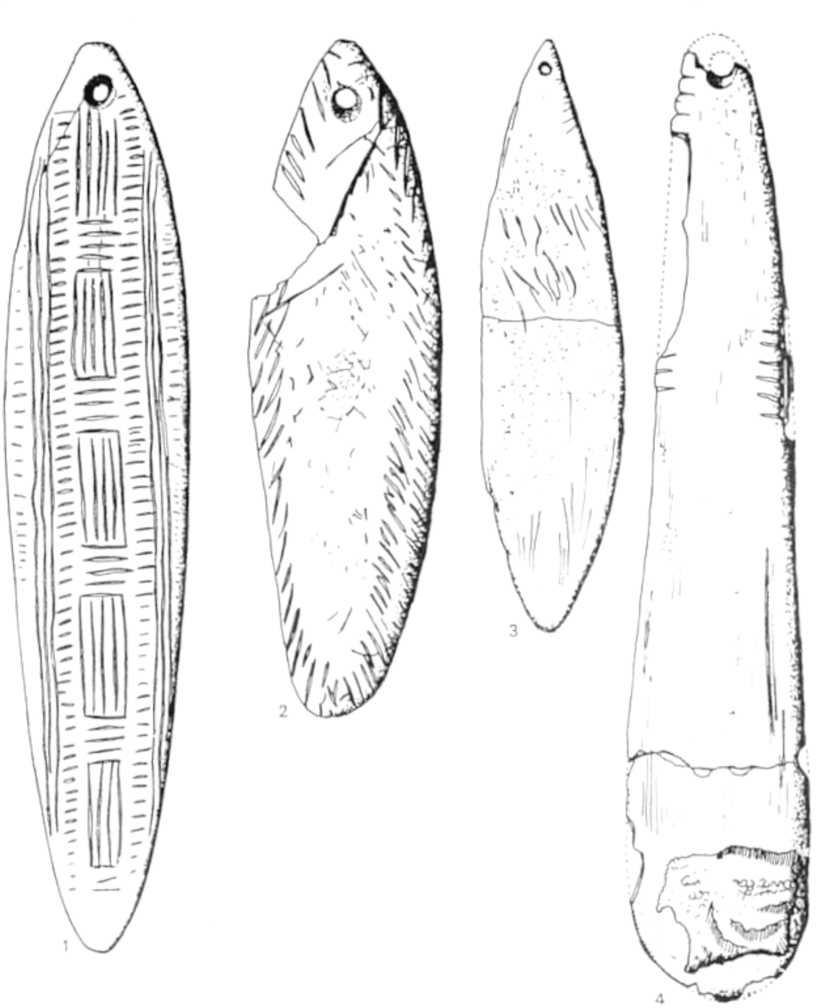

Figure 3.2: "Bullroarers".

1. Magdalenian; La Roche, Lalinde, Dordogne, see above. Length, 180mm.
2. Magdalenian; Abri de Laugerie Basse, Les Eyzies-de-Tayac. Length, 107mm.
3. Solutrean; Lespugue, Haute Garonne. Length, 90mm.
4. Solutrean; Badegoule, Dordogne. Length, 190mm.

(*from Dauvois, 1989, p.10*)

by the twine or tendon when used as a bullroarer, which can then be compared with any on Palaeolithic perforated bone fragments. One would expect such wear to be different from that caused by the threading of the object as a pendant or as a fishing weight.

Such research could be extremely valuable in adding confidence to the interpretation of pierced bone artefacts, especially more ambiguous ones; nevertheless, in the case of the Solutrean and Magdalenian examples described above, interpretation of them as bullroarers remains by far the most likely attribution of their purpose.

As with pipes, rasps and percussive instruments (see below), any surviving bone examples are likely to represent only the tip of the iceberg as regards the original providence of such items, as all could be (more) easily made from wood (as illustrated by many ethnographic examples, see Chapter 2), which would not survive to the present day.

3.2. Rasps ("scraped idiophones")

Another type of instrument possibly used by early humans is the scraped idiophone, or rasp (Huyge, 1990; 1991). A rasp can be a piece of wood, bone or stone with grooves cut into it perpendicular to its length, which are then rubbed with another object to create a staccato vibration. In the case of this type of sound-producer, there are not only possible artefacts, but also some renditions of items which may be rasps in Upper Palaeolithic engraving.

Huyge (1990) examines a Middle Palaeolithic (Mousterian) mammoth bone from Schulen, Belgium, dated by context to 50-40 Kya. This object was found in context with 18 Mousterian tools plus Lavallois cores and flakes, and has a number of parallel striations along one side (see Figure 3.3) which, Huyge suggests, may indicate its original use as a rasp. The bone is broken at the end bearing the striations, so it is impossible to tell how many were on the object before it was damaged, but at least twelve remain clearly visible (Huyge, 1990). The surface and edges of the bone are polished and smooth, which may be due to wind or water polishing after deposition, as is the case with the lithics found in context. Apart from the twelve evenly-spaced parallel grooves, there are no other lines or grooves visible on the bone, on the inside or outside surfaces.

D'Errico (1991), in response to Huyge's paper, suggests that the striations are instead the product of intense gnawing action, and the rounded edges due to salivary action, both of which features increase towards the broken end of the bone. Although this is certainly a possibility, in this instance this explanation is not wholly convincing. The grooves appear to be very regular in spacing, angle and depth, and are the only evidence of damage on the bone, apart from the break at the end, which also shows no other evidence of possible carnivore activity since it was broken. That the lithics found in context with the

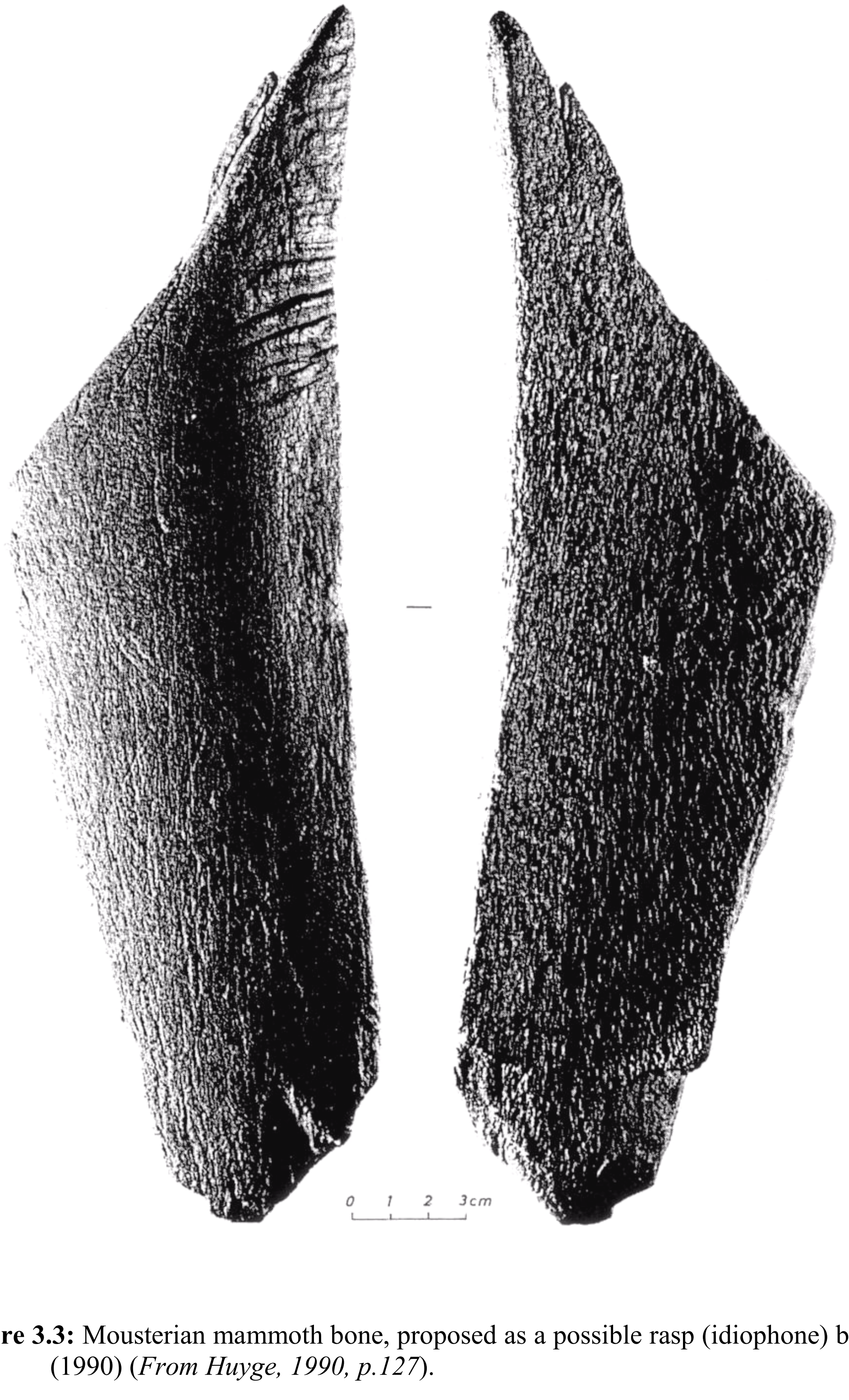

Figure 3.3: Mousterian mammoth bone, proposed as a possible rasp (idiophone) by Huyge (1990) (*From Huyge, 1990, p.127*).

fossilised bone also show polishing and smoothing suggests that weathering by wind and rain are a perfectly reasonable explanation for these traits on the bone.

In fact, the regularity and condition of the Schulen bone is in particular contrast to other claimed Middle Palaeolithic rasps (Capitan and Peyrony, 1912, and Vincent, 1988, quoted in Huyge, 1990), which have highly irregularly spaced grooves of inconsistent length, and do look far more like the product of carnivore activity than human intention (see Figure 3.4). Instead, the appearance of the striations more closely resembles those of the Magdalenian rasp-like object from Pekárna cave in former Czechoslovakia, which is made on reindeer antler, as well as very similar Magdalenian examples from Abri Lafaye Bruniquel and Mas d'Azil (see Figure 3.5), and an incised rib from La Riera cave, Spain, from a layer dated to 19,000-18,310 +/- 260 B.P. (González Morales, 1986 – although no speculation is offered there as to its purpose). For comparison, see also a Bronze-age example from Syria (Figure 3.6) (Huyge, 1990). The human agency in the construction of these later artefacts is not in question, but their use is still open to debate. They clearly have the potential to function very well as rasps, but there is always the possibility that the grooves on these objects serve some other purpose. Marshack (1972) has suggested that incised objects may serve as a tally or calendrical notation, or it may be that they serve to improve grip on the object, which was then used for something else; the same consideration applies, of course, to the Schulen bone.

In the instance of the Magdalenian examples from Pekarna, Abri Lafaye Bruniquel and Mas d'Azil an explanation of them as calendrical notation is not adequate; the grooves on each artefact are deep, evenly sized and spaced notches and not mere incised lines on the surface of the antler (which is all that would have been required for notation). The cuts are also such that they actually form a serrated edge to the antler, and are so consistent in form on each object that it is difficult to imagine that each cut was not made on the same occasion.

These artefacts also share a feature of a loop (or what appears to be the remains of one) at one end. It is possible that the serrations were created to improve grip on the tool while the loop at the end was used for some other purpose. Conversely, if they were used as rasps as is suggested by Dauvois (1989) and Huyge (1990), then they would presumably have been held using this loop, and the serrated edge would have provided the idiophone surface. It is possible that held like this, the serrated edge might also have served some other purpose, but it is hard to imagine what purpose a serrated-edged antler might have fulfilled that a lithic tool could not do better. It is conspicuous too that Mas d'Azil, Bruniquel and Pekarna have each also produced several examples of other artefacts

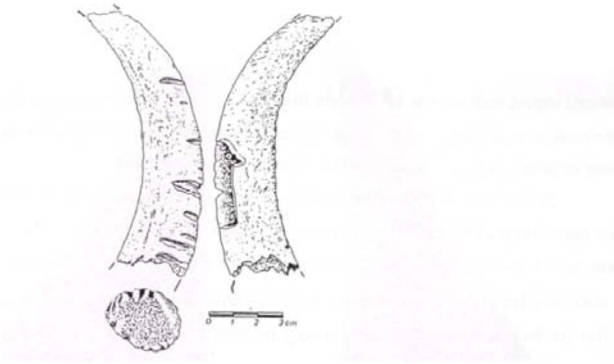

Figure 3.4: Middle Palaeolithic bone from Grotte Vauffrey. Claimed as rasp by Vincent, 1988. carnivore activity more likely (*from Huyge, 1990, p.130*).

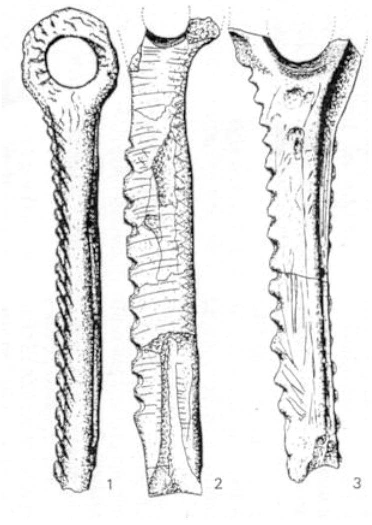

Figure 3.5: Magdalenian bone implements interpreted as idiophones:

1. Pekarna, Moravia (Czech Republic). Length, 205mm.
2. Abri Lafaye Bruniquel, Tarn-et-Garonne. (France). Length, 236mm.
3. Mas d'Azil, Ariège (France). No length given.

(*from Dauvois, 1989, p.11*)

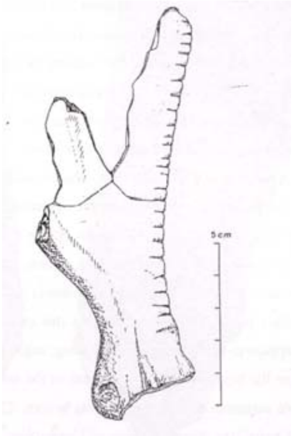

Figure 3.6: Bronze age incised bone, interpreted as idiophone, from Syria. (*from Huyge, 1990, p.130*).

interpreted as flutes, pipes and phalangeal whistles made on reindeer and bird bone, from Gravettian through to Magdalenian contexts (see below, and Table 1, Appendix, objects 050 G/P-061 G/P, 093 Ma, 094 Ma, 099 Ma, 119-20 UP, and Table 2, Appendix, objects W027-29 G/P, W063 Ma, W073 UP, W074 UP and W086-87 UP; see section 3.3, below, for an explanation of these tables). If these sites really did form a focus of activity for large groups of people, incorporating musical activity, then this may add to the plausibility of the interpretation of the antler artefacts as scraped idiophones.

Of particular interest is the bas-relief known as the “Venus of Laussel” or (“Dame à la corne”) from Laussel, France, and dated to 25,000 years ago (Huyge, 1991). This image consists of a corpulent woman holding an incised horn (generally accepted as that of a bison) (Huyge, 1991). The horn bears thirteen regularly-placed incisions along its length (see Figure 3.7, overleaf) and is being held up at shoulder height by the female figure. It has been proposed (by the original discoverer of the relief, G. Lalanne, quoted in Huyge, 1991) that this may represent a tally of hunting kills, given that the context of the rest of the bas-relief implies a possible association with hunting. As well as three females, the relief features a mutilated image of a masculine figure who appears to be about to throw a lance at a deer (Huyge, 1991; Stockman, 1986). It has also been suggested that it may be a record of (crescent) moons in a year (Marshack, 1972).

However, the horn very closely resembles idiophones made on bovine horn from Mexico and Bonaire (see Figures 3.8 and 3.9). The possible hunting context, if correctly interpreted as such, need not detract from the musicological interpretation of the object, given the many of associations of music and hunting represented in the ethnographic literature (see, for example, Chapter 2) and in the art of Çatalhöyük (Stockman, 1986; discussed further below). Although the purpose of the horn is impossible to identify with certainty, the relative simplicity of the rasp idiophone and the close resemblance of the image to existing examples makes an idiophone a very reasonable explanation.

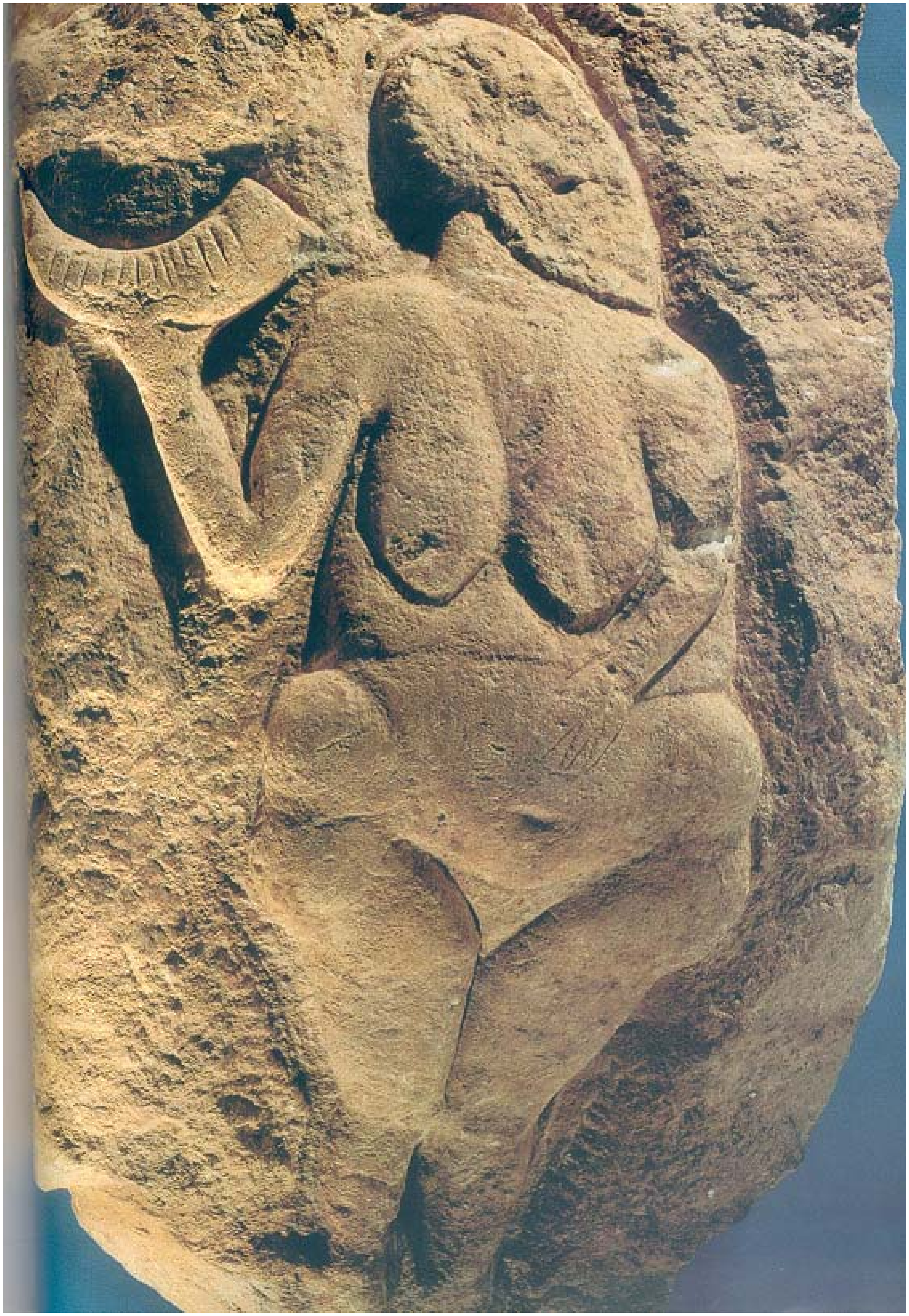

Figure 3.7: "Venus of Laussel", or "Dame a la Corne", Laussel, Dordogne. Height 440mm. She holds an incised bison horn, proposed as a representation of an idiophone (Huyge, 1990) (*from Bahn, 1997, p.113*).

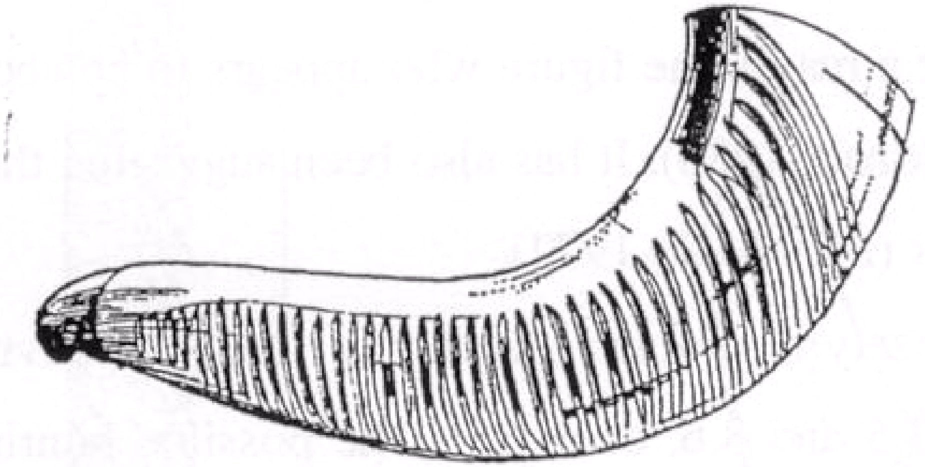

Figure 3.8: Incised bovine horn idiophone, from Mexico (*from Huyge, 1991, p.15*).

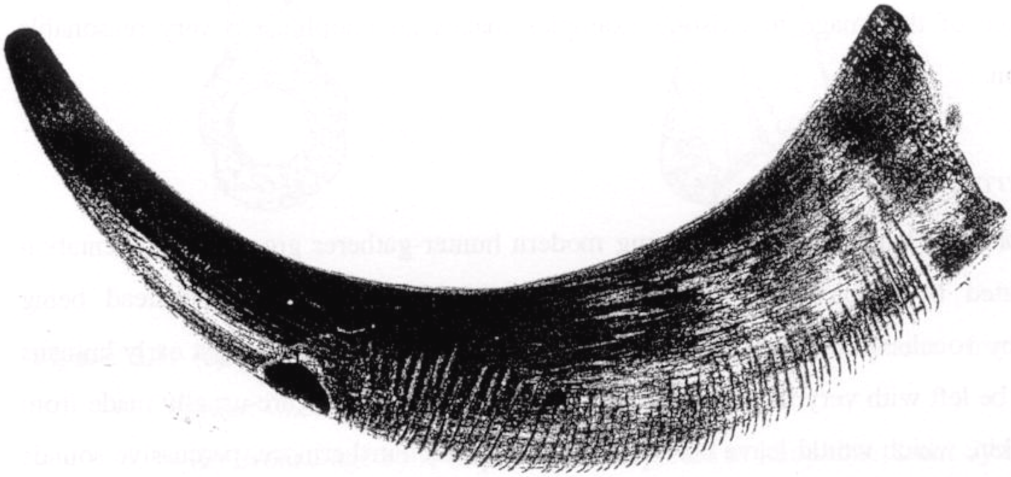

Figure 3.9: Idiophone on bovine horn, from Bonaire (*from Huyge, 1991, p.15*).

3.3. Pipes and flutes

3.3.1. The earliest pipes and flutes

Sections of animal bone with or without holes in their shafts, naturally hollow after the marrow has decayed or been removed, have been found in Upper and Middle Palaeolithic contexts since the 19th Century (Scothern, 1992; Lawson *et al.*, 1998). Many of these have been described as pipes or flutes (the two terms are used interchangeably) or whistles, and the catalogue of these items is now quite large. It is these objects that have received by far the most attention in considerations of Palaeolithic music. Scothern (1992), as part of her Doctoral thesis, carried out an inventory of these items, listing 106 sections of bone which had been variously described as flutes, pipes or whistles, plus around 90 punctured phalanges interpreted as "phalangeal whistles", all of which had come from Palaeolithic contexts. A further 21 were attributed to Mesolithic contexts. By those aware of it (it is unpublished), this has generally been considered to be the most complete inventory of these items (Lawson *et al.*, 1998). Until Scothern's work, the most complete catalogue of Palaeolithic flutes was that of Fages and Mourer-Chauviré (1983), who listed 30 (Dams, 1985; Lawson *et al.*, 1998), and it is this list that is still most frequently cited by other authors discussing Palaeolithic instruments. The current author has since corrected, expanded, revised and updated the information from Fages and Mourer-Chauviré (1983) and Scothern (1992) to the extent that the complete database now includes around 225 reputed flutes or phalangeal whistles (122 of the former and over 100 of the latter). These, and all available information about them, are detailed in the Appendix to the dissertation, in Table 1 for the "flutes" and Table 2 for the "phalangeal whistles". The latter are discussed in detail below. These tables have as their foundation the inventory prepared by Scothern (1992), and owe much to it, but as noted above, have been considerably expanded, corrected and updated. The reputed artefacts are organised alphabetically by site within each period, and each piece is identified by a numerical code designating its position in the table, and a letter designating its period; e.g. 011 A is the eleventh item listed, and is from the Aurignacian. The entries in Table 2 are preceded by a W (designating that they are reputed "phalangeal whistles"). Each entry contains notes on its original location, current location (where known), period, age (where known), a description of the object, and its status as a sound producer, if such information is available. It also notes the principal sources in which each piece is featured. It is not intended that all this detail should be replicated in the body of the text here, as the table provides the most concise and digestible form of this information, but some especially salient examples will be commented on.

The majority of these objects come from contexts associated with Aurignacian (c.38-28 Kya), Gravettian (c.28-22 Kya) and Magdalenian (c.17-11 Kya) technological complexes in Europe, and there are a few from Solutrean (c.22-17 Kya) contexts. There are also several dating to the Mousterian (c.200-40 Kya). These are particularly interesting (and have been subject to greater controversy) because the Mousterian pre-dates the time generally asserted for a so-called "Upper-Palaeolithic revolution" (e.g.: Mellars, 1989; Stringer and Gamble, 1993), after which we supposedly find the first evidence for art and symbolic thought. The implications of these items, if they could be shown to be intentionally produced, would be great in relation to the cognitive abilities of early modern humans and the ancients (archaic *Homo sapiens* and *Homo neanderthalensis*), prior to the arrival in Europe of anatomically modern humans.

There is an important distinction to be made between "flutes and pipes" and "phalangeal whistles". Flutes and pipes consist of a hollow(ed) bone, which may or may not have been perforated along its length. A pipe with no perforations would produce a single tone, and each perforation made in the tube would allow an additional tone to be made by covering or uncovering the holes, effectively varying the length, and thus wavelength, of the resonating column of air inside. The closest modern approximation is a recorder or flute. Phalangeal whistles, on the other hand, consist of a single pierced phalanx, usually punctured with only one hole. A tone is produced by blowing across the top of this hole, in much the same way as one would if producing a tone by blowing over the top of a wine bottle. This would produce a high-pitched sound, due to the small internal volume of the phalanx. Table 2 deals with the phalangeal whistles, and they are discussed in detail below (section 3.4).

These items, of both types, have been subject to considerable debate over the years. Foremost is the issue of whether a given item can be confidently identified as an *intentional* sound-producing object; i.e.: not merely an object which is capable of producing sound, but one which was manufactured with the intention of it being used to produce sound. Second, as noted above, is debate regarding the implications of such objects for the possibility of symbolic thought amongst their manufacturers, and whether they could have been produced only by *Homo sapiens* or by Neanderthals or even *Homo heidelbergensis*.

A high proportion of the objects detailed in Table 1 were recovered from excavations in the latter part of the nineteenth century and first half of the twentieth century. As a consequence their provenance is often unclear, with site reports often nonexistent or lacking any mention of specific pieces, exact descriptions of them and of the materials used in their construction are absent, and the basis for their interpretation as sound-producers often seems to be merely upon their appearance and the absence of any directly contradictory evidence. Frequently the main publication detailing them was written years after their discovery, on the basis of examination in museum collections (e.g. Seewald, 1934; Megaw, 1960). In consequence, many of the (often widely cited) examples are of dubious authenticity as genuine sound-producers, and many others which are more likely to be genuine instruments are rarely, if ever, cited. Table 1 includes comments on the status of such examples.

As has been discussed above, the anthropic (man-made) origin of many perforated bone artefacts has been called into question by the work of d'Errico and Villa (1997) and d'Errico *et al.* (1998). D'Errico and Villa also concur with Davidson (1991) that the perforated length of bone from Haua Fteah in Libya (002 M in Table 1), dated by context to the Mousterian, is most likely the product of carnivore teeth on the basis of the morphology of the hole and the lack of any evidence of human workmanship. This was originally proposed as a flute (McBurney, 1967) and has been reported as such since (e.g. Scothern, 1992, albeit amidst debate regarding its correct age).

Fortunately, the features which result from carnivore activity are possible to recognise if the objects in question are analysed closely. Generally the surface of the bone is pitted with tiny indentations (microconcavities) and the edges of the bone are thin and sharp. Holes have cylindrical sections with straight edges and depressed margins, rather than the more conical or biconical (hourglass) shaped sections of holes deliberately produced with a borer or burin (d'Errico and Villa, 1997). It is clear that many of the original reports of bone flutes need to be viewed with some caution, and that many of the artefacts may require careful re-examination and re-interpretation using d'Errico and Villa's criteria.

The issue of the existence of "modern", instrumental, musical behaviour amongst Neanderthals has been a source of great debate. There are four objects from Mousterian contexts that have variously been interpreted as pipe or flute sound producers. As noted above, the possible anthropogenesis of the example from Haua Fteah, Libya (002 M), has been effectively dispelled by d'Errico and Villa (1997), leaving an example from Ilsenhöhle, Germany (004 M), which is of doubtful stratigraphy, a hare bone from Kent's Cavern, England (003 M), which is unlikely to be a sound-producer according to Scothern's (1992) analysis, and the so-called "Neanderthal flute" bear femur from Divje babe I cave in Slovenia (001 M).

This latter piece has recently received a great deal of publicity and academic attention due to its potential importance with regard to the origins of instrumental behaviour in Neanderthals, and as the oldest reputed musical object to date. It is notable that in general "flute" finds appear to elicit more attention and excitement than other reputed instruments, there being what could almost be described as a "flute fetish"; when the object in question is reputed to be both the oldest, and from contexts associated with a different species (or sub-species) of hominid it generates even more passionate debate. Several reconstructions of this item have been attempted, for example, by Jelle Atema (Anon., The Times, 21/02/2000; Ahuja, The Times, 23/02/2000) and Fink, (1997; 2000). The potential importance of this item and the polarisation of the debate regarding it warrant a closer examination of the issues here.

3.3.2. Mousterian musicianship? The case of the Divje Babe I bone

A thorough review of the literature regarding this object reveals that the debate about its origins is characterised on all sides by an odd blend of thorough research and well-reasoned arguments on one hand, and passions, inconsistencies, misquotes and mis-understandings on the other. It also reveals that, contrary to initial expectations, the issue is far from easily resolved, with some questions impossible to answer at this stage. The literature falls into three main groups; those who believe the features of the Divje babe I bone to be the product of carnivore activity (e.g. d'Errico *et al.*, 1998; Chase and Nowell, 1998), those who believe that it is undoubtedly a Mousterian flute (e.g. Fink, 1997, 2000; Otte, 2000) and those who advocate it being a flute, consider the possibility of carnivore action, and have carried out experimental analysis of both possibilities (e.g. Turk *et al.*, 1997; Bastiani and Turk, 1997; Kunej and Turk, 2000).

The bone in question is the diaphysis of the left femur of a one to two year-old cave bear, found securely locked in context in breccia near a Mousterian hearth. The layer (layer 8) is radiocarbon dated to 43,100 ± 700 yrs B.P. (Nelson, 1997) (This date is in contrast to dates sometimes erroneously quoted of up to 82,000 yrs; e.g. Fink, 1997). The bone features two artificial holes in the posterior side, and semicircular features at the proximal and distal ends of the posterior side which have been suggested to be the remains of a further two holes. The bone has the following dimensions:

Length: 113.6mm (original length of the diaphysis plus metaphysis: c.210mm)
Width at narrowest point: 23.5mm and 17.0mm (due to oval section)
Maximum diameters of the two holes: 9.7mm and 9.0mm
Distance between the centres of the two holes: 35mm

(Kunej and Turk, 2000). See Figure 3.10 (p. 59).

The principal arguments for this object being a flute have relied upon musicological analyses of reproductions (Fink, 1997; Kunej and Turk, 2000) and the unusual features of the damage (Turk *et al.*, 1997; Bastiani and Turk, 1997; Kunej and Turk, 2000). The main arguments against its being a man-made object rely mainly on comparisons with other bones that have been the subjects of carnivore activity (d'Errico *et al.*, 1998) and with further observations of the bone itself (Chase and Nowell, 1998). Responses to these arguments from the advocates of an anthropic origin for the object have pointed out that the damage on the Divje babe I bone is quite atypical of bones that have been subject to carnivore activity, and shows damage that is difficult to attribute to carnivore action.

Musicological analysis:

Whilst there remain only two complete holes in the bone, commentators have frequently asserted that there may have originally been three, four, or even five holes made in the bone, e.g. Otte, (2000) says "the instrument consists not of two perforations (as Chase and Nowell indicate) but five ("like five fingers of a hand"): four on one side, one on the opposite side… the fifth hole appears at the base of the opposite side, at the natural location of the thumb" (p.271). Fink (1997) observes that if there were four (or five) holes in the bone before it was damaged, they appear to be spaced so as to reproduce a portion of a diatonic scale (me, fa, so, la); the probability of such a set of hole spacings having occurred by chance exceedingly small (Fink, 1997). Whether the holes do reproduce these notes as a portion of a scale depends, though, on the total length of the original bone.

However, Kunej and Turk (2000) state that, due to the young age of the bear cub, the complete bone would not have been long enough to reproduce the length of air column necessary for Fink's (1997) reconstruction. Consequently the issue of the chance occurrence of a diatonic scale spacing for the holes becomes a non-issue. Fink, in updates to the pages, posits the idea that the portion of the bone may have been extended with the addition of a further section of bone to create the required length to produce the "me, fa, so, la" sequence of notes. However, as outlined below, as far as Turk *et al.* (1997) are concerned, the damage at the ends of the femur, including the semi-circular hole damage, is consistent with carnivore damage, and is paralleled by several other cave-bear femurs from the site (Fig 11.19, p.173, Turk *et al.*, 1997), only one of which displays clearly any other tooth-mark damage. This, in all probability, leaves under question only the agency responsible for the two complete holes in the bone.

Hole spacing:

The spacing between the holes (35mm) does not match the spacing between known carnivore canines (the smallest found were 45mm), so the holes are likely to have been made on separate attempts (Turk *et al.*, 1997). The morphology of the edges of the complete holes differs from that of the (incomplete) suspected hole on the anterior side and one of the (incomplete) suspected holes on the posterior side (Kunej and Turk, 2000), which as noted above, seem to be typical of carnivore damage (Turk *et al.*, 1997). The putative fifth hole (often cited, e.g. Otte, 2000) is dismissed by Turk *et al.* (1997), saying that the distal anterior fracture "…constantly leads to the idea that there was another hole here. However, it is not so, or the hole was more rectangular" (p. 157). Kunej and Turk (2000) say that "In the flat, posterior side [of the bone], two holes had been chipped. Perhaps a third hole was here; however, it was not made in the same way as the first two, or contemporaneously with them" (p.249). They also say that the "hole" in (damage to) the anterior side of the proximal end was made in the same way and at the same time as this "third" hole (in fact, they must mean the reputed hole in the anterior side of the *distal* end, as there is no such damage at the anterior proximal surface).

With regard to this semi-circular "third" hole, a little further confusion is added by the following inconsistency: Kunej and Turk (2000) state that the flute "was not damaged during excavation, although it was located deep in breccia" (p.240), but in Turk *et al.* (1997) it is stated that "The semi-circular notch on the proximal part may be the remains of a third hole, which was additionally slightly damaged during excavation" (p.160). Whether they are saying that the notch *was* damaged during excavation, or that it *may* have been damaged during excavation is unclear.

As for the semi-circular notch at the distal end, "The morphology of the edges of the notch corresponds to the morphology of the rims of both complete holes. However, it is highly probable that a carnivore actually made the notch, and that it is not the remains of a hole or an artificial mouthpiece" (p.249). It is curious that they make this statement, as they elsewhere imply that the morphology of the edges of the complete holes is reminiscent of damage caused by chipping out the holes (Bastiani and Turk, 1997; Kunej and Turk, 2000); in contrast, by implication, the previous statement would suggest that the morphology of the edges of the complete holes was also in all probability caused by carnivores, or at least that damage resembling chipping can, without any recourse to human agency, accompany damage caused by carnivore gnawing. Whether such damage occurred at the same time as gnawing damage or at a later date is open to speculation.

Otte (2000) asserts that the evidence of carnivore activity at the ends of the bone are overlying evidence of human activity. There is, however, no reason to believe that there was any human agency in the production of the semi-circular holes at the ends of the diaphysis; these can be explained perfectly well by carnivore gnawing at the ends of the bone. As Turk *et al.* (1997) state, “there are quite a number of such examples on limbs at the site” (p.174) (see also preceding paragraph, above). In fact, Chase and Nowell (1998) state that the carnivore damage to the ends of the bone is a “textbook case” of the damage caused by carnivores when they remove the nutritious cancellous tissue from the ends of the bone; since the object would have been non-functional as a flute before the removal of this tissue, this implies that if there is any evidence of human agency on the bone it would have to post-date the damage to the ends of the bone, not pre-date it. Therefore the carnivore activity cannot be said to overlie evidence of human activity in these areas of the bone.

In summary, it would seem that there were probably only ever two complete holes in the bone, that the shape of the ends of the bone can be ascribed to carnivore action, and that damage resembling chipping can accompany carnivore gnawing action. Whilst it does remain likely that there were a number of factors and processes responsible for the features of the bone as we find it today, in the case of the ends of the bone, human activity need not be invoked to explain these features.

Bone damage and the mechanics of piercing:

The complete holes in the diaphysis are more difficult to explain as the product of carnivore activity. The force required to pierce the bone cortex completely with a wolf or hyena tooth (canine, carnassial or precarnassial) would be likely to crack a juvenile bone; it would also be expected to leave traces of tooth marks on the opposite side of the bone to the piercing, and adjacent to the piercing. Such traces do not appear to be present on the bone (Kunej and Turk, 2000).

However, Turk *et al.* (1997) do say the anterior surface has fossil damage in a number of places: slight “indentations” and at least two shallow “cuts”, but due to poor preservation of the surface it is impossible to assert how the damage was caused. Turk *et al.* (1997) also say that “Bones which have one straight wall and the other convex, normally give under pressure on the straight or concave side. So there are twice as many holes and indentations on straight or concave surfaces than on convex ones. Similarly, holes and punctures are created mainly where the bone compact is thinnest. On concave surfaces and thickened (side) walls we have only slight impressions which heavy weathering or polishing of the bones can erase. So it is not crucial for holes which have been made by carnivores to have traces of teeth on the opposite side" (p.172). Indeed, Turk *et al.* (2001) state that of seven juvenile cave bear bones from Mokriska jana, Slovenia, pierced on the proximal side as is the case with the Divje babe I specimen, only one shows any other signs of chewing. They go on to say that "The form and size of the holes [thus produced] does not differ from those on the suspected flute" (p.172). This archaeological evidence directly contradicts a statement made the year before by Kunej and Turk (2000), saying "opposing and neighbouring teeth should have made an impression with such a powerful bite force as is required for the tooth to pierce the thick, compact bone of the diaphysis" (p.244). In any case, as Chase and Nowell (1998) point out, and as is clearly visible in good photographs of the bone, it is "in large part covered by calcareous encrustations which may or may not hide other traces of carnivore teeth" (p.551).

As for the cracking of the bone under pressure, Turk *et al.* (1997) established by experiment that fresh bear bone is so elastic that they could puncture holes in it at any point without its smashing as fossil bone does. Having said this, the force required to puncture the middle portion of the diaphysis was so great as to lead the authors to conclude that such holes could only rarely be made by carnivores, and then only by those with very strong jaws and pointed carnassials (see 4, below).

In summary it would seem that the bone need not have shattered under the pressure required to puncture it, and that it would be more likely to puncture on the flat side of the bone than the convex side, the latter which need not show any tooth marks (if one accepts Turk *et al.*, 1997), though there are some unidentifiable cuts and indentations on the anterior surface. It is the flat (posterior) side, in fact, where we find the holes in the Divje babe I specimen. The issue of the probability of the holes occurring in line with each other is considerably reduced in magnitude by virtue of the fact that three of the putative five holes were almost certainly never holes, and the remaining two are in the side of the bone most likely to puncture, after the weaker epiphyses.

There does still remain the issue of the difficulty of puncturing the thick wall of the diaphysis at this point, though. These middle two holes are in the thickest part of a femur, which is an unusual place for carnivore gnawing, particularly twice on the same bone. Nevertheless, as discussed above, if this area of the diaphysis is to be punctured at all, it is most likely to occur in the area (the flat side) where it did so. The shape of the holes reflects that of a canine tooth (producing nearly circular holes) rather than a carnassial or precarnassial (producing oval or rhomboid holes) (Kunej and Turk, 2000). However, it is unlikely that canine teeth could exert the force necessary to pierce 3-4mm of juvenile cortical bone (1300-1900 Newtons; Kunej and Turk, 2000), and if they had been used, one would expect to see tapering in the outer sides of the hole rather than a straight-sided puncture. Turk *et al.* (1997) consider that holes with a diameter greater than 5-9mm (the hole diameters on the bone are 9.0mm and 9.7mm) could not have been produced by a wolf, the predominant carnivore represented at the site. They state that lower P_4 and upper P^3 of hyena are very suitable candidates for the damage, in size and shape, but hyena are so far unrepresented at the site. It should be noted, however, that they are common in the region, including the neighbouring cave (Turk, 1997). Upper P^3 of cave lion or leopard are also suitable, and these fauna are represented well at the site, including in the same layers as the flute. However, cave lion and leopard do not devour bones.

The damage features on the holes – possible causes:

If human agency for these two complete holes is asserted instead, they could theoretically have been made by boring, punching or chipping the hole. Boring is unknown in the Mousterian (Turk *et al.*, 1997), appearing only in the Aurignacian. In any case, this is easily recognisable in artefacts by funnel-shaped holes. The holes in the Divje babe bone do not have this shape (Turk *et al.*, 1997).

Punching, when attempted experimentally by Turk *et al.* (1997) on the fresh thigh bone of a young brown bear, proved impossible with all tools with the exception of an iron punch. Holes made by punching or piercing have a characteristic irregular cone shape created in the direction of the operation of force (i.e. on the inside). Turk *et al.* (1997) state that “With punching or piercing, the bone is fractured conically on the opposite side. Our example also has such fractures” (p. 171). They go on to say that distinct bevelling along the *inner* edges of the holes was a characteristic feature of the supposed flute (p.177).

Chipping out a hole, when attempted experimentally, produced a characteristic type of damage including a slightly conical shape on the outside. Turk *et al.* (1997) state that the find does not feature such damage or a conical shape. However, having said in 1997 that the damage morphology of the holes does not resemble that experimentally reproduced by chipping (Turk *et al.*, 1997), the opposite observation is made in 2000 (Kunej and Turk, 2000). In contrast, they suggest that the edges of the (central) holes on the bone have chipped edges and transverse striations; these, they say, *did* resemble the damage created by using a stone tool to chip and bore a hole in a modern bear femur. Kunej and Turk (2000) actually make several contradictory statements regarding the chipping of holes in bones. Initially, they say that the technique of chipping a hole in bone with a stone tool, supposedly producing a characteristic shape of hole closely resembling those in the Divje babe I bone, is “not recognised in the Upper Palaeolithic and is so far completely unknown” (p.245). However, they go on to say that “unusual, mostly single holes in bones… were present in large numbers in Slovenia at the beginning of the Upper Palaeolithic and then completely disappeared. Some are almost identical to the holes in the Divje babe I bone” (p.248). So, are there other examples of bones with holes resembling those of the Divje babe I bone, or not? And if so, are they chipped? The answers to these questions remain unclear.

Kunej and Turk (2000) also say “Among tools [found at the site] a great many are suitable for making holes. Some are damaged in such a way as to suggest their use for chipping bone, but they do not show the specific damage that occurs if used for that purpose” (p. 239). This statement seems to be an oxymoron; they do appear to be saying, however, that the damage on the tools found at the site, though possibly caused by their use for chipping bone, does not closely resemble the damage on experimentally produced tools when used for that purpose. Elsewhere (e.g. Fink, 2000) this statement is interpreted in the opposite way, namely, that the use wear reproduced experimentally on stone tools in chipping holes in bone closely resembles that on tools found at the site. Again, this remains unclear.

Raw material use:

Otte (2000) suggests that the abundance of cave bear bones at the site and the difficulty of other provisioning due to the altitude of the cave would have resulted in the Mousterians having a “natural and spontaneous tendency” to use these bone remains for whatever they were making at the site. Were this the case, however, there would surely be more evidence of human agency in the cave bear bones present at the site; in fact, there are only two other possible cases of human action on all of the bone from the site, and the subject femur is the only one of 600 cave bear femurs to bear any such possible traces (Turk, 1997).

Otte also suggests that the lack of other flutes from Mousterian layers is not due to their absence, but lack of preservation, the majority of flutes being made of wood rather than bone prior to the Upper Palaeolithic. The issue of the use of wood as a raw material in the Middle and Lower Palaeolithic is certainly a very important one, particularly with regard to instrumentation given that the majority of musical instruments are (traditionally) made from biodegradable matter. However, in this instance it seems unlikely that this was an adaptation onto bone of an existing technology previously used on wood, if the holes were, as Bastiani and Turk (1997) and Kunej and Turk (2000) suggest, chipped out - which is, they assert, the only way they could have been made.

Summary

The principal arguments and problems presented by the advocates of the flute are answered quite effectively by the close analyses of the object carried out by its finder (and sometime advocate), Turk, and colleagues. The strongest arguments for the object being a flute had hitherto been *arguments from design*; namely, the features, appearance and sound reproductive properties of the object. However, as outlined above, many of the "design features" of the object are easily explained by other means, and do not appear to have been made contemporaneously with each other. There does remain one significant problem, however. If the two holes in the diaphysis *are* the product of carnivore activity, it is very unusual carnivore activity. Whilst the collections of cave bear bones examined by d'Errico *et al.* (1998) do show similar shaped and damaged holes, and a similar incidence of such holes, none of these occur in the diaphysis of a femur. The best candidate they can suggest is an unspecified "large carnivore with strong jaws and very pointed carnassials" (p.76). The damage is certainly not easily attributable to any of the carnivores present at the site, without making significant allowances for unusual behaviour on the part of the carnivores. Herein lies a central issue of the debate, because according to received wisdom, the counter-argument also relies on making significant allowances for unusual behaviour on the part of Neanderthals. It is this situation that has caused some of the most heated responses to the object, and the analyses of it; for example, Otte (2000) is incredulous at what he sees as the prejudice inherent in interpretations of objects found associated with Mousterians and the *a priori* assumptions made about Neanderthal aptitude. Hopefully the above points illustrate, however, that the debate need not make any recourse to human aptitude or prejudice about Neanderthals in order to address the major issues associated with the creation of the features of the bone.

The balance is swung against hominid agency, the current author believes, by the abundance of evidence of *other*, typical, carnivore activity on the bone, whilst the direct evidence for human agency is ambiguous, and there is a lack of evidence of *other* possible human workmanship on the bone.

3.3.3. Upper Palaeolithic pipes and flutes

The reliability of the interpretation of such objects has generally been considered to be greater in the case of those of more recent origin, as a consequence of the existence of greater numbers of pieces and generally better preservation of individual specimens. It does have to be said though, that with the exception of the work of Lawson and d'Errico (2002) on the most complete Isturitz pipes, none of these objects from Upper Palaeolithic contexts have been subject to the same level of analysis as has the Mousterian object from Divje babe I. In appraising these objects one must be aware of a sometime inclination on the part of authors to impose a musicological interpretation on the evidence *ex silentio* (c.f. Otte's (2000) position as outlined above). Scothern's (1992) analysis, however, went a good way towards addressing the objective analysis of many of the reputed artefacts.

The earliest known widely-accepted pipe-type sound-producers found to date are made on a pair of swan ulna or radii, from Geissenklösterle, in Germany (see figure 3.11, p. 59) (005 A and 006 A, Table 1, Appendix). These are dated to 36,800 +/- 1000 years BP, and were found in context with Aurignacian II split bone points (Hahn and Münzel 1995; Turk and Kavur, 1997; Richter *et al.*, 2000)**.** One of the bones has 3 holes preserved (diameters 5.3 x 3.4mm, 3.5 x 3.0mm, 2.8 x 2.4mm), with the distance between holes being 30-40mm. At least two of the holes have the conical shape of a hole bored with a tool (Turk and Kavur, 1997). It is suggested that it would originally have been 18-19cm long, but is now 12cm long, with a surviving bore of 8-9cm. The second specimen has 1 hole remaining (Hahn and Münzel 1995; Turk and Kavur, 1997; Goldbeck, 2001).

A musicological analysis of a reconstruction in white resin of the most complete of the two artefacts has been carried out at Tubingen University by "archaeological engineer" (Archäo-Techniker) Wulf Hein and musician Bernadette Käfer. Hein carried out the reconstruction at the original location of the find, which took only an hour, where it was subsequently tested by Käfer. It was discovered that in practice, two different air-columns can be brought into vibration to create sound in this artefact, one between the mouth and lips (as in many brass wind instruments) and one at the sharp edges of the holes, as in a conventional end-blown pipe or flute. Hohner (recorder) company also carried out a frequency-analysis of the reconstructed instrument, which showed that it was capable of producing other pitches in addition (Goldbeck, 2001).

By far the richest source of intentionally-produced sound-makers is Isturitz, in the French Pyrenees. This site shows evidence of habitation from Aurignacian through to Magdalenian contexts as a social focal point of contact, and has produced a variety of art and bone-working. Seventeen bone flute-like objects have been retrieved, from throughout the period of occupation of the site, and several of these show deliberate signs of working, with truncated ends, holes bored and then smoothed, incised lines, and even some block and duct examples (Scothern, 1992). Graeme Lawson and Francesco d'Errico (2002) have carried out extensive analyses of the most complete examples (Lawson and d'Errico, 2002; d'Errico *et al.*, 2003), and suggest that at least two of the Aurignacian examples seem to have been designed to be played as end-blown trumpet- or reed-voiced pipes, rather than as flutes. There seems to be a great consistency in their manufacture, and they closely resemble numerous Mayan and mediaeval examples. It is the various Isturitz examples that are most frequently cited and discussed in the literature regarding Palaeolithic instrumentation, and they are particularly significant in appearing to indicate large scale communal musical activity or, at least, musical activity at large scale communal gatherings.

Whilst they are undoubtedly important, the reverent status sometimes accorded to some of the Isturitz examples should be tempered, however. Several of the flute-like sections of bone are extremely fragmentary or seemingly unfinished. Scothern nevertheless seems very keen to accept them all as sound producers. Despite commenting that the bone that she calls SP52D (038 G/P in Table 1) "…represents a complete and intact artefact [which] is therefore very puzzling, since it does not give any indication of sound production" (Scothern, 1992, p. 94) and features of bone No.77142 (012 A in Table 1) indicate "…that the inner bone has not been worked in order to enlarge its internal bore and, therefore, influence its acoustic properties" (Scothern, 1992, p.94), she still concludes that "the seventeen bone fragments from Isturitz must surely now be interpreted as sound-producers" (Scothern, 1992, p.101). Although some, especially those examined by Lawson and d'Errico (2002), have undoubtedly been worked, and very finely, others perhaps should be interpreted with more caution. Some of the Isturitz bones are separated by many thousands of years in age. To assign them all the same qualities and origins on the basis of sharing the same site (albeit a very important one artistically and culturally) seems ill-founded.

Nevertheless, many of the Isturitz flutes provide a real indication of the level of sophistication of which their manufacturers were capable, and are indicative of a level of importance to musical activity only hinted at elsewhere. The style and craftsmanship of the true Isturitz flutes continued through the Magdalenian, albeit with improvements in hole-making techniques with the invention of burins. As noted, the design seems to have remained unchanged up to at least Mediaeval times in Europe and the Americas.

The earliest definite example of a "block and duct" flute comes from a Gravettian context at Goyet, in Belgium (032 G/P). This 10.5 centimetre-long bone has the first example of an intentionally produced "sound window" (Scothern, 1992), a feature designed to direct the breath against a block in the inside of the bone, creating the vibrating air column necessary to produce sound. The block in the middle of the bone might be made of wood, or resin, as is the case in one of the examples from Dolni-Vestonice, Moravia (110-113 UP), where the resin plug is still, remarkably, preserved. Modern recorders (or vertical flutes) still use the same block and duct principle, essentially unchanged for 25,000 years.

As for other less certain artefacts, Scothern (1992) concludes, from the literature and personal observation, that the early Aurignacian pieces from St. Avil (024 A), Istállóskö (011 A), Bukovac (008 A) and Potocka Zijalka (022 A) are more likely to be the products of natural processes than the result of human agency. Although all can produce sound, it is impossible to demonstrate that the modifications are intentional or man-made. Later Aurignacian and Gravettian artefacts from Spy (025-31 A) and Maisieres Canal (044 G/P) are more likely to be intentional sound-producers according to Scothern (1992). Where commentary and assessment is available, the reputed artefacts in Table 1 include notes regarding their likely status and capabilities.

Representations of musical instruments in Palaeolithic art are sparse, as are any representations of people in parietal art before the Holocene. Apart from the possible representation of a rasp in the hand of the Venus of Laussel (section 3.2, above), there is one example of an illustration of an anthropomorphic figure with an object protruding from the face area which may be a nose flute. This engraved image is at Grotte de Trois Frères, Ariège, and depicts a figure with the head of a bison, who appears to be dancing, and from whose nose appears to protrude a long linear object. It has been suggested that this may represent a nose flute (Dauvois, 1989; Bahn, 1997). Bahn's suggestion is accompanied by exclamation marks, which belie the fact that nose flutes are a well-documented instrument in many ethnographic groups (see, for an overview, Sadie, 2001). Whilst this is a possible interpretation of this picture and, potentially, of some artefacts, in this instance the picture is so ambiguous and stylised that it is difficult to be certain as to what is intended to be represented. If the image is to be interpreted literally, then it is possible that it depicts a person wearing a bison head, with a pipe to the nose of the bison; in this case it is difficult to imagine that is would be possible to play the pipe. Alternatively, it may be that it is a pipe and the image is supposed to depict it merely being held to the face of the bison mask, or it may be something else entirely. On the other hand, if it is an abstract and symbolic image representing, for example, a "bison spirit", then it is difficult to make any confident assertions about the meaning of any of the extraneous lines and features of the engraving.

3.4. Phalangeal whistles

Another form of aerophone sound-producer that is almost as prolific as the reputed pipes and flutes is the "phalangeal whistle". These objects consist of a phalanx bone pierced, in most cases, with a single hole at the proximal end of the distal surface (Harrison, 1978).

The product of this is a small hollow vessel-flute; by placing the proximal end of the phalanx against the lower lip, one can blow over the top of the hole to produce a clear tone. Some examples pierced nearer the centre of the diaphysis may have been played by placing the bone lengthwise against the lips (see Figure 3.12 for examples). The tone produced depends upon the size of the phalanx and the dimensions of the hole (Harrison, 1978; Dauvois, 1989). Known details of these reputed artefacts are outlined in Table 2, Appendix.

Many such objects have been found in Palaeolithic contexts dating from the Mousterian through to the Magdalenian, the Mesolithic, and beyond to recent history. There are records of their manufacture and use in various ethnographic contexts (e.g. amongst North American Indians; Sollas, 1924) as an aid to signalling in hunting, or in shamanistic rituals (Harrison, 1978). Dauvois (1989) has compared their use to the whistling language of the shepherds of the Vallée d'Aas in the Pyrenées-Atlantiques.

Of those phalanges of which the faunal origin is stated, the overwhelming majority are from reindeer (e.g.: Megaw, 1960; Harrison, 1978, Dauvois, 1989), and consist generally of the first phalanx, although the smaller second phalanx is also sometimes used (Dauvois, 1989). Harrison (1978), who carried out a series of experiments regarding the process of creating punctured phalanges (see below), found that the creation of a phalangeal whistle using a stone tool took on average between three and four minutes.

That such whistles have had an anthropogenic origin and important roles in human society at various points is not in doubt, and many of those phalangeal whistles found are clearly the result of human agency. However, the interpretation of many of these objects as intentionally created sound producers is complicated by the fact that damage resembling deliberate piercing can occur naturally very easily under the process of deposition (Harrison, 1978), or as a consequence of carnivore action (Chase, 1990). Although a phalanx itself will not provide much sustenance, they can be swallowed by a carnivore in the course of consuming neighbouring areas, and suffer damage through chewing and digestion. Chase (1990) illustrates this with an example of a phalanx found in the coprolite of a coyote, which closely resembles many punctured phalanges found in archaeological contexts (Enloe *et al.*, 2000). The area at the posterior surface of the proximal end of a reindeer phalanx is very fragile relative to the rest of the phalanx, being about half the thickness of the bone elsewhere, and when pressure is exerted upon a phalanx by a surrounding gravel matrix it is invariably this area that punctures, whether the bone is fresh or old (Harrison, 1978). Furthermore, the hole that results is not dependent upon the shape of the stone that pierced it, but upon the properties of the bone itself, meaning that very

Figure 3.10: The Divje babe I bone: pierced femur of juvenile cave bear, length 113.6mm (*from Turk, 1997*).

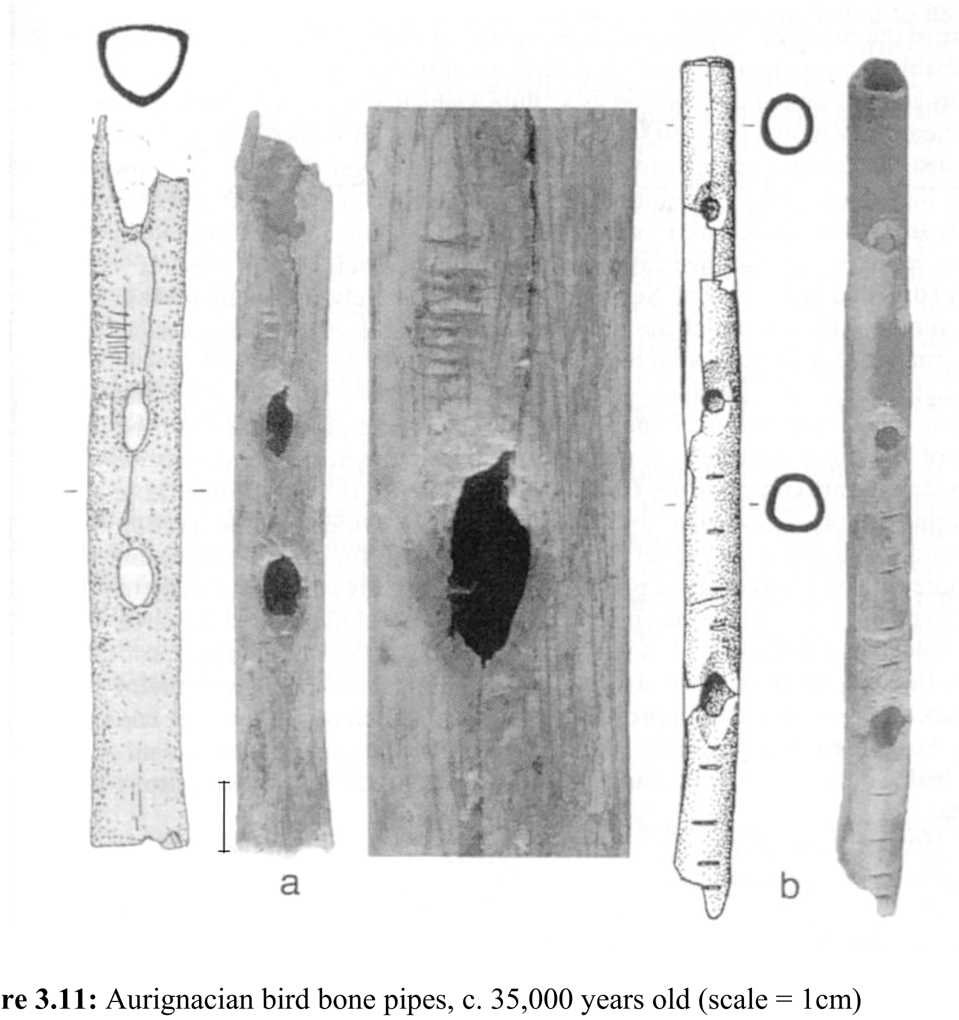

Figure 3.11: Aurignacian bird bone pipes, c. 35,000 years old (scale = 1cm)

a) From Isturitz, France, vulture ulna
b) From Geissenklösterle, Germany, swan ulna

(*from d'Errico* et al.*, 2003*)

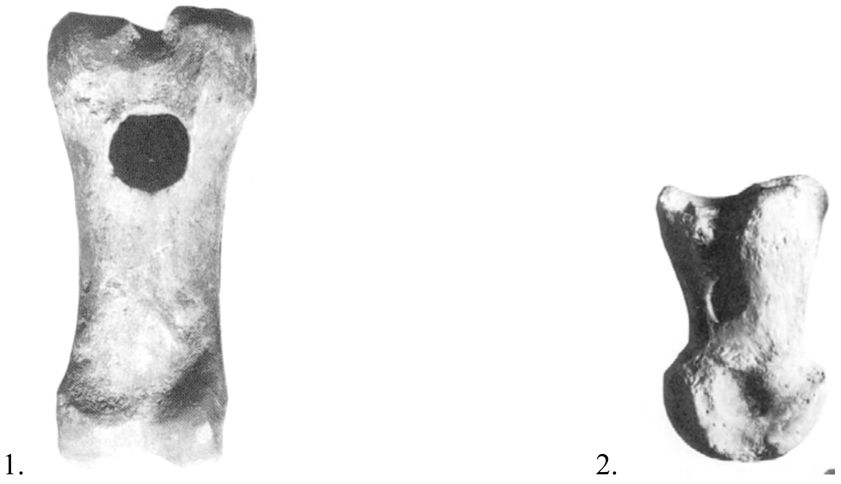

Figure 3.12:

1. Perigordian pierced 1st phalanx of reindeer, Laugerie Haute, Les Eyzies-de-Tayac, Dordogne (France). Length 53.2mm (*from Dauvois, 1989, p.9*).
2. Magdalenian pierced 2nd phalanx of reindeer, Saint-Jean-de-Verges, Ariège (France). Length 38.6mm (*from Dauvois, 1989, p.6*).

consistent-shaped holes are produced under different conditions, dependent only upon the age of the bone.

Harrison (1978) found that fresh bone punctures consistently, resulting in a hole with a smooth curved outline, the only irregular section being where the puncture abuts the superior epiphysis. As much as seven-eighths of the circumference of the hole may be smooth and round (p. 10). Such damage has the appearance of a deliberately created hole, and punctured phalanges created naturally in this way also tend to function well as whistles.

An old bone under these conditions tends to puncture in a more fragmentary way, producing a rough-edged hole more easily recognisable as natural damage. However, there is conversely the possibility that genuine anthropogenic phalangeal whistles would suffer the same type of damage upon deposition if they were old when deposited.

The implications of Harrison's (1978) work are several. In terms of determining the authenticity of punctured phalanges as artefacts, there are three main problems:

1) Reindeer phalanges puncture easily and consistently at the proximal end of the distal surface, and naturally-produced punctures in fresh bone look intentionally-produced
2) Intentionally-produced punctures may be damaged in deposition so as to be unrecognisable as such

3) Both naturally-produced and intentionally-produced punctured phalanges tend to be capable of functioning well as whistles. Thus, functionality cannot be taken as an indicator of anthropogenesis.

There are some important indicators as to authenticity that can be applied, however:

1) Phalanges punctured in deposition will tend to include pieces of the bone wall inside the cavity; an intentionally-produced whistle would have had these removed as they would impede its function.
2) Intentionally-produced punctures often have smooth edges where they have been smoothed by the rotation of the punching tool or burin, and are often slightly bevelled as a consequence.
3) Any smooth hole that is *not* located at the proximal end of the distal surface is likely to be intentionally produced, as the bone diaphysis does not puncture easily anywhere else.

It is on the basis of these criteria that those examples subjected to close scrutiny, detailed in Table 2, have been judged, in particular by Harrison (1978), Scothern (1992) and to a lesser extent by Dauvois (1989). Note that not all of the items covered by Table 2 have been subject to this level of scrutiny and, as with the pipes/flutes, several have not been re-examined closely since their original discovery and/or publication.

Whilst the status of many individual reputed phalangeal whistles must be subject to question, their existence and use by Palaeolithic humans is not at issue; there are numerous examples of which their anthropic origin is well-established. Reindeer phalanges were a very readily available raw material throughout much of the Pleistocene, easy to work, and would have been plentiful in living environments as reindeer formed a major subsistence staple in both the Middle- and Upper Palaeolithic. For what they were used is open to greater question.

Whilst many of these objects can be confidently classified as intentionally-produced sound-producers, they should not necessarily be classed as musical instruments as there are many potential uses for a small whistle. In his analysis Dauvois (1989) focused more on the acoustic properties of various French examples of phalangeal whistles, than on their provenance. He noted that the whistle-language of the Pyrenéan Vallée d'Aas shepherds uses frequencies between 2000 and 2650 Hz – sounds within these frequencies travel well over long distances, and the human ear is most sensitive to sounds between 1500 and 4000 Hz. It is especially sensitive to sounds around 3800Hz as this is the frequency range at which the *auditory meatus* (ear canal) resonates (Campbell and Greated, 1987), and at which the ossicles (middle-ear bones, *malleus*, *incus*, and *stapes*) operate most effectively. Dauvois analysed the frequencies produced by various examples of phalangeal whistles to see how they compared with this range. Of the 35 examples he tested, 21 produced tones within the range of the shepherds' whistle-language, and all 35 fell within the 1500-4000Hz range of sensitive human hearing. Harrison (1978) also carried out tests of the phalanges' sound-producing capabilities as part of his analysis. Some of the examples he tested (e.g. one from La Madaleine, W044 Ma in Table 2) were audible with 100% reliability over distances up to 1.25km, and clearly have the potential to be used as communicating devices between individuals separated in the environment. Those with small holes proved poor at making sounds that carried over long distances (Harrison, 1978), although often small holes have been intentionally produced. An explanation other than long-distance communication or hunting co-ordination must be sought for such artefacts.

3.5. The use of bone for instrument manufacture in the Middle- and Upper Palaeolithic

As noted in the preceding sections, the material most commonly used for the production of the reputed bullroarers, rasps and phalangeal whistles is reindeer, either antler or bone. The bone of which the Palaeolithic "flutes" consist varies. Of those instances in which the faunal origin of the bone is specified (and in fifty of the approximately 122 reputed pipes or flutes it is not), the majority are from birds, although there are a few examples of other animals such as bear, hare, elk horn (Megaw, 1960), chamois (Absolon, 1936) and pig. When the fauna used for the reputed sound producers in Table 1 is examined, it becomes clear what a high proportion of those of which the faunal provenance is known are made on bird bone. It is also conspicuous the proportions of these that are considered likely to be genuine sound-producers, as opposed to others made on bear or ungulates, for example, of which many more are of doubtful anthropogenic origin or musical function. There is also a geographical split in the use of raw material for the manufacture of the flutes/pipes. The majority of the bird-bone artefacts come from France and western Europe; whereas those made from other materials such as ungulates, or more commonly bear bone, come from central and eastern Europe. This is clear from Table 1, and is also observed by Leocata (2001).

In most instances the birds' species is not specified, but of those that are, there are eagle (Piette, 1874, 1875, 1907; 077 Ma, 078 Ma), vulture (Passemard, 1923; 012 A), swan (Seewald, 1934; Bergounioux and Glory, 1952; Hahn and Münzel 1995; 005-006 A, 105 Ma, 116 UP) and greylag goose (Absolon, 1936; 060 G/P) (Scothern, 1992). Forty-eight

(40%) of the approximately 122 reputed pipes and flutes from the Upper Palaeolithic are of bird bone, and of these 18 are from Isturitz. Indeed, as described above, the earliest widely-accepted pipe-type sound producers found to date are made on swan ulna, the Geissenklösterle examples.

It can be hypothesised that cultural and cognitive differences explain the apparent absence of such artefacts amongst Neanderthal populations, in which the use of bone as a raw material is relatively rare and the use of bird bone unknown, but why do we see musical instruments associated with anatomically modern humans for the first time in Europe, rather than elsewhere? Is the availability of suitable faunal resources the determining factor in the incidence of instruments in the Upper Palaeolithic - is it that avian faunal resources were available in Europe that were not available elsewhere? Is it representative of a cultural or creative revolution when modern humans reached Europe? Or is it the result of differential preservation elsewhere?

3.5.1. Raw material availability?

Homo neanderthalensis seems not to have trapped or hunted birds; the technologies and behaviours, or environmental demands, necessary to do so appear to have come to Europe with modern humans. Stiner *et al.* (2000) suggest that the use of avian fauna (and small game) was an innovation especially undertaken by modern humans in times of limited large game resources and population/resource pressure. This is echoed by observations, both archaeological and ethnographic, by Gotfredson (1997), Lefèvre (1997) and Serjeantson (1997) that the importance of birds as a subsistence resource depends upon the availability of other resources, and that this is dependent largely upon season and latitude. Consequently, avian fauna seem to be most important to subsistence at high latitudes, under arctic and sub-arctic conditions, although they can also form a major part of the diet in coastal and lakeside locations where large numbers of seabirds accumulate. Stiner *et al.* (2000) hypothesise that factors of environmental stress and human demand for resources resulted in increased use of relatively small avian food resources, such as partridge and quail, at the start of the Upper Palaeolithic; they cite major use of these avifauna as a subsistence resource in early Upper Palaeolithic in Italy and Turkey.

However, there is also evidence of large avian fauna being used as a main subsistence resource themselves. Also in Italy, Grotta Romanelli shows that at the end of the Upper Palaeolithic (Epigravettian), 40 species of birds were being used to supplement the main diet of large mammals, but five large species were represented in numbers great enough to indicate their forming a subsistence diet in their own right. These were white-fronted goose (*Anser albifrons*), bean goose (*Anser fabalis*), brent goose (*Branta bernicla*), great bustard (*Otis tarda*) and little bustard (*Otis tetrax*). In addition, swans, ducks, eagles, hawks, owls and ravens also all displayed cut-marks on the bones, and in some cases, decoration (Cassoli and Tagliacozzo, 1997).

Such use of large avifauna is not restricted to northern latitudes. Whooper swan, greylag goose and white-tailed eagle (all of which have been used for creating sound-producers during the Upper Palaeolithic, see Table 1, Appendix) have been found in Ohalo II in Jordan at the last glacial maximum (the site is dated 19,000-23,000 b.p.) (Simmons and Nadel, 1998), indicating that these large avian species were readily available outside continental Europe under these environmental conditions. The use of large avian species for subsistence results, as a by-product, in the ready availability of large, relatively easily worked, hollow bones. In addition, species may be selected specifically for their potential as a source of raw material for tools, and such activity is documented in later archaeological and ethnographic contexts. For example, Van Wijngaarden-Bakker (1997) concludes from the study of Dutch Neolithic sites that large avian species were actively selectively hunted primarily as a raw material, rather than subsistence, resource; at Aartswoud for example, the five bone tools retrieved were made from white-tailed sea eagle and crane, neither of which species are found in the food remains of that site.

Cassoli and Tagliacozzo (1997) point out that during moulting, ducks, geese and swans cannot fly, so are relatively easy to approach and kill “using simple sticks or clubs” (p. 308). These birds moult and become flightless during the summer months, from May in the case of geese, through July and August in the case of swans and ducks, and would at this time provide a relatively accessible and substantial resource for food and bone. It is worth noting here, perhaps, that at none of these sites described in the preceding paragraphs have bird bone flutes or pipes been found, despite the presence of suitable raw material, suggesting that even when such materials were readily available, their use for the manufacture of sound-producers was rare (as is indeed the case amongst many hunting and gathering societies of recent history, as discussed in Chapter 2).

So it would seem that the use of avian bone as a raw material is unlikely to be dependent upon differential availability outside of continental Europe in times of glaciation at least, although greater subsistence pressures at more northern latitudes may increase the demand to make use of such resources. It seems also that the procurement of substantial food (and tool) resources from large birds need not be dependent upon sophisticated tool use (such as projectiles for example), especially during the summer months when geese, swans and ducks are moulting; clubs and nets would be quite adequate.

This leaves the possibilities of a behavioural change in terms of choice of avian resources for subsistence, or issues of archaeological preservation and recovery to explain the change in the record of the use of avian bone amongst anatomically modern humans. There is the possibility that the apparent increase in the use of bone in the archaeological record of modern humans between the Middle Palaeolithic/MSA and the Upper Palaeolithic is a consequence of differential recovery of such materials. Upper Palaeolithic Europe has been subject to such a level of intensity of excavation in comparison to areas outside of Europe, and generally comparatively favourable conditions of preservation, that it seems distinctly possible that future excavations outside Europe pre-dating the Upper Palaeolithic will result in the discovery of avian faunal use for subsistence and tools.

3.5.2. Cultural revolution?

From the earliest occupation of Europe by *Homo sapiens*, there is evidence of the use of bone for the creation of tools and other objects (amongst the earliest being the Geissenklösterle flutes; Richter *et al.*, 2000), suggesting that the use of this raw material was a technology brought with these earliest modern humans to Europe, and not something "discovered" upon arrival. The recent finds from Blombos cave in South Africa confirm that the use of bone as a raw material has a long time-depth in modern humans; these examples of spear points and awls date to around 70,000 years ago, and are made on fragments of ungulate long bones and mandible (Henshilwood *et al.*, 2001; d'Errico *et al.*, 2001). Whilst this technology may have been lost and regained many times during the timespan of *Homo sapiens*, its early incidence indicates that the use of bone to make other items is not dependent upon a single cognitive revolution amongst modern humans. This evidence does not undermine the possibility that the choice to use bone, and in particular, avian bone, to produce *musical instruments* was something to arise in Europe, but if so, this was an innovation virtually contemporaneous with human arrival in Europe.

Traditionally, if Palaeolithic music archaeology is mentioned at all it tends to be in the context of the emergence of other art and culture. Certainly there is a revolution and an explosion of creative activity in the archaeological *record* of Europe at the time of the arrival of *Homo sapiens*, and is clearly in contrast to the record of *Homo neanderthalensis*, but this does not necessarily mean that it was at this time that there was a revolution or explosion of creativity in the modern human *brain*; it may merely indicate the arrival in Europe of such behaviours with modern humans, not the time of their arising in modern humans.

Alternatively, it could indicate the spread and popularisation of such behaviours, and this need not be indicative a change in the cognitive capacity for the behaviours; there are many people in the world today with the cognitive *capacity* to use or program a computer, for example, but who will never do so, along with the rest of their contemporaries in their culture. Likewise, few would argue that a citizen of the Roman empire lacked the cognitive capacity to drive a motor car. As amply discussed by McBrearty and Brookes (2000), an increasing amount of evidence such as that from Blombos suggests that the idea of the sudden emergence in *Homo sapiens* of an *entire suite* of artistic and symbolic capabilities upon their arrival in Europe is unlikely to be accurate, and that at least the foundations for those behaviours were being exhibited by *Homo sapiens* considerably prior to their arrival in Europe. The capacity can be there without it being used to fulfil that purpose. So if there is a revolution, it can be a cultural one without being a cognitive one.

Such models of a creative explosion had hitherto frequently led to the equation of musical behaviours with other artistic behaviours, supposedly being part of a package of creative and artistic capabilities and sensibilities that emerged simultaneously in modern humans. Further discussion of the validity, or otherwise, of such an equation are discussed in Section IV of this thesis, regarding evolutionary rationales for the development of musical capacities. The evidence considered in Section III illustrates that aspects of the physiological and neurological foundations of musical behaviours were in place considerably before the advent of anatomically modern *Homo sapiens*, and it is very likely that all such foundations were in place long before the European Upper Palaeolithic.

A revolution we may be seeing, however, is in the bone resources used. The Upper Palaeolithic appears to show a great increase in the working of bone as a raw material for the production of other items, such as containers, needle-holders and needles, pipes for blowing paint, as well as sound-producers. Scothern (1992, p. 84) suggests that the first true sound-producers occur at this time as a consequence of a marked increase in avian fauna during the Aurignacian. As noted, bird-bones, especially long wing bones, are often hollow, and she suggests that the action of blowing down them to clear any debris may have led to the first realisation of their potential as whistles. Perforating them would, she says, have been a logical progression.

Whilst indicating an apparent increase in the *use* of avian faunal resources, and whilst such resources provide an easily-worked material for the manufacture of sound-producers, caution should be applied before concluding that these artefacts necessarily represent the genesis of anthropogenic sound-producers. It has been noted above that the incidence of the use of bone in the manufacture of sound-producers in contemporary hunter-gatherer societies is rare, and that the vast majority of such objects are made using other materials. Furthermore, pipe-type instruments form a small proportion of those used. There is no reason to conclude that the pipe-type sound-producers on bone that we see in the Upper-Palaeolithic of Europe represent the first musical instruments, and even less so, the first incidence of musical behaviour. What seems more likely is that the record we have is representative of a change in the production in instruments, either in the form of the raw materials chosen for the purpose, the capability to work bone materials or to catch certain fauna whose bone could then be used, or by a change in musical behaviour that resulted in an increased dominance of melodic instruments (relatively rare in modern hunter-gatherer groups, as illustrated in Chapter 2).

What they do frequently show, however, even from the earliest examples, is an already well-developed and sophisticated mode of production, superior to many mediaeval examples. What they are most valuable for, in contrast to identifying the earliest incidence of musical behaviour, is to tell us about the manifestation of apparently well-developed musical behaviours in anatomically modern humans in the Upper Palaeolithic.

3.6. Percussion

Ethnographic evidence shows that among modern hunter-gatherer groups instrumentation is dominated by percussive rather than melodic instruments, melody instead being provided by vocalisation (discussed in detail in Chapter 2). If this were also the case amongst early humans we would be left with very little evidence of it, as these instruments are usually made from wood or skin, which would leave no archaeological trace. Furthermore, percussive sounds are often made by striking body-parts, which requires no instrumentation at all.

Some possible evidence for the use of percussive instruments by prehistoric humans does, however, exist. A set of six mammoth bones from Mezin, Ukraine, dated to 20,000 years ago, appear to have been deliberately and repeatedly struck, and were found in context with two beaters and a variety of ivory “rattles” (Bibikov, 1978). Also found in the same settlement were piles of red and yellow ochre, and other mammoth bones which had been incised and painted. The “orchestra” bones (including a shoulder blade, thigh bone, jaw bones and skull fragments) produce a selection of tones when struck with beaters (Bibikov, 1978).

While Bibikov’s interpretation of these objects as a collection of musical instruments was seconded by a selection of Soviet archaeologists, forensic scientists and medical experts, some other archaeologists and musicologists have remained cautious or sceptical. While the artefacts do show wear from percussion, Scothern (1992) and Lawson *et al.* (1998) point out that this does not necessarily merit a musicological explanation, as many daily activities involve percussive actions.

The fact that these bones occurred in context with a reindeer antler mallet and beaters, as well as what appear to be rattles, counts in favour of the musicological interpretation, though. The criticism that the percussive traces on them could have been caused by a different activity is one that could be levelled just as easily, if not more so, at features of many of the artefacts readily accepted as flutes; in the Mezin bones case, however, the human (as opposed to carnivore) agency in the artefacts is not doubted. Until the original objects are available for analysis by other researchers outside the former Soviet Union, it is likely that debate will continue.

Albeit Neolithic rather than Palaeolithic (dating to around 9,500 years ago), the excavation of the city of Çatalhöyük in Turkey has produced some wall paintings representing what appear to be dancing and hunting activities, accompanied by horn-shaped, bow-shaped and drum-shaped objects (Stockman, 1986). The occupants of the city subsisted through a combination of hunting and agriculture for 700 years or more, from at least the middle of the seventh millennium B.C.. Stockman observes that the material artefacts and wall paintings from Çatalhöyük indicate that the activities of the hunter-peasants of the city were “deeply rooted in existing traditions, known already from Upper Palaeolithic and Mesolithic times” (Stockman, 1986, p.12), but that they also were quite creative, trying out new possibilities of tools and materials.

As regards music, the images so far excavated are dominated by what may be idiophones and percussion instruments, most of which, Stockman says, can be traced back to Palaeolithic times. These include figures holding what appear to be horns, bows (often two such items, one in each hand, which questions their use as hunting bows), beaters and sticks (Stockman, 1986). These paintings occur on the walls of a room known as “the hunting shrine” and the many figures depicted are almost always in poses of dynamic movement, surrounding a large representation of an animal; in one instance this is a bison, another shows a stag and a boar. As the paintings feature no perspective, it is unclear whether the animals are supposed to be standing or lying down on their sides. This allows for a couple of interpretations of the pictures: the people could be running, participating in an active hunt, or they may be dancing around a kill.

The latter interpretation, although initially perhaps seeming less intuitive, could be more likely for a couple of reasons. Firstly, as mentioned, several of the figures are carrying *two* bows or *two* sticks or beaters, which suggests that they were not being actively employed in hunting at the time represented in the image. Secondly, the animals depicted are not in dynamic poses of movement; in contrast to the humans, the positions of their legs are those of an animal that is still. This interpretation may be reinforced by the fact that both the bison and the stag seem to have their tongues hanging out, an image evocative of a dead animal.

If these interpretations are correct, of the objects as instruments and the activity as dancing, then this provides further evidence for the prehistoric use of music in association with celebratory activities, in particular related to hunting, as well as a prehistoric musical tradition dominated not by melodic instruments but by percussive ones. This is not only valuable in itself, but also goes some way to justifying analogy between the musical activities of modern hunter-gatherers and prehistoric peoples.

3.7. Caves and lithophones

The Venus of Laussel and the Trois-Frères bison-man are the only images encountered in the literature interpreted as potentially illustrating musical instruments, but Palaeolithic art has been connected with music and the use of sound in less direct ways.

Reznikoff and Dauvois (1988) carried out an extensive analysis of the acoustic properties of three caves featuring Palaeolithic painting, in the Ariège area of the French Pyrenees. They made vocalisations throughout the caves, identifying "points of resonance"; places that resonated to particular notes. They then compared the locations of these points with the locations of the cave paintings, which have been dated to around 20 Kya (Scarre, 1989). They found that there was remarkably consistent correlation between the locations of the paintings and the places of particularly great resonance. Most of the cave paintings were within a metre of a point of resonance, and most of the points of resonance were accompanied by a painting of some sort. Further, they conclude that the location of some of the paintings can *only* be explained by their relationship with a point of resonance, as they are often not easily visible or accessible. In fact, some of the locations are marked only with red dots, where there was no room for a full figure (Reznikoff and Dauvois, 1988, quoted in Scarre, 1989).

It would seem that the acoustic properties of the cave were at least as significant to the painters as the art itself, as the position of the art seems to have been dictated by the resonance. These paintings are of the same age as many of the Upper Palaeolithic flutes and whistles described in section 3.1.1, above, and it seems quite likely that the painting activities, and possibly whatever other activities were carried out in the caves, were accompanied by sound-producing of some kind. The authors suggest that the full potential of the caves' resonance would only be elicited by the great range of the human voice, rather than the somewhat limited ranges of the flutes and whistles. It is not possible to be certain of the nature of the sound-production of the peoples' responsible for the art in these caves, but what does seem to be true is that they were very aware of the acoustic potential of features of the world around them, and considered such properties to be highly significant.

Dauvois (1989) also studied the acoustical properties of features within the caves which bear the marks of being deliberately struck. Several of the stalactites and other natural calcareous depositions bear chips and impact marks where they have been deliberately struck. Dauvois reports that these features are capable of producing several distinct tones, depending on where they are struck, and that the damage marks correspond with these tone-producing locations (quoted in Lawson *et al.*, 1998). These "lithophones", like the resonant areas, also tend to correspond with locations of artwork.

The work of Reznikoff and Dauvois was not the first to observe and analyse correlations between Upper Palaeolithic art and the acoustical properties of spaces and features in caves. Five years earlier, Dams reported on similar features in the cave of Nerja, in Malaga, Spain (Dams, 1984), Escoural, in Portugal, and Roucador, Cougnac, Pech-Merle and Les Fieux in France (Dams, 1985). This work built upon observations by Abbé Glory in the 1960s (Glory, 1964, 1965; Glory *et al.*, 1965).

The cave of Nerja contains at least 19 paintings so far reported, and has excellent acoustics, such that it is now sometimes used as a venue for concerts and dance in the summer months. It also contains a feature known as "the organ", which consists of an area in an alcove where the calcareous deposits have formed into a fluted, concertina-like feature. The average height of these folds is 3-4 metres, and each fold is separated from the next by a gap of 8-15 cm. The total length of the feature is about 5 metres (Dams, 1984). In total there are over 200 folds, and they are nearly all decorated with abstract representations, as well as an ibex and a hind. Lines and dots in red and black are the most common decorations, some at ground level and others over three metres up. Most are not visible from a distance, many are impossible to see unless the observer puts themselves in awkward positions to peer in between the creases; aesthetic considerations do not seem to have been a priority in the creation of the icons. All the decorations use the same colours and are equally faded, so the whole feature was probably decorated over a relatively short period. These examples are not amongst those that have yet been subjected to direct dating (see Bahn and Vertut, 1997), but the style of the ibex and hind illustrated suggest a

Solutrean (c.22-17 Kya) age for the illustrations according to Dams (1984). Layers of calcareous deposition overlaying the illustrations confirm a very great time has elapsed since they were painted.

The folds are also worn along the edges and are chipped, apparently deliberately, at different heights. This damage is, like the paintings, covered by more recent layers of calcite deposits, which dates it to the same age as the artwork. Dams (and Glory, 1964) consider this damage to have been caused by deliberate striking of the folds with hard objects. Indeed, when they are struck with a stick or blunt flint they produce clear, harp-like tones (Dams, 1984). The deliberately damaged sections seem to have been broken with the aim of altering the tone produced, as they effectively alter the length of the ridge (Dams, 1984); the shorter the ridge, the higher the tone produced (Dams, 1985). Dams believes that this entire feature was used as a lithophone during the Upper Palaeolithic, and that the artwork on it relates directly to that function, be it abstract or as a type of mnemonic aid dictating striking-places on the rock. Similar features in other caves are also compared with the “organ” in Nerja (Dams, 1985). The cave at Roucador, France contains a very similar ridged lithophone feature, which has, like the Nerja one, been decorated with black dots and lines, and has had fragments broken off many of the ridges. If there were any doubt of ancient human agency in the damage, it is dispelled by the fact that in one instance, these broken fragments have been wedged back into the wall between two ridges, where they have since been cemented in with a new layer of calcite (Dams, 1985). The painted marks have likewise been covered by calcite layers since their creation, as have soot-marks on the rock, presumably from the torches held by the artist/performer or observers. These stalactites, like those of Nerja, are resonant, and produce a bell-like tone when struck. Roucador also contains an unpublished frieze of painted and engraved bison, horses, a mammoth, circular signs, stencilled hands and other symbols, all located six metres above the current floor level, and near to the lithophone (Dams, 1985).

Black and red dots and lines mark a similar feature of stalactite ridges at Cougnac cave (also famous for its Upper Palaeolithic art), and a free-standing stalagmite, all covered by layers of newer calcite deposits. These again give forth “pure and clear” vibrations, and are located near an engraving of the head of a bear. The Pech-Merle and Escoural caves feature lithophones too, marked with red and black lines and dots. These both also exhibit re-calcified fractures, and the ground around the former is littered with bone remains and chips from the stalagmite itself. These are, again, situated close to other art and representations of animals. The Pech-Merle example is next to engraved images of ibex and a possible mammoth, as well as cup-marks and other engravings (Dams, 1985).

The art in all of these caves has been dated to the Solutrean (Dams, 1985). It will be a valuable addition to the study of these features when they are added to the gradually growing catalogue of examples of cave art that have been directly dated by radiocarbon or AMS (Accelerator Mass Spectrometry). Whilst confirming most stylistically-based attributions of age, there have been notable exceptions, for example, Chauvet, which turned out to be 10,000 years older than expected when dated by AMS (Bahn and Vertut, 1997).

There are several things that can be asserted with some confidence from these studies. Firstly, these calcite formations have been covered with abstract decorations, which can only be viewed through close observation; the formations had specific significance for the people decorating them, but were not decorated for general viewing or aesthetic reasons. Secondly, they occur in areas established as places of considerable artistic, and possibly social, significance. Thirdly, they produce clear tones when struck with wood or stone. Fourthly, they have been chipped and worn, probably through being struck, and this probably occurred at the same time as the painting was carried out, during the Solutrean, around 20,000 years ago. Finally, they each appear to be part of a network of similar features in southern France, Spain and Portugal that were all treated and used in the same way.

Considering the level and nature of the attention these calcite formations were given by the Upper Palaeolithic users of the caves, it is far easier to believe that these features *were* used as lithophones, as Glory, Dams and Reznikoff and Dauvois suggest, than it is to believe that this acoustic quality of the features escaped the notice of these early people.

The idea that rocks and rock formations were used as sound-producing devices is not one without a considerable history, and an extensive global precedent; the lithophonic use of natural rock formations and large stones is also known from more recent prehistory and history. There are a number of assemblages of ringing stones in the Canary islands, which appear to have been used by the native inhabitants of the islands prior to the arrival of the Spanish, and possibly up until the last century (Álvarez and Siemens, 1988). These stones have obviously been struck, being chipped along the edges, and are situated variously in circles and lines in natural contexts which amplify their sound acoustically. Natural lithic features variously known as “Rock gongs” and “ringing stones” are also found in India (Boivin, *in press*), Bolivia, South-East Asia and Africa (Laura Basell, *personal communication*; Fagg, 1997, carries out an extensive global review). Further examples of “singing stones” traditionally believed to have magical properties in the production of sound are provided in Swedish historical records. These are often cup-marked Neolithic or Bronze Age stones which have continued to be known of as sound producers over many millennia, even to the present day (Henschen-Nymann 1988). Britain is not excluded from such traditions of sound-producing with rock; apart from the possibility of bronze- and iron-age cup-marked stones fulfilling this purpose, several examples of "geological pianos" were constructed in the nineteenth century from Cumbrian slate (Cross, n.d.).

Most recently, Ian Cross has begun work testing the acoustic properties of flint and other stones as potential intentional sound-producers (Cross, 1999a; Zubrow, Cross and Cowan, 2001; Cross, Zubrow and Cowan, 2002). Anybody who has carried out flint knapping will know that it requires a great awareness of the sound that the stone produces as it is struck. A skilled knapper has to be aware of the subtlest variations in the sound produced by the stone being struck, as this provides important clues as to where and how hard to strike the stone next (John Lord, personal communication). Cross proposes that the necessary awareness of the different sounds produced by stone during tool knapping may have led to its use for percussive or even melodic sound production at an early stage.

3.8. The nature and possible purposes of music in the Upper Palaeolithic

Despite the fact that a large amount of evidence has undoubtedly been lost to the archaeological record, it has been possible to ascertain some clues as to the nature of music in the Upper Palaeolithic. It is difficult to be certain of the human origin of any of the artefacts prior to around 40,000 years ago, but after that age, there are several examples of sound-producing objects, mostly bone, in which we can be reasonably confident of human agency. The contexts in which these objects occur may provide some clues as to the nature and use of music produced.

The cave site of Isturitz, which has produced many of the Upper Palaeolithic bone flutes, was an important focal point for large groups of people throughout the Upper Palaeolithic (Bahn, 1983), and seems to have been a focus for major gatherings in spring and autumn in particular. Similarities in the stone and bone artefacts with those from the Dordogne area, and the transferral of flint over distances over 100km (Gamble, 1983) suggest wide social contact (Scothern, 1992). Mas d'Azil shows similar evidence of large aggregations of people, as well as parallels with Isturitz, Le Placard and Marcamps in the bone flute design and engravings (Scothern, 1992). The flutes are commonly found inside the decorated caves of these sites, suggesting either that acoustics were particularly important, or their relation to the cave art, or both.

It is difficult to demonstrate whether the production of music in the caves was a group activity indulged by all or an activity performed only by a select few, an issue which also applies to the production of cave art. In her analysis of the Solutrean lithophones (described above) Dams (1984; 1985) notes that the lithophones occur in caves which generally also allow a number of people to congregate in nearby chambers, up to 20-30 people at Roucador, for example (Dams, 1985). It is worth noting, however, that the locations of the lithophones are dictated by natural cave features rather than human agency so their position in relation to other large chambers is not pre-meditated. What would be particularly valuable as a comparison would be to know of caves with the same calcified structures in them that *haven't* been decorated and used percussively. If such examples exist, it might suggest that the ones in the caves described by Dams were chosen selectively because they could accommodate many people.

The similarities between the features and decoration of the lithophones Dams describes suggests wide social contact between the people responsible, so the possibility exists that large groups of people congregated to experience (and perhaps participate in) the sound produced by them. Whether the activities involved many or few people, it is clear from the work of Glory (1964; 1965), Dams (1984; 1985) and Reznikoff and Dauvois (1988) that the acoustics of the caves were highly significant, and that sound production bore an important relationship with both abstract and representational art.

The representational art occurring in these caves near to the sound-producing areas depicts fauna which were the main subsistence sources for the artists. This may be evidence of a link between sound-production and hunting, as certainly seems to be the case in Çatalhöyük (Stockman, 1986). It is impossible to say with any certainty that music and the *subject matter* of the art were linked, however, only that the *location* of the art appears to be related to the sound producing ability of the caves.

A clue to the issue of the group versus solitary music use may come from the Mezin bones, reported by Bibikov (1978). The 20,000 year-old mammoth-bone "orchestra" was found in a large, open fronted, communal hut in the Mezin settlement, rather than a small dwelling. This does suggest, if it is indeed a collection of instruments, that the music was supposed to be heard by the community, and was not performed in solitary private rites. Again, it is difficult to draw any conclusions regarding the use of the music, other than as a social activity. It may have been significant to the mammoth-hunters that the instruments were made of mammoth bone; on the other hand, it may simply be that this was their most ready raw material for such artefacts. There is evidence for the trafficking of raw materials and visually-related goods between European populations and those of the Russian plains of the Magdalenian (Scothern, 1992), so it is quite possible that sound-production traditions were also shared. The possible actions of music on the human mind, and roles it may fulfil in social contexts are explored in Section IV of the thesis.

3.9. Chapter summary and conclusions

There is no indisputable evidence for musical instrumentation before the Upper Palaeolithic in Europe around 40,000 years ago. Some of the evidence, particularly earlier, has been rather too readily interpreted as intentional sound-producers, being instead the product of carnivore activity or natural forces. After 30,000 years ago in particular there is a sudden diversity, with bone flutes, rasps and percussion and the acoustic use of rocks and caves, particularly during the eras associated with the Solutrean and Magdalenian techno-complexes. These seem to be widespread, often occurring in what appear to be focal locations for large amounts of human activity, including art. Such evidence as there is suggests that musical performance, or at least appreciation, was a group activity, rather than one involving a select few. The vast majority of past authors on the subject have taken this apparently sudden advent of the use of sound-producers to indicate the beginning of the use of music. While some earlier bone objects have been interpreted as instruments with a lack of caution in the past, it is worth treating this conclusion with caution too, for a number of reasons. Firstly, the use of instrumentation and the use of music are often considered synonymous in the literature, an assumption which is unjustified, as the ethnographic evidence considered in Chapter 2 illustrated; there are many types of music which do not rely on instrumentation, and many instruments which would leave no archaeological trace.

Secondly, the first evidence that we see of instruments in Europe (the Geissenklosterle and earliest Isturitz pipes) is highly sophisticated in design. As d'Errico *et al.* (2003) put it,

> "The sophistication of the pipes' various design elements (technological and ergonomic), both at Isturitz and Geissenklösterle, suggests that such instruments must, even at around 35,000 years, be several conceptual stages removed from the earliest origins, even of instrumental musical expression, to say nothing of those universal vocal, manual-percussive and dance forms which must have existed independently of – and before – any need for such tools" (d'Errico, Henshilwood, Lawson *et al.,* 2003, p.46).

Thirdly, it is possible that the determining factor in the increase in the production of bone flutes was not that flutes were not used before that point, but that the lithic technologies were inadequate for working *bone* finely, or that suitable bone resources were not being used. Wood and cane may have been used long before, both being easier to work than bone. Alternatively, the advent of bone flute manufacture may be as a consequence of the increase in the use of avian fauna for subsistence during the Aurignacian, providing the raw material necessary for them.

Fourthly, all the evidence so far has been accrued from Europe. There is increasing evidence (e.g. Mellars, 2000; McBrearty and Brookes 2000; Henshilwood *et al.*, 2001; d'Errico *et al.*, 2001; d'Errico *et al.*, 2003) that many of the cultural behaviours such as bone-working and art hitherto associated with the "Upper Palaeolithic revolution" were being evidenced in Africa during the Middle Palaeolithic by anatomically modern humans. Vast areas of Africa still require archaeological exploration, and continued excavation of African evidence may well show that musical instrumentation was also used there prior to the European Upper Palaeolithic.

There is, finally, the issue of symbolic thought capacity. The "Upper Palaeolithic Revolution" has traditionally been seen as the dawning of fully formed symbolic thought (e.g. Mellars and Stringer, 1989), and it has been thought that music could not have been performed prior to this. The view that music need be symbolic is rather anthropocentric, however, and leaves at least 70,000 years of anatomically modern humans in Africa, and perhaps in Europe, supposedly without any symbolic thought. The issue of the requirements for symbolism (or lack of) for musical behaviours is touched upon in the ethnographic evidence considered in the preceding Section (Chapters 1 and 2), and is explored further in Section IV (Chapters 7 and 8), regarding evolutionary rationales for the development of musical capacities.

SECTION III

EVOLUTION OF THE PHYSIOLOGICAL AND NEUROLOGICAL CAPACITIES FOR MUSIC

Previous neuropsychological and developmental studies have suggested that human musical abilities have a deep evolutionary history (e.g. Wallin 1991; Fernald, 1992; Papousek, 1996; Carter 1998; Dissanayake, 2000) but this contrasts with the evidence from the archaeological record, as discussed in Section II, in which musical instruments and evidence for the possible use of acoustics do not appear until 35-40,000 years ago. This is more than 80,000 years after the first anatomically modern humans had appeared. This section aims to address this apparent disparity between the archaeological record and biological evidence by looking for evidence in the fossil record of the evolution of the physiology and neurology required for musical behaviour.

When attempting to extrapolate the nature of any musical ability prior to the presence in the archaeological record of manufactured musical instruments, one must focus on the emergence vocal and percussive methods of music production. What it may be possible to determine is at what point the various mechanisms fundamental to music perception and production began to emerge; i.e. at what points hominids developed the physiological and neurological *capacities* to produce and process melody and rhythm. To have the capacity is not necessarily to make use of it in that particular way (an issue dealt with in Section IV), but it should allow a time to be identified before which musical behaviours could not have been carried out. It should also allow the identification of the relationships between the various capacities that music relies upon and how and why they might have evolved. This section consists of two chapters. The first (Chapter 4) consists of an analysis of our knowledge from the fossil evidence of hominids regarding the evolution of the physiological features required for vocal and rhythmic production, and an examination of the physiology used in the production and processing of sound. The second chapter (Chapter 5) investigates the neurology dedicated to the production and perception of music, speech and language, and reviews the evidence from neuroscience for distinct areas of the brain relating to musical ability and their relation to those controlling speech. These chapters lead logically into Section IV of the thesis (Chapters 6, 7, and 8), which examines, and proposes, evolutionary rationales for the origins and evolution of the physiological and neurological features investigated in this section, as well as emotional response to music and its relation to the development of linguistic ability. Together, these analyses should ultimately allow conclusions to be drawn regarding when musical capabilities could have begun to emerge, and how such capabilities would have developed over time.

CHAPTER 4

The Origins and Evolution of the Physiology of Musical Production and Perception

This chapter examines the evidence for the evolution and development of the various human physiological features responsible for the production and perception of vocalisations and rhythms. As noted in the introduction to this section, before the presence of musical instruments in the archaeological record, the study of the origins of the capacities for musical behaviours must recourse to the features of physiology and neurology that are used in such activities. Vocalisation is a fundamental component of musical behaviour, and the production and perception of long sequences of tonally-varied sound sequences would be a prerequisite to any behaviour that we would recognise as musical. The first section of this chapter examines the evidence for the evolution of a variety of features of the vocal tract with a view to mapping the evolution of the capacity to produce extended tonally-varied vocalisations. This is followed by a brief discussion of the development of the physiology used for rhythmic production (more is said of the evolution of rhythmic capacities in Chapters 5 and 6). The third major section of this chapter is concerned with the evolution of the inner ear and sound-perception capabilities, and how these relate to sound-production capabilities and the evolution of other physiological features. Ultimately, it should be possible to conclude which of our hominid ancestors would have had the physiological capacities in place for musical activities, and the path by which such features evolved, and how these capacities might interrelate.

4.1. Production: the vocal apparatus and its evolution

The evolution of the physiology necessary for vocalisations has received much attention in the last thirty years, in particular since interest in language evolution has blossomed (e.g. Pinker, 1994; Deacon, 1997; Jablonski and Aiello, 1998). A consequence of language evolution being the driving force behind most of this research has been that the results have generally been presented in terms of the development of speech capabilities in hominids. Dealing as it does with the fundamental physiological features required for *controlled* and *varied* vocalisations, however, the research is equally valuable as regards hominid capabilities to produce *any* vocalisations, linguistic, melodic or otherwise. Note that vocalisations are *not* equivalent to verbalisations, despite the fact that these terms have been used interchangeably in some literature; the ability to vocalise has often been spuriously equated with linguistic speech ability with little consideration of other roles of vocalisation. Recently, some attention has been given to these features with specific reference to music as well (e.g. Frayer and Nicolay, 2000, who consider some of the evidence discussed below, albeit with greater brevity).

Vocalisations are produced in two stages, in the supralaryngeal tract and the oral cavity (see Figure 4.1). The acoustic energy of vocal sound is generated by the larynx, which modulates the airflow in the throat (supralaryngeal tract). This sound is then adjusted by the action of the tongue, lips, teeth and palate in the oral cavity (Lieberman, 1992). Unfortunately, the vast majority of the physiology responsible for vocal production (for example, the tongue, larynx and supralaryngeal tract) consists of cartilaginous and soft tissue that is not preserved in the fossil record. Consequently, when considering the evolution of vocal ability in our ancestors, it is necessary to draw conclusions about the nature of these features from the bony structures which do survive such as the jaw (mandibles), bottom of the skull (basicranium) and, in some rare and particularly valuable instances, the hyoid bone. These can provide some clue to the nature of the original oral and laryngeal features with which they were associated.

4.1.1. The larynx and basicranial flexion

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

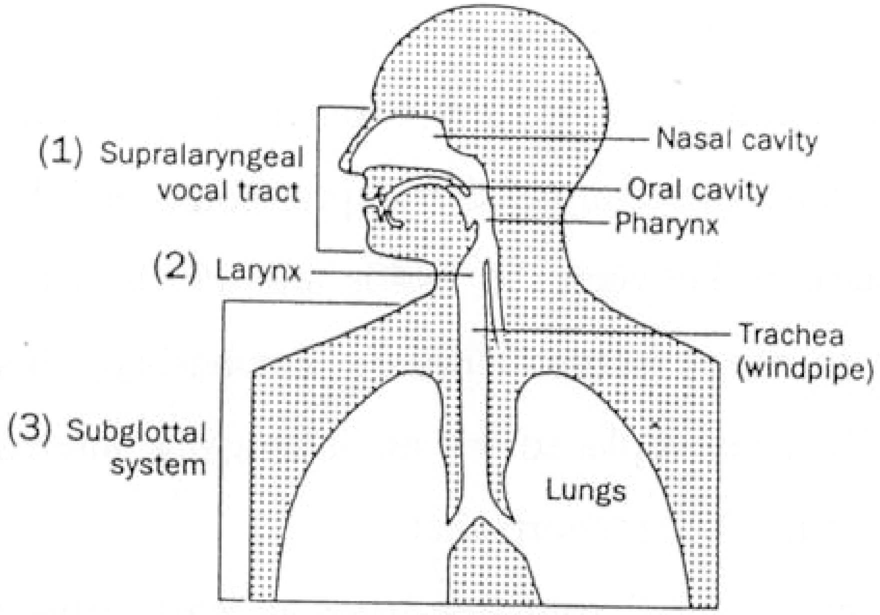

Figure 4.1: The components of human vocalisation physiology (*from Lieberman, 1992, p. 134*)

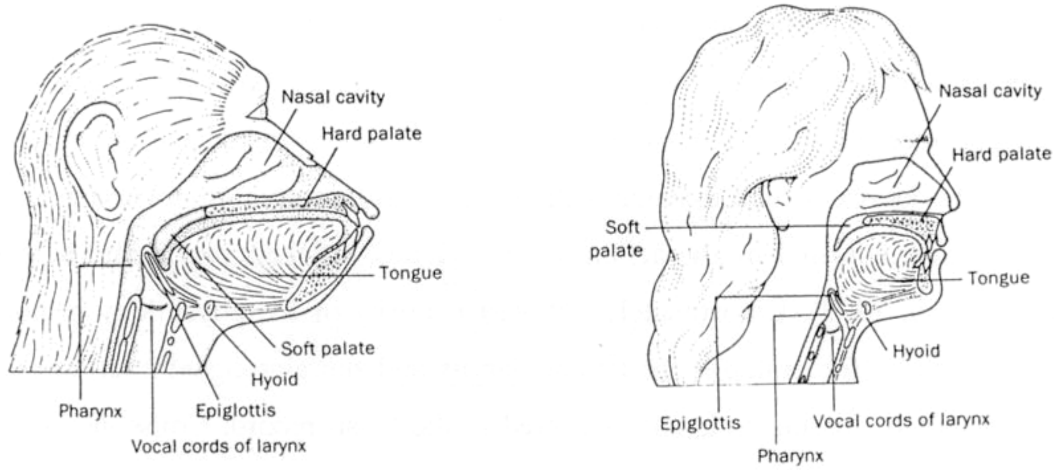

Figure 4.2: Chimpanzee and adult human speech anatomy, showing positions of larynx, pharynx, tongue and hyoid bone. (*from Lieberman, 1992, p. 135*)

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

At birth, the larynx occupies the top of the throat in humans, as in other mammals, with the same important benefit with regard to feeding and breathing. However, as the child grows up, particularly after the first year of life, the larynx descends towards its adult position. This process is only completed during adolescence (Magriples and Laitman, 1987). It seems that a similar process has occurred over the course of human evolution. Detailed analysis of chimpanzee vocal anatomy (Laitman and Heimbuch, 1982; Duchin, 1990) and other non-human mammals including baboons (Laitman and Reidenberg, 1988) shows that non-human primates have a high larynx; in contrast, humans, and only humans (Duchin, 1990, p.688), have a larynx low in the throat (see Figure 4.2, previous page). Thus, if it is possible to reconstruct the position of the larynx in fossil hominids it should be possible to draw conclusions as to the extent of their vocal abilities, in comparison with modern humans and non-human primates.

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

Laitman and colleagues carried out examinations of Plio-Pleistocene specimens such as KNM-ER 406 (*Paranthropus boisiei*) and Sterkfontein V (*Australopithecus africanus*), both approx. 3 Million years ago, or Mya (Laitman and Heimbuch, 1982), and found that these specimens show no flexing of the basicranium, as is the case with modern apes. The first evidence for the arching of the basicranium is displayed by *Homo ergaster* skull KNM-ER 3733 (from Koobi Fora, Kenya), approximately 1.75 Mya (Laitman, Heimbuch and Crelin, 1979). A fully flexed, modern-like, basicranium is first seen in archaic *Homo sapiens* (specifically, *Homo heidelbergensis*), 300-400 thousand years ago (or Kya), in specimens such as Steinheim, Broken Hill 1 and Petralona (Laitman, 1984).

Laitman suggests that *Homo ergaster* would have been able to produce the vowel sounds found in “boot”, “father” and “feet”, whereas archaic *Homo sapiens* would have been able to produce a full modern repertoire of vowel sounds.

Variations in methods and quality of reconstruction of crania can result in problems with such research, however. Researchers must use the utmost discretion in their choice of sample, as is illustrated by the initial work carried out by Lieberman and Crelin (1971), who created computer simulations of the vocal tract based on the anatomy of the La Chapelle-aux-Saints Neanderthal specimen. This led them to attribute only limited vocal ability to Neanderthals, not comparable to that of modern humans. Although ground-breaking at the time, this study has since been severely criticised on a number of fronts. First and foremost, the La Chapelle-aux-Saints cranium is that of an aged and crippled Neanderthal, who was also suffering from the bone-altering disease osteoarthritis, and so is not a good approximation for the species as a whole. Secondly, Lieberman and Crelin’s simulation was based partly on the original Boule reconstructions of the skull from 1911-12; Heim’s (1985) reconstruction showed that it in fact had a basicranium very similar to that of modern humans (Heim, 1985, cited in Scothern, 1992). Despite the criticisms levelled at this particular reconstruction, it does seem that Neanderthal basicrania are, on average, longer and flatter than those of modern humans and most *Homo heidelbergensis* (Laitman *et al.*, 1979), and this is often still used as a basis for suggesting incomplete speech production amongst Neanderthals.

Lieberman has since modified the severity of his conclusions in the light of the above criticisms of the 1971 simulations (Lieberman, 1989), shifting his emphasis towards the connections between physiology and neurology. He still maintains that “classic Neanderthal hominids appear to be deficient with respect to their linguistic and cognitive ability… they probably communicated slowly at extremely low rates and were unable to comprehend complex sentences” (Lieberman, 1989, p.391). He makes this assertion on the basis that there are important neural connections between vocal physiology and the cognitive substrates responsible for language production and perception. He hypothesises that since the physiology appears to be underdeveloped in Neanderthals, so too was the neurology.

There are still several major problems with these assertions, however. Kuhl (1988) concludes from her studies of the perceptual mechanisms and abilities of non-human mammals and pre-linguistic infants (see below, section 4.3.) that the mechanisms relied upon to process speech sounds were in place long before humans began to use speech. She concludes instead, that speech was shaped by our perceptual abilities. This means that

Neanderthals could have had the neurological mechanisms necessary to perceive speech sounds even if they *were* unable to produce those sounds themselves.

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

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

Other supralaryngeal reconstructions have been attempted. Budil (1994) compared the basicrania of 109 modern adult humans (from a total of 17 disparate ethnic groups) with those of 103 non-human primates (from 18 genera), 28 fossil hominids and the measurements from a further 16 fossil hominids as quoted in Laitman's publications. He used the evidence from all of the modern specimens (human and primate) to devise a formula relating basicranium, styloid process, hyoid, and laryngeal position. Based on this, he reconstructed the supralaryngeal tract of the Petralona specimen, a *Homo heidelbergensis* approximately 200,000 years old. His results indicate that a modern-like supralaryngeal resonance cavity was present in the Petralona specimen, and others with similarly flexed basicrania, including other *Homo heidelbergensis* such as Broken Hill and Steinheim, and *Homo sapiens* such as Taforalt 12, Cro-Magnon 1, Predmosti 3 and 4, Afalou 5 and Skhul 5. His conclusions thus agree with those of the earlier work by Laitman and Reidenberg (1988) and Arensberg *et al.* (1990, see below), that an upper respiratory tract similar to that of modern humans (and thus similar vocal ability) first appeared 300-400 Kya (thousand years ago), in *Homo heidelbergensis*, and that the laryngeal physiology of the Australopithecines resembled that of modern apes. The results for Neanderthal specimens again fall outside the range of any modern humans, despite post-dating the other specimens by many millennia, agreeing with Lieberman's analysis of the physiology. To Budil, this finding "remains enigmatic". On the basis of flexing of the basicranium, at least, the precise nature of the linguistic and vocal ability of *Homo neanderthalensis* remains unresolved, but it is also clear that there are several influences at play in determining cranial base morphology.

Arensberg *et al.* (1990) also cite evidence of craniofacial reconstructions of hominids (Solow, 1966) that basicranial flexion is negatively correlated with maxillary width and posterior height. This has the consequence of overall nasopharyngeal volume apparently remaining consistent, irrespective of degree of flexion. Thus, although the flexing indicates a change in the position of the larynx and other elements of the vocal apparatus, the nasopharyngeal soundspace may not be much affected by the process.

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

The selective influence on the change in the flexing of the basicranium seems to have been that it was a way of accommodating increasingly large brain size along with a relatively short cranial base (Spoor and Zonneveld, 1998); in extant primates the degree of basicranial flexion is positively correlated with brain size relative to basicranial length (Ross and Ravosa, 1993; Spoor and Zonneveld, 1998). As ever with physiological evolution, there are numerous factors at play. Whilst brain size increased in absolute terms, the cranial base was shortened, both relatively and absolutely, by the reduction of kyphosis of the spine with bipedalism (the shift to an upright spine) and the consequential midsaggital location of the foramen magnum (the hole in the base of the skull where the spinal cord enters), as well as the reduction in facial prognathism. All of these processes would have had effects upon the positions of the vocal tract and larynx relative to the skull, and more recent studies have begun to take these factors into account, with several other writers (e.g. Arensberg *et al.*, 1990; Kay, Cartmill and Balow, 1998) now suggesting that the basicranium is only one of many complementary features which indicate vocalisation quality and enunciation ability. Likewise, supralaryngeal space is only one of several physiological factors implicated in sound production that would be affected by laryngeal lowering. This does imply that a dismissal of Neanderthal vocal sophistication solely on the basis of basicranial morphology and supralaryngeal volume is premature; basicranial flexion is a good indicator of the level of development of vocal tract morphology, and aspects of the anatomy, but it is not *the* determining factor in vocal ability. Other factors are discussed further below.

It is worth noting at this stage that the potential individual cost of the lowering of the larynx in the throat (namely, that the individual is increasingly subject to choking on liquid) could be a great one; for the progressive lowering of the larynx to have occurred over the course of hominid evolution as is discussed above, the selective advantage to be gained from this process must have outweighed this potential cost. However, Clegg and Aiello (2000), having looked at death certificates from 1894-1994, do point out that the incidence of mortality from choking in recent history has been small, and that the cost of laryngeal lowering may have been somewhat exaggerated. Of course, it may be that the incidence of fatal choking is now low as a consequence of laryngeal lowering being offset by increased control of breathing (itself intimately related to increased vocal control, see section 4.1.4, below), allowing humans (and other hominids) to control against inhaling liquid where before the high larynx was required to fulfil this role. Alternatively, it is possible that the potential cost of the lowered larynx has been offset by increasingly sophisticated social relations, which are able to act in aid of individuals who are choking – thus reducing deaths from choking, although not necessarily the incidence of choking. Modern human food preparation also contributes to ease of eating, and possibly accordingly reduced risk of choking. A second potential cost of increased basicranial flexion and decreased prognathism, which has only recently been mitigated, is that the shorter palate often results in impacted teeth, which until the 19th century were commonly fatal (Callaghan, 1994).

4.1.2. The hyoid bone and mandible

Although the flexing of the basicranium can indicate the position of the larynx in the throat, laryngeal position alone may not be enough to draw precise conclusions about the extent of vocal ability possessed by a specimen. Fortunately, other physiological features can be considered. A rare find from Kebara Cave, Israel, of the intact hyoid bone of a hominid dated to 60 Kya has allowed more detailed reconstructions of the vocal apparatus (see Figure 4.3). This find is unique in its completeness, though Frayer and Nicolay (2000) make mention of a second, less complete one.

The cranium of the Kebara 2 specimen does not survive, but the postcranial skeleton and age indicates that it is of a very robust Neanderthal-type hominid (Arensberg,

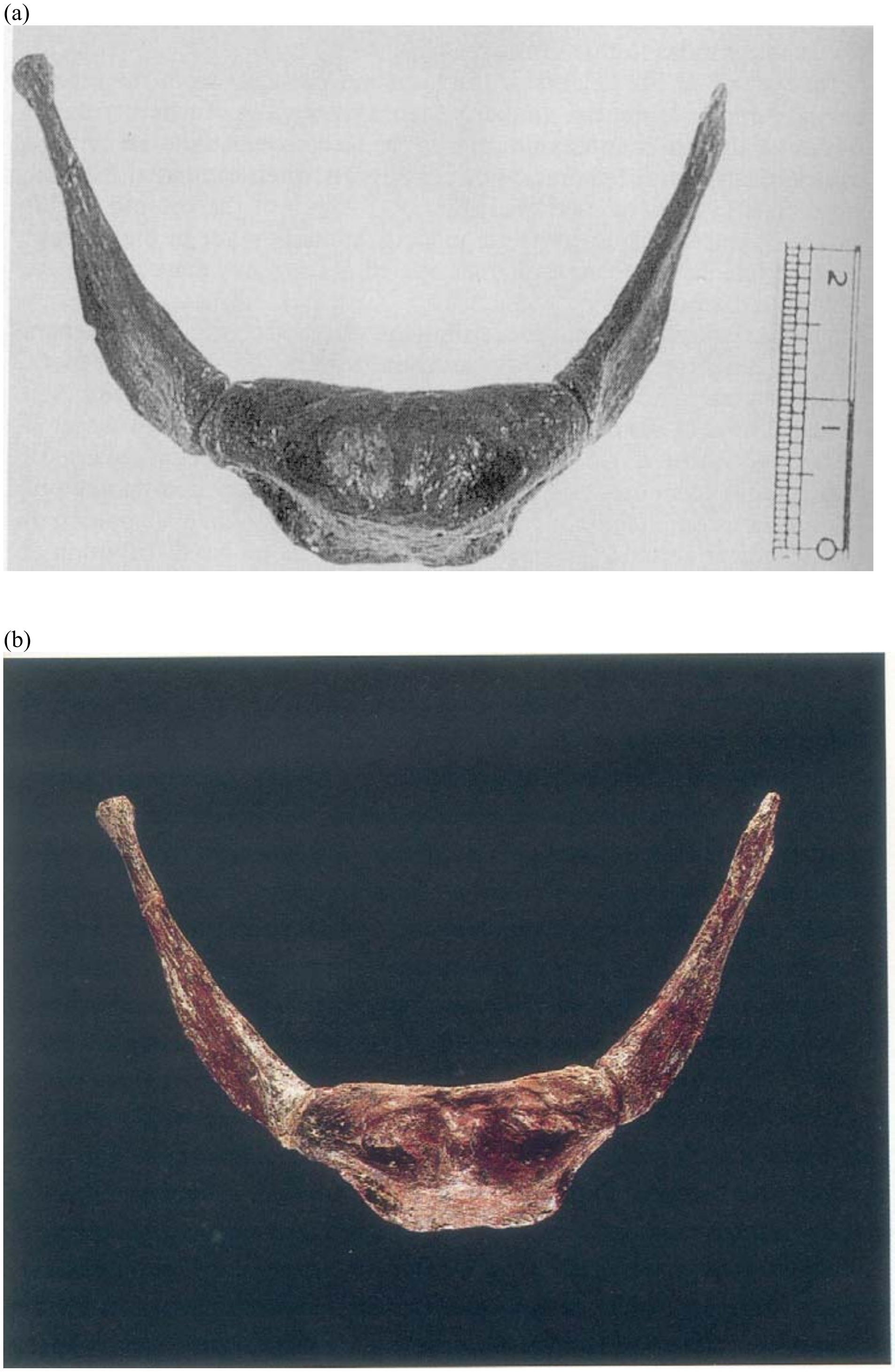

Figure 4.3: The intact hyoid bone from the Kebara 2 *Homo neanderthalensis*, dated to 60,000 years old. Its morphology and dimensions are well within the range of the hyoids of modern humans. The scale is in cm. ((a) *from Arensberg, 1989*; (b) *from Johanson and Edgar, 1996*).

1989). Arensberg *et al.* (1990) carried out a reconstruction of the vocal tract of the hominid using evidence from the muscular attachments on the hyoid, spine and mandible, and compared the hyoid bone with those of modern apes and humans. They conclude that the hyoid and vocal tract morphology of the Kebara 2 Neanderthal is virtually indistinguishable from that of modern humans, and is in contrast to that of great apes.

Arensberg *et al.* (1990) also conclude that rather than basicranial flexion, features of the mandible are a far better indicator of the position of the hyoid (and thus the larynx) and consequent vocal ability. The position of the muscular systems supporting the hyoid bone and larynx are indicated by the angle of a groove on the inside of the mandibles, called the *mylohyoid groove* (see Figure 4.2 for the position of the hyoid bone). Evidence from Middle Palaeolithic hominids, including Neanderthals and early moderns (Kebara 2, Amud 1, Qafzeh 9, Skhul V, Tabun I and II, Regourdou and Krapina J and C) show that this is identical in position and orientation to that of modern humans “suggesting a suite of muscular relations similar to modern humans must have been present” (Arensberg *et al.*, 1990: p.141). The conclusion drawn from all this is that, regardless of basicranial flexion, the hyoid and larynx of Neanderthals must have been low relative to the cranial base; they should thus have had supralaryngeal spaces approximately the same size as those of modern humans, and no significant difficulties in articulation.

4.1.3. The hypoglossal canal and tongue

Recently another hard tissue feature has been used as a source of information regarding the nature of the vocal soft tissue of human ancestors: the *hypoglossal canal* (Kay, Cartmill and Balow, 1998) (See Figure 4.4). The hypoglossal canal transmits the nerve that supplies the muscles of the tongue, and the authors hypothesise that the relative size of this canal (and accordingly the hypoglossal nerve that occupies it in life) may reflect the degree of motor co-ordination of the tongue present.

Interestingly, a further implication of a lowered larynx is that it can allow greater movement and control of the tongue body to produce sounds in the oral cavity, because of the low position of the tongue root (Fitch, 2000a); thus, evidence of greater innervation of the tongue muscle might suggest increased development of control of tongue articulation concomitant with increased vocal range. Kay *et al.* (1998) found that the relative size of the canal in *Australopithecus africanus* (Stw 19 and Stw 187) and what is possibly *Homo habilis* (Stw 53) (based on three skulls from Sterkfontein) is within the range of modern *Pan troglodytes* and *Gorilla gorilla*, whereas those of *Homo heidelbergensis* (Kabwe and Swanscombe), *Homo neanderthalensis* (La Chapelle-aux-Saints and La Ferrassie 1) and

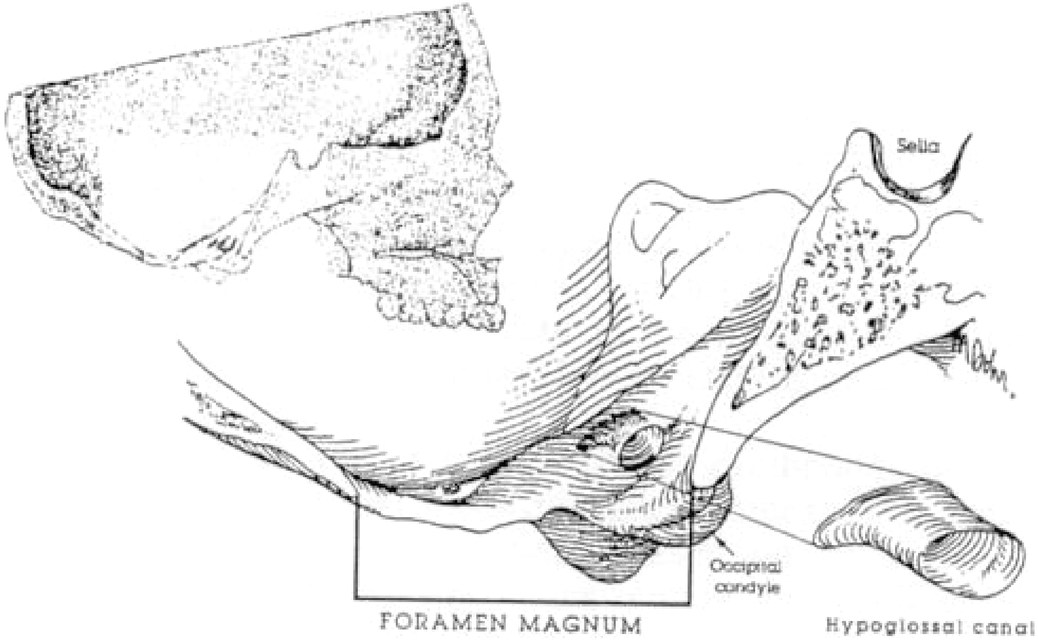

Figure 4.4: The bony hypoglossal canal in *Homo sapiens*, viewed from the interior of mid-sagitally sectioned skulls (*from Kay, Cartmill and Balow, 1998, p. 5418*).

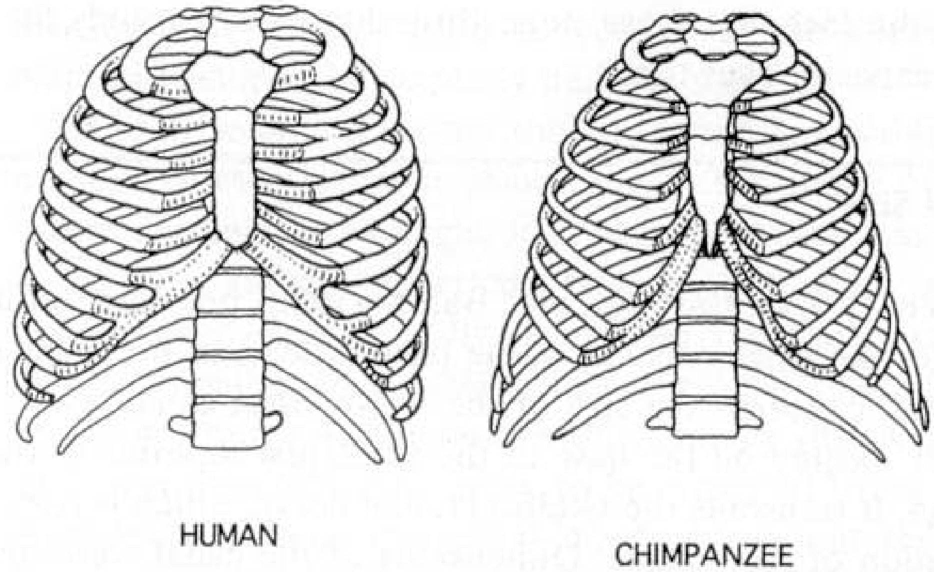

Figure 4.5: The rib cages of modern humans and of chimpanzees. The thoracic cage of great apes and australopithecines is funnel-shaped, whereas that of *Homo ergaster* and subsequent hominids, including humans, is barrel-shaped (*from Frayer and Nicolay, 2000, p.226*).

early *Homo sapiens* (Skhul 5) are well within the range of modern humans and significantly larger than *Pan troglodytes*. Given the estimated age of Swanscombe and Kabwe, Kay *et al.* suggest that a modern human pattern of tongue motor innervation and vocalisation abilities had evolved by around 300,000 years ago, in *Homo heidelbergensis*, and was present in subsequent species including Neanderthals.

These findings have more recently been questioned by DeGusta *et al.* (1999), following a thorough analysis of hypoglossal canal size in a large sample of non-human primates (prosimians, New World and Old World monkeys, and apes), modern humans and fossil hominid skulls. In their analysis they found that many non-human primates have hypoglossal canal dimensions within the range of modern humans, both in absolute terms and relative to the volume of the oral cavity, and that the same applied to fossil samples from *Australopithicus africanus* (the same three Sterkfontein specimens used by Kay *et al.*, 1998), *A. afarensis* (A.L.333-45, A.L.333-105 and A.L.333-114) and *A. Boisei* (Omo L.338-y-6). This difference in the findings of DeGusta *et al.* (1999) and Kay *et al.* (1998) seems largely to be due to the greater sample size and precision of measurement in the case of the more recent analysis. It is also evident that there is a large range of hypoglossal canal size in modern humans; this is more conspicuous in DeGusta *et al.*'s sample of 104 specimens than in Kay *et al.*'s sample of 44.

Furthermore, a study of five modern human cadavers by DeGusta *et al.* revealed no confident relationship between hypoglossal canal volume and the size or number of axons in the hypoglossal nerve (DeGusta, Gilbert and Turner, 1999). However, the issue of hypoglossal control of the tongue is not entirely a moot one; Jurgens and Alipour (2002) identified stronger and more direct neurological connections from the motorcortical tongue area to the hypoglossal nucleus in non-human primates than in non-primate mammals. They suggest that this indicates a phylogenetic trend toward stronger cortico-motoneuronal connection for tongue control in humans. They also suggest that "this might be one reason for the superior role the tongue plays in human vocal behaviour in contrast to non-human vocalisation" (p.245). One might suggest that the latter stages of such a development may be *a consequence of* the increased importance and complexity of tongue movement in vocalisations in humans. Certainly it suggests the laying of an important foundation for greater control of the tongue musculature at an early stage. Whilst this does not help the current case of trying to date the occurrence of such a development from the *fossil* record of hominids, the identification of such phylogenetic trends, apart from being useful in their own right, does suggest that with refinement of analysis technique there may yet be hope for identifying fossil physiological correlates with such a neurological development.

4.1.4. Vertebral innervation, intercostal musculature and breathing control

Another very important consideration pertaining to vocal ability is the fine control of the rib (intercostal) muscles; these muscles are responsible for aspects of breathing, through control of the ribcage, and thus are also implicated in the fine breath control required for extended and modulated vocalisations. Conclusions regarding the innervation of these muscles have also been drawn from the fossil and primate record. McLarnon and Hewitt (1999) carry out a very detailed analysis and discussion of the evidence for intercostal muscular control in primates as well as some fossil hominids.

The physiology and innervation of the thorax of extant primates and humans is quite different; the thoracic cage of great apes and australopithecines is funnel-shaped, whereas that of *Homo ergaster* and subsequent hominids, including humans, is barrel-shaped (see Figure 4.5, p. 88). Vertebral canal dimensions in extant primates (and humans) are quite well correlated with the dimensions of the spinal cord they contain, particularly in the middle and upper vertebrae, so vertebral canal dimensions can be interpreted as an analogue for the degree of spinal cord innervation.

The thoracic spinal nerves innervate the intercostal and subcostal muscles of the thorax, and the thoracic and abdominal muscles thus innervated have several functions (MacLarnon and Hewitt, 1999). As well as maintaining posture, and controlling movements such as body-rotation, they are also involved in coughing, parturition and defecation, and in the control of breathing. Direct comparison of modern primates and humans indicated that it is “only the grey matter of the modern human thoracic spinal cord that has expanded beyond the typical relative size for the order” (p. 347). In contrast, “the bulk of white matter containing nerve fibers passing through the thoracic region, to and from the hindlimbs, is of expected relative dimensions for a primate” (p. 347), so the canal dimensions do not seem to have been increased by greater innervation of the hind limbs in humans.

What is the implication of this innervation for vocal control? The respiratory muscles are responsible for controlling the air pressure from the lungs that reaches the larynx and upper respiratory tract. Of course, this is the case in all primates, but in humans they are able to maintain a constant air pressure throughout vocalisations of varying length, preventing the reduction in volume and intensity (and tonal control) that would otherwise occur over the course of a long vocalisation. Not only does this allow humans to “speak fluently in long sentences, without disruptive pauses for inspirations, and with the necessary pauses placed at meaningful linguistic boundaries” (MacLarnon and Hewitt, 1999, p. 350), but it also allows fine control of intensity, syllabic or phonemic emphasis, and pitch and intonation patterns (*ibid.* p. 350). This latter function would certainly be essential in making full use of the potential range of the modern vocal tract in both speech and melodic vocalisation, and MacLarnon and Hewitt go so far as to suggest that it is more important than the length and shape of the upper vocal tract structures in modifying loudness and emphasis.

MacLarnon and Hewitt speak of the importance of subglottal pressure control for the production of human speech "At both the larger scale of breath cycles and the finer scale of detailed features within phrases and words…" (MacLarnon and Hewitt, 1999, p. 350). Whilst this is undoubtedly the case, this description inverts the emphasis from that in which the current study is interested. They are working from *language* backwards towards the physiology rather than considering how the development of these physiological capabilities could form the basis for the complex vocalisations, the production of which they are trying to explain. Although MacLarnon and Hewitt do not say so, the function of fine control of intensity, syllabic or phonemic emphasis, and pitch and intonation patterns would almost certainly precede that of controlling long expirations. Control of tonal quality, pitch and amplitude of vocalisations seems to have a far older provenance than lengthy vocal sequences, as many sources of evidence suggest, from primatology to neurology (discussed in detail in Chapter 5). Such capabilities would be of importance earlier than "long sentences" and "linguistic boundaries", linguistic structures which would surely have evolved *in response to* or at least *in tandem with* physiological capabilities, rather than such capabilities being shaped by linguistic demands. There would almost certainly be "bootstrapping" between the physiological features and *vocal communicative* demands, but *linguistic* structures should emerge out of such already extant vocalisation capabilities.

At what stage in the fossil record do we see the earliest evidence of this increased vertebral innervation? MacLarnon and Hewitt (1999) found from their analysis of primate and fossil hominid skeletal remains that relative thoracic vertebral column diameter is increased in modern humans and Neanderthals relative to body mass, in comparison with extant higher primates, but that of Australopithecines and *Homo ergaster* is not. Australopithecines appear to have thoracic vertebral canal dimensions within the range of great apes (on the basis of measurements of AL288-1, Sts 14 and Stw 431), whereas Neanderthals (La Chapelle, Shanidar 2, Shanidar 3 and Kebara 2) and early modern humans (Skhul 4) have dimensions within the range of modern humans (MacLarnon and Hewitt, 1999). Unless this is a case of convergence, this would suggest that the common ancestor of Neanderthals and modern humans (and thus also *Homo heidelbergensis*) also had vertebral canal dimensions equivalent to those of modern humans.

At what point between Australopithecines and Neanderthals such increase occurred is more contentious, as the case of *Homo ergaster* (in particular, KNM-WT 15000) is the cause of some debate. Whilst MacLarnon and Hewitt (1999) assert that the thoracic canal arches in *Homo ergaster* fall within the range of modern primates in relation to estimated overall body mass, Frayer and Nicolay (2000) contend that they fall within the range of modern humans, if considered relative to cranial capacity. It would seem in this case that MacLarnon and Hewitt's treatment of the spinal innervation relative to body mass makes for a better grounded comparison, for a couple of reasons. Firstly, since cranial capacity has also increased disproportionately relative to body mass over the course of hominid evolution, surely human thoracic canal size *relative to cranial capacity* is *less* than that of other higher primates. Secondly, there is no reason to believe that innervation of the spinal cord should be related to neocortical development in the case of a fundamental physiological process such as breath control, unless you *presuppose* neocortical involvement, implying voluntary control, which is actually what we are hoping to identify. The null hypothesis should instead be that vertebral innervation correlates with body size and mass, in which case it is disproportionate growth in these terms that is interesting, as it does imply other processes at work.

Frayer and Nicolay (2000) also question whether thoracic canal measurement is relevant as an indicator of breath control at all. Whilst MacLarnon and Hewitt (1999), MacLarnon (1993) and Walker (1993) assert that the thoracic nerves are most significant for fine control of breathing for vocalisation, Frayer and Nicolay (2000) state in contrast that thoracic diameter is of little relevance: "most of the muscles involved in speech respiration are innervated by either cranial or cervical nerves…" (p. 226). The only *cervical* vertebra preserved from KNM-WT 15000 is C7, which does not have reduced dimensions from those expected (citing MacLarnon 1993).

In fact the authors' views are artificially polarised; both thoracic and cervical innervation is implicated in breathing control, but thoracic innervation is more important than Frayer and Nicolay (2000) acknowledge and less fundamental than MacLarnon and Hewitt (1999) assume. Whilst the thoracic intercostal nerves are indeed responsible for *some* control of breathing (Palastanga *et al.*, 2002), they seem to be implicated in additional fine control of a function that is principally controlled through cervical innervation. Cases of spinal injury in modern humans confirm this. Spinal injury below C6 still allows control of breathing without assistance, and vocalisation capabilities indistinguishable from normal, except that the patient is limited in the depth of inhalation. What the thoracic nerves seem to allow is control of especially prolonged vocalisations. The duration of vocalisations can be affected in patients with spinal injuries below the cervical vertebrae, but the *quality* of such vocalisations is not affected by the loss of thoracic control (The PoinTIS Spinal Cord Occupational Therapy site of the SCI Manuals for Providers; *pers. comm*. G. C. Morley, clinical physiotherapist).

If the cervical vertebrae of KNM-WT 15000 are indeed of modern dimensions as Frayer and Nicolay (2000) and MacLarnon (1993) assert, this indicates that there is no reason why we should not attribute to *Homo ergaster* the ability to voluntarily moderate its breathing sufficiently well to produce vocalisations controlled for pitch, intensity and contour. Limiting the variation in pitch and contour would have been the degree of development of the upper respiratory tract and, in particular, the larynx, as discussed above (which shows the first signs of development in *Homo ergaster*). What it may have lacked was the ability to produce *extended* voluntarily controlled vocalisations. It seems that this *is* an ability which has developed over the course of human evolution, indicated by thoracic canal innervation.

Other conclusions of Frayer and Nicolay (2000) should also be subject to further scrutiny. In line with their interpretation of the evidence regarding breathing control, they conclude from their analyses that by 1.5 mya, in *Homo ergaster* “both the articulatory capacity to form vowels and the respiratory capacity to maintain high-volume airflow were present” (Frayer and Nicolay, 2000, p. 233). Whilst the latter may well be the case (albeit limited in duration, see above), they base their conclusions regarding the former, articulatory capacity, on the evidence from the hyoid bone and basicranial flexion. The hyoid evidence comes from the Kebara specimen, dating to 60 kya (see above), so conclusions based on this cannot be extended to *Homo ergaster*. The basicranium, whilst showing the first evidence of flexion in *Homo ergaster*, does not achieve modern proportions until *Homo heidelbergensis*, 300-400 kya (see above) and, in any case, is only one of a suite of features that determine vocalisation ability. It should be possible to attribute to *Homo ergaster* some articulatory versatility, but not modern abilities; greater value, in developmental terms, is gained from the above evidence by identifying that it was with *Homo ergaster* that the first major steps towards modern vocalisation abilities were made.

What is especially significant is that both early anatomically modern humans and Neanderthals do show modern thoracic vertebral canal development, suggesting that such control over extended-duration vocalisation has an ancient provenance and is not a feature solely of modern humans. It would seem likely that the common ancestor of *Homo neanderthalensis* and *Homo sapiens* also exhibited this development; hopefully examples emerging from Atapuerca will confirm or refute this when they are subjected to analysis.

In summary, the foundations for the control of pitch and contour, as controlled by the cervical vertebral innervation would have been present in *Homo ergaster*, but it would probably have been limited as to the duration of such utterances. However, it is with this species that we see the first evidence for an increase in vocal versatility in the form of laryngeal development. This indicates an increase in tonal versatility and control before increased duration of utterances. It would seem likely that the requirement to control the pitch and contour of vocal sequences, and subsequently their duration, would have increased proportionally with the increase in range of pitch tones producible by the vocal tract as it evolved, as is described in section 4.1.

As MacLarnon and Hewitt (1999) point out, many primates vocalise in the form of discrete units of sound created with single air movements, but are limited in the duration of these and the order in which certain sounds can be made in the breathing cycle. They are also limited in the diversity of such sounds that they can make. This again suggests the subsequent development of long sequences of controlled vocalisations out of an initial ability to make discrete vocalisations which were first controlled for pitch and tone, the foundations for which are already evident in higher primates. From the evidence considered above, the most parsimonious scenario would seem to be that increasing control of intensity, pitch and intonation patterns of discrete vocalisations occurred initially, and increased with progressive laryngeal development, and that control over maintaining long sequences of such utterances followed. Short sequences of vocalisations controlled for pitch, contour and intensity would be communicative in their own right, and progressively more so with increasing control. On the other hand, long sequences of vocalisations with little control over pitch, intensity and contour would be far less meaningful. In the former case, as control increased, the length of sequences of such pitched and contoured utterances could increase; subsequently, the order in which the expressive vocalisations occurred could assume importance.

4.2. Production: percussive sound and rhythm

In contrast to vocalisation, which has specifically-dedicated and specialised physiological features, rhythm production does not have any dedicated physiological correlate of which there will be fossil evidence. In terms of percussive sound production we know already, from the evidence of stone tool manufacture (e.g. Wynn, 1993; Johanson and Edgar, 1996), that very early humans had the co-ordination and dexterity necessary to strike one item with another quite precisely. This would result, intentionally or unintentionally, in percussive sound (recent research by Cross, 1999b; n.d., has investigated the potential use of lithics themselves for intentional sound-making). Even without co-ordination to knock two objects together, percussive sequences could be created by foot stamping, an activity obviously physiologically possible long before the advent of the genus *Homo*. The question remaining is at what point *Homo* could have had the mental capacity for forethought and planning that would have been necessary to create these sounds in deliberate complex sequences necessary to produce a "meter or pulse". This issue, and the identification of the neurological mechanisms responsible for the control of meter and rhythm, is addressed in Chapter 5 and Section IV of the thesis.

The other question is of at what point *Homo* would have actually had the *incentive* to produce structured percussive rhythm for its own sake (rather than as a by-product of tool use, or as a product of other physiological activity or drives). Several potential evolutionary rationales for the use of music and rhythm have been proposed by various authors, and are discussed in detail in Chapter 6 and 7. As for potential selective benefits to such tendencies in adult life, McNeill (1995) reports how the use of rhythm alone may be significant in inducing group cohesiveness. He observes that when people move together in dance (or military drill) this elicits a physiological arousal in the participants, and a great sense of emotional solidarity. He calls this "muscular bonding", and proposes that such effects would have important selective benefits to any hominid group that practised them. On the basis that all surviving cultures of humans make use of music and dance of some form, he suggests that those groups that had practices of moving together rhythmically developed stronger communities and became better at co-operating together in other spheres, particularly hunting.

Whilst this is speculative, the principle of "muscular bonding" itself has a strong foundation; it is used in military circumstances the world over, and ethnographic evidence of the action of dance on evoking group solidarity is very widespread (e.g. Nettl, 1992; Locke, 1996; Johnston, 1989; see Chapter 2). Further, the very close association of the areas of the brain responsible for complex vocalisations and complex sequences of muscular movements, which seem to have evolved in tandem (see Chapters 5 and 6), may go some way to explaining the association of rhythm, dance and music in almost all cultures.

4.3. Perception: the ear and its evolution

The production and perception of sound are inextricably linked; in most mammalian species the frequency range of their vocalisations (mainly determined by their laryngeal system; see above) largely coincides with that most effectively perceived by their auditory system (Wind, 1990). Indeed, the length (25-30mm) of the human auditory meatus (ear canal) means that it resonates at about 3500 Hz, resulting in the ear being progressively more sensitive to frequencies up to this range; this is also the upper range of normal human vocalisation frequencies (Benade, 1990; Campbell and Greated, 1987), meaning that the whole of the human vocal range is very effectively perceived. A number of important changes appear to have occurred over the course of the evolution of the human auditory system with specific relation to its use for perceiving vocalisations. It would seem that sometimes the development of the auditory system has been spurred by developments in the vocalisation capabilities, at other points, vice versa.

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

Over the course of mammalian evolution, laryngeal and respiratory differentiation made more variation in pitch and amplitude possible, as well as more frequent vocalisations, while increasing auditory sophistication enabled a greater variety of sounds to be perceived and analysed. As Wind puts it, 'Mammalian vocalisations are mainly used for intraspecific communication. Therefore auditory function must have been shaped by selective pressures gearing them to vocalisation capacities'. (Wind, 1990, p.185).

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

That these "natural auditory categories" are most effectively perceived by non-human mammals and pre-linguistic humans suggests that the perceptual mechanisms relied on for the perception of speech sounds are very ancient, pre-dating humans by millions of years. As hominids developed the ability to control their vocalisations in order to communicate, there would have been strong selective pressure to be able to vocalise using these sounds that are most easily perceived by others. So, it would seem that audition, specifically the existence of 'natural auditory categories', was initially responsible for the formation of speech categories, and that the mechanisms for perceiving such categories were in place in our hominid ancestors before they were capable of actually producing articulate speech. This has important implications for the debate concerning Neanderthals' information processing abilities with respect to their supposed vocal abilities (*contra* Lieberman, 1989; section 4.1).

No matter in which order the major evolutionary developments occurred, it appears that the vocal and auditory systems must have evolved in response to each others' capabilities. A further piece of evidence indicating that the inner ear systems are intimately linked with the development of vocalisation is that the stapedius inner-ear muscles of humans (which regulate intensity of the operation of the stapes bones on the eardrum) involuntarily contract before the onset of vocalisation. This reduces the intensity of the sound of our own voice and prevents our own speech from masking other sounds from the surrounding environment, allowing us to hear other sounds even while we speak (Borg and Counter, 1989). The value of this function would presumably increase as the length of vocalisations, and the diversity of their frequencies, increased; vocalisations obscuring environmental sounds would be less of a handicap if such utterances were short, and the more extended they became, the greater the proportion of the "auditory scene" that would be lost. The same goes for the frequency range that constitutes the vocalisations: the greater it is, the greater the proportion of environmental sounds that would be obscured during an utterance. This suggests that the auditory system, over the course of its evolution in human ancestors, also faced significant selective pressure as a consequence of our vocalisation abilities.

Another major factor in the development of the human auditory system was bipedalism (Daniel, 1990; Spoor *et al.*, 1994; Spoor and Zonneveld, 1998). The labyrinthine structure of the inner ear is important for the regulation of balance, as well as hearing. Daniel (1990) and Spoor *et al.* (1994) link the anatomy of this structure with bipedal posture. In all quadrapedal mammals, the labyrinthine capsule of the inner ear (see Figure 4.6) is situated vertically, whereas in anatomically modern humans it is rotated through almost 90 degrees (Daniel, 1990). Fortunately, the bony labyrinth is often preserved inside fossil hominid skulls, which has allowed Spoor *et al.* (1994) to carry out detailed computer reconstructions of the labyrinthine system of various fossil hominids, based on computer tomography (CT) scans of the skulls. They then compared the rotation and dimensions of the labyrinths of the fossils with those of modern humans and extant great apes. Whilst there is not the same extreme of difference in rotation of the labyrinth between humans, hominids and great apes as there is between humans and quadrapedal animals, there are still important differences. They found that the labyrinths of *Australopithicus* and *Paranthropus* closely resemble those of modern great apes, whereas fully modern morphology is first exhibited by *Homo ergaster* (SK 847, c. 1.5 Mya), and take this to imply that the Australopithecines and Paranthropines were only partially bipedal, whereas *Homo ergaster* had full bipedality. More recently a more comprehensive analysis and comparison of the entire labyrinthine structure in humans, great apes and other primates has been carried out by Spoor and Zonneveld (1998), who show that there are other important phylogenetic differences in morphology of the labyrinth. Especially pronounced are differences in the relative size and some of the angles of the semicircular canals, and the angle of rotation of the cochlea. Their analysis also shows that all four of the great ape species (*Pan troglodytes, Pan paniscus, Gorilla gorilla and Pongo pygmaeus*) share the same morphological pattern, as do the South African australopithecines and the hominoid *Dryopithicus brancoi* (RUD 77; Spoor, 1996; Kordos and Begun, 1997) found in 10 million year-old deposits at Rudabanya, Hungary. These similarities strongly suggest that the differences evident in humans have emerged over the course of the emergence of the genus *Homo* (Spoor and Zonneveld, 1998).

In something of a moderation of the conclusions presented by Spoor *et al.* (1994), Spoor and Zonneveld acknowledge that the semicircular canals are responsible for a whole suite of functions in addition to balance and the coordination of body movements during locomotion, including stabilisation of gaze during movement of the body through moderation of neck and ocular reflexes. All these functions may be implicated in the development of the morphology of the labyrinth. However, it does still seem likely that the moderation of head movement with particular regard to locomotion would have been a significant selective force in shaping the morphology of the semicircular canals, and they may also fulfil a more fundamental role in the regulation of bipedal motion itself. As Spoor and Zonneveld describe, Ito and Hinoki (1991) found that in quadrapeds the vestibulo-

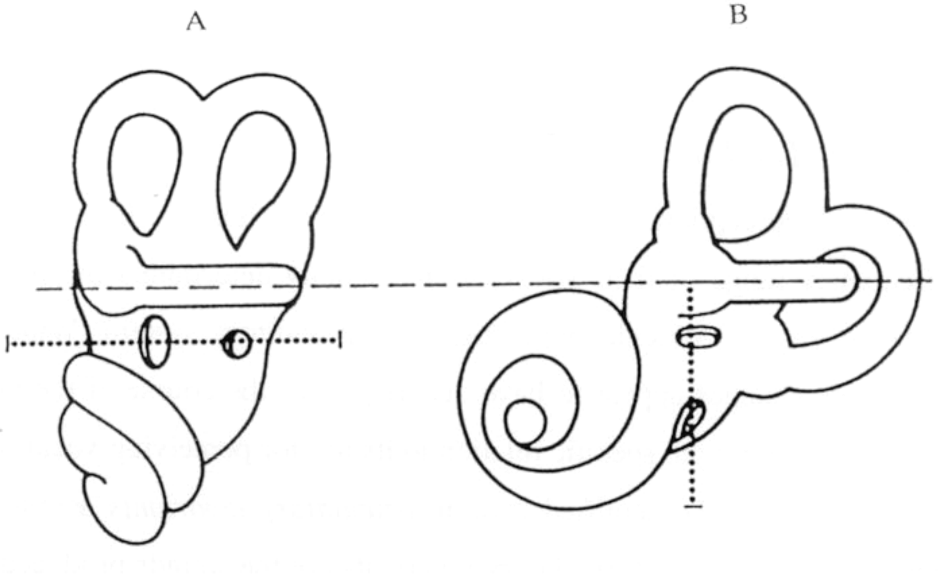

Figure 4.6: The left labyrinths of the inner ear of (A) a quadraped and (B) *Homo sapiens*. The horizontal plane is represented by the dashed line. The whole arrangement has rotated through 90 degrees, apparently as a consequence of bipedalism (*from Daniel, 1990, p. 260*).

spinal tract, which is supplied from the semicircular canals, extends only to the upper thoracic levels of the spinal cord, whereas in humans it seems that neurological impulses from the semicircular canals reach the lumbosacral cord, influencing the activity of lower- as well as upper-extremity movement.

The important question for this study remains at what point these changes in the morphology of the inner ear occurred. Spoor and Zonneveld (1998) reiterate the findings of Spoor *et al.* (1994) that "detailed analysis of hominid remains make it possible to pinpoint in what taxa and what period canal enlargement occurred and compare such findings with postcranial evidence on the emergence of habitual bipedalism" (p. 243), showing that *Homo ergaster* displays modern human canal dimensions, whilst the South African australopithecines show great ape-like morphology. Interestingly (given that *Homo ergaster* is thought to be ancestral to all later forms of *Homo*), Neanderthals prove themselves yet again to be the exception to the rule by having smaller *vertical* canal dimensions than modern humans; this does highlight the fact that there is clearly more than simply a shift from quadrapedalism to bipedalism at work in determining the morphology of the semicircular canals. Nevertheless, the fundamental changes to inner-ear anatomy (to that exhibited by modern humans) do appear to have occurred with *Homo ergaster* and not before, and the morphology of the human inner ear labyrinth is unique amongst extant primates. At least some of the changes in the features of the labyrinthine anatomy appear to be correlated with another physiological change particularly associated with *Homo*, namely, changes in the shape of the basicranium. The cranial base accommodates the labyrinthine structure, and with the basicranium shortening and flexing with increased brain size, which occurred in all subsequent hominids to some extent (see section 4.1.1, above), it seems this has had distorting effects upon the labyrinth.

Daniel relates the rotation of the inner ear anatomy in humans to various other neurological, as well as physiological, changes. In humans, rotation of the labyrinth from vertical to horizontal occurs ontogenically too, during the development from foetus to adult, which reinforces the idea that bipedal locomotion is a major cause of the rotation (Daniel, 1990). It would seem that these adjustments to the labyrinthine system as a consequence of bipedalism had important neurological knock-on effects with regard to language and learning. The vestibular neurological system associated with the labyrinth is also phylogenically and ontogenically the first to exhibit lateralisation tendencies. Daniel (1990) proposes that this, and clinical connections between vestibular dysfunction and language and learning deficits, suggest that the foundations of linguistic functions are first evidenced by *Homo ergaster*. According to the findings of Spoor *et al.* (1994), *Homo ergaster* (1.5 Mya) is the first hominid to have modern anatomy in this respect, which is further evidence that *Homo ergaster* is the earliest hominid to exhibit signs of vocal and linguistic ability different from those of higher primates. Ultimately, the conclusion is that regardless of when *language* actually emerged in humans, the beginnings of the neurological predisposition for it, and other complex vocalisations such as singing, coincided with bipedalism and the associated labyrinthine-vestibular changes at least 1.5 Mya, in *Homo ergaster*.

4.4. Chapter summary and conclusions

In discussing the earliest foundations of musical behaviours in the human lineage, one is necessarily investigating the origins of the production and processing of complex vocalisations and muscular movements. The research discussed in the preceding chapter reaches a consensus with regard to the earliest occurrence of modern-like vocal physiology and ability. Although there are clearly difficulties with carrying out reconstructions and analyses of fossil specimens, these techniques are becoming increasingly refined, and the evolution of the vocal apparatus can be tracked with reasonable confidence in the fossil record.

Beginning with *Homo ergaster* about 1.7 Mya there was a progressive increase in the degree of basicranial flexion, culminating in modern-like basicranial anatomy in *Homo* *heidelbergensis* about 300 Kya. Basicranial flexion is a good indicator of the level of development of vocal tract morphology, but it is not the sole determining factor in vocal ability. It is most valuable when considered in tandem with the development other physiological features. Neanderthals show less basicranial flexion than modern humans, and even than *Homo heidelbergensis*, but this need not imply that the nasopharyngeal soundspace of Neanderthals was significantly less, as reduction in kyphosis of the spine as a consequence of full bipedalism also results in a lowered laryngeal position relative to the cranial base. This, in combination with various other physiological developments as indicated by the mylohyoid groove and hyoid anatomy, indicate that Neanderthals would have had supralaryngeal spaces approximately the same size as those of modern humans, and no significant difficulties in articulation.

The foundations for the control of pitch and contour, as controlled by the cervical vertebral innervation would have been present in *Homo ergaster*, but it would probably have been limited as to the duration of such utterances. Also limiting the variation in pitch and contour would have been the degree of development of the upper respiratory tract and, in particular, the larynx; however, it is with this species that we see the first evidence for an increase in vocal versatility in the form of laryngeal development. This indicates that there was an increase in tonal versatility and control before there was an increased duration of utterances.

It seems that the changes in basicranial anatomy, and consequently of the laryngeal tract and supralaryngeal space, were instigated by the shift to a fully upright bipedal posture, and that the ability to control the increasing diversity of pitches possible as a consequence also emerged at that time. This ability would likely have had significant benefits for individuals in terms of their vocal expressiveness, especially in the context of increased social complexity. Subsequently, there emerged an ability to control breathing sufficiently well to produce increasingly extended sequences of such vocalizations. It would seem likely that the requirement to control the pitch and contour of vocal sequences, and subsequently their duration, would have increased proportionally with the increase in range of pitch tones producible by the vocal tract as it evolved.

By the time of both Neanderthals and early *Homo sapiens* the physiology for control over pitch and duration of utterances seems to have been fully developed. Unless this is a case of convergence between Neanderthals and modern humans, this would suggest that the common ancestor of Neanderthals and modern humans (and thus *Homo heidelbergensis* too) also had such control over vocal range and extended duration, equivalent to that of modern humans.

From the evidence considered above, the most parsimonious scenario would seem to be that increasing control of intensity, pitch and intonation patterns of discrete vocalisations occurred initially, and increased with progressive laryngeal development, and that control over maintaining long sequences of such utterances followed. Short sequences of vocalisations controlled for pitch, contour and intensity would be communicative in their own right, and progressively more so with increasing control. On the other hand, long sequences of vocalisations with little control over pitch, intensity and contour would be far less meaningful. In the former case, as control increased, the length of sequences of such pitched and contoured utterances could increase; subsequently, the order in which the expressive vocalisations occurred could assume importance.

As with the vocal apparatus, the first significant developments of auditory anatomy occur with *Homo ergaster*, 1.5-1.7 Mya, and also seem to be related to a shift to a fully upright posture. It seems likely that the moderation of head movement with particular regard to locomotion would have been a significant selective force in shaping the morphology of the semicircular canals of the inner ear, and they may also fulfill a more fundamental role in the regulation of bipedal motion itself. The shortening and flexing of the basicranium with increasing brain size had distorting effects on the shape of the labyrinth, housed as it is in the cranial base. Earlier hominids, on the other hand, exhibit auditory anatomy that is not significantly different from that of other higher primates.

Some features of auditory perception are obviously very ancient, being features of mammalian audition in general. One of these is the preferential perception of the so-called "natural auditory categories"; these are also universal features of human speech sounds. This suggests that these qualities of vocalisation were tailored to the capabilities of the auditory system: as hominids developed the ability to control their vocalisations in order to communicate, there would have been strong selective pressure to be able to vocalise using these sounds that are most easily perceived by others. So, it would seem that audition, specifically the existence of "natural auditory categories", was initially responsible for the formation of particular vocalisation properties, and that the mechanisms for perceiving such categories were in place in our hominid ancestors long before they were capable of actually producing articulate speech.

Other features of human auditory function appear instead to have faced significant selective pressure as a consequence of hominid vocalisation capabilities. The value of the function of the stapedius inner ear muscles in reducing the intensity of our perception of the sound of our own voice would presumably increase as the length of vocalisations, and the diversity of their frequencies, increased. Vocalisations obscuring environmental sounds would be less of a handicap if such utterances were short, and the more extended they became, the greater the proportion of the "auditory scene" that would be lost; the same goes for the frequency range that constitutes the vocalisations: the greater it is, the greater the proportion of environmental sounds that would be obscured during an utterance.

In summary, it seems that important changes toward modern auditory anatomy occurred in *Homo erectus* with full bipedalism, that the vocal system is set up to produce sounds which are particularly well-perceived by the auditory system, and that the human ear contains features which are particularly useful in an animal which carries out extended tonal vocalisations.

Chapter 5 examines in depth the neurology used for these activities, and the fossil and primate record, for more clues as to their development and how they are interrelated.

CHAPTER 5

The Origins and Evolution of the Neurology of Musical Production and Perception

Chapter 4 examined the evolution of the physiology required for music production and perception. The physiological features are only one aspect of musical capacities, however; in fact, they tell us when the *capability* to carry out such behaviours may have become developed, but little about the form of the behaviours which may have used those capabilities. Having the physiological mechanisms does not necessarily imply possession of the cognitive capacity to use them for the production or perception of rhythm and melody. It is important to identify the areas of the brain responsible for the production and perception of vocalisations, rhythm and melody, and how these functions interrelate. This chapter examines the evidence from a wide range of disciplines, including music psychology, neuropathology, primatology and developmental psychology, with a view to answering these questions. It aims to determine which of these capacities fundamental to musical behaviour are unique to humans and which share foundations with higher primates, which are innate and which are learned, which appear to be specifically dedicated to music and which are related to other behaviours.

The first section of the chapter reviews evidence from neuropsychology, neuropathology and music psychology identifying the areas of the brain responsible for the production and processing of music, speech and language. The second likewise examines the evidence for the locations of the production and processing of rhythm. The third part of the chapter looks at human and primate developmental studies for evidence of which elements of vocalisation, melody and rhythm ability are innate and inherited, as opposed to culturally determined. Finally, the fourth section of this chapter consists of an analysis of evidence from the fossil record of endocasts and neurological studies of primate vocalisation capabilities, allowing conclusions to be drawn regarding the phylogenetic emergence of these capacities.

5.1. The locations of the production and processing of music and language.

As singing and speech make use of the same physiological mechanisms for their production and perception (namely, the vocal chords and ear) one might expect that the two are closely linked, in evolutionary terms and in terms of the neurological mechanisms they use. Neurologically the parallels between music and language function are more complicated. Research relating musical functions (and deficits) to neurological study is reviewed comprehensively by Marin and Perry (1999), and earlier by Borchgrevink (1982). Such research has a long history (almost half of Marin and Perry's 290 references pre-date 1980) and this body of research continues to grow. Dealing as it does predominantly with the western musical tradition, much of this research and literature is concerned with technical instrumental ability, the naming of pieces of music and reading and writing of notation. The current chapter is concerned with musical behaviours at a more fundamental level, of course, pre-dating such elements of musicality as written notation and the naming of tunes, and there are two main purposes to be fulfilled here by examining the neuropsychological literature regarding the location of musical and other vocal functions. One is to be able to relate this information to such evidence as we have of neurological development in the fossil record, the other is to attempt to identify interrelationships and differences between the neurology responsible for musical and linguistic functions, in attempting to propose a path for their development. Consequently, of greatest interest are those investigations exploring prosodic, tonal, rhythmic and lexical-semantic comprehension and production in music and language.

5.1.1. Evidence from pathology

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

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

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

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

Marin and Perry (1999) go on to state that examples of aphasia without amusia "decisively contradict the hypothesis that language and music share common neural substrates" (p.665). By this, given their previous observations, they are using the term "language" to refer only to verbal, *lexical* language functions, and not the affective tonal elements of speech. Such cases of aphasia without amusia only indicate that music and the *lexical, syntactic* elements of speech do not share common neural substrates.

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

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

Further, that the comprehension and production of the prosodic elements of language often exist intact with the comprehension and production of the prosodic elements of music, and that when one of these functions is damaged, so too is the other (Mazzucchi *et al.*, 1982; Peretz, 1993; Peretz *et al.*, 1994; Patel *et al.*, 1998), indicates instead that there *are* shared common substrates between elements of speech and musical functions, namely, those concerned with prosodic intonation and melodic contour. Patel *et al.* (1998) conclude from their study of two amusics that their findings “support the view that the perception of speech intonation and melodic contour share certain cognitive and neural resources, as do the perception of rhythmic grouping in the linguistic and non-linguistic domain” (p. 136). Snow (2000) goes even further, with regard specifically to the emotional expression in the intonation of utterances, concluding that “linguistic analysis suggests that even intonation patterns traditionally described as nonemotional have their underpinnings in the speaker’s emotions” (p. 1) and goes on to observe that pathological studies “indicate that intonation patterns described as either linguistic or emotional are mediated by right-hemisphere substrate specialised for emotional experience” (Snow, 2000, p. 1). This suggests, then, that whatever the content of speech, its production, specifically its intonation, unavoidably involves some input from neurological systems concerned with affective function.

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

So, the evidence from pathologies resulting in aphasias and amusias suggests that we have a common neurology for musical and linguistic vocalisations at a "lower" emotive prosodic level, and contralateral localisation of neural processing at "higher" linguistic semantic levels. This would suggest that the specialised human neurology dedicated to the higher linguistic and musical functions emerged later out of a set of common neurological substrates concerned with (complex) vocal emotional expression and comprehension. Indeed, following their review of a large body of research, Marin and Perry (1999) propose that "The close correspondence between the networks of regions involved in singing and [linguistic] speaking suggests that [linguistic] speech may have evolved from an already-complex system for the voluntary control of vocalisation. Their divergences suggest that the later evolving aspects of these two uniquely human abilities are essentially hemispheric specialisations" (p. 692).

5.1.2. Evidence from brain activity

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

Some specifics of these specialisations are suggested by Benson (1985). Benson observed P.E.T. (positron emission tomography) scans of the activity in the brains of subjects while they carried out various tasks relating to the production and perception of speech. His results show that the left hemisphere appears to be dominant with regard to semantic verbal meaning and syntactic sequencing and relationships, whilst the right hemisphere is dominant in prosodic melody and semantic visual images (Benson, 1985, Table 3, p. 201). For the other aspects of gestural, prosodic, semantic and syntactic language, both hemispheres participate.

This accords well with the findings of Borchgrevinck (1980, 1982) with regard to rhythm and melody production and perception. Borchgrevink selectively anaesthetised either the right or the left hemisphere of his subjects by administering a local anaesthetic (sodium Amytal) via the carotid artery (on either the right or left side of the neck, respectively). Anaesthetisation of the right hemisphere resulted in loss of control of pitch, whilst left-hemisphere anaesthetisation arrested both speech and singing (Borchgrevink, 1980). Subsequently he observed the performance of the subjects at various tasks relating to the production and perception of rhythm and melody, with one or other hemisphere anaesthetised. He found that the right hemisphere seems to be responsible for controlling pitch and tonality of singing (cf. prosodic melody in speech, above), whilst musical rhythm appears to be controlled by the left hemisphere (cf. syntactic sequencing, above) (Borchgrevink, 1982). These links between the systems of linguistic and musical prosody and the systems of linguistic and musical rhythm are further reinforced by findings by Peretz *et al.* (1994), whose patients with bilateral lesions of the superior temporal cortex exhibited simultaneously impaired perception of prosody in speech and perception of melody in music, whilst having preserved perception of the rhythmic elements of speech and music. Whilst one would perhaps not expect to see rhythmic capacities affected by temporal lobe damage, this serves to highlight the link between prosodic elements of speech and melody, and their neurological separation from rhythmic elements speech and melody.

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

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

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

5.1.3. Analysis of sound

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

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

Distinctions between the processing required for linguistic, musical and environmental sounds are also observed as a result of neural pathologies. In cases of auditory agnosia (disruption of the perception of speech, animal noises, music and other sounds), when the brain lesion is in the left hemisphere only, the perception of non-speech sounds (including music) tends to be largely intact, whilst speech perception is disrupted (Marin and Perry, 1999). In contrast, there are also cases of unilateral right hemisphere lesions in which preservation of normal speech perception exists but the perception of non-verbal sounds (including music) is disrupted (Fujii *et al.*, 1990). In the latter case, and in generalised auditory agnosia (usually the result of bilateral damage), the perception of timbre and complex sounds, as well as rhythm and temporal information in the stimuli, is often severely disrupted (Marin and Perry, 1999).

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

In summary, the right hemisphere appears to be responsible for processing and production, in both melody and speech vocalisation, of prosodic melody, pitch control, tonality of singing, timbre processing and voice recognition. The left hemisphere appears to be implicated in production and processing of semantic verbal meaning and syntactic sequence, as well as rhythmic and analytical function.

Thus, both hemispheres are required in the production and processing of both music and language; some of the fundamental elements of music and language production and perception are shared, using the same neurological mechanisms, and some have subsequently become specialised. Musical functions as a whole are less clearly lateralised than language functions, but tasks relating to pitch and pitch discrimination do seem to be right hemisphere dominated. Linguistic functions seem to be most detrimentally affected by left hemisphere lesions; most musical functions seem to be impaired in some respect by damage to either hemisphere. Combinations of aphasia and amusia tend to result from left hemisphere damage, whilst cases of sensory amusia without aphasia tend to be related to right hemisphere lesions.

5.2. The locations of the production and processing of rhythm

As regards the location and the development of the areas of the brain responsible for the production and processing of rhythm and meter, there are also a number of important recent pieces of work, in addition to the implications of the various research discussed in the preceding sections. Petsche *et al.* (1991) report a number of pathological cases illustrating the importance of the left frontal hemisphere in rhythm perception and production. Their own work using EEG scans illustrated that, in fact, many areas in both hemispheres are stimulated when processing rhythm, and that the two halves of the brain are functionally closely connected whilst doing so, more so than “during any other sensory function” (Petsche *et al.*, 1991, p. 326).

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

It would also seem that auditory processing of rhythmic input is importantly implicated in corporeal muscular control, and that the systems are somewhat linked. Thaut *et al.* (1997) demonstrated that the use of a metronomic stimulus has a remarkable effect upon the rehabilitation of gait control in stroke patients. It seems that the external rhythmic stimulus fulfils an important cueing role in gait control, in terms of stride rate, length, and lateral (left, right) rhythm, and especially when internal mechanisms for generating such cues are no longer functional due to the pathology. It seems that it is not merely the case that the patients are moving in time with the metronomic cue, but that this is actively aiding in the stimulation of motor control; when the rhythmic stimulus is removed, motor control also degenerated. Thaut *et al.* suggest that "auditory rhythmic timekeepers may enhance more regular motor unit recruitment patterns… [this effect] underscores the existence of physiological mechanisms between the auditory and motor systems. The ability of auditory rhythm to effectively entrain motor patterns and also influence nontemporal parameters such as stride length, may help to assign rhythmic auditory stimuli a larger role in motor control than previously assumed." (p. 211). Such associations between locomotor control and rhythm lend weight to a suggestion that the shift to bipedal locomotion may have been an important factor in the development of rhythmic entrainment in humans.

Trevarthen (1999) expounds upon the evidence for what he calls the human "Intrinsic Motive Pulse", or IMP. Existing as the brain stem, basal ganglia and the limbic structures of the Emotional Motor System, the Intrinsic Motor Formation (IMF) is responsible for integrating functions of attention, learning, and physiological actions of expression, including the synchronisation and co-ordination of sequences of movements, according to Trevarthen. The physiological and functional parameters of the IMF and IMP ("The body-moving rhythmic and emotionally modulated system is the Intrinsic Motive Pulse, or IMP", Trevarthen, 1999, 160) are not clear from Trevarthen's exposition, but it does seem to be the case that there is a close interrelationship between the emotional-controlling elements of the limbic system and the areas responsible for the co-ordination of motor sequences and posture. In Trevarthen's own words, the "movement-creating reticular networks and nuclei are intricately combined with the neurochemical systems of emotion. The same activating neurones that select movements and control their energy and smoothness also cause changes in the emotions felt and the intensity and 'colour' of consciousness" (p.161).

This system is active automatically in all types of interpersonal interaction, but is made deliberate use of in musical activities. Trevarthen believes that this system has been able to reach its current state of versatility and expressiveness as a consequence of our bipedal posture, which has freed up large proportions of the body for spontaneous movement that would otherwise be occupied by locomotion.

That developed melodic and rhythmic sequences in music are not processed simultaneously in the brain by the same mechanisms (above; Peretz and Kolinsky, 1993) would suggest either that they have separate functional and behavioural roots and have come to be used together subsequently, or that whilst they might share roots in muscular control, they have come to be specialised functions subsequently. In fact, it would seem that the elements of vocal sound production that rely upon planning and execution of fine orofacial muscular movements involve the use of left hemisphere mechanisms also implicated in rhythmic sequences and planning of such sequences. On the other hand, the elements of vocalisation which rely purely on laryngeal control, and the perception of such sounds, seem to involve predominantly right hemisphere mechanisms which are also implicated in melodic production and perception. At a fundamental, non-analytical level, however, there are important links between affective and rhythmic-motor activity (c.f. Trevarthen, 1999); this is discussed further in the following sections and Chapter 6.

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

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

It also seems that infants have an innate ability to accompany the emotive-tonal prosodic infant-directed speech (discussed in considerably greater depth in Chapter 6) known as "motherese" (or more accurately, "parentese"), or in other circumstances singing, with rhythmic body movements which co-ordinate with the rhythm of the vocalisation (Trevarthen, 1999). This can occur even in congenitally blind infants, who cannot possibly be imitating the movement of the parent.

It would be a particularly valuable addition to the above infant-based research to investigate the responses of infant *Pan* (chimpanzees) to the same stimuli, to ascertain whether they display the same abilities and responses. Whilst there is always the issue that *Pan* have also been subject to 6 million years of selective pressures since the speciation which ultimately led to *Homo*, it would at least suggest whether the ability to pick up subtle affective information from vocalisations has been selected for since then or whether it was already inherent in our last common ancestor.

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

Similar findings were made by Elowson *et al.* (1998b), observing infant pygmy marmosets; response from a caregiving adult was far more likely to be given to an infant when vocalising than when it was not. The only known study to date of non-human primate "babbling" behaviour (a form of early vocal behaviour, pre-linguistic in human infants) has identified considerable parallels with human infant babbling behaviour, on a number of counts (Elowson *et al.*, 1998a). Elowson, Snowdon and Gros-Louis (1998a) review and outline seven key features of human infant babbling:

1) Babbling is universal and frequent irrespective of the infant's cultural background;
2) Babbling is rhythmic and repetitive;

3) Babbling begins between six and ten months, peaking at seven months of age;
4) Babbling comprises a subset of the phonetic sounds found in adult speech;
5) Babbling has well-formed units with the consonant-vowel structure of adult speech;
6) Babbling lacks apparent meaning with respect to how syllables are used by adults in language;
7) Babbling is instrumental in molding the caregiver-child bond in the infant's first year. (p. 33, box 2).

In their long-term study of Amazonian pygmy marmosets *Cebuella pygmaea*, Elowson *et al.* (1998a) observed that "(1) both humans and pygmy marmosets have evolved a family-based social unit of cooperating individuals; (2) both humans and marmosets have an open, plastic system of communication where subtle changes of vocal production occur over the life-time of the individual; and (3) both go through a period of babbling as part of their vocal development" (p. 36).

In particular, "the long, complex strings of vocalisations produced by the infant monkeys" (p. 32) have much in common with the babbling of human infants outlined above, such as "universality, repetition, use of a subset of the adult vocal repertoire, recognisably adult-like vocal structure and lack of a clear vocal referent" (p. 31).

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

For Elowson *et al.* (1998a) the most interesting element of their findings is that it indicates that similar processes underlie (at least some) primate vocal learning to those that underlie human vocal learning, and they accordingly propose that pygmy marmoset babbling has a relevance to understanding the evolutionary processes of vocal development. In the context of the current research the most interesting implication is a phylogenetic one or an evolutionary one. If this is a form of early vocal behaviour and learning that is phylogenetically ancient and genuinely shared with other non-human primates, then it is interesting in that right alone, indicating that we could have the expectation that such behaviours were carried out by intervening ancestral hominid species. If, on the other hand, it is a type of behaviour that has emerged by convergence, it is interesting as it could inform about the circumstances in which such behaviours are selectively useful to the individual and the group. In either case, it indicates that lower primates have the cognitive capacities to develop and use vocalisations in such a way, that such vocalisations have an important social function at an early stage, and that there is no issue with foundations of such behavioural capabilities having an ancient provenance. There is no reason for us to believe that any of our hominid ancestors could not have also exhibited such behaviours. Whether they would have had the selective pressure to do so is discussed in Chapters 6 and 7. If so, differences from species to species would presumably be in terms of the vocal range, control and capabilities, which would ultimately bear upon the complexity of the utterances possible, and normally used.

5.4. Evidence for the evolution of the neurology

How far back in our evolutionary history do we have to go to see evidence of these systems? There are several difficulties associated with making inferences regarding the development of hominid neuroanatomy from the archaeological evidence. Not least of these is that there is no *direct* fossil evidence for the nature of early human brains, because brains do not fossilise. Nevertheless, the fossil record can provide some valuable information.

5.4.1. The fossil record

Some light can be shed on the anatomy of early brains by the study of endocasts (see Figure 5.1). These are casts taken of the inside of fossil skulls, which can show the shape and features of the brain which occupied them (e.g. Kochetkova, 1978; Falk, 1992). The extent to which one may extrapolate about brain function from brain physiology is debated, however. Firstly, an endocast only shows the outer surface of the brain, as it is a cast of the imprint left by the brain on the inner surface of the skull; it can tell us nothing about the internal brain structure and little about connections between different areas (Leakey, 1994).

Figure 5.1: Fossil *Australopithecus* endocasts (*from Johanson and Edgar, 1996, p. 81*).

Endocasts can show the *shape* of the brain, though, and this has allowed the identification of development in Broca's area (important in language production; see Figure 5.2) of hominid brains (Holloway, 1983). This identification has led to the assertion that *Homo habilis* (skull KNM-ER 1470, from East Turkana) had some language abilities, as it shows evidence of development in that area of the brain (Holloway, 1983).

There are some problems with this inference, however. As discussed above (section 5.1), language ability is not located exclusively in Broca's area, but in many other areas too (e.g. Borchgrevink, 1982; Benson, 1985). Furthermore, areas around Broca's area are also used for processes other than language (Calvin, 1996). For example, complex sequences of muscular movements and fine motor muscular co-ordination of oral and facial muscles are controlled in this region (Calvin, 1996). This includes musculature necessary for production of facial expressions and vocalisations, as well as sequences of movements for rhythm production. Note that *vocalisations* are not the same thing as *language*. Thus development in Broca's area and the surrounding areas may only indicate the development of the ability to produce complex sequences and finely controlled vocalisations, and not have any bearing on language ability in the sense of syntax and grammar. In fact, Peterson *et al.* (1988) conclude from P.E.T. studies of the brain that the classic Broca's area (Brodmann's areas 44 and 45) is not concerned with semantic associations of words at all, but exclusively with motor functions; it seems to be the nearby Brodmann's area 47 that is responsible for semantic associations. Despite the fact that uncertainty about the precise functions of Broca's area may have led to the attribution of language abilities to hominids who probably had none, the identification of Broca's area in endocasts is still very useful to this study, as it gives important information regarding capabilities for vocalisation and fine vocal control.

Before about 2 Mya, fossil hominid brains show no development in the area of endocasts corresponding to Broca's area (Falk, 1992). The first evidence for any growth in this area this comes from KNM-ER 1470, *Homo habilis* (Falk, 1992). Between *Homo habilis* and *Homo ergaster* there is a disproportionate level of growth in this area, in comparison to the rest of the brain's growth (Kochetkova, 1978). While there is a general increase in brain size of around 50% between *Homo habilis* OH7 (at around 674cc; Johanson and Edgar, 1996) and *Homo ergaster* KNM-WT 15000 (estimated adult cranial capacity 909 cc; Johanson and Edgar, 1996), there is an "especially intense" growth of the lateral tuba, corresponding to Broca's area (Kochetkova, 1978, p. 200). This suggests that there was a significant development of fine motor control of vocalisation as part of the

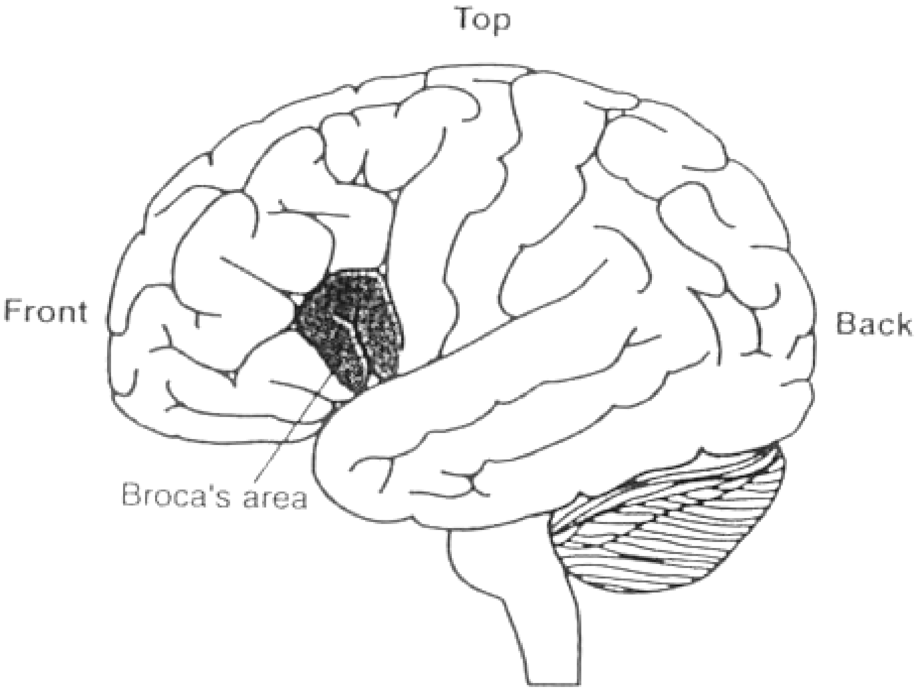

Figure 5.2: The location of Broca's area in the human brain (*from Carlson, 1994, **p. 6***)

emergence of the new species. By *Homo neanderthalensis* (Djebel Ihroud I, 1,305 cc; Holloway, 1981), this area is "very protuberent" (Holloway, 1981, p. 387).

5.4.2. Primate evidence and neurological connections

Other methods must be used to investigate connections between different areas of the brain, as these are unlikely to show up in endocasts. Studies of primate vocalisation neurology can inform as to phylogenetic origins of such neurological systems. Neuropathological studies have identified the anterior limbic cortex as being responsible for the production of voluntary emotional vocal expressions in all primates (Jurgens, 1992). Destruction of this area in humans does not affect the ability to produce involuntary vocal reactions to external stimuli but does result in an inability to *voluntarily* produce joyful exclamations, angry curses or pain outbursts. Without the use of the anterior limbic cortex voluntary vocal utterances, which would normally be highly emotionally communicative, sound more or less monotonous, with very flat intonation (Jurgens and von Cramon, 1982; Jurgens, 1992). This is an evolutionarily ancient area of the brain, fulfilling the same function in rhesus monkeys (and probably at least all primate species since), with the same effects when destroyed (Aitken, 1981; Kirzinger and Jurgens, 1982, reported in Jurgens, 1992).

On the basis of many years researching the neurological roots of vocalisation, Jurgens (1998) considers that "Squirrel monkey vocalisation can be considered as a suitable model for the study in humans of the biological basis of nonverbal emotional vocal utterances, such as laughing, crying and groaning" (p. 376), and much of his research has been carried out on these primates. Critical in the production of non-verbal vocal emotional utterances is the periaqueductal grey matter (PAG) of the midbrain, which acts as a critical relay station for such utterances (Jurgens and Zwirner, 1986). The PAG and the laterally bordering tegmentum of the midbrain collects the various auditory, visual and somatosensory stimuli that trigger vocalisation, as well as motivational-controlling inputs and volitional impulses from the limbic structures (mentioned above) and the anterior cingulate cortex, respectively (Jurgens, 1998). Interestingly, recent findings show that the PAG is only implicated in the origination of the limbic *nonverbal emotional* vocal utterances, and not *learned* utterances with a neocortical origin; a second neurological pathway not involving the PAG seems to be responsible for learned vocal patterns (Jurgens and Zwirner, 1998), although these would also be integrated at the nucleus ambiguus where control of the vocal folds occurs (Vanderhorst *et al.*, 2001).

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

So it seems that these foundations for fine laryngeal and orofacial control necessary for vocalisations have a long phylogenetic history, and that these foundations have since been built upon during the *Homo* lineage. Jurgens (1992) reports that in contrast to the primates studied, modern humans also have a direct connection to the nucleus ambiguus (site of the laryngeal motoneurons) from the primary motor cortex (responsible for planned motor control) (Jurgens, 1976; Jurgens, 1992). This connection is not present in monkeys, and probably not in any non-human mammals (Jurgens, 1992), and would seem to be what allows humans to carry out planned structured vocal utterances. Such an ability is not present in other higher primates. As Lieberman (1994) reiterates, although apes can produce many vowel sounds and phonetic features of human speech, these are "bound" into stereotyped calls based on affect, and they cannot produce novel voluntary motoric sequences. The evidence considered in Chapter 4 concerning cervical and thoracic vertebral innervation suggested that controlled vocalisations of extended duration became possible between *Homo ergaster* and the common ancestor of *Homo neanderthalensis* and *Homo sapiens*. It seems reasonable to suggest that primary motor cortex - nucleus ambiguus connection for planning such utterances developed at the same time.

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

5.5. Chapter summary and conclusions

In summary, there are a variety of sources of evidence for the development of neurological capacities relating to vocalisation and music. In terms of the fossil record, development in Broca's area and the surrounding areas suggests the development of the ability to produce complex sequences and finely controlled vocalisations. The first evidence for any growth in this area this comes from KNM-ER 1470, *Homo habilis* (Falk, 1992). Between *Homo habilis* and *Homo ergaster* there is a disproportionate level of growth in this area, in comparison to the rest of the brain's growth (Kochetkova, 1978), suggesting that there was a significant development of fine motor control of vocalisation as part of the emergence of the new species. By *Homo neanderthalensis* (Djebel Ihroud I, 1,305 cc; Holloway, 1981), this area is "very protuberent" (Holloway, 1981, p. 387).

It has already been seen that there was a gradual increase in the variety of sounds that could be produced vocally with the lowering of the larynx in the throat, from *Homo ergaster*, until modern capabilities were reached by *Homo heidelbergensis*. It seems likely that these two phenomena are connected, and that there was a 'bootstrapping' (self-perpetuating) effect between them. As the vocal anatomy evolved, allowing an increasing range and diversity of utterances, the neurology also developed allowing increasingly complex planned vocal muscular sequences.

Laryngeal muscle control is thought to be mediated by a neurological pathway from the PAG via the nucleus retroambiguus to the nucleus ambiguus. This pathway is critical to the production of vocalisations; as well as being responsible for integrating vocal fold control, the nucleus ambiguus is also crucial for expiratory control, orofacial muscular control and overall control of the laryngeal system (Jurgens, 1998), and is directly adjacent to Broca’s area (Carlson, 1994). All primates have the ability to make voluntary emotional vocal utterances, and these are controlled by a specialised area of the brain; what separates us from other primates, however, is the neurology which allows vocal behaviour involving voluntary control over the *structure* and *complexity* of vocal utterances, and the degree of the social functionality of these. Only humans have the neurological connection allowing them to use their vocal apparatus for extended *planned sequences* of such utterances. Our ability to make planned emotional vocal sequences, including melody, developed alongside the ability to produce an increasing range of sounds; the ability to actually plan and construct melody was probably proportional to laryngeal and thoracic canal development.

The greater the analytical and temporal ordering demands of either music or language, the greater the involvement of the left hemisphere, and the greater the demands in terms of melody, tonal changes and prosody, the greater the involvement of the right hemisphere. It seems likely that the production and processing of rhythm and meter in speech and of rhythm and timing in music use related neurological mechanisms, in the left hemisphere and in the reticular and emotional limbic systems. On the other hand, the prosody of speech, pitch perception, melodies and emotional sounds seem to be controlled and processed by neural substrates that are less unilaterally localised, although much evidence suggests that important elements of these functions are predominantly controlled by the right hemisphere. These mechanisms probably predate those in the left hemisphere, and the evidence suggests that the specialised human neurology dedicated to the more analytical “higher” linguistic and musical functions emerged later out of a set of common neurological substrates concerned with vocal emotional expression and comprehension. It could be that the emotional limbic system and the IMF proposed by Trevarthen (1999) (see

Chapter 6) constitute the uniting neurology between the tonal emotive vocalisation elements and the rhythmic-motor elements of interaction and musicality.

There appears to be a great deal of overlap between speech vocalisations and singing vocalisations in young children. This is not to suggest that ontogeny is recapitulating phylogeny, but highlights once again the shared elements of such vocalisations. It seems likely that these capabilities provide the foundations for the more specialised, culturally determined, acquisition of specific language and musical abilities later.

Studies of infant capabilities and their development show that the capacity to perceive melody and rhythm and tonal vocalisation and to gain affective information from them is inherited and innate at birth. Such vocalisations have an important social function at an early stage in humans; social use of vocalisation is prevalent in other primate species and especially important for infants in at least some, suggesting that the social use has been a fundamental driving force in the development of vocalisation control and complexity. There is no reason for us to believe that any of our hominid ancestors could not have also exhibited such behaviours; differences from species to species would be in terms of the vocal range, control and capabilities, which would ultimately bear upon the complexity of the utterances possible, and normally used. Evidence considered in Section IV considers the nature of these vocalisations, and proposes reasons why such behaviours would have been important.

SECTION IV

EVOLUTIONARY RATIONALES FOR MUSIC

> "Music is our oldest form of communication, older than language or art; it begins with the voice, and with our overwhelming need to reach out to each other".
> (Menuhin, 1979)

It was possible in Chapters 4 and 5 to determine when various important neurological and physiological changes important to the use of music occurred. Whilst the neurological evidence has provided some clues as to the functions that these capacities seem to fulfil, especially through the ways in which they are interdependent, more investigation is required to build up models of the behaviours that these capacities supported and why they might have developed. This section consists of three chapters. Chapter 6 unites evidence from a great diversity of disciplines to gain insight into the complex interrelationships of the various capacities that underlie musical behaviours, and how their development relates to that of social behavioural complexity and linguistic capacities. Chapter 7 examines rationales for the evolution or perpetuation musical behaviours as a whole. Chapter 8 concludes this section by looking at how the development of musical capacities fits within some existing models of human cognitive evolution.

CHAPTER 6

Rationales for the Evolution of the Capacities for Music

As discussed in the preceding chapters, the fact that music and language use the same physical mechanisms for production and perception - namely, the vocal apparatus and the auditory apparatus – has led to various proposals that language and music are closely related in terms of their origins. Some authors believe language to be the precursor of music, and that music makes use of mechanisms originally selected for by language (e.g. Pinker, 1997; Calvin, 1996) and some vice versa (e.g. Livingstone, 1973; Vaneechoutte and Skoyles, 1998*)*. A few others believe that neither of these is the case, but that language and music have a common foundation (e.g. Scherer, 1991; Brown, 2000).

The evidence examined in Chapter 4 showed that human ancestors gradually evolved the apparatus necessary to make complex vocalisations over a period of around 1.5 million years, between *Homo ergaster* and *Homo heidelbergensis*. By 300-400,000 years ago, *Homo heidelbergensis* had a brain of 1,220cc (Petralona 1, Johanson and Edgar, 1996, p.200) and a vocal apparatus that was essentially modern. It is not generally suggested that *Homo* had modern language by 300,000 years ago, however, in terms of developed lexical complexity and syntax; in fact, the predominant view has traditionally been that there is little evidence for fully modern language use before we start to see cultural artefacts produced by anatomically modern humans (e.g. Mithen, 1998). If they did not have modern language or symbolic thought, what were these ancestors to modern humans, and modern humans themselves, using their large brains and fully developed vocal tracts for? What was the selective pressure responsible for vocal ability being fully developed perhaps 300,000 years before it was used for fully-developed language?

This chapter draws upon a great diversity of evidence from different disciplines in its aim of building up a picture of the interrelationship and co-evolution of the various capacities that underlie musical behaviours. The first section consists of an investigation of the development of vocal complexity and versatility, including rationales for the evolution of the vocal tract, and relationships between emotive content in linguistic and non-linguistic vocalisation, including in infant-directed speech. The next part of the chapter looks at social vocalisation and the origins of complex and melodic vocalisation. This section goes on to propose and discuss rationales for a shared heritage between elements of speech and music. Part three consists of a detailed discussion of emotional expression and elicitation by music, and social, empathic and contextual factors in emotional content in music. Section four of the chapter looks closely at the interrelationship of vocal production and control of corporeal and manual movement, and the relationship between vocalisation and gesture, through a review of evidence from speech disorders and developmental psychology. This then leads to an examination of the possibility of an evolutionary link between complex vocalisation structure and gesture. The final, fifth, part of the chapter consists of an investigation of the relationships between rhythm, corporeal control and emotional expression and evocation. All in all the aim of this chapter is to explain the complex interrelationships of these fundamental elements of musical behaviour, how they came to constitute musical behaviour, and to identify their development in relation to linguistic and social capacities.

6.1. The evolution of vocal versatility and complexity

There have been several explanations suggested for the development of the versatility of the modern vocal tract, coming from research in a variety of disciplines, including primatology and zoology, developmental psychology (child development), palaeontology, linguistics and neurobiology.

6.1.1. The vocal tract

Lieberman *et al.* (1969, 1972) proposed that the evolution of the modern form of the vocal tract was selected for to support articulate speech, and this proposal remains very influential. On the basis of developmental research, Dissanayake (2000) has suggested that vocal versatility and perception arose to support infant-directed (ID) speech, and Fernald (1992b) has suggested that modern vocal capabilities are derived from those originally used for ID speech (these ideas are discussed further below). Following studies of primate and other mammalian vocalisation, Fitch (1999, 2000a, 2000b) has pointed out several important areas of comparison between the vocalisations of humans and other animals, including an ability in some animals to lower their larynx (temporarily) during vocalisation. This has led him to suggest that the lengthening of the vocal tract through laryngeal lowering might have taken place to lower the formant frequencies produced. Each of these rationales is discussed in the following section.

In Fitch's (2000a) opinion, the best explanation for the permanently lowered larynx in humans is still provided by the hypothesis of Lieberman *et al.* (1969, 1972) (and many others since, it should be noted): that it is used for producing articulate speech. He proposes incorporation into this hypothesis of the ability, at an early stage, to temporarily lower the larynx to produce lower formant frequencies and greater tongue versatility; he then suggests that in a species that spends a lot of time vocalising, it may be more energy efficient to have a permanently lowered larynx than to have to lower it before each vocalisation (using the strap muscles, sternothyroid and sternohyoid). A low larynx has implications for tongue control (a fact relevant to all discussions of vocal tract development, not just this particular scenario). The low position of the tongue root concomitant with a low larynx allows for greater movement and control of the tongue body to produce sounds in the oral cavity. In Fitch's words, "…having the skeletal support for the tongue body in a permanently low, stabilised position may provide significant advantages for producing the rapid, precisely controlled vocal tract movements that characterise modern human speech" (Fitch, 2000a, p. 215). Formant frequency, on the other hand, determines the characteristic pitch of vowel sounds; the sound quality of a vowel is normally determined by a combination of formant frequencies. In humans, dogs and monkeys, vocal tract length is positively correlated with body size, and in turn, formant frequencies are closely related to vocal tract length (Fitch, 2000b). Low formant frequencies are produced by large animals with long vocal tracts, and formants play a role in individual identification, being a sort of vocal signature, and certain animals have the ability to temporarily lower their larynx during vocalisation to produce lower formant frequencies (see below).

In other words, Fitch suggests that a capacity to lower the larynx temporarily to produce lower formant frequencies and tongue control could have formed an important early component of articulate vocalisation, and if the larynx became permanently lowered this would have been more efficient in a vocally very active species. For this rationale to hold together, it requires that low-formant-frequency vocalisations with a temporarily lowered larynx constituted a selectively important component of the vocalisations of a highly vocally active species, before sophisticated control of pitch contour and vocal control in the supralaryngeal tract developed. What it is important to remember is that the temporary lowering of the larynx by the strap muscles does not allow a greater *range* of controlled vocalisations to occur – it simply allows a *different*, *additional* vocal sound to be made. In both cases this sound could be moderated by tongue and orofacial movements, but this is not the same as an increase in range of tonal vocalisations. The above rationale requires the existence of a hominid ancestor which was very vocally active using this facility, prior to the development of a wide vocal range, versatility and control. So in this scenario a lowered larynx essentially formed a pre-adaptation (exaptation) to complex controlled vocalisations in a creature that was already very vocally active.

All this leads Fitch (2000a) to propose a three stage process for this development:

> “In stage one, some early ancestor used a standard mammalian vocal gesture to produce calls, but introduced tongue body perturbations during larynx lowering to produce a wider range of formant patterns (and hence a greater diversity of discriminable ‘calls’ or phones)… In stage two, the dual use of degrees of freedom of the tongue body was consolidated into the communication system, with a variety of vowel-like sounds and formant transitions being produced. However, these would be made with a temporarily lowered larynx, and the larynx would be returned to the nasopharynx during resting breathing. Finally, in stage three, the larynx would have assumed a permanent low resting position during ontogeny, as it does today, giving these hominids less effortful speech and perhaps more vocal control” (Fitch, 2000a, p. 216).

This seems a very plausible scenario for the beginnings of the process of vocal tract evolution, and the early stages of complex vocalisation. Into stage one of this model Fitch has subsequently integrated his “size exaggeration” hypothesis (Fitch 2000b), dealt with below, although it is by no means necessary to the model, with the other arguments of increased tongue control and tonal range. The vocalisations outlined in stage two would have been of limited duration; this fits with the fossil evidence described in Chapter 4 and neurological evidence of Chapter 5 which suggests that the ability to plan and control vocalisations of extended duration is something that arose after the beginnings of vocal versatility in *Homo ergaster*. However, as pointed out above, the model provides little explanation of the development of the fine tonal and pitch control, muscular control and complex vocalisations that use the whole of the supralaryngeal tract; these are the subject matter of the subsequent sections. These elements of vocal sophistication would have to have developed during and subsequent to stage three of Fitch’s proposed scheme; in this respect the model accounts for the beginnings of some articulatory control but can provide an account only of the earliest stages of vocal tract evolution and pitch control. It is at stage three of the model that fine control of vocal tone could become important and selected for, and although Fitch’s model seems to imply that the lowering of the larynx occurred in one fell swoop, the evidence considered in Chapter 4 belies this; further lowering of the larynx would also have occurred subsequently, with concomitant increasing control over the necessary musculature.

As noted, Fitch (2002b) has elaborated on the first stage of his model by suggesting that the selective force for the initial lowering of the larynx may have been to create an exaggerated impression of size through lower frequency vocal formants. A lowered larynx increases vocal tract length and decreases the formant frequencies produced, allowing an animal with a lowered larynx to duplicate the vocalisations of a larger creature, thus exaggerating the impression of size conveyed by its vocalisation. Fitch outlines several precedents for the size-exaggeration role of vocalisations elsewhere in the animal kingdom. Many birds have elongated tracheae, which lowers the formant frequencies that they produce. Fitch (1999, 2000b) suggests that this might be highly effective at exaggerating apparent size “at night or from dense foliage” (Fitch, 2000b, p. 264). Another example, this time in mammals, is provided by red deer and fallow deer males. These have the ability to lower their larynx temporarily to produce their loud roar with low formant frequencies, used to intimidate rivals and impress females.

These observations lead Fitch to suggest that this ability might have been the principal selective influence on the evolution of a lowered larynx in hominids, and that our primate ancestors might have used formant frequencies as an estimate of body size from vocalisations. He then suggests that this might have formed a pre-adaptation for "vocal tract normalisation", "…a crucial feature of speech perception whereby sounds from different sized speakers are normalised to yield equivalent percepts" (Fitch, 2000b, p. 263). He does not elaborate, however, as to how a physiological feature that was supposedly selected for in order to distinguish and differentiate individuals might have formed a pre-adaptation for a crucial feature of human speech *perception* which apparently acts to the opposite purpose, standardising a percept. This seems a significant problem with this argument for this form of vocalisation as a precursor to human speech vocalisation.

Whilst illustrating that size exaggeration vocalisation apparently has precedents in the animal kingdom, the examples given provide little analogy with human vocalisation, or for how the elements of vocalisation which are exclusive to humans must have developed. In the case of the birds (and perhaps with the size exaggeration hypothesis in general), Fitch's assertion that size exaggeration might be especially effective at night or in dense foliage should perhaps read that it might *only* be effective at night or in dense foliage (i.e. in situations of limited visibility, when the vocaliser can not be clearly seen). The first evidence for lowering of the larynx in hominids occurs with *Homo ergaster*, at a time when, it is thought, hominids were moving away from arboreal habitats and into more open environments. If size exaggeration were the selective force behind laryngeal lowering one would expect to see it at a time when hominids were living in habitats of limited visibility, not moving away from living in such environments. It is conceivable that size exaggeration might have been useful at night in intimidating predatory animals on the savannah, but all the examples of size exaggeration in nature given by Fitch (1999, 2000b) confer sexual selection advantages in mating and contest with rivals.

In the case of the laryngeal lowering of the red and fallow deer bulls in their roar to intimidate rivals and impress females, this illustrates a further potential use of lowered formant frequency. However, the mechanism is no analogy for the human vocalisation ability – the deer can only lower its larynx temporarily to roar, whereupon it returns to its original position. It has no control over the intervening frequency ranges. In contrast, the human larynx is permanently lowered, and we have very fine control over the range of frequencies it is possible for the larynx to produce.

It may be that Fitch is correct in suggesting that the ability to temporarily lower the larynx formed a *precursor* to a permanently lowered larynx, and there is a variety of evidence to suggest that formant frequency is still used as an indicator of body size by humans (Fitch, 2000a). In this sense, the size exaggeration hypothesis provides an interesting model for vocal tract use in our ancestors which has otherwise received little attention, and describes a potential *additional* role of the vocal tract. It also provides a possible explanation for the second lowering of the vocal tract at puberty that is especially prominent in adolescent males, an effect of hormonal change at sexual maturity (Fitch, 2000a, 2000b). However, given that vocal frequency is generally an indication of body size in all species, one would expect to see a sexual difference in vocal range emerge at the age where sexual dimorphism of body size becomes prevalent in humans. Admittedly the lowering of the male voice at puberty is quite marked, but it is perhaps enough to suggest that this is an exaggerated signal of body size, rather than a signal exaggerating body size; i.e. it functions to remove ambiguity about body size (and sex), rather than to mislead about body size.

The size-exaggeration hypothesis seems to have little to tell us about the evolution of the human-exclusive properties of the vocal tract. It provides no explanation, or logical path, for the evolution of fine vocal control of the wide diversity of frequencies that the human larynx is able to produce, or how the larynx came to be permanently lowered; the latter is accounted for by the 3-stage model proposed by Fitch (2000a) (above), but the former is not. It also fails to explain why such fine control of intervening vocal range apparently co-evolved with the lowering of the larynx (as suggested by the evidence considered in Chapter 5), and why we are so sensitive to the prosodic content (i.e. frequency variations) of utterances. It is these properties of the vocal tract (and their evolution) with which this thesis is principally concerned.

If increasing versatility and control over complex vocalisations is taken as the main selective force for the development of the vocal capabilities of melody and speech that we now exhibit (as is overwhelmingly suggested by the neurological and developmental evidence considered in Chapter 5), it follows that the muscular control (and cognitive capacity) necessary must have evolved in tandem with the vocal tract. What would have been the driving forces for the development of such control?

6.1.2. Infant-directed speech, music and vocalisation

As a form of vocalisation which exaggerates the prosody of normal speech (Trainor *et al.* 2000), infant-directed (ID) speech (or “motherese”, or “parentese”) has received an increasing amount of attention in recent years, in the context of language evolution, the development of linguistic capacities, and the evolution of music (e.g. Fernald, 1992a; Fernald, 1992b; Fernald, 1993; Werker *et al.* 1994; Papousek, 1996; Lewkowicz, 1998; Dissanayake, 2000; Mang, 2000; Trainor *et al.*, 2000; Falk, *in press*).

ID speech is used most intensively toward children between the ages of 3 and 5 months, but continues to be used up to around the age of 3 years (Falk, *in press*; Stern *et al.*, 1983), and is characterised by particular frequency ranges and contours. It has a higher frequency (F_0), and thus pitch, than adult-directed (AD) speech, pitch contour is exaggerated in utterances which have varied pitch, it has a larger pitch range, slower tempo, and is more rhythmic than typical AD speech (e.g. Fernald, 1992b; Trainor *et al.*, 2000). In particular, when soothing an infant, infant-directed vocalisations have a low pitch, with a falling pitch contour, when engaging attention and eliciting a response, a rising pitch contour is usually used, and when attempting to maintain the attention of the infant, a bell-shaped contour is usual (Fernald, 1992b). These properties have a high correlation across cultures, and this use of exaggerated vocal range and stereotyped patterns seems to be a universal of human parental behaviour, having been observed in cultures in Europe, Africa and Asia, in tonal and non-tonal languages (Greiser and Kuhl, 1988; Fernald, 1992b; Fernald, 1993; Werker *et al.* 1998; Kitamura *et al.*, 2002). For example, English and Mandarin mothers use very similar ID speech to speak to their babies (Greiser and Kuhl, 1988), and English and Cantonese infants aged four and a half to nine months both showed preference for ID over AD speech, irrespective of whether it was coming from English or Cantonese speakers (Werker, *et al.*, 1998); these findings suggest a particular significance to the properties of ID speech, irrespective of cultural origin.

In addition to these attention-gaining or -maintaining vocalisations, there is also great consistency across cultures in the form of ID vocalisations communicating approval, disapproval/prohibition, or calming/soothing; essentially emotive vocalisations communicating the affective state of the speaker, or attempting to elicit such a state in the receiver. Across English, French, German and Italian, these are highly consistent in form (Fernald, 1992b). Infant-directed vocalisations of approval have a high mean F_0, a wide F_0 range, and a rise-fall contour. Prohibition and disapproval vocalisations, in contrast, have a low mean F_0, narrow F_0 range, are short with an abrupt onset. Comforting vocalisations also have a low F_0 and narrow F_0 range, but they are longer, have a softer onset, and are described by Fernald as having a “legato” quality.

At a rudimentary level, one would expect precisely this difference between prohibitive and calming vocalisations, in terms of onset and duration: the former takes the form of a sudden, arousing, shocking sound, eliciting a state of sensory arousal which must be similar to that elicited by any loud and unexpected noise from the environment, and when coming from the infant’s own parent is perhaps particularly potent as a consequence of coming from a source normally associated with comfort and security. The latter, calming vocalisations, act in exactly the opposite way: gentle onset would result in little autonomic arousal, and a narrow frequency range over a longer duration would maintain this low state of positive arousal with no unexpected stimuli. A continual sensory input of a consistent form is comforting precisely because it contains nothing unexpected, and the fact that it is coming from a source already associated with such calm and security must add to this effect.

As noted, accompanying onset and duration in importance are cues from contour and average pitch. These properties of exaggerated contour and raised frequency seem to fulfil important roles in addressing the initial limits of an infant's perceptual and cognitive abilities. Infants are more sensitive to sounds at higher pitch ranges, and continuity in pitch contour is an important cue in attending to a single speech source; exaggerating these contours makes it easier for an infant (who cannot rely on linguistic content for this purpose) to attend selectively to the sound source (Fernald, 1992b). The signal is thus "high in perceptual salience and relatively easy to process" (Fernald, 1992b, p.419).

The overall F_0 of the utterance is clearly also important in communicating affect, with high frequency consistently being associated with positive sentiment, such as expressing approval and stimulating and maintaining attention. This is consistent with the findings of Scherer (1985, 1986) that increases in mean F_0 and F_0 range are typical of vocal expressions of enjoyment and happiness. In contrast, F_0 decreases, especially with a harsh voice, are typical of expressions of irritation and anger. This appears to be consistent not only across human cultures, but across various other species too. Morton (1977) found that across a diversity of animals high tonal vocal sounds were associated with appeasement, submission, friendliness or fear whereas low sounds were associated with threat. This has led Trainor *et al.* (2000) to suggest that another factor in the higher frequency of ID vocalisations might be that vocalisations directed towards an infant naturally underplay aggression.

Whilst various reasons might be suggested for this association between pitch and level of aggression, there is a significant factor influencing the frequency produced by the action of the vocal chords, which does not rely on any external environmental explanations: facial expression.

Orofacial musculature has a fundamental role in shaping the upper vocal tract and modulating the output of the vocal chords (see Chapter 4), in all speech, ID and AD. It is also responsible for the control of facial expression. It is thus perhaps unsurprising (though highly significant) that there is a correlation between given facial expressions and characteristics of vocalisations made at the same time. Facial expressions of disgust, happiness, sadness and anger are not learned and culturally determined, and the capacity to produce and recognise them seems to be innate and universal (e.g. Ekman and Friesen, 1971; Ekman, 1980; Carlson, 1994). In humans, Tartter (1980) showed that vocalisations made whilst smiling were of a higher mean frequency than those made with neutral expression due to the effect of smiling on altering the shape of the upper vocal tract; smiling increases second formant frequency (Tartter and Braun, 1994). Tartter and Braun also showed that vocalisations made whilst frowning had lower formant frequency and longer vowel duration than speech with a neutral expression; listeners were able to discriminate speech made whilst frowning, smiling and with neutral expression from each other with no visual input. Furthermore, they were able to do so in both vocalised and whispered speech, suggesting that the same effects of facial expression on elements of affect of vocalisation would be applicable to vocalisations even in the absence great vocal chord versatility (which accords well with Morton's (1977) findings regarding vocalisations in other animals). Given the universality of certain fundamental facial expressions and the correspondence between these and characteristics of vocalisations, we can also expect characteristics of particular emotional vocalisations to be universal and innate too.

This correlation between facial expression and vocal quality also apparently has an ancient provenance, being shared by our nearest relatives. Chimpanzees also frequently couple particular vocalisations with particular emotional facial expressions, the vocalisations being moderated by alterations of the size and shape of the mouth, and thus the resonating upper vocal tract (Falk, *in press*). Amongst bonobos, utterances always occur along with facial expressions, gestures and tactile communication (Falk, *in press*; Bermejo and Omedes, 1999), which is true for human ID speech too (Falk, *in press*). Humans also use facial affect and vocal affect to inform judgement about the affective content of each other; they seem to be interdependent systems in both production and perception (DeGelder and Vroomen, 2000), with the former having had a significant impact on the development of the nature of affective communication in the latter.

It is very important to reiterate at this stage that infant-directed vocalisations are not something completely different from other vocal behaviour, they are a variation of it. Whilst ID speech is easily distinguishable from AD speech, especially non-emotional AD speech, the prosodic contours of ID speech are an exaggerated version of those that occur in normal speech, and are not something independent and unique (Trainor *et al*. 2000). Trainor and colleagues demonstrated that ID speech essentially reflects free expression of emotion, in comparison with typical AD speech, in which emotional content is more inhibited (although still present): “When AD speech does express emotion, the same acoustic features are used as in ID speech” (Trainor *et al.*, 2000, p. 188). They found that the emotions “love-comfort”, “fear” and “surprise” were equally discriminable across ID and AD samples, with few differences between the properties of ID and AD samples, but significant differences between emotions expressed. Observations such as “through the melodies of the mother’s voice, infants could gain early access to her feelings and intentions” (Fernald, 1992, p. 423) are equally applicable to adult speech prosody, for adult listeners. The difference is that ID speech is tailored to the emotional and perceptual needs of infants. The fact that the prosodic features of vocalisations (in particular pitch, pitch contour and tempo) are exaggerated in ID speech allows infants to attune to them better (see above), but they are not fulfilling a role fundamentally different from the role of prosody in AD speech.

Falk (*in press*) proposes that since, in contrast to chimpanzee mothers, “human mothers continually produce affectively positive vocalisations to their infants”, this behaviour derived from an initial evolution of prosodic and instructional vocalisations in early hominin mothers. Falk argues that as hominin babies were born increasingly early (due to increases in cranial size), and thus increasingly helpless, they would have been unable to cling to their mothers (as is the case with modern human infants). They would thus have relied on their mothers to carry them, which would have left the mother unable to use their hands to carry out the manual tasks in which they would normally have been engaged. Hominin mothers would thus have often been required to temporarily put the infant down whilst engaging in their manual tasks.

For the infant that has been put down and separated from its mother, emitting prosodic emotional cries would be a good way of attracting attention of the mother, and eliciting care. From the point of view of the mother, uttering emotive vocalisations to reassure the infant that the mother is near and solicit its attention would also be beneficial, certainly in stemming its cries, the utterances being “disembodied extensions of mothers’ cradling arms” (Falk, *in press*).

From this scenario, Falk argues that this situation of “parking the infant” led to the development of the characteristics of ID speech, and that this form of vocalisation then formed the foundation for the development of later, complex linguistic vocalisations; “Over time, words would have emerged in hominins from the prelinguistic melody (Fernald, 1994: 65) and become conventionalised” (Falk, *in press*).

Whilst Falk’s scenario of the necessity of setting infants down seems likely, and emotive vocalisations would almost certainly have become increasingly important in this situation (as they are amongst other primates when infants are unintentionally separated from their mothers), there is no reason to believe that these should have formed the *basis* for affective vocalisations. Likewise it seems quite reasonable that protolanguage and then more complex language emerged from affective vocalisations (an idea explored in much greater detail below), but no reason to believe that it was from this particular situation. Almost all of the situational examples that Falk uses might apply equally well to interaction between *any* individuals within a hominin group. For example, the idea that distance between emotionally conjoined individuals would elicit emotive vocalisation might apply equally to any removal of physical proximity in a species which is used to relying on grooming activity to build and reinforce affective relationships. The foundations for such vocalisations are used across all members of a group, not just between mothers and infants, in higher primates and modern humans, as has been illustrated in the preceding sections, and discussed further below.

Even the premise that human mother-infant interaction is unusual in that it is characterised by almost continual affective vocal interplay is overstating the case of mother-infant interaction relative to that of other interpersonal interaction. It is a challenge to consider any situation where *any* two emotionally conjoined humans would sit in silence not verbally interacting (with the possible exceptions of reading, watching television, or in a theatre, all of which are very recent developments in evolutionary terms; even in these situations it is relatively unusual to sit in silence, and it could be argued that the characters in the book, program or play are providing surrogate interaction anyway). It might be counter-argued that this is because our adult vocalisation behaviours are derived from these mother-infant interactions, but considering that these themselves appear to be derived from vocal behaviours which span all members of higher primate groups, this seems a less likely explanation than that emotive vocalisations of all forms are derived from interpersonal vocalisations between all members of a group. It seems more likely that mother-infant vocalisations are a specialised perpetuation of a sophisticated form of non-linguistic interpersonal interaction which was used earlier between all individuals.

This said, the current author believes that the evidence that Falk cites, and that investigated here, concurs with her penultimate paragraph, which is not specific to ID vocalisations:

> "It is reasonable to speculate that by the time individuals across social groups began to originate and conventionally share simple instructive utterances, protolanguage was in the process of emerging from the prelinguistic melody. Whatever its precise nature, however, protolanguage and the other languages that

> eventually evolved would, forever after, retain some of that melody. Thus, rather than being totally separate from language (Burling, 1993), tone of voice represents a signature from its very origin that, as transpired for the cosmic microwave background signature left over from the big bang, should be recognised and investigated" (Falk, *in press*).

So what significance does all this investigation of protolanguage and prosody have for musical evolution? Terms such as melody, tempo and pitch are conspicuously prevalent in descriptions of prosodic affective vocalisations and ID speech, and the evidence analysed in Chapter 5 illustrated that prosody and melody are produced and processed by closely related or shared neural structures. Tellingly, ID singing also appears to be a human universal, with shared characteristics (such as gentle dynamics, gradual pitch contours and repetitive motifs) across cultures (Trehub *et al.*, 1993); lullabies and playsongs also convey emotional rather than linguistic meaning, and ID speech and singing share many of the same features and characteristics, in terms of pitch and tempo (above; Fernald, 1989b; Papousek *et al.*, 1991; Trainor *et al.*, 1997; Trehub *et al.*, 1993).

A particularly interesting observation regarding ID speech is its level of apparent "redundancy", namely repetition. In the words of Fernald (1992b), "…mothers repeat over 50% of their phrases when interacting with 2-month-old infants. Prosodic repetition is common too (Fernald and Simon, 1984), often with slight melodic variations, which keep these repetitive runs interesting as well as highly predictable for the infant (Stern, 1977)" (p. 419). Of course, because of this, the repetition is actually in no way "redundant", but instead important in fulfilling the role of the vocalisation, namely, the communication of affect, modulation of arousal, and eliciting of attention and affective response. These properties of ID speech, repetition with variation, along with highly variable pitch contours and high rhythmicity, are characteristic features of the melodic content and structure of music too. It is perhaps unsurprising that music frequently has these same effects on the listener/participant.

It is from these parallels that much can be gained regarding the origins and foundations of musical behaviour. What it is important to note is that these are parallels not just with ID speech but with the prosodic affective elements of all emotional vocalisation, whether accompanied by linguistic structure or not. Interestingly, Trainor *et al.* (2000) observe that "Because the emotions music conveys to, and induces in, listeners constitute the meaning of music (Meyer, 1956), speech that adopts musical features might also be expected to be good at communicating emotional information" (p. 194). Leaving aside other definitions of meaning and emotion in music for the moment (dealt with in section

6.3, music and emotion), in considering that (at least some) speech *adopts* musical features, Trainor *et al.* evidently consider this type of speech to post-date these elements of music. Consensus is that these tonal elements of speech certainly pre-date the linguistic, syntactic elements. Given the precedents for the importance of tone and affect in many other primate vocalisations, perhaps Trainor *et al.* should really be read that speech *shares* musical features, rather than *adopts*. Whilst the vocal interaction between parent and infant is often described as "musical" (e.g. Fernald, 1992a; Papaeliou and Trevarthen, 1998; Trainor *et al.*, 2000) Lavy (2001) points out that

> "what the interplay between mother and baby is actually very much like is positively valenced, emotionally-charged vocal utterance; it is referred to as musical because *many musical sounds are very much like the vocal utterances that humans have an innate ability to produce and perceive* [emphasis in original]. Moreover, parameters along which music operates are the same as those which encode emotional cues in acoustic signals and are perceived and quite probably responded to as such." (p. 51).

It seems that the best explanation for the above phenomena is that the shared prosodic pitch and tempo-related properties of emotional speech (ID and AD) and music are not borrowed from one to the other, in either direction, but are, and always have been, a shared fundamental component of both.

Dissanayake (2000), like Papousek (1996), has proposed that the musical perceptual abilities of infants were selected for because they are used in the non-verbal elements of vocal communication between mother and infant. Whilst ID speech does have lexical (linguistic) content, which plays a role in the linguistic development of the child (Burnham *et al.,* 2002), this seems to be secondary in role and motivation to the prosodic elements. As discussed above, these utterances rely on tone, inflection and intonation, and have a wide vocal range; their principal role is to communicate information about well-being, emotional status, needs, approval and intention, as is evident from their cross-cultural, pan-linguistic similarity. Dissanayake hypothesises that this mother-infant behaviour emerged as a consequence of hominid infants being born progressively earlier as brain size increased, and thus more helpless, increasing the need for emotional conjoinment at this early stage of life (presaging Falk's arguments, above, by several years). This would, she suggests, have become particularly important at the time of archaic *Homo sapiens* (*Homo heidelbergensis*), which exhibited a large increase in brain size. She proposes that these vocalisations then came to be adopted increasingly into daily life and ritual, eventually becoming the basis for musical behaviour, because they "were found by evolving human groups to be emotionally affecting and functionally effective… to emotionally conjoin and enculturate the participants" (Dissanayake, 2000, p. 401).

The benefits of such social-emotive interaction between individuals (emotional conjoinment and enculturation) is well-argued and supported by much of the evidence considered here, and the connection between musical behaviours and affective vocalisation is an important one (as illustrated by the evidence considered above and in Chapter 5). However, it is a little difficult to envisage hominids beginning to use a behaviour in other situations which was originally evolved as a "baby-talk", because they realised it "promoted affiliation and congruence in adult social life" (Dissanayake, 2000, p. 401). As discussed previously, there are precedents for the importance of affective-communicative elements of such vocalisations in adult vocalisations, as well as in primates without language (see below).

Additionally, the fact that the vocal tract of *Homo heidelbergensis* was effectively fully modern (see Chapter 4) would suggest that selective pressure for its use for wide-ranged tonal expression had come much earlier than Dissanayake proposes. This form of communication might gradually have evolved into music as Dissanayake suggests, or at least provided shared foundations, but it could also have been the basis for language amongst all of a population (c.f. Aiello and Dunbar, 1993; Scherer, 1991; Brown, 2000; below). It may be that the use of this mode of emotive utterance for communication between (pre-linguistic) infants and adults today is a vestigial use of this earliest form of social communication; after all, it is potentially still selectively useful at least until the child gains full linguistic abilities, and is still inherent in tonal elements of *any* emotional vocal communication (Trainor *et al.*, 2000). Thus, as mentioned above, an alternative, and perhaps more parsimonious, explanation to Dissanayake's suggestion is that this (social-emotive vocalisation) was a form of communication that came to be used *throughout* the social group at a much earlier time, without preference, both adult-adult and infant-adult, but is now perpetuated, in this predominantly non-lexical form, in adult-infant interactions, and in the prosodic content of adult speech.

A shared heritage between this type of vocalisation and singing-type behaviours is further suggested by the observations of Mang (2000). Mang's study of the vocalisations of pre-school children over a 42-month period concluded that in young children who were at the initial stages of gaining vocal control, it was often very difficult for judges to make distinctions between song and speech vocalisations. As the children started to make their own clear distinctions between singing and speaking they would sometimes *purposefully* alternate between singing and speaking vocalisation "to communicate in novel forms of contextual intermediate vocalisations" (p. 116). The immediate implication of this is expressed by Mang (2000): "…the 'fuzzy' boundary between preschool children's song and speech brings out the question as to what constitutes a conception of singing and speaking both acoustically and contextually" (p. 116).

6.2. Social vocalisation and the origins of melodic vocal behaviour

Such evident overlap between the form and function of proto-musical and proto-linguistic vocalisations has led to suggestions that they share an evolutionary, functional, heritage, and thus a common ancestor (Scherer, 1991; Brown, 2000). Underlying such suggestions should be a consideration of the nature, purposes and advantages associated with such proto-vocalisation abilities.

6.2.1. Proto-music/language: rationales for a shared ancestry

Apart from the important shared neurological and physiological mechanisms for the production and perception of linguistic vocalisations and melody (discussed in Chapters 4 and 5), language and music also share several structural and functional features which reinforce this apparent shared heritage. From the perspective of emotive (affective) communication, the possibility that the earliest forms of language and music might have had a common foundation has also been suggested by Scherer (1991):

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

A common ancestral "musilanguage" has also recently been proposed by Brown (2000), who suggests that music and language are "reciprocal specializations of a dual-natured referential emotive communicative precursor, whereby music emphasizes sound as emotive meaning and language emphasizes sound as referential meaning" (Brown, 2000, p. 271). Brown points out that the major parallels between linguistic speech and melody are combinatorial syntax and intonational phrasing. The latter is based on scales of discrete pitch levels in both melody and speech (although in relation to a fixed pitch in western music). In both music and speech, phrase is the basic unit of structure and function, and both speech phrases and musical phrases are "melodorhythmic" structures, with expressiveness of the phrasing being very important.

According to Brown, the melodic and rhythmic elements of melodic and linguistic vocalisation are derived from:

1) acoustic properties of the fundamental units of music and speech; pitch sets, intensity and duration in the case of the former, and phonemes and phonology in the latter,
2) sequential arrangement of these properties in a given phrase,
3) expressive phrasing mechanisms that modulate the basic acoustic properties of a phrase.

These acoustic properties constitute a "phonological level" of meaning to melody and speech, an "acoustic mode" involving emotive meaning and interpretation. There is additionally, of course, a level of higher-order interpretation, or "meaning level", involving referential meaning. Both music and language have this "meaning level" (or "vehicle mode" as Brown calls it), but it is at this level that they diverge from their parallels, with the nature of the meaning level being determined by the culturally and contextually-specific elements of music and language.

Brown thus suggests that the common features of music and language evolved before the distinct features; namely that lexical tone, combinatorial syntax and expressive intonation were shared ancestral features of music and language. More specifically, he suggests that initially, ancestral to both music and language, there was a "simple system involving a repertoire of unitary lexical-tonal elements" which evolved into a "less simple system based on combinatorial arrangements of these lexical-tonal (and rhythmic) elements" (p. 290). This system of lexical-tonal units being strung together into sequences results in the creation of "simple, unordered phrases having higher-order meanings" (p. 285), where the phrases as a whole have a global emotive and/or pragmatic meaning, due to the overall contour of the phrase. Examples in modern speech are surprise and question intonations. Such phrases have a rhythmic as well as a melodic structure (due to the temporal arrangement of the constituent units); because of the relationship between the individual units, they communicate on a compound basis as well as the global basis described above.

It is interesting that Brown considers that the unitary elements of such utterances would have been *lexical*-tonal, and does not mention *emotive*-tonal utterances, when he originally highlights the emotive-communicative elements of music and language as being a major parallel. He describes lexical-tone as referring to the use of pitch in speech to convey semantic (lexical) meaning, but makes no mention of pitch and tone to communicate emotional state and reactions. Perhaps he means that the lexical referent of such utterances is emotive, but it is not clear from his explanation that the units are not instead fulfilling the role of nouns. As discussed in some detail below (section 6.4), a single discrete vocalisation can, through tonal contour, express a great diversity of information regarding emotive state and reaction, without being part of a larger sequence, and without having a specific referent other than being an expression of personal state. This does not undermine Brown's suggestion of the emergence of the importance of lexical-tonal sequences, or of the importance of tonal contour in the phonological meaning ("acoustic mode") of such utterances as a whole, but it seems most likely that the use of lexical-tonal units would only have grown out of emotive-tonal units. In these, tonal contour is already important, and there is no additional requirement for an explanation as to how the importance of tonal contour of a phrase could come to develop out of discrete units which were not themselves tonally contoured. Tonal contour is already important, at a discrete utterance level, and the combination of tonally-communicative utterances would lead to pitch contour being expressive at a global-phrase level.

The emergence of the use of vocal contour *is* something that Brown discusses, however. Indeed, the importance of the production and perception of pitch in vocalisations is central to the musilanguage model. Brown considers that since most primate vocalisation systems rely heavily on unpitched grunts and pants, a pitched vocalisation system (shifting from one pitch to another) must have emerged at some stage in the development of human music and language. In fact, there is little need to argue for the emergence of an entirely new system of pitch control. Brown's description underplays the use of pitch in primate vocalisations: his emphasis on "unpitched grunts and pants" implies a lack of laryngeal engagement in vocalisation in chimpanzees and monkeys, which is not the case. Whilst grunts and pants are important vocalisations in higher primates, it is not the case that they (higher primates) do not use pitch moderated by the larynx (the various forms and roles of these vocalisations are discussed below, section, 6.2.2). What *is* the case is that these are of limited range, and do not tend to shift in pitch much over the course of a single vocalisation (with the exception of the pitch-glides of gibbon "song"). What must have emerged in the course of the evolution of pitched-contoured vocalisations is the integration into vocalisation of an *increased range and control* of pitch contour, but this needn't have emerged as a new system of vocalisation; it could instead have been building upon the limited pitch control used for emotive-tonal-social vocalisation already used by primates (as discussed below).

This is in fact more in keeping with the cross-modal "sentic modulation" (Clynes, 1977) which Brown takes as an underlying mechanism of emotional expression in humans and other animals. This is the idea that expression of emotion across the media of tone of voice, facial expression, dance step and musical phrase (for example) can all be seen as equivalent along three spectra: tempo modulation (slow-fast), amplitude modulation (soft-loud) and register selection (low-high pitch). Brown interprets this as "a general modulatory system involved in conveying and perceiving the *intensity* of emotive expression across a continuous scale" (emphasis in original) (p. 287). This appears to be invariant across modalities of expression in humans (discussed in considerably greater depth below) and, as mentioned previously, seems to function similarly in emotive expression in other animals (Morton, 1977, 1994). It is particularly interesting in the light of the aforementioned more recent observations of the strong correlation between vocal affect and orofacial expression (e.g. Tartter, 1980). This idea of cross-domain expression and action closely parallels Donald's (1991, 2001) conception of "mimetic" cognition, in which a single controller is responsible for expression across facial, vocomotor, manual and corporeal media. Mimesis describes a system of abilities which manifest themselves in expression and control of emotion through public gesture and action, and allows the sharing of attention and knowledge "by means of gesture, body language and mime, any of which can communicate an intention quite effectively, without words or grammar" (p. 263). It thus forms the foundations for complex social interaction and structure, and the enactment and re-enactment of events and actions. The implications of such developments in cognitive ability are considerable, and are discussed further in Chapters 7 and 8.

Regarding the divergence of music and language from a musilanguage ancestor, Brown (2000) suggests that the common ancestral form of *language* was most likely to be tonal, as is the case with Chinese and, in fact, the majority of the world's languages (p. 281), rather than non-tonal (like intonation languages such as English). As far as Brown himself is concerned,

> "The model's principal contribution to the study of language evolution is to provide a new chronology for the development of language's structural features: language evolved out of a sophisticated referential emotive system; phonological syntax preceded propositional syntax; tone languages preceded intonation languages; speech could have evolved early, due to its exploitation of lexical tone instead of enlarged segmental inventories; lexical tone, combinatorial syntax, and expressive intonation were ancestral features of language that were shared with music; broad semantic meaning preceded precise semantic meaning; and language's acoustic modality preceded its representational modality." (p. 294).

In terms of music evolution, its contribution is to highlight the roots of melodic vocalisation, and melody in general, in this interpersonal affective communication. Brown seems to consider that the entrainment of rhythm as a feature of musical behaviour is something that was integrated into it afterwards, following the divergence of music and language.

All-in-all, Brown's musilanguage model is very convincingly argued and, although he cites virtually none of the same evidence as is dealt with in this thesis, the model accords extremely well with the diversity of data examined here. It does not, however, include any scenario for the original use of such a musilanguage, any relation of it to skeletal or cognitive development, any *rationale* for its development into the "reciprocal specialisations" of music and language (although he does describe the possible mechanism in depth), or any discussion of the timing of such developments. These are elements, amongst others, of the origins and evolution of music which this thesis aims to elucidate. The following sections examine in particular the extent to which musical and linguistic expression of emotion, and interpersonal interaction, are interrelated.

6.2.2. Social vocalisation in primates

In the words of Fernald (1992b), "…human symbolic communication builds on our primate legacy, a foundation of affective communication established in the preverbal period" (p. 423). There is a strong precedent for the use of vocalisations in social interaction and bonding activity displayed by a variety of non-human primates, as has been noted in several contexts in the preceding sections. As Seyfarth and Cheney (1997) observe, "In the wild, nonhuman primate vocalisations signal the presence of different predators, provide information about the group's location and movement, facilitate friendly interactions, and lead to reconciliation between individuals who have recently exchanged aggression" (p. 249). The former case has recently received much attention as a possible form of symbolic referential behaviour and as a precursor to syntax (e.g. Seyfarth and Cheney, 1992) but it is perhaps to the latter two cases to which we would be best to look for evidence of the roots of complex vocalisation in inter-individual interaction.

As noted above, amongst bonobos, utterances always occur along with facial expressions, gestures and tactile communication (Falk, *in press*; Bermejo and Omedes, 1999), and chimpanzee mothers also vocalise softly whilst examining their infants (Falk, *in press*; Nicolson, 1977). Interestingly, Seyfarth and Cheney (1997) conclude from their analyses of the circumstances under which such socially-facilitative vocalisations occur that there is no indication that those individuals interacting have any knowledge of their conspecifics' mental state, knowledge, beliefs or desires. Instead they suggest that in the case of reconciliation calls, recipients have learned from experience that individuals producing reconciliation-type calls rarely follow them with aggressive, but instead friendly, behaviour. Even if they are correct in this belief, the significance of such behaviour should not be underestimated – it is not merely "stimulus-response" Pavlovian conditioning. Whether or not the individuals involved have knowledge of complexities of their conspecific's mental state, they clearly have a set of expectations about how the interaction is going to proceed following such an utterance. That such utterances come to be a predictive precursor to friendly interaction establishes them firmly in a repertoire of inter-individual behaviour. That they are then followed by activities further consolidating such interpersonal relations suggests a group or repertoire of related behaviours based on expectations about the other's behaviour that cannot easily be viewed simply as a sequence of stimulus-response reactions.

Gelada baboons also use a series of vocalisations for "social grooming" purposes, to establish vocal relationships and social bonds with conspecifics (Richman, 2000). Interestingly, these are similar in form and vocal detail to human vocal formulae; in both humans and geladas friendly vocalising is produced in units of an average length of 9-10 syllables, about 5 syllables per second, 3 or 4 strong beats per unit, with an intonation contour, and the end of the vocal unit being signalled by tonal change (Richman, 2000). The findings of Elowson *et al.*, (1998 a and b) and Gros-Louis, (2002), regarding primate infant vocalisations, also illustrate an early and important social function to vocal behaviour; such vocalisations can be very important in eliciting care responses from mothers to infants in various primate species, in instigating affiliative activities with others in the group, and operate in a variety of very specific contexts. It seems that vocalisations form a very important component of repertoires of social inter-individual behaviour in both infant primates and adults.

6.2.3. Evolutionary rationales for complexity of vocalisation: proto-music, proto-language and social vocalisation

Aiello and Dunbar (1993) note that there is a proportional relationship amongst the higher primates and humans between degree of encephalisation (cortical development) and group size; as group size increases, so does neocortical development. This strong correlation between group size and neocortex size exists not only in primates, but in carnivores and cetaceans too (Kudo and Dunbar, 2001). Indeed, links between neurological development and group size are well attested by other studies of primate neurology. Neocortex size and the size of the striatum increase with the mean number of females in a group; the amygdala and the number of neurons in the parvocellular lamina also increase with overall group size (Mondragon-Ceballos, 2002). Furthermore, lesions of the prefrontal cortex in primates lead to extreme social apathy; it is clear that cortical development is implicated in social function (ibid.). Treating the relationship between group size and neocortical development as a direct one, Aiello and Dunbar (1993) calculated the probable group sizes of the various Palaeolithic hominids on the basis of their brain-size.

Amongst primates such as chimpanzees, grooming is a very important function in maintaining an individual's social network, alliances and coalitions within a group; such a network is important in providing support in power, mate or food resource contests. However, beyond a certain group size, it would become impossible to maintain a social network effectively through manual grooming alone, as the time taken would be too great (Dunbar, 1998). Evidence considered already, above, has illustrated the social importance of vocalisation in many primates, in particular in instigating relaxed proximity and grooming. The importance of sensitivity to emotional content of vocalisations has also been discussed. Dunbar (1998) suggests that physical grooming began to be supplemented with utterances that increased its efficiency, and that as group size increased there would be a selective pressure for the "grooming" utterances to become progressively more efficient and expressive. This could then feed back allowing group size to increase further, and so on. As noted above, vocalisations already form an important part of inter-specific interaction and affiliation in primates, and the integration of such affective vocalisation into grooming activity would not be a great step. Note that it is not asserted that neocortical development is the determining factor in group size, but that it places a constraint on group size. The major influences on group size are ecologically imposed local costs and benefits, but it is hypothesised that the cognitive mechanisms that allow individuals to live together in coherent stable groups have evolved to support the group size imposed (Kudo and Dunbar, 2001).

The proposed direct relationship between encephalisation and group size is the possible weak link in Aiello and Dunbar's (1993) otherwise well-supported hypothesis. It may be the case that this relationship is not a straight line at all, and that it is impossible to accurately predict the group size of earlier hominids from it, in which case it would not be possible to assert at what point physical grooming would have become impractical without the presence of "vocal grooming". However, more recent research by Kudu and Dunbar (2001) has further reinforced the connection between grooming and neocortical development in primates. In examining a wide cross section of primates (including modern humans) they found that coalition size/grooming clique size and neocortical development are strongly correlated, and that coalition size/grooming clique size is proportional to overall group size. It seems that there is consistently a strong relationship between cortical development and social group size, and that social relationships are an important stimulant to cortical development. Hopefully, future fossil evidence will add to the record of data regarding the group size of hominids; the "pit of bones" (Sima de los Huesos) site at Atapuerca, Spain, contains the remains of more than 32 individuals apparently part of the same community (Arsuaga, 1997) so may eventually provide clues as to the group size of *Homo heidelbergensis*. In any case, the evidence that the use of language to establish and maintain social networks far outweighs its use for "higher" discussion provides important support for the idea that social use was one of the fundamental foundations of language – for example, Dunbar *et al.* (1997) and Emler (1992) both found through analysis of natural conversations that 60-70% of all conversation time is devoted to social information exchange. The evidence discussed in Chapter 5 and elsewhere in this chapter also reinforces the proposed early importance of the associations between vocal behaviour and interpersonal social-emotional expression (versus lexical or signalling content).

If it is indeed the case that social-affective content was initially the most important component of vocal communication, then users of this system would have derived considerable selective benefit from the development of increasing vocal agility, because as the variety of sounds they could make increased, so too would the variety of sound sequences (and thus expressiveness); with greater expressiveness, social grooming would become more efficient, and the individual's social network could become larger and more cohesive. An individual able to maintain a large cohesive social network that provides support in power, mate or food resource contests would have considerable selective advantages over its peers.

It is also interesting to note that the development of complexity in social vocalisations as an adjunct to grooming behaviour provides an intriguing explanation for the compulsive interrelationship between vocalisation and manual movement, specifically, between fine laryngeal, orofacial and manual muscular control. If complex vocalisations developed as an adjunct to grooming behaviour, then one would actually expect the action of the hands (in particular, precision movements) and vocalisation to be intimately related. As vocalisations increasingly substituted for manual grooming, the movements of the hands might become more and more structured in their accompaniment of the vocalisations; out of this structure they could form the foundations of syntactic structure as hypothesised by Armstrong *et al.* (1994) (see section 6.4.3) which then came to be incorporated into vocalisations too.

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

That the establishing of pair and group bonding is an older requirement than full linguistic speech is evident from many animal behaviours, but this is not a function of vocalisation that has been lost since the development of full linguistic capacities; it is important to note that language also fulfils important functions with regard to aiding social behaviour. In fact, this social bonding property is a very important common feature of linguistic and musical behaviour. There is plenty of evidence of importance, or even dominance, of the social function of language – as noted above, Dunbar *et al.* (1997) and Emler (1992) both found that around two thirds of all conversation time in natural conversations is devoted to social information exchange. It seems that social use forms a major component of linguistic behaviour.

As discussed in Chapter 2, ethnographic evidence from a wide diversity of hunter-gatherer and other "traditional" cultures parallels this finding for music. Many of the most important roles of music amongst Plains Indians (Nettl, 1992; McAllester, 1996), African Pygmies (Kisliuk, 1991; Turino, 1992; Ichikawa, 1999), Yupik and Inuit Eskimos (Nettiez, 1983; Johnston, 1989) and Australian Aborigines (Breen, 1994; Myers, 1999) are social, interactive and integrative, and the performers themselves often see these as the most important consequences of the activity. Blacking (1995) also observes, from his studies of the Venda peoples, that performance of a musical pattern "may announce social situations, recall certain feelings and even reinforce social values" (Blacking, 1995, p.39). Blacking himself believed that social bonding was not the limit of the function of music, but that there is actually a relationship between music and social structure, with particular types and elements of music being used by different groups within an integrated culture.

There is an alternative explanation to Vaneechoutte and Skoyles' suggestion. This is that the reason that language can today function without full tonal use of the vocal chords is that lexical complexity and syntax have removed ambiguity from our vocal utterances, allowing meaning to be expressed at an additional linguistic level. This would not have been the case in a proto-language without modern lexical complexity and syntax (similar to that proposed, for example, by Sperber (1996), Aiello and Dunbar (1993), or Brown (2000); see above), and would not be applicable to melody. These *would* rely on the full use of the vocal tract. It seems most likely, then, that the vocal tract developed to the extent that it did not to support song before language, as Vaneechoutte and Skoyles (1998) suggest, but instead to support tonal prosodic affective utterances that were a precursor to both language and melodic music (as Brown, 2000, suggests), *and were useful as social-emotional communication*. In this case, it is only since complex lexicon and syntax developed that language has been able to be used without recourse to the entire human vocal range.

Interestingly, the findings of Watt and Ash (1998), that music has an action on the mind similar to the action of interacting with a person, and that people attribute human-like qualities to music, also complement the idea that melody in music could have been developed from the use of a socially-based proto-language. Thus the central problem expressed at the start of this chapter of the capacity for complex *vocalisation* existing long before complex *language* is not problematic if we accept a hypothesis that this capacity was used by progressively increasingly complex proto-language based on tone-dominated affective social utterances. Aiello and Dunbar (1993; Dunbar, 1998) have proposed that early vocal utterances were initially used to facilitate and then replace social grooming; the evidence of Chapter 5 and the preceding sections suggests that such utterances would have been prosodic tonal emotive vocalisations, and these would indeed seem most likely to be effective in facilitating social grooming. The social hypothesis, along with the observations and findings relating to infant-directed and emotional vocalisation, suggest that human use of non-verbal song has far more in common with emotional, social and emotive-social vocalisation than with the song of birds, which is territorially or reproductively orientated rather than expressive and socially bonding (Slater, 2000).

In summary, it seems likely that both language and music grew out of non-verbal vocal emotive utterances, expressing emotional state, which were initially made to increase the efficiency of physical grooming activities, and which gradually came to be used in broader spheres of behaviour. Exactly these utterances could also be used to accompany rhythmic percussion and dance, thus increasing the emotive response and social bonding experienced as a consequence of the activity. Emotive vocables (non-lexical vocal syllables) are used in exactly this way by Plains Indians (Nettl, 1989; Nettl, 1992; McAllester, 1996) and African Pygmies (Locke, 1996) (see Chapter 2).

6.3. Emotion and communication in music

The associations described above between musical, melodic and non-linguistic social and emotional vocalisations, in human development and in other primates, would lead us to expect to see emotional and social associations in our experience of music. That music and emotion are strongly associated is without doubt, and the nature of this relationship has been the focus of debate and research since classical times. Theories about the interrelationship between emotion and music often seem to have been, in the past, characterised by impositions of mutual exclusivity on various explanations, when in fact there are a number of likely causes for the eliciting of an emotional response to a given piece of music, which can operate together at different levels within the listener. In recent years the study of the relationship between music and emotion has expanded considerably, encompassing the disciplines of psychology, neurology and philosophy, as well as musicology, to the extent that there is now a dedicated volume giving good coverage of the approaches and conclusions from each of these disciplines (Juslin and Sloboda, 2001). As discussed above, there are important connections between infant-directed, emotional and melodic vocalisations in communication of emotional state, and the evocation of emotion in the listener; the implications of this are discussed further below. However, this is only one of many connections between music and emotion.

Discussing emotion and music is potentially confusing, as the relationship can be considered along a number of dimensions which interrelate and overlap; separating those dimensions for the purposes of discussion is difficult and even potentially undesirable (see Lavy, 2001). In terms of the actual *processing* of the musical input, Lavy (2001) points out that the various rigid distinctions made between different categories of emotional content associated with music are somewhat artificial, bearing little resemblance to our understanding of emotion in other domains. Nevertheless, in discussing the *derivation* of such content they are useful. Whilst considering the relationship between emotion and music along these dimensions, hopefully the following sections will illustrate the extent to which these classifications are interdependent.

Music may have *intrinsic* or *extrinsic* emotional content (Juslin and Sloboda, 2001), which may be *iconic*, *indexical* or *symbolic* (Dowling and Harwood, 1986), and may *represent* and/or *elicit* emotion (arousal theory vs. expression theory, Davies, 2001). The expression of emotion and eliciting of an emotional response by music can occur simultaneously and may, but need not, correspond – for example, whilst a given piece of music might be perceived by two listeners to be *representing* a particular emotion, the emotion(s) *elicited* in each listener need not be the same, either as that being expressed or as that elicited in the other listener. This is because emotional responses are elicited in the listener for a variety of other reasons too.

6.3.1. Intrinsic and extrinsic emotional content of music

Sloboda and Juslin (2001) and Sloboda (2001) categorise the roots of emotional effects of music in terms of *intrinsic affect* and *extrinsic affect*. The intrinsic emotional properties of the music are those which elicit in the listener an emotional response as a direct consequence of structural properties of the music itself (i.e. with no outside referent). These are the *symbolic* properties of music, its syntax and style (Dowling and Harwood, 1986). Into this category fall characteristics such as the fulfilment or violation of expectations (Meyer, 1956), intensity, and perceived tension and release. *Extrinsic* affect is the emotional response elicited as a consequence of association of properties of the music with previous events or experiences (its *indexical* properties) or through the music's resemblance to other phenomena (its *iconic* properties).

Suggestions that the form of emotion experienced when listening to music is not the same as a "genuine" emotion are belied by physiological responses and by observations of neural activity during music listening. Whilst intrinsic emotional properties of music have to date received less laboratory examination than extrinsic properties, that which has been carried out suggests that there are genuine physiological responses experienced in response to these stimuli within the structure of the music. Features such as syncopations, melodic appoggiaturas and enharmonic changes create, maintain, confirm or disrupt musical expectations (Sloboda and Juslin, 2001), and can elicit physiological reactions such as changes to breathing rate, heart rate, blood pressure, temperature, galvanic skin response (conductivity due to perspiration) (Krumhansl, 1997), "shivers" (piloerection) and weeping (Sloboda, 1991, 1998; Panksepp, 1995; Panksepp and Bernatsky, 2002).

For example, Sloboda (1991) reports that specific properties of music seem to evoke particular emotional and physiological responses, regardless of the aesthetic type of the music (classical or pop, for example). He asked subjects to report at which points of a variety of pieces of music (from classical to contemporary pop and jazz) they experienced physical sensations such as racing heart, goosebumps or shivers down the spine, or tears. He found that there was a distinct correlation, regardless of type of music, between emotional-physiological response elicited and particular types of sequence in the music. Tears were correlated with melodic or harmonic sequences, downward harmonic movements, appogiaturas and suspensions. Shivers were associated with harmonic, textural or dynamic discontinuities, and racing heart with syncopation and "other forms of accentual anticipation" (Sloboda, 1998, p.27). Research using brain scans is beginning to give insight into the neurological areas in use. Such work is starting to elucidate correlations between the processing of these elements of the music and the processing of aspects of speech and motherese which elicit the same responses (Panksepp and Bernatsky, 2002 give a good overview of this neuro-affective research; see also Chapter 5 for examination of the neurology used in the perception of prosody and emotional content).

Expectations about the progress of a melody may be due to learning of conventions through personal experience (through exposure to particular forms of music), or may be innate expectations grounded in gestalt laws of cognition, such as the expectation that movement in a given direction will continue in that direction (applicable to an ascending scale, for example). It seems that some of these expectation-violations cannot be overridden by subsequent familiarity with a given piece of music, being "hard-wired", such that irrespective of the number of times the piece is heard it may still "surprise" those gestalt perceptual mechanisms, and thus stimulate arousal in the listener.

A problem with this explanation of emotional response to music is that it doesn't account for the feelings of joy, elation or weepiness (for example) in response to the subversion of expectations. In explanation, Sloboda and Juslin (2001) suggest that intrinsic emotional properties are responsible largely for the intensity, or *valence*, of the affect (i.e. a particular level of emotional arousal), whilst extrinsic emotional properties of the music are largely responsible for determining the emotional *content* of the music.

As noted above, these *extrinsic* emotional properties of music are its *indexical* and *iconic* properties. Iconic properties may just represent emotional content, or may elicit emotional response as well. The indexical properties are likely to elicit, rather than represent emotion, as they refer to personal and cultural associations. These are called extrinsic emotional properties of the music because such emotional content exists by virtue of its reference to events or phenomena outside the music itself.

Confusing things somewhat, Sloboda and Juslin (2001) classify iconic properties of music as extrinsic, whilst Sloboda (2001) classifies them as intrinsic. The current author concurs with Sloboda and Juslin’s position, in that iconic properties of music definitively have a referent other than the structure of the music itself; i.e. they simulate something else to some extent, be it a thunderstorm, large animal, or properties of human emotional expression (see below for examples of these), and thus can elicit in the listener a response appropriate to such a stimulus. Sloboda (1991) claims that “a necessary consequence of iconic recognition is a cognition such as ‘this is happy music’”, and whilst this may lead to the music making the listener feel happy, it need not do so. Whilst the latter assertion is true, the former is not - the response to the iconic stimulus may be essentially instinctive and not involve cognising of the stimulus as referring to something else. This is perhaps where the difficulty in classification arises - if this property of the music elicits an emotional reaction without any cognition on the part of the listener as to its referent, should it be considered to be an inherent property of the music, or should it still be taken to be iconic of something else? In this situation the iconic stimulus actually has more in common with symbolic, intrinsic, characteristics of the music. This may especially be true in the case of music with properties similar to human emotional expression, as these properties may be an inherent foundation of music, as well as seeming to have human emotional expression as a referent; they thus have the possibility of being both *iconic* and *symbolic*. This idea is discussed further below.

6.3.2. Ecological considerations, and the human factor

As an example of the ways in which music may elicit or express emotions, consider a listener sitting in a movie theatre. This may seem a curious example to choose in a discussion seeking to elucidate evolutionary connections between emotion and music, but it is a good illustration of the number of levels at which emotion and music can be associated. The music accompanying a particular scene in the film could have the following emotional associations:

1: as index, as a consequence of association with past circumstances in which that music has been heard,

2: as symbol, in tensions and release in structures of the music itself, or

3: as icon, being loud and fast, and thus suggesting energetic emotion such as excitement (Sloboda and Juslin, 2001), or resembling a thunderstorm, for example.

So, in terms of its iconicity it can represent a physiological state, or ecological environmental stimuli. As will be seen in subsequent sections, physiological state and emotion are fundamentally interrelated, and at a deeper level than mere iconic resemblance. However, at a more literal level, the iconic representation of ecological stimuli also has evolutionary implications, as it may resemble the types of sounds that would be threatening in an ecological environment; an example is outlined by Cross (2003a):

> "For example, a passage which incorporates a fairly rapid crescendo from *pp* to *fff*, combined with an upward expansion of tessitura has the same properties as a sound produced by something large in the real world that is either approaching us (hence getting louder and with an increasing amount of higher-frequency energy), or that is staying in the same location relative to us but is increasing in the energy it is expending. Hence we might experience arousal not through any breaching of expectation but simply because of the 'thrill' that is likely to accompany any real-world encounter in which a large sound-producing entity approaches us, or when we are in close proximity to a large sound-producing entity that is becoming rapidly more energetic."

This could, in theory, apply equally to ecological sounds that are pleasing rather than threatening. This example illustrates well the earlier point that the listener need not actually cognise a resemblance between the music and an ecological stimulus in order for it to elicit the appropriate emotional response; i.e. the listener need not think "this sounds like a large sound-producing entity approaching me, how exciting" in order for the music to elicit the appropriate thrill reaction. They may, in fact, make such a rationalisation *after* their reaction to the stimulus.

There is also another layer of *contextual* emotional association with the music: it is directly associated with a scene in the film, the emotional content of which is connected with their simultaneous experience. It is also associated with the context of the viewer themselves, who might have their hand entwined with that of their romantic partner in the dark at the back of the theatre. This is neither an intrinsic or extrinsic property of the music itself, being determined by the context in which the music is experienced (although such associations may subsequently constitute indexical extrinsic properties of the music when it is experienced again and they are re-awakened). These contextual associations may feed into each other, forming part of the initial experience of emotion elicited as the various incoming sensory stimuli are associated, and are likely to affect the extent to which emotional response is elicited in the listener. In evolutionary terms, we have until recent history never been presented with circumstances where our personal context and state is dissociated from the visual and auditory stimuli with which we are faced. Situations where these inputs *are* engineered to be dissociated (such as watching theatre, when one's personal situation is divorced from that which is the principal sensory input) are a relatively recent historical invention. Consequently we naturally assimilate sensory stimuli simultaneously with our mood and level of arousal, each sensory input either influencing or being strongly associated with each other and with the affective state consequently experienced. Such contextual associations may feed back into future experiences of the same piece of music too, becoming part of its *indexical* communication of emotion (as described above).

There is also the possibility of being aware of emotional expression on the part of the composer or performer of the music (*expression theory*), which may or may not elicit the same emotion in the listener. Expression theory views the expression of emotion not as an act of the music itself, which is not sentient, but to occur as a consequence of either an expression of emotion by the composer, performer, or some perceived persona that is represented in the music itself (Davies, 2001). Whilst an emotion can be elicited in a listener in a variety of ways (as discussed above), when it is a response to an emotion being *expressed* in the music, it may be useful to view it in terms of sympathy and empathy. We can be sympathetic to another's emotional state without experiencing it ourselves. However, if we empathise with another's emotional state, we experience (at least something of) the emotion that they are experiencing themselves. The same might be said of music. We can be quite aware of the emotion being expressed, and in many circumstances not experience that emotion ourselves; in this case, we are in sympathy with the music. When we find ourselves experiencing the emotion being expressed, then this is an empathic response to the music.

The idea that music is naturally interpreted by listeners as having human-like properties has been strongly supported by experiments carried out by Watt and Ash (1998). Watt and Ash conclude from their experiments that music has an action on the mind similar to the action of interacting with a person. Their experiments involved subjects hearing a selection of pieces of music, and attributing a variety of properties to them, selected from a list. Some of the adjectives were people-traits (e.g.: male/female), some people-states (e.g.: gentle/violent), some described movement (e.g.: leaden/weightless) and some were adjectives that are rarely applied to people (e.g.: sweet/sour). A summary of the research cannot do justice to the experimental controls which Watt and Ash imposed on their procedure and analysis, but they found that between subjects there was strongly statistically significant agreement with regard to person-like attributions given to the different pieces of music, in contrast to the non-person-like attributions. They conclude, from this and other analyses, that music has an action on the individual similar to the actions of a person, and that when music is perceived, it is assigned attributions that would normally be assigned to a person, including trait-qualities, such as age and gender, and state-qualities such as emotions.

Although apparently unaware of Watt and Ash's research, Davies (2001) observes that "Registering music's expressiveness is more like encountering a person who feels the emotion and shows it than like reading a description of the emotion or than examining the word sad" (p. 30). Having made this observation, it is slightly odd that Davies considers dubious Kivy's (1989) suggestion that expressive instrumental music recalls the tones and intonations of emotional content in speech. Whilst acknowledging that instrumental music can sometimes imitate singing styles, and that singing styles sometimes recall ordinary [vocal] occasions of emotional expressiveness, Davies states that "For the general run of cases, though, music does not sound very like the noises made by people gripped with emotion." (p. 31).

This, however, is not the assertion being made by Kivy (or here). Music can *recall* the intonation of the emotional (prosodic) content of speech without *resembling* the noises made by people *gripped with emotion*. Firstly, the noises made by people gripped with emotion are not the same thing as the emotive prosodic contour of speech. Secondly, to recall something is not necessarily the same as to resemble something. The use of the term *resemble* in this context implies that the listener is cognizant of a similarity between the musical stimulus and human vocal emotional expressive sounds. In contrast, the key is that it is at least partly (as demonstrated in Chapter 5) being processed by the same cognitive mechanisms, and thus eliciting, in the listener, similar responses; one need not cognize a resemblance in order for it to stimulate the same response. Lavy (2001) succinctly observes that "Humans have a remarkable ability to communicate and detect emotion in the contours and timbres of vocal utterances; this ability is not suddenly lost during a musical listening experience." (p. v).

Interestingly, considering his above position, Davies (2001) goes on to say that "if music resembles an emotion, it does so by sharing the dynamic character displayed either in the emotion's phenomenological profile, as Addis (1999) maintains, or in the public behaviours through which the emotion is standardly exhibited" (p. 31). This has been called *contour theory* (Davies, 2001), and holds that our interpretation of music as expressive of emotion is due to the resemblance of its contours and forms to physical expressions of emotional state. This falls closer to an arousal, rather than expression model, and parallels previous suggestions of associations between qualities of music and gestural and other physical expressions of emotion (Clynes, 1977; Scherer, 1991). Cross (2003a) elaborates Davies position:

> "Music, it is claimed, operates at the emotional level by some system of resemblances. Listening to 'sad' music simply leads to an emotional or affective state in the listener (by virtue of the some correspondence between features of the music such as a descent and features in the real world that would be interpretable as expressing an emotion such as downturned corners of the mouth and a slackness of posture that makes it appear that the body is somehow more earth-bound than usual)"

Such physiological manifestations of emotional state are an extremely important stimulus in interpersonal interaction, sensitivity to them and interpretation of them are essential social skills. However, to explain emotional response to music as Davies (2001) does, as occurring solely in terms of *iconic* resemblances between features of the music and physical expressions of emotion, is under-representing the complexity of the response to this aspect of the music. It unnecessarily separates physical and vocal expression of emotion. Amongst these physiological manifestations of emotional state is tonal-vocal expression of the emotion, which has, after all, a physiological basis, in laryngeal and orofacial muscular control. Resemblances, in the iconic sense that Davies uses the term, are insufficient to account for the response to tonal qualities of a melody, which seems to occur at a more fundamental level with less conscious cognizing required than making direct associations between features. A down-turned mouth, slackness of posture, lower voice and reduced pitch contour in vocalisation, for example, are all part of a gamut of physiological responses associated with sadness and/or depression. Lavy (2001) summarises the findings of Kappas *et al.* (1991), for example, that "boredom is characterised by low fundamental frequency and a low intensity [of vocalization]; irritation leads to an increase in fundamental frequency and intensity, coupled with downward intonation contours; sadness and dejection lower the fundamental frequency significantly; by contrast, fear evokes a vast increase in fundamental frequency." (p. 41). Moreover, tonal-vocal qualities manifested (such as timbre, frequency and tempo) and the physiological properties manifested (such as posture) are intimately interrelated, and dictated by levels of autonomic nervous system (ANS) arousal (Wagner, 1989; Lavy, 2001). The properties of music which seem to express sadness (for example) do not merely *resemble* these expressive physiological and vocal expressions of emotional state, but are part of the same system and are processed by some of the same mechanisms as the tonal-vocal elements (see above, and Chapter 5). In this sense, this property of music is more *symbolic* than iconic. Although varying in intensity, these vocal-tonal and physiological manifestations of emotional state rarely occur in isolation in natural situations, and we interpret one as indicative of the presence of others. For example, if we see someone standing in the corner of a room, with slumped shoulders and downturned mouth, we have very strong expectations about the pitch and contour of their voice when they speak to us; likewise, if we hear a voice on the radio at a relatively low pitch and with little variation in pitch contour, we have strong expectations, and probably a mental image, regarding what their facial expression and posture would be. The same applies for other emotional states too of course, such as elation, excitement and anger, for example. We do not need to resort to discussing *resemblances* between the dynamic character of public physiological expression and musical contour; they are part of the same system. This is consistent with Clynes' (1977) theory of sentic modulation and Donald's model of mimetic cognitive systems (discussed below and in Chapter 8). This also overcomes the problem of the cross-cultural applicability of contour theory, because it relies on a universal human mechanism of emotional processing rather than resemblance between emotional expression (universal) and musical features (culturally specific).

This explanation does not imply that we treat the *performer* (or composer, or other imaginary entity inherent in the music) as expressing to us this state in him/herself (as suggested by *expression theory* - although this is not to rule this out as a *different* possible cause of attribution of emotional reaction to music). Instead, it suggests that this auditory cue is interpreted *in the same way* as physiological, and corresponding auditory, expressions of emotional state. We thus do not need to resort to positing the creation by the listener of a "persona" within the music, to whom is attributed the cause of the emotional content being perceived (as is a suggestion of the *expression theory*). It may be useful in this case to make the distinction that Matravers (1998) does between a *feeling* and an *emotion* elicited in the listener, in that an emotion has a focus; it is object- (or individual-) directed, and thus motivates us to action in response to the emotion (Davies, 2001). A feeling, on the other hand, has no direct object as its focus and so does not elicit a drive to act on their feelings; as Davies (2001) puts it, "For instance, the response feels like sadness or pity, and this makes it true that the music expresses sadness, but the response is not an object-directed cognitively founded emotion" (p. 39). The listener's reaction "does not take the music or any other thing as its emotional object" (p. 39).

Whatever the philosophical or semantic concerns, the implication of much of the preceding evidence remains that it appears that we process emotion in musical sounds in the same way as in vocalisations (and other expressions) of affective state, and whilst the vast majority of music in traditional cultures is vocal rather than instrumental (see, for example, Chapter 2), this appears to apply also to our perception of instrumental music.

The acoustic signals used in the production and perception of emotion in instrumental music appear to be the same as those used in vocal utterances, such as high intensity and tempo and harsh timbre for anger, low intensity and tempo and slow vibrato for sadness, whilst variation in timing and intensity typifies fear (Gabrielson and Juslin, 1996; Lavy, 2001).

6.3.3. Context: the social factor

The associations between music and emotion considered so far have focused mainly on the direct relationship between the listener and the music. It has been noted that, in addition, stimuli from the listener's immediate context may feed into the emotional state experienced at the time, and thus come to be associated with the music itself. A contextual cue that can be very important in the personal experience of emotion is the emotion apparently being experienced by the people in the immediate environment. We use other peoples' emotional reaction to a situation in a variety of ways to influence our own reaction, both consciously and unconsciously. As Sloboda and Juslin (2001) elucidate, infants frequently look to adults for a cue as to how to react to a fall, being far more likely to cry if the parent seems worried or upset than if they seem unconcerned. It seems likely that such cuing as to how to respond to a given circumstance occurs in other situations too, and is an important part of a child's learning the appropriate (or inappropriate, as the case may be) way to respond to events or interactions. As we get older we also may ask ourselves how a given person *in absentia* (e.g. a parent or other role model, such as a literary or screen character) would respond to the situation we are in and try to respond in the same way. At a less conscious level, emotions are also "contagious" (Hatfield *et al.*, 1994) in that we respond to a prevalent mood (Kraut and Johnston, 1979), which also influences our interpretation of our own arousal (Schacter and Singer, 1962).

As well as influencing our own emotional state (Kraut and Johnston, 1979), being sensitive to the emotional state of others is also an extremely important skill in that it allows us to know how to respond to them (Schmidt and Cohn, 2001; Sloboda and Juslin, 2001). This can assist in the avoidance of harm at the hands of an angry individual, allow us to anticipate and prevent harm to third party (Sloboda and Juslin, ibid.), as well as being fundamentally important in more positive situations related to social bonds, networks and procreation, such as providing comfort when appropriate and pursuing and maintaining interest from the opposite sex. These are very sensitive capabilities, and through its drawing upon many of the same capacities as interacting with a person (as discussed in the preceding sections), music draws upon these abilities. In addition, there is the aforementioned role of social contextual cues in the experience of a response to a given stimulus. Whilst the precise emotional reactions engendered by a given piece of music can be individual-specific, as discussed in the preceding sections, much of the value of listening to music in a communal situation (for example, at a concert or club) is that you are not listening in isolation. It is a shared experience. One knows, as a member of an audience, that even if one's own reactions to the music are quite personal and self-specific (and it is from this perception of them that they gain some of their value), the surrounding people are similarly experiencing such reactions. Thus the experience of a strong *individual-specific* reaction is also a *shared* experience, with the knowledge that the reaction in others, even if subtly different, is being elicited by the same stimulus. It is noteworthy that in the majority of cultures (including Western ones until relatively recent history) musical behaviour is a communal and participatory activity, rather than a solo one.

It should be pointed out that this particular rationale is not specific to music, but applies to any communal situation involving the experience of strong emotions. For example, when watching a comedy show or riding on a rollercoaster one's reaction to the situation tends to be stronger, and more visibly and audibly expressed, in the presence of others, although the salience of emotion is individual-specific. Indeed, Kraut and Johnston (1979) found that people in situations that were likely to make them smile were more likely to do so in the company of other people than when they were on their own.

That music contains important social-emotional information is further evidenced by studies of performance of music by sufferers of autism, and of Williams' syndrome. A well-known feature of autism (and Asberger's syndrome) is an inability to emotionally relate to or communicate socially with other human beings; typical autistics appear to have no empathic ability (Davison and Neale, 1994), and an aversion to social interaction (Huron, 2001). Sloboda *et al.* (1985) report on the case of a young autistic man, Noel Patterson. Whilst able to reproduce musical phrases and sequences of notes with a high degree of accuracy after they had been played to him, he reproduced them with no expression, playing the sequences of notes mechanically. Similarly, Sacks (1996) describes Temple Grandin's consideration that music is "pretty", but her lack of emotional response to it (Huron, 2001).

In complete contrast to the autistics' experience is that of sufferers of Williams' syndrome; Williams' sufferers are typified by a gregarious sociality and highly developed verbal capacities, despite mental retardation. They also exhibit a great tendency to undertake musical activities, especially communally (Levitin and Belugi, 1997; Huron, 2001). There appears to be a correlation between sociality, empathic and other emotional capacities (or lack thereof) and the experience of music. This does not appear to affect musical *ability* (autistics can be technically excellent, and Williams' no better than average), but it appears to have a profound influence on their *experience* of music, and consequent incentive to undertake such activities.

The mechanical lack of emotional expression exhibited by autistics and Asberger's syndrome sufferers parallels closely the findings reported by Jurgens (1992) concerning the anterior limbic cortex (see Chapter 4). Destruction of this area in humans results in an inability to produce *voluntary* joyful exclamations, angry curses or pain outbursts. Without the use of the anterior limbic cortex voluntary vocal utterances, which would normally be highly emotionally communicative, sound more or less monotonous, with very flat intonation (Jurgens and von Cramon, 1982; reported in Jurgens, 1992). It is not known precisely the extent of neurological damage in autistic patients (although parts of the cerebellum are underdeveloped in 14 of 18 autistics; Davison and Neale, 1994), but it could be the case that the areas of the brain responsible for voluntary emotional exclamations also control voluntary emotional expression in music.

Panksepp and Bernatsky (2002) suggest, in their review of their own neurological research, that "the emotional impact of music is largely dependent on both direct and indirect (i.e. cognitively mediated) effects on subcortical emotional circuits of the human brain that seem to be essential for generating affective processes" (p. 137). Ultimately their research, and that of others, leads them to state that

> "Our overriding assumption is that ultimately our love of music reflects the ancestral ability of our mammalian brain to transmit and receive basic emotional sounds that can arouse affective feelings which are implicit indicators of evolutionary fitness. In other words, music may be based on the existence of the intrinsic emotional sounds we make (the mammalian prosodic elements of our utterances), and the rhythmic movements of our instinctual/emotional motor apparatus" (p. 134).

As for the cognitive foundations of these prosodic elements, they hypothesise that "music was built upon the prosodic mechanisms of the right-hemisphere that allow us affective emotional communications through vocal intonations" (p. 136). This is all very much in keeping with the findings of Chapter 5 and the preceding sections of this chapter, including Brown's "Musilanguage" model, with the proviso that as far as the present author is concerned, it is not as *indicators* of evolutionary fitness, but as a fundamental *component* of evolutionary fitness that these mechanisms are important. That is to say that the production and processing of the prosodic elements of vocalisation is a social skill which is fundamentally important to individual and group survival. Music uses these same mechanisms.

6.4. The interrelationship of vocal control and corporeal control

Whilst the preceding sections have been principally concerned with vocalisation ability as an obvious prerequisite of melodic behaviour, manual movement is also a fundamental component of much musical behaviour and, indeed, of vocal behaviour and expression of emotion. Gesture and vocalisation appear to be, in some respects, if not entirely, contiguous and interdependent systems, with neurological associations apparently existing between the fine control of orofacial and laryngeal musculature, and that of manual gesture. Further, investigation of the relationship between manual control and vocalisation may help to shed light on the apparent strong association between rhythmic and melodic behaviours in music. For these reasons it is important that gesture should receive investigation here. The involvement, often involuntarily, of manual gestures in vocal expression will be familiar to most people. This situation has led to a body of research examining the interrelationship of these functions (e.g. McNeill, 1992, McNeill (ed.), 2000), and hypotheses that manual gesture and vocal control are functionally linked and share important neurological and/or evolutionary foundations. Some authors have suggested that manual gesture formed the foundation for syntactic elements of language (e.g. Hewes, 1973, 1992; Corballis, 1992; Armstrong *et al.*, 1994; Stokoe, 2000), whilst others have carried out experimental observation of interrelationships between manual and vocal control (e.g. Petitto and Marentette, 1991; Locke *et al.*, 1995; Feyereisen, 1997; Mayberry and Jaques, 2000; Locke, 2000).

6.4.1. Vocal content and manual gesture

Research by Mayberry and Jaques (2000) regarding gesture in normal speech and stuttered speech casts considerable light on the interrelationship between vocalisation and gesture. They found that when disfluency occurs in stuttered speech it is accompanied by fewer gestures than fluent speech – in fact, at the time of stuttering, gesture ceases entirely. This is in contrast to disfluency in normal speech, in which case at a pause gesturing tends to increase along with vocalisations like “uhm” and “uhr”. These vocalisations actually maintain the elements of prosody in normal speech disfluencies, and gesture, likewise, is unaffected. In stuttering, on the other hand, the whole of speech prosody and rhythm cease, as does gesture.

So, in normal speech disfluency there is no disruption of the *motor* and *prosodic* elements of speech – these are maintained in ums and ahs, non-lexical vocalisations – the disfluency originates in a lexically-induced pause as one seeks the correct word or phrase, which does not affect gesture. This is not the case in stuttered disfluency. In the words of Mayberry and Jaques (2000): "…gesture and speech are an integrated system in language production. When speech stumbles and stops as a result of stuttering, the hand always waits for speech so that the meanings being expressed by the hand in gesture always coincide with the meanings being expressed by the mouth in speech, even when the gesture must wait for a very long time. Gesture and speech are clearly not autonomous" (p. 208).

To exclude the possibility that this correspondence of gesture and vocalisation cessation in stutterers was caused by motor shutdown, Mayberry and Jaques carried out an experiment in which stutterers were required to carry out a button-pushing task whilst describing a cartoon. That they proved able to carry out the motor task of button pushing with their arm (which clearly uses the same motor systems as gesturing) uninterrupted throughout bouts of stuttering indicates that motor shutdown is not at the root of the usual co-occurrence of gesture ceasing with stuttering. In fact, stuttering individuals are able to carry out all manner of manual non-gesture movements whilst stuttering, but do not execute gestures that are speech-related. Clearly it is possible for vocalisation and manual movement to operate independently of each other, but when the hands and arms are not otherwise occupied, they are preferentially engaged with vocalisation, and when this occurs it seems to be at a deep and early stage of the process.

Mayberry and Jaques observe that "the fact that the temporal concordance between gesture and speech execution is always maintained throughout stuttered and fluent speech suggests that the complex neuromotor patterns of gesture and speech are coordinated and integrated prior to their production in extemporaneous expression" (p. 209). In fact, the concordance between speech and gesture does not appear to be instigated by the lexical components of the speech, but by cyclical motor control. Franz *et al.* (1992) found that when subjects were required to produce repetitive movements with the finger, forearm, and jaw, and the repetition of a syllable, there was significant correlation within subjects of the cycle duration of each task; i.e. each subject reached a "default" timing cycle for the repetitive muscular action, irrespective of which task was being performed. Franz *et al.* conclude that common timing processes are involved not only in movements of the limbs, but also in speech and non-speech movements of oral structures, and suggest this indicates a governing cognitive "muscular timing module" responsible for instigating all rhythmic cyclical muscular activity. This accords well with the findings of Alcock *et al.* (2000)

(described in Chapter 5) that the capacity to perform rhythms, both manually and verbally, forms an important foundation of oral/praxic ability. Franz *et al.* (1992) also found that simultaneously produced finger, arm and oral movements concur whilst carrying out repetitive non-linguistic motor repetition, which Mayberry and Jaques (2000) suggest could be the mechanism by which gesture-speech co-expression occurs. Whilst the concordance of gesture and speech control appears to be instigated on a motor-control basis rather than a lexical basis, the content of an utterance clearly has some influence on the nature of the gesture. As Mayberry and Jaques observe, the harmonized complex motor patterns of the gesture and speech system must ultimately, subsequently, be integrated with the output of the conceptual linguistic systems.

The question still remains of what cycles in vocal control are being integrated with the gestural cycles. Mayberry and Jaques suggest from their research, and from that of McClave (1994) and Nobe (1996) that it is the prosodic patterns of speech that contain the oscillatory cycles of muscular control. McClave found that it is with the nuclei of tone groups that gesture co-occurs, rather than with syllable stress as was commonly thought, and Mayberry and Jaques' (2000) observations of the onset of gesture and vocalisation in normal speakers and stutterers accord with those of McClave.

The implication of all this research is that the cross-modal coordination of gesture and speech doesn't require a central representation or linguistic input initially; such integration occurs whether utterances are linguistic or not. It seems that there is a cognitive rhythmic motor coordinator which instigates such muscular sequences irrespective of the musculature that is used, and that the complex patterns of gesture (finger, hand, arm, shoulder and joint musculature) and vocalisation (orofacial, laryngeal and respiratory musculature) are co-ordinated (Mayberry and Jaques, 2000). The cycles of vocalisation that are integrated with the gestural cycles are prosodic, tonal ones; in the case of speech, as opposed to non-linguistic vocalisation, linguistic meaning and narrative sentence structure are integrated into the gesture-speech system subsequently, before their physical manifestation.

6.4.2. Developmental findings

The vocal and gestural motor behaviour of infants and children accords well with this model. From the earliest stage, infants' babbling and gesture seem to be interrelated: Masataka (2000) found that in (Japanese) babies at the babbling stage their utterances were more syllable-like when they were able to move their arms than when they were not (Mayberry and Jaques, 2000). Locke *et al.* (1995) observed a strong association between the onset of babbling behaviour in infants and their exhibiting of lateralisation of motor control. They found that in children who had not yet begun to babble, and in those who had been babbling for some time, there was no evident dextral bias in their use of a rattle; however, in children who had recently begun to babble, their use of a rattle was much greater if placed in their right hand than in their left (see below for a description of babbling behaviour). Locke *et al.* suggest that this is as a consequence of a greater involvement of the left hemisphere (responsible for right-hand control) in the production of repetitive vocal-motor activity, such as is the basis for babbling; the development of the skill and coordination of the vocal sequences resulting in the increased development of the coordination of the manual gestural activity.

It is interesting to note the congruence of these findings with those of Trevarthen (1999) described below, that even congenitally blind infants make manual gestures in time with the rhythm of parental vocalisations (which, incidentally, are predominantly tonal with no linguistic meaning to the infant – see section 6.1.2, above). It seems the perception, as well as production, of vocalisation can be linked with gesture.

The gestures made by babbling infants are not iconic; they are rhythmic and emotionally determined (Trevarthen, 1999; Falk, *in press*) movements which accompany the vocalisations, but do not add meaning or symbolism to them. As language begins to develop, the gestures become more intentional and instrumental, however (Messinger and Fogel, 1998). From the point when children first combine gesture with a meaningful word in the same utterance (around the age of 14-17 months), gesture–alone combinations begin to decline and synchronous gesture/speech combinations begin to increase (Butcher and Goldin-Meadow, 2000). Butcher and Goldin-Meadow (ibid.) also observed that after this integration of gesture and lexical speech had occurred, a novel type of combination emerged in which speech was combined with gesture conveying related but different information, in particular, combinations of object and action. It seems the integration of gesture and speech may provide the foundation for the combination of object and action, with either gestural object and verbal action, or gestural action and verbal object.

Furthermore, the first children to produce these gesture-word combinations were also the first to subsequently produce two-word combinations. In all the children, the former occurred before the latter. "…all of the children we have observed thus far were able to concatenate elements of a proposition across gesture and speech at a time when they were unable to accomplish this feat in speech alone" (Butcher and Goldin-Meadow, 2000, p. 254).

Shifting the emphasis on this observation, perhaps it is the case that gestural-vocal integration must be in place before the ability to integrate two words, and more specifically, noun and verb or adjective, can emerge. At least, it seems that the integration of two concepts (e.g. object and action) occurs across two media before it occurs within a single medium, and it is possible that the former is an important stepping-stone toward the latter.

This evidence suggests that the interrelationship between manual gesture and vocalisation, both production and perception, is a deep one. That the two functions not only seem to co-occur in development but can also interfere with each other further reiterates this relationship. Feyereisen (1997) found that (adult) subjects carrying out a manual gestural task had a longer vocal onset time when asked to concurrently carry out a verbal task than those who were not carrying out a manual task at the same time, supporting the idea that there is a competition for shared resources between gestural and speech-production systems (Feyereisen, 1997). Interestingly, no delay to manual initiation time was observed when subjects were concurrently carrying out a verbal task, perhaps suggesting that verbal task initiation co-opts some neurological capacity already in use for manual control, and not vice versa.

6.4.3. Gesture in language evolution

Considering the size of the body of research investigating language origins, and of that regarding interdependence of gesture and vocalisations, it is surprising that the twain have not more often met. There have been some such connections made, however, leading some authors to suggest that manual gesture formed the foundation for syntactic elements of language (e.g. Hewes, 1973/1992; Corballis, 1992; Armstrong *et al.*, 1994). It is not the purpose of this thesis to explore the origins of fully developed language *per se*, so this issue will not occupy as much space as it undoubtedly could; it has been, and continues to be, extensively explored and debated elsewhere across a number of disciplines. However, as illustrated above, connections between vocalisation, manual and corporeal movement are important in any form of vocal behaviour, and probably have a long evolutionary history, pre-dating modern linguistic behaviour. Furthermore in suggesting shared neural and behavioural foundations for music and language in the form of emotive vocalisations, it would be remiss not to explore how each specialised form of vocalisation could have emerged from that foundation.

Clearly, manual, digital and corporeal movement can communicate important information, either as an adjunct to vocal language, or on their own. As Hewes (1994) observes, "Gestures of the hands, fingers and other upper body parts, even when they are not elements in a developed sign language, can indicate locations and directions of movement, the shapes and sizes of things, and the effects of actions on other things in ways that vocal sounds not part of a vocal language cannot match" (p. 361). This quality of gesture, and its co-occurrence with vocalisation, has led a few authors to suggest gestural origins to elements of linguistic behaviour. A particularly well-received paper of recent years is that of Armstrong, Stokoe and Wilcox (1994).

Armstrong *et al.* (1994) propose that, in addition to the development of vocal communication, the syntactic referential elements of language may have their foundations in gesture. Their thesis is really to identify a mechanism for explaining the introduction of sentence formation into the evolutionary model that already includes vocal behaviours; as they put it, it was with the origin of syntax that "naming" was transformed into language. There does seem to be an inherent presupposition that "naming" was the original role of vocalisations, rather than communication of affective state, i.e. that utterances were associated abstractly with real objects or phenomena, rather than with emotional expression. In fact, both these elements of the content would have been important; as the primatological and neurological cases discussed above illustrate, emotional content would certainly have preceded abstract reference in importance in vocalisation, but the latter would likely be an element of speech before syntax could become important.

Armstrong *et al.* argue that (some) gestures can be construed as either words or sentences, and point out that a single gesture can represent an entire sentence's meaning. An example they give is of the American Sign Language (ASL) gesture for "seize": "…the upper arm… rotates at the shoulder to bring the forearm and hand across in front of the body until the moving hand closes around the upright forefinger of the other hand" (p. 355).

This is an iconic gesture, actually representing the activity being communicated, but Armstrong *et al.* caution against missing the cognitive importance of such communication: the single gesture can have a subject, a verb and a direct object, or agent, action and patient, in semantic terms, and to produce or interpret such a gesture requires cognitive foundations that can process these relationships. They propose that such "word/sentences, in the form of visible gestures, could have provided the behavioral building-blocks associated with neuronal group structures for constructing syntax incrementally, both behaviourally and neurologically" (p. 356).

They propose that gestural complexity, rather than vocal complexity is a better candidate for instigating such development for a couple of reasons. Lieberman (1991) has suggested that the neural mechanisms for syntax might have had their origins in the evolution of rapid coherent vocal articulatory movements. Whilst they acknowledge this connection between syntax and fine control of muscular complexes, Armstrong *et al.* (1994) propose that instead syntax arose from analysis by members of a hominid species of complex *gestural* muscular actions, as they have a greater "perceptual salience" (p. 356), i.e. a greater availability for analysis than do vocal muscular movements.

The second reason that they propose for a greater antiquity to gestural rather than vocal muscular movements as origins of syntactic capabilities is that they claim we see a large brain in hominid lineages before we see a modern vocal tract. In their words "if language-cum-syntax has been built up from gestural embryo word/sentences, then it becomes easier to explain the evolution of a large brain prior to the appearance of the current configuration of the vocal tract. Explaining the large size of the brain and its gradual increase prior to the appearance of anatomically modern *H. sapiens* is a major problem if the capacity for syntax arrived by mutation at that time" (p. 356).

There are issues with both these rationales. In the case of the first, the perceptual salience of gestures for analysis, why rely on the *interpretive* aspect of the gesture to explain the development of the neurological complexes? For a gesture to have perceptual salience, it must first have some expressive purpose. Surely production (and the neurological complexes for it), even at a most rudimentary level, must occur before perception can begin to be useful. It is true that for a gesture or vocalisation to be meaningful it must be both produced and perceived, but this is not a chicken and egg situation. In evolutionary terms, it seems much less likely that an individual would develop an ability to analyse a type of signal that is not being produced by its contemporaries (and for that form of signalling to subsequently occur), than for an individual to develop the ability to attempt to signal in some way, and for that form of signalling to subsequently come to be understood by its contemporaries.

If the assertion were that syntax arose as a consequence of the *production* of manual gesture, because manual muscular movements are somehow more salient than are vocal muscular movements in inducing that neurology to arise, then the argument would be more acceptable. However, this is not the way that the case is put by Armstrong *et al.*. They seem to propose that syntax became established out of *interpretation* of such movements. If the assertion were that syntax would be more likely to be understood initially in gestural proto-word/sentences than it would in vocalisations, with both occurring contiguously, then this too might be reasonable; ultimately, this is their implication, in that they are justifying why manual control should have taken precedence over vocal control as the instigator of syntax. Despite this criticism of their rationale, there is much of value in the evidence and analysis; an interpretation of Armstrong *et al.*'s evidence can be proposed that avoids these issues. Perhaps, in the co-occurrence of manual and vocal communication, any emergence of syntactic structure in either vocal or manual system during human evolution would have been most likely to become perpetuated and developed in the manual medium initially, as this is the more perceptually salient of the two. In other words, when the neurological foundations for syntactic structure in the *planning* of complex muscular movements emerged, they would have manifested themselves in both vocal and manual sequences, which co-occur. Such syntactic capability would then have become, at least at first, stimulated into further elaboration in the manual medium as this is the medium most available to interpretation by others. This may seem very similar to what Armstrong *et al.* propose, but the difference is that they appear to argue that the neurological foundations *emerged* initially out of the perception rather than the production of such sequences. There would seem to be no problem, on the other hand, of proposing that "perceptual salience" resulted in gestural syntax subsequently leading to the ability being *developed* further in that medium.

As for the second point, regarding brain size and explaining the evolution of a large brain prior to the appearance of the current configuration of the vocal tract, this is a non-issue. The evidence considered in Chapter 4 suggests that, like the brain, there has been gradual (or at least incremental) evolution of the vocal tract, until essentially modern configuration was reached perhaps around 300,000 years ago with *Homo heidelbergensis*. That is to say, there is plentiful evidence for vocal behaviour of a potentially high complexity before the emergence of modern language ability. Far from there being a problem of explaining the appearance of a large brain before a modern vocal tract, the problem may well be the reverse. At the least, they appear to have developed in tandem. In any case, if one does not hold with a Chomskyan sudden emergence of syntax (and Armstrong *et al.* profess not to hold that position), it is not just the current configuration of the vocal tract that is relevant to the discussion of the evolution of such a capability, its intervening stages are just as relevant.

Neither of these criticisms undermine the possibility that gestural communication played an important, or even dominant, role in the development of syntactic neurological structures; the evidence for the importance of gesture in communication is convincing. However, the criticisms do (in the current author's opinion) undermine two of Armstrong *et al.*'s rationales for giving gestural muscular control *precedence* over vocal muscular control at this stage of development, and their underestimation of the possible complexity of vocal communication prior to linguistic vocalisation.

It is worth noting too that the capacity of gestures to act as sentence/words is a property that is also shared by some vocalisations. Another example of a gesture representing a whole sentence is given by Armstrong *et al.* (1994). In ASL, when a signer responds to another by bending their elbow and touching their forehead with their fingertips, this means “I know that” (not just simply the verb “know” as many ASL dictionaries imply) (Armstong *et al.*, 1994). This can also be true of non-linguistic vocal utterances, however. Such utterances can also function as an “embryo word/sentence” in this sense, and perhaps could have done so during hominid evolution. For example, a particular tonal contour of “mmm” (usually involuntarily accompanied by certain body language, though not essentially so to its communication) would communicate “I know that” as opposed to “I agree with that”, “I understand that”, “I’m not sure about that”, “I grudgingly accept that”, “what did you say?”/”I didn’t hear that” or “I like that”. The number of such meanings potentially attributable to an utterance would increase with greater vocal tract versatility and control. In expressing a state of mind, or reaction, such an utterance expresses what can, using lexical language, only be communicated in a sentence using several words.

That such different expressions of “mmm” are also accompanied by quite specific body language, involuntarily on the part of the speaker, betrays their affective origin in a system in which vocalisation and corporeal expression, or to put it another way, vocal and corporeal gesture, are intimately linked. As Blount (1994) puts it, “Vocalised phonetic ‘gestures’ are accompanied by a priming of the motor system, producing coordinated movements of other parts of the body. Perhaps the capacity to align the components is what evolved as underlying the capacity for gesture, vocal or otherwise” (p. 359) (cf. Clynes’ (1977) “sentic modulation” and Donald’s (1991) “mimesis”, discussed in Chapter 8). This accords well with the evidence discussed above regarding the interaction of gesture and vocalisation; it would also be at least equally applicable to body movements accompanying musical vocalisation, and vocalisation accompanying dance. The fact that seeing body language is not *essential* to understanding an utterance (although potentially helpful in disambiguating it) is perhaps an indication of the degree of sensitivity to vocal contour, and of control over it, that we have since developed.

In reality our spoken language is a combination of lexical and prosodic modes of communication. The tonal contour of an emphatic “I *know* that!” seems very similar to the tonal contour of an “*mmm*!” communicating the same thing, and the same can be said of a sympathetic or a placatory expression of the same. The lexical form of this type of utterance is most useful in disambiguating it in a sensorally deprived context, such as over a distance. In contrast, it seems that the more intimate the circumstance, the more inclined we are to use the non-linguistic, essential, tonal, efficient “mmm” in preference to the linguistic equivalent. This may be evident, for example, when tasting (or smelling) fine food, when one might utter “*mmm*!”, in preference or precedence to *saying* “I *like* that!”. The same can be said of the familiar vocal response to a firework display. Such vocalisations are visceral, and can be more immediate, more expressive, and consequently a more valuable currency of communication; furthermore, they often cross linguistic and cultural boundaries.

Of course, expressing personal state is only one of the many uses to which *lexical* language is put. It may be noted that all the various meanings attributed to the utterance “mmm”, above, express relations (in particular, affective reaction) between the first person and an object or concept, “I ____ that”, and this is a limitation of that form of utterance. Interestingly, whilst certain non-linguistic vocalisations seem to be able to communicate relations between self and object/other, they cannot do the same for other and object/self. It is very difficult to express any equivalent statement of “you like that”, “you know that”, “you agree with that” etc. in a single tonal utterance in the same way as described above for first person statements of the same. The closest we can get is a tonal contour of “mmm” which expresses “you like that?” or “you understand that?”. These are not the equivalent of statements of the relation between other and object/self, but are invitations to the other to utter their own such statement about their own state.

So whilst these non-linguistic vocal expressions can act as statements about relationships/reactions between the self and other, they are limited to that purpose. However, it is quite reasonable to suppose that such utterances formed the first *vocalised* form of *concepts* relating the self to object/other, and that this ability to vocalise such concepts was subsequently built upon by more complex vocal and gestural sequences. Again, this does not undermine Armstrong *et al.*’s (1994) illustration of the importance of syntactic properties of gestural communication, but it shows that, within a limited repertoire, simple non-verbal vocalisations can communicate similar relations between self and other. Hopefully it is also a reminder of quite how expressive such simple non-linguistic vocalisations can be. A fundamental difference between these vocalisations and the gestures of which Armstrong *et al.* speak, however, is that such utterances do not have separate components that are being interrelated; for Armstrong *et al.* (1994) the important feature of gestural signs is that they have a potential for decomposition into meaningful agent/action subunits. As such it seems quite plausible that they could have bridged the gap between the vocalisations described above, and multi-part vocalisations with syntactic structure; indeed, this seems to be what occurs ontogenically (Butcher and Goldin-Meadow, 2000), as described above.

It is important to note that Armstrong *et al.*'s (1994) thesis is actually more moderate than it might at first seem. Although they underestimate the potential versatility and communicative power of complex vocalisations prior to full language, they are at pains to reiterate that their proposal is not for a *gestural origin of language*, but that gesture may have played an important part in the development of the cognition necessary for modern language. They recognise the importance of social vocalisation from the outset in hominid evolution, and see no reason for positing a phylogenetic discontinuity between primate vocalisation and speech.

Armstrong *et al.* (1994) also propose that language had multiple causes and multiple purposes. What they ultimately suggest is that there was an evolutionary stage at which visible gestures took the lead "with respect to flexibility of output and, critically, the elaboration of syntax." (p. 358). Subsequently, "the cognitive apparatus necessary to support these developments would then have been available in spoken language as well." (p. 358). Underlying this is a belief that vocal "gesture" and manual "gesture" are essentially part of a continuum of finely-controlled body movements "broadcasting information to the (social) environment" (p. 358). It seems likely that vocal and manual fine muscular control were always contiguous, but that (in linguistic speech) the balance later shifted towards a vocal dominance. What Armstrong *et al.* propose is that there was an intervening period where the balance shifted toward manual gesture. This is in contrast to Hewes (1973/1992) who argues that "a preexisting gestural language system would have provided an easier pathway to vocal language than direct outgrowth of the 'emotional' use of vocalisations characteristic of non-human primates" (p. 65). This latter rationale makes an artificial distinction between gestural and vocal communication, and doesn't acknowledge the joint importance of both syntactic and prosodic elements to vocal language. What seems more likely from the evidence considered in the preceding sections is that both were used together, but that gestural complexity, and the integration of iconic gesture with vocalisation, could have led to the development of syntactic elements of language, whilst the prosodic components have their root in affective vocalisations. This scenario, and ultimately that of Armstrong *et al.* (1994), seems a reasonable proposition. Findings such as those of Alcock *et al.* (2000) (described in Chapter 5) that the capacity to perform rhythms, both manually and verbally, forms an important foundation of oral/praxic ability, provides further support for this idea, as do the findings of Franz *et al.* (1992) and Mayberry and Jaques (2000). Indeed, whilst all other primate vocalisations have tonal affective content, human vocalisations are unique in their rhythmic content – even the 26 other species that “sing” do not have rhythmic content (Geissman, 2000). Together these findings indicate that this feature of human vocalisation is something that has emerged along with planned complex sequences of vocalisation and related motor control. They highlight not only shared neurological foundations between the ability to execute linguistic and manual gestural sequences, but also the link between vocal and manual *rhythmic* capability.

6.5. Rhythm, corporeal movement and emotion

Whilst rhythm and melody are apparently processed in neurologically specialised areas of the brain which are somewhat independent of each other (as discussed in Chapter 5), there is also clearly some important integration of these systems, especially in the case of the interrelationship of rhythmic physiological movement, vocalisation and emotional state. The relationship between the instigation of vocalisation, emotional expression and corporeal muscular control extends beyond manual gesture accompanying speech. The evidence discussed above for the integration of vocal control and corporeal control provides several clues as to reasons for the interrelation of melodic, rhythmic and corporeal activity. One manifestation of this interrelationship is dance.

As Mitchell and Gallaher (2001) observe, music and dance are historically interdependent developments, with many common features: they are both temporally organised, and described in terms of rhythm, tempo, beat, pace, movement, choppiness or fluidity, for example. In fact, one rarely exists without some form of the other. Each of these properties can be observed and experienced across numerous modalities, in vision, audition and kinaesthesis (the proprioceptive feedback associated with voluntary body-movement).

Particular pieces of music can elicit consistent emotional responses in listeners, and movements that often occur in response to music, which are kinaesthetically experienced, can also be a part of emotional experience. Body posture and emotional state are strongly interrelated, as discussed in section 6.3. Our posture and movements can express a great deal about our emotional state, intentionally and unintentionally, and others’ body-posture and movements thus provide important cues as to their emotional state. As well as being able to observe such cues, we can empathically experience something of their emotional state in mirroring them with our own bodies.

Musicality and rhythmic movement, Trevarthen (1999) points out, involve deliberate control and sequencing of this system. This can result in a self-directed feedback from movement into emotional state and, importantly, feedback and interaction between individuals, in terms of synchrony of movement and of emotional state.

As discussed in Chapter 5, Trevarthen (1999) has explored a functional system existing as the brain stem, basal ganglia and the limbic structures of the Emotional Motor System, which he calls the Intrinsic Motor Formation (IMF). This is responsible for integrating functions of attention, learning, and physiological expression of emotion, including the synchronisation and co-ordination of sequences of movements; it seems to be the case that there is a close interrelationship between the emotional-controlling elements of the limbic system and the areas responsible for the co-ordination of motor sequences and posture. This system is active automatically in all types of interpersonal interaction, but is made deliberate use of in musical activities.

Findings relating to facial expression and proprioceptive freedback may have important implications here. As discussed in section 6.1.2, facial expression and emotional content of vocalisation are closely related, with particular orofacial musculature configurations resulting in particular vocal qualities (and the production of certain vocal tones and contours being dependent upon the instigation of certain facial expressions). Furthermore, core emotional facial expressions seem to be human universals (e.g. Ekman and Friesen, 1971; Ekman, 1980). What is especially interesting is that feedback from facial expression actually *effects* our emotional state. Levenson, Ekman and Friesen (1990) found that asking subjects to carry out various facial movements to create the expressions of fear, anger, surprise, disgust, happiness and sadness caused distinct changes to the activity of the autonomic nervous system, such as changes in heart rate, galvanic skin response (skin conductivity due to perspiration), and temperature. The subjects were not informed as to the facial expression being replicated, they were simply told to carry out a sequence of muscular movements which resulted in the formation of a given expression, for example, to raise their eyebrows, then pull them together, whilst tightening the skin under their eyes and stretching their lips horizontally (producing an expression of fear). This particular expression resulted in increased heart rate and reduced skin temperature, whilst a facial expression of anger increased both heart rate and skin temperature, and a happy espression decreased heart rate whilst leaving skin temperature unaffected (Carlson, 1994).

Furthermore, humans have an innate tendency to imitate the facial expression of others with whom they are interacting. Even at the age of 36 hours, infants tend to imitate facial expressions that they see (Field *et al.*, 1982); clearly this is an innate, and not a socially conditioned, tendency, and results in some "contagion" of the emotion being expressed. Wild, Erb and Bartels (2001) found that even when exposed to images of another's facial expression (showing sadness, disgust, happiness, anger, surprise, fear or pleasure) for as little as 500 m/s, subjects reported corresponding changes in emotional state. Carlson (1994) suggests that "…imitation provides one of the channels by which organisms communicate their emotions. For example, if we see someone looking sad, we tend to assume a sad expression ourselves. The feedback from our own expression helps put us in the other person's place and makes us more likely to respond with solace or assistance." (p. 351). The findings of Kraut and Johnston (1979) have already been outlined regarding the increased tendency of people to smile in the presence of others, and the considerable implications for selective advantages of the ability to empathise with conspecifics have also been elaborated (Schmidt and Cohn, 2001; Sloboda and Juslin, 2001).

The implications of all this for the present discussion are threefold:

Firstly, the production of particular vocal tones normally relies upon adopting certain facial expressions, which correlate with the emotion being expressed;

Secondly, the production of a particular facial expression whilst producing a particular vocalisation will result in some degree of feedback which actually affects emotional state;

Thirdly, there is a natural inclination to mimic such expressions and to feel such associated emotions; i.e. such physiological-emotional feedback may occur not only during production, but also during *perception* of such a stimulus.

The non-verbal corporeal communication of affective state is not limited to facial expression, but incorporates the whole body and posture; the work of authors such as Trevarthen indicates that these above findings of feedback and "contagion" of emotion, and their implications, apply equally to whole-body expressions of affective state. This may go a considerable way towards explaining physical and emotional response to rhythmic and dance stimuli, both proprioceptively and visually.

Other research that points in this direction, with specific reference to music and dance, has also been carried out. Krummhansl and Schenck (1997) found that subjects who listened to a piece of music and different subjects who observed a dance to that piece of music, showed great concordance both within and between the two conditions as to their interpretation of the timing of tension and emotion in each, and the emotion that they experienced at given points. That the dancer interpreted the music in such a way that crescendos in the music corresponded with high movements, high notes with leaps and staccato and legato sequences with matching physical movements perhaps renders that congruence in interpretation between the two less surprising, given the literality, or *iconicity*, of the interpretation; it nevertheless remains interesting that such physical movements are the (apparently) obvious way of physically representing such auditory phenomena, and that interpretation of them as such occurs. The main implication of the findings relevant to the current discussion is that both media equally represent tension, release and particular emotions, underlining the cross-modality of such affective expression and interpretation. The affective content is apparently interpreted equivalently in visual, auditory and kinaesthetic media.

As with facial expression, this is not restricted to perception, but is also a two-way process. Mitchell and Gallagher (2001) review a considerable body of research which illustrates that “music prompts kinesthetic (motor) responses in both children and adults that often match some aspect of the music” (Mitchell and Gallaher, 2001, p. 66), and that “kinaesthetically experienced movements that sometimes occur in response to music can also be a normal part of emotions (*ibid.* p. 66/67). This is a phenomenon with which most people are familiar already, of course, which is not to diminish the value of such research, but to add to it. That this co-ordination of emotional communication and of rhythmic behaviour is innate is evidenced by studies of infant-parent interactions. As mentioned in Chapter 5, infants can accompany emotive-tonal prosodic infant-directed speech, or in other circumstances singing, with rhythmic body movements which co-ordinate with the rhythm of the vocalisation (Trevarthen, 1999). This is an innate capacity, as it can occur even in congenitally blind infants, who cannot possibly be imitating the movement of the parent. In terms of potential selective benefits to such tendencies in adult life, as also noted in the preceding chapter, McNeill (1995) reports how the use of rhythm alone may be significant in inducing group cohesiveness.

Merker (1999) has suggested an alternative, or additional, selective scenario. He suggests that co-ordinated synchronous rhythmic and melodic group behaviour could be derived from synchronous chorusing activities carried out to attract female hominids from neighbouring territories. This is on the premise that, like chimpanzees, Lower- and Middle-Palaeolithic populations may have been exogamously organised (females leave their natal population group and join another where they settle to raise young). In this situation, males can make distance-calls to attract prospective mates from neighbouring groups; Merker points out that if this activity was carried out by several individuals at once, synchronising their calls (as is the case with several insect species such as crickets and, visually, glow-worms), then the area reached by the sound from the call increases proportionally. Having attracted females from neighbouring groups in this way, it would then be in the interest of individual members of the calling group to have an individual distinct calling activity to attract females to themselves rather than to the rival males in their group.

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

6.6. Chapter summary and conclusions

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

6.6.1. Tonal and emotional content of vocalisation and melody

In the emotional state that they express, some vocal sounds and frequency changes are fundamental, invariant across cultures, and even species. Part of the reason for this is that facial expression has a fundamental influence on vocal quality, as orofacial musculature helps determine properties of vocalisation such as frequency and vowel duration. Given the universality and innateness of certain fundamental facial expressions and the correspondence between these and characteristics of vocalisations, we can also expect characteristics of particular emotional vocalisations to be universal and innate too.

This correlation between facial expression and vocal quality also apparently has an ancient provenance, being shared by our nearest primate relatives, and similar correlations between vocal sound and emotional expression are also exhibited by several other species. We use facial affect and vocal affect to inform about the content of each other, inter-dependent in both production and perception.

There is an additional factor that comes into play in the relationship between emotion, vocalisation and facial expression: the production of a particular facial expression whilst producing a particular vocalisation results in some degree of feedback which actually affects emotional state. Furthermore, there is a natural inclination to mimic such expressions and to consequently feel such associated emotions; i.e. this physiological-emotional feedback may occur not only during production, but also during *perception* of such a stimulus. This means that in producing, and even to an extent in perceiving, a particular sound we generate some emotional response in ourselves due to the kinaesthetic feedback from the physiology required to produce that sound.

The universality of the vocal sounds and frequency changes that express particular emotions is especially evident in infant-directed (ID) speech, as the exaggeration of these elements of the vocalisation is a characteristic feature of ID speech. ID vocalisations can tell us a great deal about the nature and role of the prosodic elements of speech and their relationship to musical melodic behaviour. It should be noted that the characteristic features of ID speech are also characteristics of the tonal (non-linguistic) elements of adult-directed (AD) speech, and they apparently share the same foundations and roles in emotional expression.

Many of the properties of ID speech are shared with music. There are numerous parallels, in terms of variable pitch contour, high rhythmicity, repetitive motifs, and the communication of affect, modulation of arousal, and eliciting of attention and affective response. Vocalisations of pre-school children in particular are often difficult to classify as either linguistic or musical.

It seems likely that infant-directed vocalisations are a specialised perpetuation of a sophisticated form of non-linguistic interpersonal interaction which was used earlier between all individuals, and that the shared prosodic pitch and tempo-related properties of emotional vocalisation (ID and AD) and music are not borrowed from one to the other, in either direction, but are, and always have been, a shared fundamental component of both.

Music elicits emotional responses in listeners for a variety of reasons. In eliciting an emotional reaction, music can generate genuine physiological responses such as changes in respiration, heart-rate, temperature and tingling. Some emotional responses can be elicited as a consequence of learned association with particular circumstances, for example, others as a consequence of direct resemblance to ecological phenomena to which we have instinctive or conditioned responses. A third form of emotional response occurs as a consequence of processing by, and stimulation of, auditory and kinaesthetic mechanisms associated with interpersonal interaction. We can react to and process music in many of the same ways as we do interaction with a person; it shares many of the properties of such interaction, and is apparently processed by the same mechanisms. Consequently, we can (although don't necessarily) emotionally empathise with music, if it elicits the same emotion as is being expressed. The context in which we experience music is also very important in determining the emotion, and intensity of emotion, experienced. Especially important in this respect is the social context – the extent to which the experience is shared with others.

6.6.2. Vocalisation, corporeal expression and rhythm

Gesture and speech are inter-dependent. Both are affected simultaneously in stutterers, and gestural and vocal behaviours are interrelated from the earliest babbling in infants. They can operate independently, but when the upper limbs are otherwise unoccupied they are sequestered into speech-related gesture. It seems that there is a cognitive rhythmic motor coordinator which instigates such muscular sequences irrespective of the musculature that is used, and that the complex patterns of muscular movement of gesture (finger, hand, arm, shoulder and joint musculature) and vocalisation (orofacial, laryngeal and respiratory musculature) are co-ordinated. The concordance between gesture and speech is instigated early in the vocalisation process, by cyclical motor control, with gestural movements being associated with the nuclei of tone groups - prosodic rhythm - rather than the lexical elements of speech. This is also evident in the gestures accompanying infant vocalisations. Greater lateralisation of gesture occurs after babbling begins, implying left-hemisphere development at this time in the integration of the two actions. These earliest gestures are emotive and rhythmic rather than iconic (corresponding with the above finding that gesture corresponds with prosodic rhythm rather than lexical content) and only start to be used iconically and in combination with words when lexical behaviour has started to develop. In the case of speech, (as opposed to other, non-linguistic, vocalisations), linguistic meaning and narrative sentence structure are integrated into the gesture-speech system after the integration of gesture with prosody, but before their physical manifestation.

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

Production of complex vocalisation relies on priming of the whole motor system. Particular non-lexical vocal utterances (and non-linguistic content of speech) are accompanied by quite specific involuntary body-language; they share an affective origin in a system in which vocalisation and corporeal expression, or to put it another way, vocal and corporeal gesture, are intimately linked. Note that this is at least equally applicable to body movements accompanying musical vocalisation, and vocalisation accompanying dance. There are not only shared neurological foundations between the ability to execute vocal and manual gestural sequences, but also a link between vocal and manual *rhythmic* capability. Humans are the only primates to carry out vocalisations with rhythmic content, and the capacity to perform rhythms, both manually and verbally, forms an important component of oral/praxic ability – detriments to one result in detriment to the other. This integration occurs whether utterances are linguistic or not.

Whilst rhythm and melody are apparently produced and processed in neurologically specialised and distinct areas of the brain which are somewhat independent of each other (as illustrated in Chapter 5), there is also clearly important integration of these systems, with rhythmic muscular movements being coordinated with prosodic elements of vocalisation in their production. Specifically, they are inter-dependent in the planning and execution of sequences of muscular movement associated with instigation of vocalisation, rhythmic physiological movement, and expression of emotional state in these media. The production and perception of tonal *content* in vocalisations does not appear to require any input from rhythm-controlling systems in the left hemisphere (see Chapter 5) but the planning and execution of sequences of such vocalisations does.

6.6.3. The role of this system in evolution

There is evidence that group size increased over the course of hominid evolution, and it has been suggested that complex vocalisation and language became increasingly important as a consequence. The role of complex vocalisation of affect, and sensitivity to the prosodic content of vocalisations, has already been elaborated; there is much evidence to suggest that full language also remains important in this role.

Linguistic speech and melody share common features in the form of intonational phrasing and combinatorial syntax, and share a “phonological level” of meaning, an “acoustic mode” involving emotive meaning and interpretation (Brown, 2000). Whilst music often does use the full range of the vocal tract, full language does not need to use the whole range of the vocal tract to communicate effectively, as linguistic structures disambiguate meaning. This would not have been the case for a pre-syntactic, pre-lexical proto-language.

Discrete tonal-contoured units expressive of affective state would have become combined into progressively larger globally-contoured units. What must have emerged in the course of the evolution of pitched-contoured vocalisations is an *increased range and control* of pitch contour, allowing greater vocal versatility, expressiveness, and thus efficiency, in proto-linguistic vocal affective communication. This needn’t have emerged as a new system of vocalisation initially, however, instead building upon the limited pitch control already used for emotive-tonal-social vocalisation by higher primates. The selective advantages associated with the possession of such capabilities resulted in the continual evolution of the vocal tract, and control over it, through the lowering of the larynx and increased innervation of the associated laryngeal and upper vocal tract musculature (see Chapter 4).

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

Whilst extremely important in the type of information they express, non-verbal vocal utterances, in meaning, are limited to acting as expressions of personal state and relationships between the self and other. In terms of the emergence of language from such vocalisations, however, it seems likely that such utterances formed the first *vocalised* form of *conceptualisations* relating the self to object/other, and that this ability to vocalise such concepts was subsequently built upon by more complex vocal and gestural sequences. In fact, it is at that stage that the syntactic potential of the coupling of vocal and *iconic* gestural communication may have become fundamentally important, bridging the gap between non-lexical vocalisations of affect, and multi-part vocalisations with syntactic structure and iconic associations.

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

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

Why "musical" behaviours specifically should have emerged onto these foundations, and why they might have persisted, is the subject matter of the following chapter.

CHAPTER 7

Rationales for the Evolution of Musical Behaviours

The preceding three chapters have been concerned with the evolution of the capacities which form fundamental components of musical behaviours and how and why these might have come to have formed the foundations for such behaviours. A separate question concerns the perpetuation of music as a fully-developed behaviour. Are there, or were there at some point in the past, selective advantages to musical behaviour as a whole? Does music only make use of functions and capabilities of the human anatomy and neurology that were already in place, or did the use of music, or early music-like activities, contribute to shaping these capabilities?

As mentioned at the start of Chapter 6, there have been several proposals in the past that language and music are closely related in terms of their origins, with some authors believing language to be the precursor of music, and that music makes use of mechanisms originally selected for by language (e.g. Pinker, 1997; Calvin, 1996) and some vice versa (e.g. Livingstone, 1973; Vaneechoutte and Skoyles, 1998*)*. The authors in the former category generally assert that musical abilities could not have been selected for in their own right. The preceding chapter has aimed to illustrate that neither of these is the case, but that language and music have a common foundation, from which both subsequently developed.

7.1. Non-adaptive origins of musical behaviours

Views advocating non-adaptive roots for musical behaviours have been prominent for much of the last 50 years (Brown *et al.* 2000). A couple of influential voices, one an (evolutionary) psychologist (Pinker, 1997) and the other a psychologist of music (Repp, 1991), have recently expressed this opinion. Repp takes a position typical of the dominant opinion of the post-war years. He states that

> “There is no reason to believe that there is a universally shared, innate basis for music perception. Although the possible survival value of music has often been speculated about (e.g., Roederer, 1984), music has not been around long enough to have shaped perceptual mechanisms over thousands of generations. Clearly, music is a cultural artefact, and knowledge about it must be acquired. Moreover, in

> contrast to speech, this knowledge is acquired relatively slowly and not equally by all individuals of a given culture.” (Repp, 1991, p.260)

When Repp uses the term “music” it would seem that he is actually referring to “the culturally-specific manifestations of musical activity”. If this phrase is substituted into his statement, then it becomes a little more credible, though no more useful. Equating “music” and musical ability with the culturally-determined aspects of musical behaviours is spurious, and in terms of the origins of such behaviours, not useful (which is, inevitably, the point that Repp ends up making). If Repp’s (1991) statement is taken to refer to music *per se*, and capacities for music, there are a number of issues to be raised with his position. Firstly, Repp’s assertions that there is no evidence for a universally shared, innate basis for music perception and that music is clearly a cultural artefact about which all knowledge must be acquired, are fundamentally contradicted by the evidence considered in the Chapters 4, 5 and 6. Whether or not these abilities were selectively favoured specifically for proto-musical behaviour, they are clearly fundamental components of musical behaviour, emphasising that there is more to musical capability than the cultural aspects. The ethnogenic elements, such as particular styles and instrumental skills, are only part of what defines musical ability and performance, and, as with specific languages and cultural traditions, shape the manifestation of what are biological capacities.

Secondly, there is an explicit assumption that music has not been around for long - there is no acknowledgement of the existence of proto-musical behaviours, or their potential importance. Repp must thus believe that there was no music before instrumentation, and that music came into existence (presumably in the Upper Palaeolithic) making use exclusively of perceptual mechanisms that had already been shaped through use for other purposes. However, the point of most research regarding the evolution of musical behaviour must, *by definition*, be concerned with the time *before* we see the first evidence for instrumentation. The assertion that it “has not been around long enough” for it to shape perceptual mechanisms disregards the possibility that there was any period of development of musical behaviour prior to when we see evidence for instrumentation in the Upper Palaeolithic. If the use of music had had any influence on the shaping of perceptual mechanisms it would have had to have happened before fully-realised musical behaviour occurred. If musical behaviours had come into existence fully formed as Repp’s argument suggests, the perceptual mechanisms could not have *needed* to accommodate it anyway, which makes that part of his argument rather circular.

Steven Pinker’s position is somewhat different from that of Repp. Pinker is a strong advocate of adaptive foundations for many human behaviours, but takes the view that music simply makes use of functions and capacities of the human anatomy and neurology that were already in place. Pinker believes that music is a "technology" that we have created, to make stimulating use of existing innate faculties such as language, emotion and fine motor-control, and that these faculties evolved independently through other selective pressures.

However, as has been illustrated in the preceding chapters, whilst some of the mechanisms used by music indeed evolved under other selective pressures, many were functionally inter-dependent and the contiguous use of them could in itself result in the development of behaviours which came to be musical. Further, in terms of selective advantages to such behaviours, these may not be limited to the shaping of the mechanisms of perception and production: there is also the possibility that the use of certain faculties in the combinations required by musical behaviours might confer selective advantages in their own right. For example, Cross (1999b) suggests that the use of the mechanisms that form the foundations of musical behaviours, in the combinations demanded by musical production and processing, could itself have selective advantages for an individual - such use would give these faculties periodic "work-outs". In short, if we are looking for reasons why music is still used, or why its use might have spread, then we can indeed look for such reasons since the time from which we know it existed.

7.2. Possible selective reasons for the use of music

It has often been remarked that music is hardly essential for survival, the implication being that there is no reason to believe that it could have been selectively important. However, this is to misunderstand the processes of evolution by selection, be it natural or sexual selection. A behaviour or trait need not be *essential* for survival to confer a selective benefit, it need only confer a slightly greater likelihood of survival to procreation, or greater rate of procreation (thus perpetuating that trait), than would otherwise be the case.

As has been noted, in talking about selection for musical behaviours we can either be talking about selection for music itself, as a behavioural package, which could occur at a group or an individual level, or we can be talking about selection for each of the foundations of musical abilities, which could have other selective pressures acting upon them as a factor of other functions that they fulfil. Previous writings on the subject of music evolution have often not made this distinction, and it is not always clear which process is being argued for.

The elements of the latter case, the evolution of capacities for music, were the subject matter of the preceding chapter, in which it was argued that musical behaviours can be valuable because they use mechanisms that are socially valuable, grounded in, and fundamentally important in interaction and the formation of social relationships. That musical and proto-musical behaviours use mechanisms that have been selected for prior to the existence of music does not diminish the importance of musical behaviours, from an evolutionary perspective. Music is not simply *making use of* existing cognitive mechanisms that already existed, selected for already, it is *a development of* those mechanisms. Whilst those foundations may initially have been selectively favoured as a consequence of their fulfilment of particular purposes, music developed within the context of those uses, fulfilling some of those same purposes.

Further, such a suite of related capacities could continue to develop in tandem, with interdependence increasing between them, whilst still fulfilling other functions. Such co-use of mechanisms previously related in their fulfilment of different functions could then unite them functionally in this new behavioural manifestation; subsequently they could be selected for in tandem as part of a behavioural system, changes in one "bootstrapping" changes in the other. As Huron (2001) puts it, "If music is an evolutionary adaptation, then it is likely to have a complex genesis. Any musical adaptation is likely to be built on several other adaptations that might be described as premusical or protomusical. Moreover, the nebulous rubric *music* may represent several adaptations, and these adaptations may involve complex co-evolutionary patterns with culture" (Huron, 2001, p. 44).

In addressing the former question, of selection for musical behaviours themselves, it remains to ask what reasons there could be for their perpetuation, and the development of their complexity from early forms of music or proto-music. The idea that music could have evolutionary origins and selective benefits was widely speculated upon in the early part of the twentieth century, in the light of increasing bodies of ethnographic research and the development of Darwinian theory. This approach fell rapidly out of favour in the years after the Second World War, for political more than practical reasons, with the rejection of biological and universalist ideas in anthropological and musicological fields (Brown *et al.*, 2000).

7.2.1. Group cohesion

Roederer (1984) was one of the first authors to re-open the field of investigation, suggesting several potentially selectively beneficial traits of music, which have been explored further by several other authors (with or without reference to Roederer) since. He pre-empts Papousek (1996) and Dissanayake (2000) by over a decade, suggesting that music may be developed from mother-infant communication. This type of interaction, he suggests, could have selective benefits by strengthening emotional bonds between mother and infant, and by providing a sort of pre-training for the ability to extract information from the musical components of speech, such as vowels, inflections, or pitch-cues found in oriental languages. The idea of the development of music from infant-mother interaction has been discussed at length already. Roederer also suggests that music has an important ability to transmit emotional information to whole groups of people at once, equalising the emotional state of the group, which results in a bonding effect between the group-members. The foundations, and interrelation, of musical vocalisations and rhythmic activity in interpersonal emotional interaction and fine muscular control have been explored in considerable depth in Chapter 6, as well as the effect of experiencing such stimuli under larger group circumstances. It should hopefully be evident from those investigations that the feelings and physiological changes that can be evoked by music draw upon, and stimulate mechanisms of interpersonal emotional empathy, amongst other consequences. The potential benefits of these effects are explored further in the following sections.

There is a variety of evidence from other disciplines to support the hypothesis that music can have direct effects on the nature and efficaciousness of social interactions. Ethnographic and ethological evidence shows that musical activity is predominantly social. As explored in Chapter 2, the vast majority of musical activity in modern hunter-gatherer societies, and other traditional societies, is communal – a group social activity; until recent history the same applied to Western musical activity too.

At a biological, rather than behavioural, level, musical experience has been linked with the release and action of various hormones. Freeman (1995) reports that the hormone oxytocin seems to aid in the formation of strong positive emotional memories, and in the supplanting of negative emotional memories. Its effects seem to be strongest during major activations of the limbic system, such as during trauma or ecstasy. It is also released in females whilst lactating, and in males and females following sexual orgasm, and seems to thus fulfil an important role in interpersonal bonding (both pair-bonding and mother-infant bonding). Freeman (1995) suggests that another circumstance under which oxytocin is released is whilst listening to music. As discussed in Chapters 5 and 6, the emotional limbic system seems to be particularly activated during learning, and during emotional and rhythmic motor activities. Given this, Freeman's suggestion seems likely, and not only during the perception, but especially the production of musical behaviour. If oxytocin is indeed released during musical activity, this provides a good neurological reason why music can fulfil important roles in the formation of social bonds, both in interpersonal interaction and in group musical activities, such as crowd chants (Huron, 2001).

In addition to clearly being able to regulate arousal in large group situations, Fukui (reported in Huron, 2001) has found that subjects listening to their own choice of music experienced reduced levels of testosterone; lower testosterone levels correlate with reduced aggression, competitiveness, and conflict. As Huron points out, further experiments that manipulate the music listeners are exposed to (rather than it simply being their favourite) would add considerably to the applicability of these findings. In fact, research into hormonal responses to music appears to be in its infancy at this stage, and is clearly an area of investigation that could prove most fruitful in the future regarding connections between musical and other interactive activities.

The idea of a social basis for music has also been proposed by Sloboda (1985). He suggests that all cultures require organisation for survival, and that "modern" cultures have "many complex artefacts that help us to externalise and objectify the organizations we need and value" (Sloboda, 1985, p. 267), but that in non-literate societies the organisational structures must often be expressed in terms of the ways that people interact with each other. He proposes that music can be used as a mnemonic framework for the knowledge of a community and as a way of expressing the structure of social relations.

Sadly, Sloboda does not cite any ethnographic examples to affirm these assertions, but an example of the former is certainly provided by the Australian Aborigines' musical tradition, and by some of those of the Yupik, described in Chapter 2. It is also true that the musical performances observed in the Plains Indians and African Pygmies (also discussed in Chapter 2), whilst communal, do nevertheless tend to stress roles in the performance for different genders and individuals (e.g. hunters or medicine men) so could be reinforcing social organisations or roles. Whether these uses have a selective benefit for a group is difficult to test empirically, as it would be impossible to remove music from the peoples and observe the consequences. Nevertheless, it seems reasonable to suggest that musical activities do fulfil these roles, and could likewise confer on the participants the advantages associated with such reiterations of social roles and relationships.

7.2.2. Music and sexual selection

Charles Darwin originally proposed that music in humans had its roots in courtship songs, as he believed that the vocalisations with greatest pitch changes made by apes tend to be produced by males when soliciting mates (Darwin, 1874). The courtship basis for musical performance has been echoed by a few authors since, including Jelle Atema (quoted in

Ahuja, 2000), who reconstructed the so-called "Neanderthal flute" from Divje babe I (see Chapter 3 for an analysis and critique of this object). He proposes that Neanderthals might have wooed their prospective partners with musical performances.

However, interestingly, there is actually little analogue for courtship songs in primate societies; most examples consist of homologues in distantly-related species such as birds. There is some evidence, however, of correlations between "singing" and pair bonding in some primate species. *Indri*, *Tarsius*, *Callicebus* and *Hylobates* (Geissman, 2000) are the only primate genera to have some members who sing, and in all cases both males and females sing. In most cases, duetting in bonded pairs also occurs. It is noteworthy that all known singing primate species are thought to have a monogamous structure, and this applies to those bird species that duet too (Geissman, 2000). To Geissman, "This suggests that the evolution of singing behaviour in primates and of duet singing in general are somehow related to the evolution of monogamy" (p. 112); i.e. to the maintenance of a strong pair bond with a single other individual of the opposite sex. Because the four groups of primates that sing are not closely related, Geissman states that it is likely that this trait evolved four times, independently, and suggests that the same happened in hominids subsequently.

These singing behaviours do not seem to be used to "woo" prospective mates, but occur between the members of an established mating pair. It does seem, however, that such activities *are* correlated with activities that increase pair bonding; Geissman and Orgeldinger (1998) observed that in ten siamang groups, "duetting activity was positively correlated with grooming activity and behavioural synchronisation, and negatively correlated with interindividual distance between mates" (Geissman, 2000, p. 111), suggesting that the activity is indeed related to the strength of pair bonds. In contrast, there is little evidence to suggest that our nearest relatives, the great apes, have specific sex-calls (Williams, 1980) and there is also no contemporary human ethnographic evidence of song being used in personal courtship. When music is used in relation to sex and relationships, it is usually within the context of the social relations of an entire social group (Sloboda, 1985).

Musical behaviours do, however, have various traits which could allow them to act as an indicator of fitness, a factor which may have shaped their development. A major advocate of a sexual selection hypothesis for the evolution of music in recent years has been Miller (2000a, 2000b). Miller argues that musical behaviours can indicate sexual fitness for a number of reasons. Indicators of sexual fitness illustrate selectively important traits such as status, age, physical fitness and fertility. Miller proposes that various features of music can act as good indicators of these traits, and that it is a good medium for their display. He suggests that dancing reveals aerobic fitness, coordination, strength and health, voice control may reveal self-confidence and status, rhythmic ability may indicate the "capacity for sequencing complex movements reliably", whist virtuosic performance *per se* "may reveal motor coordination, capacity for automating complex learned behaviours, and having the time to practice", the latter of which may also indicate sexual availability, as it suggests lack of parenting demands. These properties of such displays could lead to aesthetic preferences for particular forms of those behaviours. This leads Miller (2000a) to propose that "…any aspect of music that we find appealing might also have been appealing to our ancestors, and if it was, that appeal would have set up sexual selection pressures in favour of musical productions that fulfilled those preferences" (p. 342).

It should be noted, however, that this logic does not necessarily follow – it implies that *any* musical trait for which there is any preference will subsequently be selected for by sexual selection. In fact there is an important qualifier that should be imposed: selection, sexual or otherwise, for a particular trait can only occur if that trait can arise by mutation. This is not the same as the perpetuation of behaviours. Also, if sexual selection was actually *responsible* for the various traits that humans find aesthetically appealing about music then we would expect a great deal of convergence in the form of musical behaviours, and which aspects of them are considered aesthetically appealing. In fact, whilst musical behaviours are universal and share features that are recognisable in all cultures, the aesthetic preferences associated with them are quite culturally specific.

Miller admits that such rationales are speculative at this stage, and in fact his thesis is really a call for empirical testing of such ideas in the way that they have received attention in other biological fields. Such research would indeed be a very welcome addition to the field, as some of Miller's ideas of the fitness-display properties of music do seem intuitively to make a great deal of sense. In terms of explaining the foundations of the appeal of musical stimuli, they add nothing to the biological, neurological and evolutionary evidence considered so far in the preceding chapters; however, they do provide legitimate mechanisms by which *musical behaviours* may have become refined, perpetuated and spread. His theory is more an explanation of why they might have been developed into the forms we know and then propagated across the species, rather than an explanation of the form and appeal of the foundations of musical behaviours.

Miller says "If one can perceive the quality, creativity, virtuosity, emotional depth and spiritual vision of somebody's music, sexual selection through mate choice can notice it too" (p. 355). This is true, but it relies upon the extant existence of value judgements on creativity, virtuosity, and emotional depth to such a behaviour. To say that sexual selection is responsible for these value judgements begs the question – there has to be a reason why such values are applicable to music before selection can act upon their manifestation. As has been argued elsewhere in this thesis, the "quality" of the musical behaviour (its creativity and virtuosity, perhaps, in developed forms) *is* its "emotional depth", i.e. the extent to which it elicits an emotional response as a consequence of the mechanisms through which it is perceived.

In fact, the above outline of Miller's (2000a) ideas represents the full extent of his exposition, with the rest of the paper consisting of explanation of sexual selection, and his arguments for why he considers group selection or benefits to be inadequate in accounting for the development of musical behaviours.

Miller cautions against equating behaviours that are made *in* groups for behaviours that are made *for* groups; i.e. just because a behaviour is carried out in groups does not mean that it is carried out for the benefit of the group. Miller points out that the fact that music is most often performed in groups has led to a group-level function for music usually being considered to be its most important function, when this need not be the case. This is a salient observation, and the attribution of any explanation should be treated with rigour; nevertheless, Miller's objections to the attribution of group selection benefits to musical behaviours have some flaws themselves.

Miller claims that authors applying group selection models to specific situations make two errors. The first, he claims, is that they are often under the impression that it is a more friendly and humane form of selection than individual selfishness, so people are inclined to favour it as an explanation. Miller argues that, in fact, group selection is no more humane than individual selection, as the friendly groups are still out-competing and exterminating other groups, replacing "the logic of murder with the logic of genocide". There are two reasons why this "error" in applying group selection models to human behaviour is not an error. Firstly, authors' motivations in attributing an explanation do not undermine the credibility of that explanation's mechanism. Secondly, group selection mechanisms do not necessarily rely on neighbouring groups exterminating each other; group selection only has to rely on a group performing better in its own environment than other groups perform in their environment. The adverse conditions that they find themselves overcoming with the aid of a behaviour need not have anything to do directly with another group; they only have to survive when other groups do not.

The second error that Miller (2000a) believes that authors make in attributing group selection as an explanation is a failure to consider "free-riding" by members of the group who benefit from a group's behaviour without participating in it themselves, and its attendant costs. He says that if music had group-selective benefits, the situation could arise in which "wallflowers" emerge among the group. They benefit from the group-level advantages of the dancing/music behaviour that is undertaken, but do not "pay the enormous time and energy costs of dancing all night" (p. 352). Thus, the wallflower mutation would spread throughout the group within a few generations, and music/dance would be lost. Again, there are two problems with Miller's objection. The first is that it presupposes that for music/dance behaviours to be beneficial at a group level they involve "enormous time and energy costs" and "dancing all night". This need not be the case. The second problem is that, if music did have a group-selective advantage, the group that loses music and becomes a community of "wallflowers" would die out eventually, whilst the groups that survive would be those in which the balance between wallflowers and dancers has been, through selection at an individual level, kept at an optimum balance for the group, whatever that balance may be. This is not necessarily to advocate a group-selective benefit for musical behaviours, but to point out that Miller's objection does not undermine group selection for music; in contrast, it provides an explanation for a way in which it might operate, if musical behaviours *are* advantageous at a group level.

In fact, Miller (2000a) points out that "if music did have individual-level benefits, such as courtship benefits under sexual selection, it may be possible for group selection to reinforce them with group benefits" (p. 352). This thesis has so far argued for a similar mechanism. Essentially, a skill of interpersonal emotive-tonal vocalisation used for forming and maintaining networks and loyalty is one that is beneficial at an individual level; the foundations for such individual behaviour, namely the ability to produce such vocal signals, and sensitivity to their content, may also allow a group activity using those behaviours to be beneficial at a communal level as a consequence of simultaneous contiguous emotional experience. This is in addition to the potential benefits inherent in having cohesive internal networks.

7.2.3. Music and group selection

The potential role of music in selection at the level of the group has been hinted at in preceding sections. The nature, and existence, of mechanisms of selection at the group level has been the subject of considerable debate, so requires some more detailed consideration here.

Shennan (2002) has carried out a comprehensive evaluation of models of selection as they apply to human prehistory, and iterates that selection can occur at numerous levels, including that of the group. It is, of course, always the individual who is the perpetuator (or otherwise) of genes, and who thus forms the crux of selection. Models of individual self-sacrifice for the benefit of the group or species as a whole have been shown to be unworkable. Within groups, competition between individuals is generally far more important, selectively, than co-operation: "Mathematical analysis [has] demonstrated that in most circumstances, selection for interests at the individual level will override interests at the group level" (Shennan, 2002, p. 29). It is this model of selection for group level interests, with individuals undertaking behaviours detrimental to themselves but potentially beneficial to the group, that has been shown to be unworkable.

This is not the same, however, as "group selection" in the sense that it has been used in the preceding sections, which is in terms of group behaviours having an effect on the social environment in which individuals are living, feeding and breeding. As Shennan puts it, "all theoretical schools, including those that are sceptical about other levels of evolutionary process than that of individual inclusive fitness, recognise that such [individual] interests may often be served by cooperating rather than competing with other individuals of the same species" (p. 213). Gene theory, and the famous "Prisoners' dilemma", shows that if individuals have similar aims they are frequently better off modifying them slightly (if necessary) in order to cooperate with others, than to compete with them. They do so on a tit-for-tat basis making only as much self-sacrifice as they perceive, from experience, that they are likely to receive in turn. Furthermore, these behaviours are likely to conform to "pro-social norms" that emerge within a population, as a consequence of frequent interaction with the same people; adherence to these norms can have beneficial effects on the well-being of members of the group in terms of the recompense that they receive for other behaviours that they undertake (Bowles and Gintis 1998). To put it another way, behaviours which constitute the optimal behaviour for an individual can be determined culturally as well as ecologically. In a social species the likelihood of an individual surviving to procreation, or of having a high rate of procreation, depends upon their "cultural fitness" too – the way they behave in relation to their surrounding culture.

The potential value of behaviours that contribute to "group cohesiveness", as musical behaviours can, is that they may make cooperative behaviours more likely to occur within the group than would otherwise be the case. This can benefit reproductive success at an individual level, and thus also increases the likelihood that the culture practising them will perpetuate them subsequently too. Bowles and Gintis (1998) have also demonstrated that "populations whose interactions are structured in such a way that coordination problems are successfully overcome will tend to grow, to absorb other populations, and to be copied by others" (Shennan, 2002, p. 216). It can be seen that the emergence of musical behaviours as a pro-social norm which thus leads to the overcoming of coordination problems within a group could lead to the growth of such groups and the spread of those behaviours. Furthermore, not only could musical behaviours constitute a behavioural norm in their own right, because of their foundations in social cognition, those individuals with well-developed musical capacities could also be those best at identifying and conforming to other norms of social interactive behaviour.

Hull (1980) has used the terms "replicator" and "interactor" to refer to genes and the individuals in which those genes reside, respectively. An interactor is defined as "an entity that interacts as a cohesive whole with its environment in such a way that replication is differential" (Shennan, 2002, p. 65). These terms can also be applied to selection at the level of the group, with the "replicators" now being the individuals and the "interactor" being the group in its environmental context and in contact with other groups. In this circumstance, "the replicators whose propagation is affected will be genetic and cultural traits at the individual level, as well as group-level traits" (Shennan, 2002, p. 253). Thus, if a group develops or adopts a structural and/or behavioural innovation which results in proliferation of that group (as described above), individuals within the group may gain greater reproductive success as member of that group than they would otherwise (despite possibly having to pay some cost associated with the innovation). The group and individual stand or fall together in terms of reproductive success.

Musical behaviours could fit well within these models of selection at the individual and group level, and these have the potential to provide a model for the development and spread of musical behaviours. The following sections provide some illustrations of more specific ways in which music may fulfil the social roles described.

7.2.4. Music and dance as a coalition signalling system

Hagan and Bryant (*in press*) suggest that rather than music and dance *causing* social cohesion, they *signal* social cohesion that was achieved by other means. Hagan and Bryant's overall thesis is that "for humans and human ancestors, musical displays may have… functioned, in part, to defend territory (and perhaps also to signal group identity), and that these displays may have formed the evolutionary basis for the musical behaviours of modern humans" (p. 5, *in press*).

They justify this position, and reject other explanations, on a number of grounds. They argue that musical behaviours cannot contribute directly to the cohesion of a group, as they are not a good indicator of an individual's ability to contribute to the group's survival. They state that group cohesion exists on the basis of the perception of benefits to be offered and gained by individuals, and shared mutual goals. Musical behaviour, they say, cannot allow an individual to communicate their fitness interests or abilities.

However, this is viewing group cohesion purely in terms of perceived cost/benefit and the perceived value of group membership, but ignores all factors of emotional bonding and loyalty engendered by mutual emotional experience. An individual may already have established their credibility within a group, in terms of their *ability* to contribute to its survival, but this provides no indication of their *likelihood* of doing so, or to whom they will do so. Hagan and Bryant's criticism could be levelled equally at higher primate grooming behaviours, which do not have as their foundation an indication of the ability to hunt, for example, any more than does musical behaviour. The ability to instigate, express and maintain emotional bonds and loyalty is fundamentally important to the maintenance of the networks that exist within a group, be they reproductive, familial, parental, or relating to the hierarchical structure of the group as a whole. Any group which fragments under adversity with individuals withdrawing their skills to themselves will be less likely to survive than a cohesive group under similar conditions.

Hagan and Bryant (*in press*) also argue that music has no useful role in establishing group goals and coordinating group actions in the face of adversity (specifically, outside aggression – which is far from the only form of adversity that a group may face), a role which language performs adequately. This may well be true, but quite apart from the fact that such behaviours might well pre-date linguistic ability anyway, what music *can* do is to signal, and strengthen ingroup (vs. outgroup) membership. Participating in such activities may allow individuals to firmly identify themselves with the particular group with which they perform, as a consequence of the shared experience. This (musical) experience is not merely a passive one, but is, in almost all societies, an actively participative one. Apart from the potential of any shared interactive activity to create a sense of inclusiveness, as has been illustrated in Chapter 6, musical behaviours themselves appear to rely on cognitive mechanisms grounded in emotional interaction. One would expect that their ability to perform such a role would be even greater than other mutual activities.

The value of shared emotional experience in creating strong bonds of loyalty, especially in adversity, but in positive emotional experiences too, is well known. It is the basis of the loyalty and friendship associated with members of military units years after the events themselves, and is the rationale for initiation rites to any manner of other groups in all societies.

Hagan and Bryant highlight that a major function of vocalisations in primates, both on an individual basis, and to a lesser extent on a coordinated group basis, is territory defence. They point out that the possibility of territory defence having been an early role for song has been neglected relative to considerations of its possible roots in mate attraction, social- or pair-bonding, and individual or group recognition. All of these are roles of the vocalisations of other animals, and Hagan and Bryant propose that territorial considerations may have been the driving role of musical development.

It seems likely that, having developed, complex vocalisations would have come to be used in all of these contexts in hominids. However, unlike social-interactive and bonding vocalisations, territorial and warning vocalisations require only limited range and control to be effective (this is with the exception of birdsong, which often consists of complex sequences used for territorial purposes; this, however, does not occur at a group-cohesive level, so is not analogous to the situation being described by Hagan and Bryant). The characteristic feature of the human (and latter hominid) vocal tract, however, is its range, versatility and the control over extended sequences of such vocalisations that seems to have emerged in tandem. In contrast to social-emotional vocalisations, such properties would be superfluous in warning and territorial vocalisations which can (and still do) function most effectively in the form of abrupt vocalisations with little contour. Such vocalisations seem a very poor candidate for the roots of musical vocal behaviours.

So neither Hagan and Bryant's rejection of a social cohesion basis for musical interaction, nor their justification for a territorial basis of musical vocalisation are particularly convincing. It seems unlikely that these factors formed the evolutionary basis for musical behaviours in modern humans, as they propose. This, however, is not to say that musical behaviours *could not* have come to be used in such a context, only that it seems rather unlikely that such a context was a selective force for their development. The problem is that no distinction is made between an "evolutionary basis for musical behaviours" and "reasons why such behaviours may have been perpetuated", two distinctly different propositions. Whilst coalition signalling should be rejected as an explanation for the former, it may well be a good explanation for the latter. As Hagan and Bryant say, "Territorial defense and alliance formation both require communicating credible information about group capabilities to non-group members, information that would deter intruders but attract allies" (p. 5, *in press*).

They propose that music and dance act as good indicators of group stability and the ability to carry out complex coordinated actions. This seems a perfectly reasonable proposition; an example of it fulfilling exactly that role is provided by the New Zealand

All Blacks famous *haka*, incidentally. Hagan and Bryant propose that this is the case because the amount of time needed to create and practice music and dance corresponds to the quality of the coalition performing them, in indicating how much time they have in which to prepare them.

In fact, such indirect rationalising is unnecessary to explain this value of the behaviour. Such a display indicates a more direct and fundamental fact about the group – its level of cohesion and cooperation, and as Hagan and Bryant originally say, its stability and ability to carry out complex coordinated actions. It indicates these things precisely because it *engenders* these things. The ability of music to act as a forum for the practice of integrated, complex, coordinated group activities resulting in a powerful sense of membership and cohesion is one of its major values, and provides a very coherent explanation as to why such behaviours persisted at a group level. One of the manifestations of this role may indeed have been "coalition signalling", and this may even have led to its perpetuation, but this is unlikely to have been the underlying selective force for music's development.

This potential of music to act as a forum for the execution of other social behaviours has been explored further by Cross (1999, 2001).

7.2.5. Music's multiple meanings, and cognitive and social development

Whilst emphasising that musical capacities appear to be based on universal and innate human biological capacities, Cross (1999a, 2001, 2003) highlights the importance of the fact that the manifestation of musical behaviours based upon those foundations has a great deal of cultural specificity. Although the cultural elements (for example our learning and interactions with people and objects during infancy, childhood and adolescence) may themselves be *constrained* by evolutionary forces, they are not *determined* solely by such factors. Music is different things to different peoples in different places and at different times – it has "transposable intentionality" (Cross, 2003b). It is from this quality that Cross believes it derives its principal biological value.

In terms of contributing directly to survival, Cross (2001) observes that "…in itself, music does not seem capable of being a material cause of anything other than a transient hedonic encounter. It seems to be inefficacious." (p. 99). However, precisely because of its "non-efficaciousness and multiple potential meanings", Cross believes that "music can be both a consequence-free means of exploring social interaction and a "play-space" for rehearsing processes necessary to achieve cognitive flexibility." (p. 99). The only issue to be taken with this description concerns the efficaciousness of musical or proto-musical production and perception: the fact that they function within the media of social interaction actually *gives* them efficaciousness – their value comes from stimulating and exercising those mechanisms; i.e. a transient hedonic encounter may itself be efficacious and ultimately have survival benefits. In fact, Cross' conception of efficaciousness is simply in terms of immediate and direct survival benefit as a consequence of the action, and the efficaciousness of longer-term consequences of the activity are not outside his model. It seems, from the evidence considered in the Chapters 4, 5 and 6, that musical activities do not simply provide an opportunity for the carrying out of social interaction, but actually use the mechanisms of social interaction and expression – thus it could be valuable not simply as a means of exploring social interactions, but actually as a vicarious stimulus and exercise of those capacities, in a more abstract way.

As for the multiple meanings of music, Cross suggests this can allow for the development of cognitive flexibility in human infants because musical or proto-musical behaviour has the potential to make use of several domains of intelligence at once. It thus provides the opportunity to practice the integration and control of biological, psychological, social and physical systems, and what is more, can do so in a context of limited risk. So not only can it provide the opportunity to develop "capacities for flexible social interaction", building upon innate social intelligence, it can also contribute to the development of cross-domain cognitive flexibility.

It can be seen that these are actually two separate functions that can be fulfilled by music, and the former does not depend upon the latter. The value of proto-musical behaviours in providing a forum for the stimulation of social interactive capacities could predate any requirements for cognitive flexibility, whilst musical activities have the *potential* to use several domains of intelligence at once, due to their non-specificity of meaning.

Cross (2001) suggests that "If evolution has shaped the human mind, it has probably selected at the level of infant predispositions, and culture can be thought of as particularising – shaping into specific and distinct forms - the expression of those predispositions." (p. 98). So should we consider music to be solely the perpetuation into adulthood of a behaviour that is selectively useful in infancy and childhood? This is not Cross's implication, and the weight of the evidence considered so far suggests strongly otherwise. Music is built upon behavioural foundations that are useful at all stages of life-history, and at all levels of the group. That selection may occur at the level of infant predispositions does not mean to say that behaviours are selected for because they manifest themselves usefully in infants – this is not selection for infant behaviours (*contra* Falk and others) – it is selection for infant capacities which give a predisposition to develop certain behaviours in adulthood. For example, infants are not born with linguistic behaviours intact – they cannot speak language – however, they appear to be born with the mechanisms and predisposition to develop this important behaviour into adulthood. That the capacity to develop an adult behaviour is being selected for does not mean to say that the developing skill is not useful in the meantime, of course. So the above considerations do not undermine Cross's suggestion that the existence and development of musical capacities in infants and children may themselves be selectively important. As Cross (2001) points out, however, whilst these capacities and predispositions have been biologically selected for, the forms that the behaviours take when they are developed are culturally determined. Exactly the same applies to music.

This means that "In effect, the mature musics of a culture are, in part, constituted through a persistence into the adult world of childhood patterns of thought and behaviour" (Cross, *in press*). Cross suggests that this trait of musical behaviours may be connected with the process of altricialisation in hominids. Altriciality refers to the amount of time spent by a species in a pre-reproductive state, relative to its lifespan. Humans are especially altricial, and it appears that this trait correlates with neocortical development in other primates too, and with social complexity. Joffe (1997) has suggested that a greater time spent in a juvenile state allows for the fine-tuning of the social skill repertoire required for increasingly complex societies, as well as maintenance of access to flexible patterns of thought and behaviour which may confer benefits in terms of behavioural adaptability (Cross, *in press*). Despite that fact that such a process must confer considerable survival disadvantages as well, such as an increased duration of dependence and greater likelihood of death before procreation, evidently just such a process has occurred in hominid evolution; between *Homo habilis* and *Homo ergaster* (Foley and Lee, 1991), there appears to be an increase in the duration of the pre-reproductive state, and a considerable marked increase in juvenile duration between *Homo erectus* and Neanderthal and modern humans (Dean *et al.*, 2001). The advantages conferred must have exceeded these potentially high costs. In reviewing this evidence, Cross (*in press*) suggests that this process may have allowed for the perpetuation and development into adulthood of proto-musical behaviours too. Given that the evidence considered in Chapter 6 strongly suggests a foundation for such proto-musical behaviours in social-emotive vocalisation and vocal, corporeal and manual gesture (useful in mediating increasingly complex social relations), it indeed seems likely that an increased period of development of these capabilities could lead to the development and refinement of proto-musical behaviours.

The evidence considered in Chapter 4, (according with that of Dean *et al.*) places a major step in this development with *Homo heidelbergensis*. Whether this can be applied to both modern humans and Neanderthals remains to be demonstrated; this would depend upon whether the *Homo heidelbergensis* concerned were ancestral to both, or just to *Homo sapiens*.

As for the "transposable meaning" of musical behaviours, Cross suggests that this may have implications for human cognitive evolution as a whole. He suggests that as well as stimulating cognitive flexibility ontogenically, it may lie at the root of the development of cognitive flexibility in phylogenic terms, and thus the development of what is probably the defining characteristic of human cognition over that of other animals.

> "…given that it seems feasible that music plays a role in the development of cognitive and social flexibility for modern human infants, it could be that the emergence of proto-musical behaviours and their cultural actualisation as music were crucial in precipitating the emergence of the cognitive and social flexibility that marks the appearance of *Homo sapiens sapiens*" (Cross, 2001, p. 100).

How music might fit into models of the emergence of cognitive flexibility and cross-domain intelligence in the evolution of modern humans is the subject matter of the next chapter.

7.3. Chapter summary and conclusions

Past discussion of the potential values of musical behaviours as survival traits have frequently failed to distinguish between rationales for the evolution of musical capacities, and rationales for the perpetuation of developed musical behaviours; further, there has also often been a failure to distinguish between evolutionary selection for musical behaviours, and non-selective reasons why they may be perpetuated. The preceding three chapters have been concerned with the evolution of underlying capacities for musicality; this chapter has aimed to identify whether there are any reasons, selective or otherwise, why musical behaviours may have become developed and perpetuated within hominid (and/or human) groups.

It is possible that musical behaviours provide a good medium for the expression of various traits related to sexual fitness, and that aspects of those behaviours might be "fine-tuned" by preferences exhibited under such circumstances. However, this mechanism cannot account for those preferences' existence, or for their being applicable to music to start with. Miller's (2000) assertion that any aspect of musical behaviours that was appealing would have set up sexual selection pressures for that trait is also not valid: for sexual selection to act on a behaviour, that behaviour must have a genetic basis. This is not to say that *behaviours* cannot be perpetuated as a consequence of sexual preference – this may, conceivably, occur, but unless the behaviour is a consequence of a genetic trait, such perpetuation will be through emulation and not Darwinian sexual selection.

One way – with a genetic basis - in which music could be a particularly potent expression of sexual fitness lies in its roots in the ability to communicate emotionally and effectively, to empathise, bond and elicit loyalty. An individual who is talented in these respects may well be more appealing to the opposite gender, because they are more likely to be able to form strong social alliances, and strong pair and family bonds. Good musical ability may vicariously indicate such abilities, as the cognitive capacities relied upon are in many respects shared.

The fact that people are frequently strongly drawn to same-gender music groups (as Hagan and Bryant, *in press*, point out) actually does not undermine the above appeals of music, as forming strong alliances is not a gender-specific activity; music can also fulfil the role of engendering strong feelings of empathic association and group membership. Hagan and Bryant's assertion that music cannot *engender* group cohesion, but can only *indicate* it is wrong in this respect, in that they only dismiss the ability of music to directly indicate an individual's ability to contribute to a group in subsistence terms. Such a criticism might equally be levelled at grooming activity. They fail to account for any benefit music might have in the respects discussed above, namely as an indication of an individual's *likelihood* of contributing to a group, as indicated by their networks and loyalty, and as a mechanism for actually stimulating and maintaining those networks and loyalties. Hagan and Bryant also suggest that musical vocalisations may have grown out of territorial and warning vocalisations; this seems unlikely as musical vocalisations gain much of their value from their complexity and pitch range, whereas most territorial and warning vocalisations gain their value from precisely the opposite traits. They are unlikely candidates for the evolution of vocal complexity.

These objections do not, of course, undermine Hagan and Bryant's suggestion that group musical activities may be a good indicator of group cohesion, only their arguments for why that might be, and their point that music can't engender cohesion. As for the former, it seems quite likely that music can also form an important signal (and stimulus) of group cohesion, cooperation and loyalty, and a forum for the coordination of complex interactions; it still frequently fulfils this role. Like many authors, Hagan and Bryant fall into the trap of not making the distinction between an "evolutionary basis for musical behaviours" and "reasons why such behaviours may have been perpetuated". Coalition signalling should be rejected as an explanation for the former, but it may well be a good explanation for the latter. Group displays of musical behaviours can indicate group stability and the ability to carry out complex coordinated actions precisely because they *engender* these things.

So music could be valuable not simply as means of exploring social interactions, but actually as a vicarious stimulus and exercise of those capacities, in a more abstract way. Musical or proto-musical behaviour also has the potential to make use of several domains of intelligence at once, relying on the integration and control of biological, psychological, social and physical systems; furthermore it gives the opportunity to practice and develop these integrated skills in a context of limited risk. Mechanisms of group and individual selection (Shennan, 2002; Bowles and Gintis, 1998) may prove fruitful in generating models of mechanisms through which musical behaviours may have come to develop and spread, through the above advantages, conferred on a group and individuals within it.

This may seem like an attempt to assimilate all the possible explanations for music's value in social groups, and an unwillingness to reject any. This is not the case – the rationales given to date for several of the above models have been shown not to be without flaws; however, this does not mean that the perceived value itself is groundless. Whilst the different rationales for those selective roles of music have often been mutually exclusive, the roles themselves are not. What actually seems to be the case is that all of the above rationales manifest advantages that are facets of the social interactive value of musical behaviours, as this role, ultimately, underlies them all.

CHAPTER 8

Music and Cognitive Evolution

Chapter 6 has explored the evolution of the capacities that underlie musical behaviours, and Chapter 7 has looked at possible reasons why musical behaviours may have been advantageous. This chapter examines how the development of musical capacities fits within various models of hominid cognitive evolution that have been proposed. Writings on cognitive evolution in the past have typically focused on the emergence of language and symbolic behaviour as the defining characteristics of modern humans; these discussions have generally neglected to mention musical behaviours at all, or if so, they are subsumed under the umbrella of art and symbolism. The evidence in the preceding chapters suggests that symbolic and "artistic" associations with musical behaviours are not fundamental to those behaviours; nevertheless, there would have been a time in hominid evolution before which musical behaviours as a whole *could not* have emerged, due to certain cognitive capacities not being in place. How musical capacities fit within models of the development of human cognition is the focus of this chapter.

8.1. Cognitive modularity and symbolic thought

In reviewing the literature regarding the musical perceptual abilities of infants, Kogan (1997) concludes that almost all the skills of musical perception are present in children under the age of one. He suggests consequently that these abilities are innate, and because they apparently co-occur, indicate that there is a single domain in the brain responsible for these skills which must have been selected for during the brain's evolution. The idea that human cognition consists of a "Swiss Army knife" of functional modules (or domains), which have been naturally selected for, has been popularised in recent evolutionary psychology by Barkow, Cosmides and Tooby (1992). According to Kogan (1997) the music-dedicated domain that he proposes would likely have been selected for as a consequence of the ability of communal music-making to promote group morale and identity.

It should be pointed out that such a selective benefit, if being considered part of a single module, must be applied to all ancestral forms of musical behaviour, if it is to be asserted that the mechanisms that music makes use of were selected for *by music* from the outset. This means that, if they were selected for as domain-features specific to music, proto-musical behaviours must have had these social-bonding functions from their earliest occurrence.

Whilst also considering that music appears to make use of a specific cognitive domain or module, Sperber (1996) believes that this is one that music has "hijacked", and that it was not originally selected for by music. In this he concurs somewhat with the view of Pinker (1997; see Chapter 7), although without diminishing the significance of musical behaviour itself in the way that Pinker does. The original use of this domain, he proposes, was to process "complex sound patterns discriminable by pitch variation and rhythm" which would have been used in early human communication when the vocal tract was limited. This is indeed what is suggested by the evidence considered in Chapters 4, 5, and 6. He does not consider, however, the possibility that speech and music may have shared these utterances as a common ancestor, and thus shared the use of this domain, as has been discussed previously.

It seems unlikely, from the evidence considered in the preceding three chapters, that musical behaviours are controlled by a dedicated "music module". In contrast, whilst the capacities that it uses are undoubtedly finely tailored to that purpose to work in tandem, and are innate abilities, they serve other fundamentally important purposes too, which the use of music has emerged from, and built upon.

Mithen (1996) suggests that the creation of art-forms require that the "barriers" between specific functional cognitive domains (such as Social Intelligence, Technical Intelligence and Natural History Intelligence) be broken down. He suggests that this occurred in the "creative explosion" of around 40,000 years ago, with the advent of general-purpose language, and along with it evidence of visual art and representation (and musical instruments, see Chapter 3). This raises the question of whether this cross-domain fluidity (and with it symbolic thought) would have been required for the production of music.

The convention of attributing symbolism to music, and treating music as an art-form, may be a consequence of our own modern use of it. When we hear music it evokes emotional response, and with it often comes associated imagery. This may, indeed, be an exclusive product of the cognitive fluidity of the modern human mind; as a result music has come to be *used* to evoke particular imagery. However, the ethnographic evidence presented in Chapter 2 shows that imagery and symbolism are not inherent in many types of music, and that it is probably largely the modern western usage which has come to result in music being seen as requiring symbolic thought. It seems to be a western convention to categorise the performing arts and the visual arts in the same oeuvre.

Further, as discussed at length in Chapter 6, the evocation of an emotional response by music is not necessarily reliant upon symbolic associations, but upon properties of our perceptual system; music can affect our moods and create emotional response without any direct symbolic associations being made at all. This is epitomised in the findings of Watt and Ash (1998) that music has an action on the individual similar to the actions of a person; this results in listeners most readily assigning to music human-like attributes such as trait-qualities (e.g. age and gender), and state-qualities (e.g. emotions), but it is clearly not necessary that a *symbolic* association between the music and a person be made by the listeners; it is that they are most inclined to process the signal in the same way as interpersonal interaction. These findings complement the conclusion that melody in music could have been developed from the use of a form of interpersonal vocalisations, rather than having as its foundation abstract symbolic associations.

Despite the fact that abstract symbolic associations are not necessary to the production and perception of musical activity, such activity can exercise several domains of intelligence at once. As noted above, Cross (1999, 2001, 2003) has suggested that, since music appears to provide an excellent medium for the development of multi-domain cognitive and physical behaviour in infants, it may also lie at the root of the development of cross-domain flexibility in *Homo sapiens*. Music can play a role in the development of cognitive and social flexibility in modern human infants because it is built on foundations that fulfil that process; for proto-musical behaviours using those mechanisms to develop in modern humans may simply have required that they be a useful manifestation of those mechanisms.

That music has "floating intentionality" marks it out as not being restricted to one symbolic domain – it has a symbolic ambiguity – but it also does not require an ability to have cross-domain flexibility, in the sense of transferring knowledge and expertise from one domain to another and making analogical connections. Having had its foundations in the social intelligence domain, music apparently became manifest in its domain-ambiguous form, where it can mean different things to different people, on the basis of the emotions and interactions stimulated. It can also gain meaning from specific associations, but does not need to in order to be valuable to its participants. So music does not (have to have) have specific symbolism associated with it, and much of its current value comes precisely from its non-specificity. Thus the emergence of this form of proto-musical behaviour does not have to post-date the emergence of "symbolic thought" (analogical, cross-domain thought) as is generally thought to be the case for linguistic and visual artistic behaviour, with direct or abstract symbolism. But it does require the cognitive flexibility for the skills associated not to be firmly rooted in a single domain. Whilst it has social intelligence as its foundation and it is in social context that it is most useful, its execution and perception are not limited to being literally experienced as social interaction; they allow interaction to take place at several levels, simulated and literal, and provide a medium for the exercise of such interactions.

It seems reasonable from the above to suggest that ancestral forms of music using only rhythm and/or voice needn't have required any domain of cognition other than the social, if they were used as a group or interpersonal activity rather than as an individual's own piece of creative performance. Subsequent developments of musical behaviour may have benefited from increasing cross-domain flexibility, and as Cross (2001) suggests, musical behaviours may even have engendered some aspects of such cognitive flexibility. Mithen (1996) suggests that a sophisticated social intelligence (involving a theory of mind) was probably well developed as early as *Homo* (or *Australopithecus*) *habilis*, allowing it to live in relatively large social groups. The expansion in brain size between *H. habilis* (750cc, KNM-ER 1470; Johanson and Edgar, 1996, p.162) and *H. ergaster* (909 cc, estimated adult volume of KNM-WT 15000; ibid. p.182) would suggest this social intelligence domain would have become even more sophisticated, particularly if the correlation between brain size and group size proposed by Aiello and Dunbar (1993) is to be accepted. The capacities which later came to form the foundations of proto-musical behaviours must have been becoming increasingly sophisticated at this time.

8.2. Mimesis, culture and cognition

An alternative model of human cognitive evolution, not based on changes to modularity but instead on developments of mechanisms of cognition related to representation and interaction, has been proposed by Merlin Donald (1991, 2001).

Donald believes that hominid cognitive evolution occurred in three main transitions, between four stages that he calls *episodic*, *mimetic*, *mythic* and *theoretic*. Theoretic culture describes the most recent developments in human cognition, which are a consequence of the storage of large quantities of information in external symbolic media; this development Donald places around 40,000 years ago. The first stage, episodic culture, describes the mode of cognition of apes, in which their behaviours are unreflective and largely responsive to the prevailing situation – "their lives are lived in the present, as a series of concrete episodes" (p. 149), albeit with the potential to be mediated by episodic memories of past events. Episodic cognition and culture is constrained largely by memory systems.

There appear to be three main systems of memory in animals, *procedural*, *episodic* and *semantic*, with modern humans apparently being the only creatures to posses a semantic memory system. Procedural memory is concerned with the learning of sequences of actions, such as catching an object in flight. It preserves the general principles for action, rather than the specifics of a given situation. This type of memory system is possessed by most animals. In contrast, episodic memory seems to be unique to birds and mammals, and appears to be more evolved in higher primates than in other species, in its subtlety. Episodic memory contrasts with procedural memory in that it is concerned with particular episodes in the life of the organism, specific events and their associated stimuli. Episodic and procedural memory appear to use separate neural systems and to have evolved separately, as one can be damaged without detriment to the other. It is this system which is the foundation, and constraint, of higher primate cognition and social systems. Semantic memory is, according to Donald, possessed solely by humans, and is concerned with symbolism, analogical associations and mental manipulations, and "knowledge" in the form of facts and propositions. Note that humans retain, and use, all three systems, but the concern is which of them constitutes the defining mode of cognition for their culture.

As noted, apes have a highly developed form of episodic memory, but apparently semantic memory capability is not developed. They are highly adept at event perception and episodic storage, have some self-perception and problem-solving skills, and have social structures that require the maintenance of complex relationships. What they lack, however lies in their representation: "Animals excel at situational analysis and recall, but cannot re-present a situation to reflect on it, either individually or collectively….Semantic memory depends on the existence of abstract, distinctively human representational systems." (Donald, 1991, p. 160). At some point in hominid evolution, the transition from one governing memory system and associated type of culture, to the other, occurred.

For Donald (1991, 2001) there is an important intervening stage between ape and modern human cognition, the "missing link" in this transition. This is what he calls "mimetic" culture. Mimesis is the consequence of a cognitive development which allowed voluntary control and self awareness of the expression and mimicry of action. This includes corporeal, gestural, vocal and facial expression of emotion, and action. As demonstrated in Chapter 6, these are intimately related systems in many animals, and evidence suggests that there is a single controller responsible for expression in these different media; the difference, for mimesis, is the emergence of *voluntary control* over this system. Evidence discussed in Chapter 5 (e.g. Jurgens, 1992, Davis *et al.*, 1996) indeed suggests that the emergence of the neurological systems for such voluntary control, of vocalisation in particular, occurred between apes and the common ancestor of Neanderthals and modern humans.

The emergence of mimesis has several implications, and four main by-products in terms of the types of behaviours it allows: mime of emotion, imitation of action, development of skill and understanding through rehearsal and integration of action, and gesture-like mime, as expression of emotion as an intentional communicative act.

This development of conscious control over action thus allowed the development of play-acting, precise imitation, gesture and body language. It also allowed, through the ability to imitate action, and to rehearse and understand skills, the refinement of many skills, such as throwing, cutting, making stone tools, and the ability to pass such information and skills on, i.e. pedagogy and practice. Finally, it allowed the making of intentional, modulated vocal sounds; i.e. voluntary control over vocal prosody "including deliberately raising and lowering the voice, and producing imitations of emotional sounds" (Donald, 2001, p. 261). For Donald, the fundamental factor at the root of the mimetic repertoire of behaviours is "the invention of culture as a collective means of accumulating experience and custom" (Donald, 2001, p. 263).

Donald places this development with *Homo ergaster*/*Homo erectus* for a number of reasons. Not least, this species shows a great increase in encephalisation (as discussed in Chapter 5), and it is with this species that we see the first evidence for widespread omnivory. *Homo erectus* also appears to be the first hominid to show the ability to occupy vastly differing ecological niches and migrate throughout the globe. Recent finds in Dmanisi, Georgia, of a relatively small-brained hominid, which has been tentatively identified as a *Homo habilis*-like *Homo ergaster* (Vekua *et al.*, 2002), now suggest that *H. erectus* was not the first hominid to migrate, although *H. erectus* was probably still the first to occupy significantly different environments.

Donald's (2001) citation of *H. erectus* as having "left behind toolmaking sites, seasonal hunting camps and continuously occupied fire sites" (p. 261) has to be treated with caution, however, as it neglects the entirety of the "Binford-Isaac" debate regarding the interpretation of the evidence for such behaviours (see, for example, Klein, 1999), and represents a position that is far from consensual in palaeoanthropology. Nevertheless, we can at least say that it is the case that hominids were forming migratory groups, had the ability to adapt to entirely new environments to develop new skills (at least initially) and to pass them on, and to coordinate group activities such as hunting (or scavenging). It seems likely that stable groups with strong social organisation would be required for these activities to be possible.

The physiological evidence considered in Chapter 4 also indicated that it was indeed with *Homo erectus* that the ability emerged to voluntarily moderate breathing sufficiently well to produce vocalisations controlled for pitch, intensity and contour, even if it lacked the ability to produce extended vocalisations. It was also with *Homo erectus* that the first developments in vocal *range* appear to occur; if it is indeed at this time that mimetic ability emerged, this explains the continued development of both vocal control and range over the subsequent speciations.

The evidence discussed in Chapters 4, 5 and 6 concurs strongly with Donald's model of the emergence of mimesis, even though the majority of it is not cited by Donald himself. The capacities that Donald describes, their interdependence, and their foundations in increased social complexity form a coherent system, the constituent parts of which must have emerged over the course of hominid evolution. Although Donald's own evidence for placing this initial development with *Homo erectus* is sometimes less than watertight, the other evidence considered in this thesis concurs with the attribution to *Homo erectus* of significant advances in these capacities, and their culmination with archaic *Homo sapiens* (*Homo heidelbergensis*). What Donald's "mimetic" stage highlights is the interrelationship of the capacities for self-aware voluntary vocal and corporeal control, and the apparently inevitable consequences of such abilities for social interaction, expression, the potential for social complexity and the practice and pedagogy of skills. It also leads to more self-directed focus of attention, towards corporeal stimuli and control, expression and action, and away from external stimuli. This, according to Donald (2001), has particular implications for the integration of rhythmic corporeal movement (including facial, gestural and vocal rhythmic control) with rhythmic stimuli, using many motor systems in a unified way, and constantly appraising them during the activity and in review. It is the repertoire of abilities that constitute mimetic cognition that are responsible for our ability to regulate our own emotional responses, on the one hand, yet also to be moved by, and participate in mimetic displays such as national funerals, military marching, group dancing, and chanting at a sports game. It results in a type of cultural interaction based on conventional non-verbal expressions expression, of body-language, eye-contact, facial expression and vocal tone. It is also, in the same way, (though Donald makes little mention of it) the foundation that underlies musical behaviours, rhythmic and vocal, perceptive and participative.

Note that this cognitive development is not one that in itself allows for great innovation in technology, but it does allow for the adaptability to accommodate new situations, ecological and social. In a sense, Donald's model of the importance of the development of mimesis is not inconsistent with ideas of modularity, as it falls largely within the domain of social intelligence. However, one of the main implications of Donald's model is that this social foundation had great ramifications for the emergence and propagation of all manner of other cognitive and behavioural developments. The behavioural and cognitive capacities inherent in mimesis themselves form the foundations for the integration and refinement of cognitive skills previously restricted in their development, including the evolution of the capacities for complex vocalisations and ultimately language.

In terms of the timing and importance of these developments in social intelligence, Donald's and Mithen's (1996) models concur; it is in terms of the domain specificity that they differ. Donald himself puts it "Note that this is all domain-general training. Virtually any cognitive module might prove relevant to cultural life, and thus the habits of social interaction tend to engage the generalist aspects of brain function" (Donald, 2001, p. 257). Increasing complexity of cultural life and the demands placed on the attendant social behaviours may itself select for cognitive flexibility. For an individual, the social environment is as much a part of the ecological environment (or EEA, environment of ecological adaptedness) as any other part. The way an individual interacts with it – responds to it and influences it – can be as important to their survival as the way they interact with the rest of the natural environment.

The next stage in Donald's model builds on this foundation. He calls it, somewhat poetically, the "mythic" stage, and this is characterised by the emergence of language, between archaic and modern humans. The emergence of fully linguistic behaviours has traditionally been given the most credence as the defining step in human cognition, responsible for much or all other symbolism, aesthetics and characteristic aspects of human cognition. For Donald, this is not the case. Whilst language is clearly hugely important in the manifestation and transference of knowledge, and allowed the externalisation of knowledge into the "distributed cognition" of a community (Donald's third – "theoretic" - transition, with modern humans in Europe) linguistic complexity was not itself the driving force: "The priority was not to speak, use words or develop grammars. It was to bond as a group, to learn to share attention and set up the social patterns that would sustain such sharing and bonding in the species" (Donald, 2001, p. 253); i.e. the language capacity was initially shaped by the social world, rather than vice-versa; increasingly complex societies lead to increasingly complex behaviours, initially - although there would be feedback of course.

This, in turn, led to the development of narrative thought, abstract symbolism and external representation of knowledge in the "mythic" and "theoretic" stages of Donald's model; Whilst the implications of these developments are very great, and there is an accordingly great body of literature to that effect, Donald’s thesis regarding mimesis is that we should not underestimate the importance of this earlier stage of cognitive development and what it allowed: “The great divide in human evolution was not language but the formation of cognitive communities in the first place”, and “symbolic cognition could not self-generate until those communities were a reality” (p. 254).

It is evident that it is in this first major cognitive development (of mimetic capacities) that the greatest proportion of the capacities for musical behaviours emerged (and were integrated) in hominids. The latter developments of full language, symbolic representation and externalised knowledge would certainly have shaped those abilities in ways that could allow the development of the diversity of forms and roles of music that we see today, but these forms and roles do not themselves constitute *capacities* for music. The attribution of symbolic associations to music, and the integration of lyrical content to add specificity to its emotional content or add to its information-carrying potential, are sophisticated add-ons, using new skills, supported by the edifice, but not the central edifice itself.

8.3. Chapter summary and conclusions

So how do these models of cognitive evolution relate to the considerations of the possible selective advantages of musical behaviours, and their emergence as behaviours in their own right? Donald’s account is most directly relevant to the considerations of the evolution and integration of the various capacities for musical behaviours that formed the subject matter of Chapters 4-6. Its relevance for the models examined in Chapter 4 is that it puts these integrated foundations in place with *Homo erectus*, and developing between *Homo erectus* and archaic *Homo sapiens* (including *Homo heidelbergensis*), as part of a continuity of the development of cognitive capacities that mitigated increased social complexity in hominid evolution. The implication of the models discussed in Chapter 4, and of the findings of Chapter 4, was that such capacities, and their physiological concomitants, had become highly developed by *Homo heidelbergensis*. It is likely that many of the potentially useful roles of proto-musical behaviours, discussed in Chapter 7, could have been fulfilled at this time by these integrated capacities. It seems likely that at this time we could see exhibited various behaviours that, whilst they may not fit our definition of *music*, we would recognise as consisting of fundamental elements of musical behaviour, and fulfilling some of the roles that musical activities fulfil today. At this stage there would still have been considerable overlap between the roles fulfilled by proto-musical vocalisation and by social-emotional proto-linguistic vocalisations. The behaviours manifest from these capacities would have been essentially socially functional, and visceral – part of the continuum of social vicarious "grooming" activities.

It would not be until the speciation of anatomically modern *Homo sapiens* that the subsequent social and cultural diversity and innovation could shape those proto-musical behaviours into the diversity of forms that we recognise – sometimes hesitantly, but inevitably – as music today. Cognitive fluidity any greater than that inherent in social intellectual capacities is not a necessary pre-requisite of musical behaviours, even today, but the development of such analogical thinking and symbolic external representation would undoubtedly have subsequently had a great influence on the shaping of the *manifestation* of musical behaviours and their cultural associations. Perhaps equally importantly, the stimulation of such integrated capacities by proto-musical and musical behaviours could have itself helped instigate the integration of such "new" cognitive abilities in modern humans. It was with modern humans that developed linguistic vocal and musical behaviours would have truly diverged from their shared heritage; it may, indeed, been with that speciation event that the limitations of the essentially modern vocal tract in meeting new demands of social and ecological complexity led to the development of lexical and semantic structure to vocalisations. However, that shared heritage of linguistic and musical vocalisations, in terms of physiology, neurology, production, perception, development and affective content, is still very much in evidence today.

CONCLUSIONS

> "Most scholars have wisely steered clear of the issue of music's origins, since clearly the enterprise is patently speculative. At its worst, proposals concerning music's origins are fiction masquerading as scholarship".
>
> (Huron, 1999)

Hopefully the preceding research and findings have provided some indication that Huron's words of caution (which in fact precede his own writings on the subject) need not apply to all considerations of the origins of music. That "music" is so often a nebulous concept does not mean that the capacities underlying musical behaviour are in any way nebulous, or that investigation regarding their development and interrelationships need be any more speculative than any other investigation in human origins. It is hoped that the preceding research illustrates that, in contrast to "wisely steering clear" of the issue, it is one that should be explored in more depth by scholars, that such investigations may be fruitful, and that it is an important issue in any consideration of the emergence of human cognition and social behaviour.

This final section of the thesis does not aim to replicate the concluding sections of each of the chapters, but to reiterate the main implications of the evidence. These are then drawn together, to build up a picture of the origins and evolution of the human capacities for musical behaviours, and their manifestation in the Palaeolithic.

A summary of the preceding sections

Section I – Ethnographic evidence for the use of music

Despite occupying four very different environments, and being united only by their subsistence method, the musical behaviours of the four hunter-gatherer groups considered have a number of important features in common.

All four groups consider themselves to have come from the land, to be akin with the other fauna present in the environment, and use music to attempt to influence the world around them. Music plays an increased role in the most difficult subsistence season, when groups meet with neighbouring communities; at this time it has important roles in ceremonial and communal social activity. Melody is predominantly vocal, with the majority of instruments being percussive. Instruments are made from naturally occurring organic materials, generally requiring little modification. Vocal melody is, in the Plains

Indians and African Pygmies, almost exclusively non-lyrical, consisting of vocal sounds rather than words; amongst the Australian aborigines lyrics are used to store knowledge, whilst amongst the Eskimos both forms are used. Music and dance can have important uses in engendering group cohesion, altering mood, as an aid to the teaching of dance, and can facilitate group interactions and communality, within and between groups. In the majority of these instances the music itself has no directly attributed symbolism, but it can be used to accompany symbolic activities.

There should thus be no expectation that we would see evidence for instrumentation as soon as humans participated in musical behaviour, or that the first incidence of instrumentation indicates the first incidence of musical behaviour. Musical behaviours need not have been dependent upon the presence of language, symbolism or instrumentation, but instead would be limited by physical and neurological capabilities.

The fundamental similarities that exist in the nature and roles of music between these diverse groups of hunter-gatherers could imply either convergent development of these behaviours in the different environments in which we find these peoples today, or a shared cultural heritage (or both). The former (convergence) would suggest that there are important evolutionary driving forces towards those common behaviours, either as a consequence of subsistence method or of human biology, these being the principal common factors between the groups. The latter situation (shared heritage) would indicate a very ancient tradition of musical behaviour, as some of these groups are probably separated by at least 50,000 years.

Section II – Palaeolithic music archaeology

At the time of writing there is no indisputable evidence for musical instrumentation before the Upper Palaeolithic in Europe around 40,000 years ago. Some of the evidence, particularly earlier, has been rather too readily interpreted as intentional sound-producers, being instead the product of carnivore activity or natural forces. The earliest bone pipes about which we can be confident of their status as intentional sound-producers date to around 36,000 years old, and are already sophisticated in design. This evidence indicates that when modern humans arrived in Europe they were already carrying out developed instrumental behaviours; it seems likely that these instruments are considerably temporally removed from the earliest instruments, and even further removed from the earliest musical behaviours. All of the pipes about which we can be confident of their status as intentional sound-producers are made on bird bone; the apparently sudden use of this resource for pipe manufacture may be due the adoption of avian fauna as a subsistence resource in Upper

Palaeolithic Europe. Differential preservation of bone over other organic materials is almost certainly a factor in the profile of the record that has been uncovered to date; another factor is its focus on Europe over the rest of the old world occupied by anatomically modern humans at that time and before. After 30,000 years ago in particular there is a sudden diversity of evidence for musical activity, with rasps, percussion, many more bone pipes, and the acoustic use of rocks and caves, particularly during the eras associated with the Solutrean and Magdalenian techno-complexes. These seem to be widespread, often occurring in what appear to be focal locations for large amounts of human activity, including art. Such evidence as there is suggests that musical performance, or at least appreciation, was a group activity, rather than one involving a select few.

Section III – Evolution of the physiological and neurological capacities for music

It would seem that from *Homo ergaster* (c. 1.73 Mya) to *Homo heidelbergensis* (300-400 Kya) there was a co-evolution of the various features of the vocal apparatus, allowing a progressively greater range of sounds to be produced, and of progressively greater duration. In the case of vocal tonal range, as indicated by laryngeal lowering, this appears to have been a continuous change over this time; in the case of the control of extended vocalisation, as indicated by thoracic canal innervation, it is impossible to say at this stage whether this was a gradual process or a single event at speciation. It is likely that increasing control over tonal quality and range preceded increase in duration of utterances. In any case, by *Homo heidelbergensis* (or archaic *Homo sapiens*), a set of vocal apparatus that was essentially fully modern was in place, with the ability to produce a modern range of sound frequencies with the larynx, and to articulate them with the oral anatomy. By the time of Neanderthals and modern humans (and quite possibly their common ancestor) there was the ability to control vocalisations of extended duration.

It seems that the evolution of the vocal chords and ear must have occurred in tandem, for perceptual reasons, with feedback occurring between the two processes of development; indeed, *Homo ergaster* was the first hominid to begin to develop a modern human vocal apparatus, and modern inner ear structure is also first exhibited by *Homo ergaster*, c. 1.5 Mya. This is as a consequence of a variety of other physiological changes, including the development of a fully bipedal posture. Tied in with this labyrinthine evolution is the development of neurological structures which may relate to language learning, and the production and perception of complex vocal sequences such as would form the basis for melodic vocalisations.

Neurological evidence confirms the association between the emergence of vocal versatility and the development of the ability to plan and execute structured utterances of extended duration. The development of Broca's area and surrounding areas responsible for integrating vocal fold control, expiratory control and orofacial muscular control (and thus overall control of the laryngeal system) suggests that our ability to make planned emotional vocal sequences, including melody, developed alongside the ability to produce an increasing range of sounds; the ability to actually plan and execute melodic vocalisations was probably proportional to thoracic canal and laryngeal development.

Shared neurological mechanisms are responsible for the production and processing of prosody of speech, pitch perception, melodies and emotional sounds, and are bilaterally localised, although important elements of these functions are controlled predominantly by the right hemisphere. These functions probably predate those controlled by the left hemisphere, concerned with the production and processing of rhythm and meter in speech and of rhythm and timing in music. The evidence suggests that the specialised human neurology dedicated to the more analytical "higher" linguistic and musical functions emerged later out of a set of common neurological substrates concerned with vocal emotional expression and comprehension. It could be that the emotional limbic system and the intrinsic motor formation constitute the uniting neurology between the tonal emotive vocalisation elements and the rhythmic-motor elements of interaction and musicality. Such integration must have emerged with the development of the left-hemisphere functions described.

The capacity to perceive melody and rhythm and tonal vocalisation and to gain affective information from them is inherited and innate at birth. Such vocalisations have important social functions, and similar capabilities in primate infants suggests that they have a long phylogenetic history. There is no reason for us to believe that any of our hominid ancestors could not have also exhibited such behaviours; differences from species to species would have been in terms of the vocal range, control and capabilities, which would ultimately bear upon the complexity of the utterances possible, and normally used.

Section IV – Evolutionary rationales for music

Some vocal sounds and frequency changes are fundamental, invariant across cultures, and even species, in the emotional state that they express. In producing, and even to an extent in perceiving, a particular sound we generate some emotional response in ourselves due to the kinaesthetic feedback from the physiology required to produce that sound.

There are numerous parallels between speech and musical melodic vocalisation, in terms of variable pitch contour, high rhythmicity, repetitive motifs, and the communication of affect, modulation of arousal, and eliciting of attention and affective response. It is proposed that this is because the shared prosodic pitch and tempo-related properties of emotional vocalisation (infant-directed and adult-directed) and music are not borrowed from one to the other, in either direction, but are, and always have been, a shared fundamental component of both.

As a result, one of the ways in which music elicits emotion is as a consequence of processing by, and stimulation of, auditory and kinaesthetic mechanisms associated with emotional expression and interpersonal interaction. We can react to and process music in many of the same ways as we do interaction with a person; it shares many of the properties of such interaction, and is apparently processed by the same mechanisms.

Corporeal gesture and vocal gesture are interrelated. It seems that there is a cognitive rhythmic motor coordinator which instigates such muscular sequences, and that the complex patterns of muscular movement of gesture (finger, hand, arm, shoulder and joint musculature) and vocalisation (orofacial, laryngeal and respiratory musculature) are coordinated. This concordance between gesture and speech is instigated early in the vocalisation process, by cyclical motor control, with gestural movements being associated with the prosodic rhythm of vocalisation.

Production of complex vocalisation relies on priming of the whole motor system, and affective expression is multi-modal, being manifested in posture, gesture, facial expression and vocal quality. Humans are extremely sensitive to such expression, in production and perception, and feedback can occur between the different media of expression, and as a consequence of perception of such expression too – as well as being able to observe such cues, we can empathically experience something of their emotional state in mirroring them with our own bodies. Musicality and rhythmic movement involve deliberate control and sequencing of this system. There seems to be a close interrelationship between the emotional-controlling elements of the limbic system and the areas responsible for the co-ordination of motor sequences and posture, driven by the intrinsic motor formation; i.e. there can be self-directed feedback from movement into emotional state and, importantly, feedback and interaction between individuals, in terms of synchrony of movement and of emotional state.

So, whilst rhythm and melody are produced and processed in neurologically specialised and distinct areas of the brain which are somewhat independent of each other, there is also important integration of these systems, with rhythmic muscular movements being coordinated with prosodic elements of vocalisation in their production. Specifically, they are inter-dependent in the planning and execution of sequences of muscular movement associated with instigation of vocalisation, rhythmic physiological movement, and expression of emotional state in these media. This is a capacity that has emerged since the separation of the hominid lineage from that of chimpanzees, and the evidence of Section III suggests that its emergence would be related to the development of the left hemisphere functions considered.

Pre-syntactic, pre-lexical proto-linguistic communication would benefit from the use of the full range of the vocal tract, in terms of expressing its tonal content; the greater the range and control over vocalisation possible, the greater the expressiveness of such vocalisation. Laryngeal lowering allowed the emergence of such increasing versatility and control, as well as greater orofacial moderation of laryngeal sounds.

In summary, musical behaviours have as their foundation the vocalisations and gestures associated with the expression of emotion and maintenance of social bonds. Musical capacities are built on fundamentally important social and physiological mechanisms and, at an essential level, are processed as such. Music uses capacities crucial in situations of social complexity: the vocal, facial and interactive foundations of these capabilities are evident in other higher primates, and such capacities would have become increasingly important and sophisticated as group size and complexity increased. Vocal emotional expression, interaction, sensitivity to others' emotional state and empathic ability would have been selectively important abilities; those individuals in which such capabilities were more developed would have been selectively favoured. Fundamentally integrated into the planning and control of complex sequences of vocalisations, and related to the prosodic rhythm inherent in such sequences, is rhythmic motor coordination. The motor system is primed in the instigation of such vocal behaviours, and corporeal gesture is consequently incorporated into the execution of the vocal behaviour.

In terms of potential selective advantages associated with carrying out developed musical behaviours, they could confer an advantage on individuals in terms of sexual selection – because of their foundations in the capacities to communicate emotionally and effectively, to empathise, bond and elicit loyalty, musical abilities have the potential to be a proxy for an individual's likelihood of having strong social networks and loyalties, and of contributing to a group. Musical behaviour also has the potential for being a mechanism for actually stimulating and maintaining those networks and loyalties; because of the stimulation of shared emotional experience as a consequence of participation in musical activities, it has the potential to engender strong feelings of empathic association and group membership.

Musical or proto-musical behaviour has the potential to make use of several cognitive capacities at once, relying on the integration and control of biological, psychological, social and physical systems; furthermore it gives the opportunity to practise and develop these integrated skills in a context of limited risk. However, musical behaviour, and thus its emergence, does not rely on the presence of a specific "music module" or the breaking down of barriers between cognitive "domains of intelligence" in the way that Mithen (1996) advocates would be required for art and religion. Whilst the ability to cognize across domains of intelligence, and to think analogically, is undoubtedly a fundamentally important aspect of human cognition, musical behaviours need not rely on such capabilities. D'Errico *et al.* (2003) astutely entitle their paper "Archaeological evidence for the emergence of language, symbolism and music…", making an important distinction between the three. Music is neither language, nor symbolism, and need not rely on the presence of either but, like language and symbolic behaviour, has foundations in communication between individuals.

A proposal for the evolution of musical capacities

The increase in manual dexterity with *Homo habilis* led to the first development of left-hemisphere neuro-anatomy related to fine manual control, approximately 2 million years ago; associated with these areas are those controlling the planning and execution of fine muscular sequences, including vocal and orofacial musculature. Thus from the outset fine vocal and manual-corporeal gesture are related. In *Homo habilis* this development was initially small. *Homo habilis* would probably have had the ability to make discrete vocal utterances of limited tonal range and duration, for emotional expression, parent-infant communication, warning and signaling, similar to those exhibited by modern higher primates. These would be controlled predominantly by neurological mechanisms that are right-hemisphere and bilaterally localised.

With the speciation of *Homo ergaster*, a number of changes took place simultaneously. The shift to a fully upright bipedal posture led to changes in cranial morphology that impacted on both the vocal tract and inner ear. The larynx began to lower in the throat, allowing a greater range of vocal sounds to be made. Whether this occurred initially as a by-product of the physiological changes associated with bipedalism, or directly as a consequence of advantages associated with versatility of vocal communication, is impossible to assert; nevertheless, the latter advantages would have soon become selectively advantageous in the situations of increased social complexity contiguous with the move to diverse environments, especially with the migrations of *Homo erectus*. The increase in vocal range achievable led to an increase in cervical vertebral innervation, allowing finer control over the greater tonal range producible with the lowered larynx. The production and perception of tonal prosodic sounds of vocalisation were (and are) handled predominantly by the right hemisphere, and by sub-cortical emotional-limbic systems, but the left hemisphere mechanisms that began to emerge with *Homo habilis* became increasingly developed, responsible for the planning and muscular execution of those sounds. In particular this is likely to have taken the form of the development of the neurological pathway from the periaqueductal grey matter to the nucleus ambiguus, which is responsible for fine laryngeal, orofacial and expiratory control. This allowed voluntary control over the *structure* and *complexity* of vocalised sequences, hitherto impossible. The integration of the prosodic/emotional and rhythmic/structural elements of vocalisation is handled by the intrinsic motor formation (including the emotional limbic system). The instigation of vocalisations is handled by a rhythmic motor coordinator which is also responsible for complex patterns of muscular movement of gesture (finger, hand, arm, shoulder and joint musculature) and vocalisation (orofacial, laryngeal and respiratory musculature); gestural movements are associated with the prosodic rhythm of vocalisation. The increasingly sophisticated vocalisations thus produced were predominantly prosodic, emotional-tonal, and used in social interaction, a form of vocal social grooming. They were, however, limited in tonal range and duration.

Selective advantages associated with aptitude in these vocal abilities led to their continued development over the next 1.5 million years. By the time of the common ancestor of *Homo neanderthalensis* and *Homo sapiens*, thoracic vertebral innervation was of modern dimensions, suggesting the emergence of fine control over expiration, and thus the ability to carry out vocalisations of extended duration. Coupled with this would need to be increased development of the ability to plan and execute complex vocal and orofacial muscular sequences and, indeed, there is further left hemisphere development by this time. archaic *Homo sapiens* or, at least, *Homo heidelbergensis*, 300,000 years ago, exhibits vocal anatomy that is essentially modern, in terms of the range of vocal sounds producible, the control over those sounds, and their duration.

In sum, by this time there is in place a system in which production and perception of the gestural (vocal, orofacial and corporeal) expression of emotion is a consequence of, and results in, priming of the rhythmic-motor and emotional systems. Rhythmic sequences, and the prosodic content and rhythm of tonal sequences, prime this system, and each other.

This results in a multi-modal relationship between, on the one dimension, rhythmic and emotive-prosodic content, and on the other dimension, vocal, auditory, visual and kinaesthetic expression of that content.

This system of vocal and kinaesthetic communication of emotion constituted proto-linguistic and proto-melodic behaviour, a "musilanguage" (to use Brown's term). This formed the foundation for vocal communication out of which later emerged semantic, lexical linguistic capabilities, and melodic, rhythmic musical behaviours. It is proposed that infant-directed vocalisations in modern humans are a specialised perpetuation of this sophisticated form of non-linguistic interpersonal interaction which was used earlier between all individuals.

Thus the foundations of musical behaviours were not selectively favoured specifically for music; however, the interrelationship between fine muscular control of vocalisation, corporeal movement, emotion and rhythmic control mean that the emergence of music-like, or proto-musical behaviours was almost inevitable in any populations with the capacities to plan end execute tonally complex utterances.

At some point between archaic *Homo sapiens* and anatomically modern *Homo sapiens* specialisations in vocalisation form began to occur, in terms of linguistic and melodic vocalisations being discrete entities. Perhaps due to the inability of vocal range to increase any further due to the full lowering of the larynx, or to some other selective pressure, syntactic structure began to emerge in vocal and gestural control, eventually coming to dominate the structure of vocalisations, in the form of lexical syntactic language. Other authors have proposed at what point in the emergence of *Homo sapiens* this occurred. In terms of the emergence of full (specialised, as opposed to proto-) musical behaviours, their foundations in social interaction, emotional expression, and fine control and planning of corporeal and vocal muscular control lends them extremely well to exercising the integration of important cognitive skills. Cross (2003b) suggests they may thus have been instrumental in the emergence of modern human cognitive abilities. In any case the execution of musical activities could become increasingly important, and beneficial, on both an individual and group level, with increasing social complexity within and between groups.

Because music production and perception is handled by the brain in ways that are complex and related to interpersonal interaction and the formation of social bonds, it stimulates many functions thus associated. It seems musical participation, even without lyrics or symbolic associations, can act on the brain in ways that are appealing to humans, because of their vicarious stimulation of fundamentally important human interactive capacities – musical production and perception stimulates parts of the cognitive system concerned with interpersonal interaction, empathy and expression. Not only does it do so on an individual basis, but also thus adds to the potency of activities which are, in any case, social and interactive. Consequently musical behaviours could be instrumental in ensuring in-group cooperation and relations, the advantages associated with which would lead to the rapid spread of such traditions within and between groups.

These are by no means music's only appeals, or only way of eliciting emotional response, both of which have many associations in the other activities and symbolisms that, in modern humans, can be culturally built up around music; however, the underlying social-emotional factor is a fundamental element of music's action, *irrespective* of those other activities and symbolisms which may be "built on" afterwards. The perception of musical behaviours as necessarily being part of a repertoire of symbolic behaviours is probably due largely to Western conceptions of music, and to the association of music with *other* activities with which music in most cultures is now contiguous. It also possibly owes something to the misconception that musical behaviour is always carried out "for its own sake", and with no direct benefit or consequence. Having said this, the emergence of symbolising capabilities in modern humans would have added immeasurably to the diversity of circumstances and forms in which musical behaviours were manifested, as well as adding many more layers of meaning and association for participants in such activities.

Modern humans apparently arrive in Europe 36-40,000 years ago with developed instrumental behaviours. Because of the melodic capabilities of the vocal tract, we can expect that melodic instrumentation would have been a relatively late development in the prehistory of music. We can also expect that rhythmic percussive activity would have accompanied complex vocalisation from an early time, because of the fundamental link between the production of planned complex sequences of prosodic vocalisations and the instigation of corporeal motor control, described above. It may be that the use of large avian faunal bone at this time for instrument manufacture marks an innovation based on the emergence of the use of a new subsistence resource. The diversity of instruments made on bone increases in the record over the next 10-15,000 years with the appearance of rasps, bullroarers and many more pipes, as well as evidence of the use of lithic and environmental features for deliberate sound production. It appears that these activities were by this time, at least in the contexts that many of them are found, associated with group and symbolic activities. Ethnography suggests that such musical activities, in the context of hunter-gatherer subsistence, would have become increasingly important in times of subsistence stress, and would have been associated with ritual perceptions of the world, supernatural in origin, powerful, and capable of influencing the people and the world around. There can be little doubt that the musical behaviours undertaken in the European Upper Palaeolithic were ones that we would recognise today.

Limitations of the current research, and proposals for future research

The focus of the research represented in this thesis has been the emergence, development and interrelationship of the *capacities* that underlie musical behaviours, and the nature of the archaeological record representing recognisable musical behaviours. Some possible mechanisms through which the capacities may have emerged as full musical behaviours have been proposed, but much remains to be investigated regarding the transition between proto-musical behaviours and fully-realised musical behaviours, and the nature of the diversification of specialised melodic vocalisations from linguistic vocalisations.

The vast majority of finds relating to musical activities from Palaeolithic contexts were discovered in the latter parts of the 19th and early 20th century, and lack contextual associations and site reports. Many have been analysed as parts of museum collections, so little can be said about anything other than the properties of the artefacts themselves. Future such finds in Palaeolithic contexts will hopefully be analysed in terms of the context and the possible uses to which the artefacts were put. Whilst almost certainly not a musical instrument, the excavation and investigation of the Divje babe I bone provides an excellent model for the future analysis of reputed instrument finds. Such finds and research will be enormously valuable in building up a more complete picture of musical behaviours in the Palaeolithic than could be presented here.

Archaeology in general, and especially research of the Palaeolithic has, in the past, focused on the visual medium, to the exclusion of consideration of the auditory environment. Current work of researchers such as Lawson, Scarre, Cross, Watson and Reznikoff, for example, is addressing this lacuna, and will add a great deal to models of the emergence of the human behaviours such as those that have been considered here.

Investigation of the origins, emergence and nature of musical behaviours in humans is in its infancy, and has plenty more to reveal. It concerns, after all, an element of human behaviour that, in contrast to Pinker's (1997) opinion, the vast majority of people would miss very much if we were suddenly bereft of it. In fact, it would be impossible to remove music without removing many of the abilities of social cognition that are fundamental to being human.

REFERENCES

Absolon, C. (1936) Les flûtes Paléolithiques de l'Aurignacien et du Magdelénian de Moravie (analyse musicale et ethnologique comparative, avec démonstrations), *Congres Préhistiore de France.* II.XI session: p. 770-784.

Addis, L. (1999) *Of Mind and Music*. Cornell University Press, Ithaca.

Ahuja, A. (2000) Why we are touched by the sound of music, *The Times (Science Supplement)*, 23rd February 2000, p. 43.

Aiello, L. C. & Dunbar, R. I. M. (1993) Neocortex size, group size, and the evolution of language, *Current Anthropology* Vol. 34, p. 184-193.

Aitken, P. G. (1981) Cortical control of conditioned and spontaneous vocal behavior in rhesus-monkeys, *Brain and Language* Vol. 13, p. 171-184.

Albrecht, G., Holderman C.-S., Kerig, T., Lechterbeck J. & Serangeli J. (1998) "Flöten" aus Bärenknochen – die frühesten Musikinstrumente?, *Archäologisches Korrespondenzblatt* Vol. 28, p. 1-19.

Albrecht, G., Holderman, C.-S. & Serangeli, J. (2001) Towards an archaeological appraisal of specimen no. 652 from Middle-Palaeolithic level D/(layer 8) of the Divje babe I, *Archeoloski vestnik* Vol. 52, p. 11-15.

Alcock, K. J., Passingham, R. E., Watkins, K. & Vargha-Khadem, F. (2000) Pitch and timing abilities in inherited speech and language impairment, *Brain and Language* Vol. 75, p. 34-46.

Alebo, L. (1986) Manufacturing of drumskins and tendon strings for prehistoric musical instruments, in C. S. Lund (ed.) *The Second Conference Of the ICTM Study Group On Music Archaeology: Volume I, General Studies*. Royal Swedish Academy Of Music, Stockholm, p. 41-48.

Álvarez, R. & Siemens, L. (1988) The lithophonic use of large natural rocks in the prehistoric canary islands, in E. Hickman and D. W. Hughes (eds.) *The Archaeology Of Early Music Cultures; Third International Meeting Of The ICTM Study Group On Music Archaeology*. Verlag für systematische Musikwissenschaft, Bonn, p. 1-10.

Anonymous (2000) Music in our bones, *The Times*, 21st February.

Arensberg, B. (1989) Anatomy of Middle Palaeolithic populations in the middle east, in P. Mellars & C. Stringer (eds.) *The Human Revolution: Behavioural and Biological Perspectives in the Origins of Modern Humans*. Edinburgh University Press, Edinburgh, p. 165-171.

Arensberg, P., Schepartz, L. A., Tillier, A. M., VanDerMeersch, B. & Rak, Y. (1990) A reappraisal of the anatomical basis for speech in middle Palaeolithic hominids, *American Journal Of Physical Anthropology* Vol. 83, p. 137-46.

Armstrong, D. F., Stokoe, W. C. & Wilcox, S. E. (1994) Signs of the origins of syntax, *Current Anthropology* Vol. 35, p. 349-368.

Arsuaga, J.L. (1997) Sima de los Huesos (Sierra de Atapuerca, Spain): the site, *Journal of Human Evolution* Vol. 33, p. 109-127.

Arsuaga, J. L., Martinez, I., Lorenzo, C., Gracia, A., Muňoz, A., Alonso, O. & Gallego, J. (1999) The human cranial remains from Gran Dolina lower Pleistocene site (Sierra de Atapuerca, Spain), *Journal Of Human Evolution* Vol. 37, p. 431-457.

Baghemeil, B. (1988) The morphology and phonology of Katajjait (Inuit Throat-Games), *Canadian Journal of Linguistics* Vol. 33, p. 1-58.

Bahn, P. (1983) Late Pleistocene economies in the French Pyrenees, in G. N. Bailey (ed.) *Hunter-Gatherer Economy In Prehistory: A European Perspective*. Cambridge University Press, p. 167-185.

Bahn, P. (1997) *Journey Through the Ice Age*. Weidenfeld and Nicholson, London.

Bahuchet, S. (1999) Aka Pygmies, in R. B. Lee & R. Daly (eds.) *The Cambridge Encyclopedia of Hunters and Gatherers*. Cambridge University Press, Cambridge, p. 190-194.

Barac, V. (1999) From primitive to pop: foraging and post-foraging hunter-gatherer music, in R. B. Lee & R. Daly (eds.) *The Cambridge Encyclopedia of Hunters and Gatherers*. Cambridge University Press, Cambridge, p. 434-450.

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

Barratt, L., Dunbar, R. & Lycett, J. (2001) *Human Evolutionary Psychology*. Palgrave Macmillan, Basingstoke.

Bastiani, G. & Turk, I. (1997) Appendix: results from the experimental manufacture of a bone flute with stone tools, in I. Turk (ed.), *Mousterian "Bone Flute"*. Znanstvenoraziskovalni Center Sazu, Ljubljana, p. 176-178.

Beaudry, N. (1978) Towards transcription and analysis of Inuit throat-games: macro-structure, *Ethnomusicology* Vol. 22, p. 261-273.

Benade, A. H. (1990) *Fundamentals of Musical Acoustics*, Dover, New York.

Benson, D. F. (1985) Language in the left hemisphere, in D. F. Benson & E. Zaidel (eds.) *The Dual Brain*. Guildford Press, New York, p. 193-203.

Bergounioux, F. M. & Glory, A. (1952) *Les Premiers Hommes. Précis d'Anthropologie Préhistorique*. Didier, Paris, p. 289-95, fig. 151.

Bermejo, M. & Omedes, A. (1999) Preliminary vocal repertoire and vocal communication of wild bonobos (*Pan paniscus*) at Lilungo (Democratic Republic of Congo), *Folio Primatologica* Vol. 70, 328-357.

Bettinger, R. L. (1991) *Hunter-Gatherers: Archaeological and Evolutionary Theory*. Plenum Press, London.

Bibikov, S. (1978) A stone age orchestra, in D. E. K. Hunter & P. Whitten (eds.) *Readings in Physical Anthropology and Archaeology*. Harper and Row, London.

Binford, L. R. (1967) Smudge pits and hide smoking. The use of analogy in archaeological reasoning, *American Antiquity* Vol. 32, p. 1-12.

Binford, L. R. (1983) *Working At Archaeology*. Academic Press, New York.

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

Blount, B. G. (1994) Comment on Armstrong, Stokoe and Wilcox, *Current Anthropology* Vol. 35, p. 358-359

Bogen, J. E. (1985) The dual brain: some historical and methodological aspects, in D. F. Benson & E. Zaidel (eds.) *The Dual Brain*. Guildford Press, New York, p. 27-43.

Borchgrevink, H. M. (1980) Cerebral lateralisation of speech and singing after intracarotid Amytal injection, in M. Taylor Sarno & O. Hooks (eds.) *Aphasia: Assessment and Treatment*. Almkvist & Wiksell, Stockholm, p. 186-191.

Borchgrevink, H. M. (1982) Prosody and musical rhythm are controlled by the speech hemisphere, in M. Clynes (ed.) *Music, Mind and Brain*. Plenum Press, New York, p. 151-157.

Borchgrevink, H. M. (1991) Prosody, musical rhythm, tone pitch and response initiation during Amytal hemisphere anaesthesia, in J. Sundberg, L. Nord & R. Carlson (Eds.) *Music, Language, Speech and Brain*. MacMillan Press, Basingstoke, p. 327-343.

Borg, E. & S. A. Counter (1989) The middle-ear muscles, *Scientific American*, August 1989, p. 63-68.

Bowles, S. & Gintis, H. (1998) The moral economy of community: structured populations and the evolution of pro-social norms, *Evolution and Human Behaviour* Vol. 19, p. 3-25.

Brade, C. (1975) Die mittelalterlichen kernspaltflöten Mittel- und Nord Europas, *GottingerSchriften zür vor und Frügeschichte* 14. Wacholz, Neumunster.

Brade, C. (1982) The prehistoric flute: did it exist? *Galpin Society Journal* Vol. 35, p. 138-150.

Breen, M. (1994) I have a dreamtime: aboriginal music and black rights in Australia, in S. Broughton *et al.* (eds.) *World Music: The Rough Guides*. Rough Guides Limited, London, p. 655-662.

Brodar, S. & Bayer, J. (1928) Die Potocka Zijalka, eine Hochstation der Aurignacschwankung in den Ostalpen, *Praehistorica* Vol. 1, p. 3-13.

Brodar, S. & Brodar, M. (1983) *Potocka zijalka. Visokoalpska postaja aurignacienskih lovcev. Potocka zijalka, eine hochalpine Aurignacjägerstation*. Slovenska akademija znanosti in umetnosti, Ljubljana.

Broughton, S., Ellingham, M., Muddyman, D. and Trillo, R. (1994) *World Music: The Rough Guide*. Rough Guides Limited, London.

Brown, S. (2000) The "musilanguage" model of music evolution, in N. L. Wallin, B. Merker & S. Brown (eds.) *The Origins of Music*. MIT Press, London, p. 271-300.

Brown, S., Merker, B. and Wallin, N. (2000) An introduction to evolutionary musicology, in N. L. Wallin, B. Merker & S. Brown (eds.) *The Origins of Music*. MIT Press, London, p. 3-24.

Buchanan, T. W., Lutz, K., Mirzazade, S., Specht, K., Shah, N. J., Zilles, K., & Jancke, L. (2000) Recognition of emotional prosody and verbal components of spoken language: an fMRI study, *Cognitive Brain Research* Vol. 9, p. 227-238.

Budil, I. (1994) A functional reconstruction of the supralaryngeal vocal tract of the fossil hominid from Petralona, in J. Wind, A. Jonker, R. Allot & L. Rolfe (eds.) *Studies In Language Origins Vol. 3*. Benjamins, Amsterdam, p. 1-19.

Burling, R. (1993) Primate calls, human language, and nonverbal communication, *Current Anthropology* Vol. 34, p. 25-53.

Burnham, D., Kitamura, C. & Vollmer-Conna, U. (2002) What's new, pussycat? On talking to babies and animals, *Science* Vol. 296, p. 1435.

Bundo, D. (2002) Mode of participation in singing and dancing performances among the Baka, in *Proceedings of the 9th International Conference On Hunting and Gathering Societies*, in press. Currently available online at http://www.abdn.ac.uk/chags9/Bundo.htm

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

Callaghan, C. A. (1994) Comment on Armstrong, Stokoe and Wilson, *Current Anthropology* Vol. 35, p. 359-360.

Calvin, W. H. (1996) *How Brains Think*. Weidenfeld and Nicholson, London.

Campbell, M. & Greated, C. (1987) *The Musician's Guide to Acoustics*. Dent and Sons, London.

Capitan, L. & Peyrony, D. (1912) Station préhistorique de La Ferrassie, commune de Savignac-Deu-Bugue (Dordogne), *Revue Anthropologique* Vol. 22, p. 76-99.

Carlson, N. R. (1994) *The Physiology of Behavior*. Allyn and Bacon, Massachusetts.

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

Cartwright, J., & Davies, R. (2000) *Evolution and Human Behaviour: Darwinian Perspectives on Human Nature*. Palgrave Macmillan, Basingstoke.

Cassoli, P. F. & Tagliacozzo, A. (1997) Butchering and cooking of birds in the Palaeolithic site of Grotta Romanelli (Italy), *International Journal of Osteoarchaeology* Vol. 7, p. 303-320.

Cernys, A. P. (1955) Une flûte paléolitique, *Kraktie soobscentija Instituta Istorii materialnoi kyltyry, Moskva* 59, p. 129-130.

Cernys, A. P. (1956) Nouvelles découvertes dans le Fuilles de la Station de Molodova 5 sur le Dneistr, *Kraktie soobscentija Instituta Istorii materialnoi kyltyry, Moskva* 63, p. 150-152, Fig. 70.

Chang, K. C. (1967) Major aspects of the interrelationship of archaeology and ethnology, *Current Anthropology* Vol. 8, p.227-243.

Chase, P. (1990) Sifflets du Paléolithique moyen(?): Les implications d'un coprolite de coyote naturel, *Bulletin de la Société Préhistorique Française* Vol. 87, p. 165-167.

Chase, P. (1989) How different was Middle Palaeolithic subsistence? A zooarchaeological perspective on the Middle to Upper Palaeolithic transition, in P. Mellars & C. Stringer (eds.) *The Human Revolution: Behavioural and Biological Perspectives in the Origins of Modern Humans*. Edinburgh University Press, Edinburgh, p. 321-337.

Chase, P. & Nowell, A. (1998) Taphonomy of a suggested Middle Palaeolithic bone flute from Slovenia, *Current Anthropology* Vol. 39, p. 549-553.

Clark, J. G. D. (1953) Archaeological theories and interpretation: Old World, in A. L. Kroeber (ed.) *Anthropology Today*. University of Chicago Press, Chicago, p. 343-383.

Clegg, M. & Aiello, L. C. (2000) Paying the price for speech? An analysis of mortality statistics for choking on food, *American Journal of Physical Anthropology* Supplement 30, p. 126.

Clynes, M. (1977) *Sentics: the Touch of Emotions*. Souvenir Press, New York.

Corballis, M. C. (1992) On the evolution of language and generativity, *Cognition* Vol. 44, p. 197-226.

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

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

Cross, I. (2001) Music, mind and evolution, *Psychology of Music* Vol. 29, p. 95-102.

Cross, I. (2003a) Music and Emotion: lecture notes, available online at http://www.mus.cam.ac.uk/~ic108/IBMandS/emotion/musicemotion.html

Cross, I. (2003b) Music and biocultural evolution, in M. Clayton & T. Herbert (eds.) *The Cultural Study of Music: A Critical Introduction*. Routledge, London, p. 19-30.

Cross, I. (n.d.) *The "Lithoacoustics" project: Musical behaviours and the archaeological record; an experimental study*. Faculty of Music, Cambridge.

Cross, I. (2003) Music and Evolution: consequences and causes, *Contemporary Music Review* Vol. 22 (3), p. 79-89.

Cross, I., Zubrow, E. & Cowan, F. (2002) Musical behaviours and the archaeological record: a preliminary study, in J. Mathieu (ed.) *Experimental Archaeology. British Archaeological Reports International Series* 1035, p. 25-34.

Dams, L. (1984) Preliminary Findings at the "Organ Sanctuary" in the Cave of Nerja, Malaga, Spain, *Oxford Journal Of Archaeology* Vol. 3, p. 1-14.

Dams, L. (1985) Palaeolithic Lithophones: Descriptions and Comparisons, *Oxford Journal Of Archaeology* Vol. 4, p. 31-46.

Daniel, H. J. (1990) The vestibular system and language evolution, in J. Wind *et al.*(eds.) *Studies In Language Origins Vol. 1*. Benjamins, Amsterdam, p. 257-271.

Darwin, C. (1871) *The Descent of Man and Selection In Relation To Sex*, John Murray, London.

Darwin, C. (1872/1998) *The Expression of Emotions in Man and Animals*. Harper Collins, London.

Dauvois, M. (1989) Son et Musique Paléolithiques, *Les Dossiers D'Archéologie* Vol. 142, p. 2-11.

Davidson, I. (1991) The archaeology of language origins – a review, *Antiquity* Vol. 65, p. 39-48.

Davies, S. (2001) Philosophical perspectives on music's expressiveness, in P. N. Juslin & J. A. Sloboda (eds.) *Music and Emotion: Theory and Research*. Oxford University Press, Oxford, p. 23-44.

Davis, P. J., Zhang, S. P., Winkworth, A. & Bandler, R. (1996) Neural control of vocalisation: respiratory and emotional influences, *Journal of Voice* Vol. 10, 23-38.

Davison, G. D. & Neale, J. M. (1994) *Abnormal Psychology (6th edition)*. John wiley and Sons, New York.

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

Dean, C, Leakey, M. V., Reid, D., Schrenk, F., Schwartzk, G. T., Stringer, C. & Walker, A. (2001) Growth processes in teeth distinguish modern humans from Homo erectus and earlier hominins, *Letters to Nature: Nature* Vol. 414, p. 628-631.

DeGelder, B. & Vroomen, J. (2000) The perception of emotions by ear and by eye, *Cognition and Emotion* Vol. 14, p. 289-311.

DeGusta, D., Gilbert, W. H. & Turner, S. P. (1999) Hypoglossal canal size and hominid speech, *Proceedings of the National Academy of Sciences of the United States of America* Vol. 96, p. 1800-1804.

Dennel, R. (1997) The world's oldest spears, *Nature* Vol. 385, p. 767-767.

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

D'Errico, F. (1991) Carnivore Traces or Mousterian Skiffle?, *Rock Art Research* Vol. 8, p. 61-63.

D'Errico, F. & Villa, P. (1997) Holes and grooves: the contribution of microscopy and taphonomy to the problem of art origins, *Journal of Human Evolution* Vol. 33, p. 1-31.

D'Errico, F., Villa, P., Pinto Llona, A. C. & Ruiz Idarraga, R. (1998) A Middle Palaeolithic origin of music? Using cave-bear bone accumulations to assess the Divje babe I bone "flute", *Antiquity* Vol. 72, p. 65-79.

D'Errico F., Henshilwood, C. & Nilssen, P. (2001) An engraved bone fragment from c.70,000 year-old Middle Stone Age levels at Blombos Cave, South Africa: Implications for the origin of symbolism and language, *Antiquity* 75, p. 309-318.

D'Errico F., Henshilwood, C., Lawson, G., Vanhaeren, M., Tillier, A.-M., Soressi, M., Bresson, F., Maureille, B., Nowell, A., Lakarra, J., Backwell, L. & Julien, M. (2003) Archaeological evidence for the emergence of language, symbolism and music – an alternative multidisciplinary perspective, *Journal of World Prehistory* Vol. 17, p. 1-70.

Dissanayake, E. (2000) Antecedents of the modern arts in early mother-infant interaction, in N. L. Wallin, B. Merker & S. Brown (eds.) *The Origins of Music*. MIT Press, London, p. 389-410.

Dowling, W. J. & Harwood, D. L. (1986) *Music Cognition*. Academic Press, London.

Donald, M (1991) *Origins of the Modern Mind: Three Stages in the Evolution of Culture and Cognition*. Harvard University Press, Cambridge, MA.

Donald, M. (2001) *A Mind So Rare: The Evolution of Human Consciousness*. Norton, London.

Duchin, L. E. (1990) The evolution of articulate speech: comparative anatomy of *Pan* and *Homo*, *Journal of Human Evolution* Vol. 19, p. 687-697.

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

Dunbar, R. I. M., Duncan, N., & Marriott, A. (1997) Human conversational behaviour, *Human Nature* Vol. 8, p. 231-246.

Dunbar, R. I. M., Knight, C. & Power, C. (eds.) (1999) *The Evolution of Culture*. Edinburgh University Press, Edinburgh.

Ekman, P. (1980) *The Face of Man: Expressions of Universal emotions in a New Guinea Viillage*. Garland SPTM Press, New York.

Ekman, P. & Friesen, W. V. (1971) Constants across cultures in the face and emotion, *Journal of Personality and Social Psychology* Vol. 17, p. 124-129.

Elowson, A. M., Snowdon, C. T. & Lazaro-Perea, C. (1998a). "Babbling" and social context in infant monkeys: parallels to human infants, *Trends in Cognitive Sciences* Vol. 2, p. 31-37.

Elowson, A. M., Snowdon, C. T. & Lazaro-Perea, C. (1998b) Infant "babbling" in a non-human primate: complex vocal sequences with repeated call-types, *Behaviour* Vol. 135, p. 643-664.

Emler, N. (1992) The truth about gossip, *Social Psychology Newsletter* Vol. 27, p. 23-37.

Enloe J. G., David, F. & Baryshnikov, G. (2000) Hyenas and hunters: zooarchaeological investigations at Prolom II cave, Crimea, *International Journal of Osteoarchaeology* Vol. 10, p. 310-324.

Epp, H. T. (1988) Way of the migrant herds: dual dispersion strategy amongst bison, *Plains Anthropologist* Vol. 33, p. 95-111.

Fages, G. & Mourer-Chauviré, C. (1983) La flute en os d'oiseau de la grotte sépulcrale de Veyreau (Aveyron) et inventaire des flutes préhistoriques d'Europe, in F. Poplin (ed.) *La Faune et l'Homme Préhistorique: dix études en hommage a Jean Bouchud. Mémoires de la Société Préhistorique Française* Vol. 16, p. 95-103.

Fagg, M. C. (1997) *Rock Music*, Pitt-Rivers Museum Occasional Papers No. 14. Pitt-Rivers, Oxford.

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

Falk, D. (2004) Prelinguistic evolution in early hominins: Whence motherese? *Behavioral and Brain Sciences*, Vol. 27 (4), p. 491-503

Falk, D. (2004) Prelinguistic evolution in hominin mothers and babies: For cryin' out loud! *Behavioural and Brain Sciences* Vol. 27 (4), p. 461-462

Fernald, A. (1989a) Emotion and meaning in mothers' speech to infants, paper presented to *The Society for Research in Child Development*, Kansas city, April 1989.

Fernald, A. (1989b) Intonation and communicative intent in mothers' speech to infants: is melody the message?, *Child Development* Vol. 60, p. 1497-1510.

Fernald, A. (1992a) Meaningful melodies in mothers' speech to infants, in H. Papousek, U. Jurgens & M. Papousek (eds.) *Nonverbal Vocal Communication*. Cambridge University Press, Cambridge.

Fernald, A. (1992b) Human maternal vocalisations to infants as biologically relevant signals: an evolutionary perspective, in J. Barkow, L. Cosmides & J. Tooby (eds.) *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. Oxford University Press, Oxford.

Fernald, A. (1993) Approval and disapproval – infant responsivelness to vocal affect in familiar and unfamiliar languages, *Child Development* Vol. 64, p. 657-674.

Fernald, A. (1994) Human maternal vocalisations to infants as biologically relevant signals: an evolutionary perspective, in P. Bloom (ed.) *Language Acquisition: Core Readings*. Harvester Wheatsheaf, London.

Fernald, A. & Simon, T. (1984) Expanded intonation contour in mothers' speech to newborns, *Developmental Psychology* Vol. 20, p. 104-113.

Feyereisen, P. (1997) The competition between gesture and speech production in dual-task paradigms, *Journal of Memory and Language* Vol. 36, p. 13-33.

Fidelholtz, J. L. (1991) On dating the origin of the modern form of language, in W. Von Raffler-Engel *et al*. (eds.) *Studies In Language Origins Vol. 2*. Benjamins, Amsterdam, p. 99-113.

Field, T., Woodson, R., Greenberg, R. & Cohen, D. (1982) Discrimination and imitation of facial expression in neonates, *Science* Vol. 218, p. 179-181.

Fink, R. (1997) Neanderthal Flute: Oldest Musical Instrument's 4 Notes Matches 4 of Do, Re, Mi Scale, http://www.webster.sk.ca/greenwich/fl-compl.htm (updated 1998, 1999, 2000).

Fink, R. (2000) Chewed Or Chipped? Who Made The Neanderthal Flute?, http://www.webster.sk.ca/greenwich/chewchip.htm

Fitch, W. T. (1999) Acoustic size exaggeration of size in birds via tracheal elongation: comparative and theoretical analyses, *Journal of Zoology* 248, p. 31-48.

Fitch, W. T. (2000a) The phonetic potential of nonhuman vocal tracts: comparative cineradiographic observations of vocalising animals, *Phonetica* 57, p. 205-218.

Fitch, W. T. (2000b) the evolution of speech: a comparative review, *Trends in Cognitive Science* 4, p. 258-267.

Foley, R. (1992) Studying Human Evolution by Analogy, in S. Jones, R. Martin & D. Pilbeam (eds.) *The Cambridge Encyclopedia of Human Evolution*. Cambridge University Press, Cambridge, p. 335-340.

Foley, R. A. & Lee, P. C. (1991) Ecology and energetics of encephalisation in hominid evolution, *Philosophical Transactions of the Royal Society of London, Series B - Biological Sciences* Vol. 334, p. 223-232.

Frayer, D. W. & Nicolay, C. (2000) Fossil evidence for the origin of speech sounds, in N. L. Wallin, B. Merker & S. Brown (eds.) *The Origins of Music*. MIT Press, London, p. 217-234.

Franz E. A., Zelaznik, H. N. & Smith, A. (1992) Evidence of common timing processes in the control of manual, orofacial, and speech movements, *Journal of Motor Behavior* 24, p. 281-287.

Freeman, L. G. Jr. (1968) A theoretical framework for interpreting archaeological materials, in R. B. Lee and I. DeVore (eds.) *Man The Hunter*. Aldine, Chicago, p. 262-267.

Fujii, T., Fukatsu, R., Watabe, S., Ohnuma, A., Teramura, T., Kimura, I., Saso, S., & Kogure, K. (1990) Auditory sound agnosia without aphasia following right temporal lobe lesion *Cortex* Vol. 26, p. 263-268.

Gabrielsson, A. & P. N. Juslin (1996) Emotional expression in music performance: between the performer's intention and the listener's experience, *Psychology of Music* Vol. 24, p. 68-91.

Gamble, C. (1983) Culture and society in the Upper Palaeolithic of Europe, in G.N. Bailey (ed.) *Hunter-Gatherer Economy In Prehistory: A European Perspective*, Cambridge University Press, p. 201-211.

Gardner, P. (1991) Forager Pursuit of Individual Autonomy, *Current Anthropology* Vol. 32, p. 543-572.

Geissman, T. (2000) Gibbon songs and human music in an evolutionary perspective, in N.L. Wallin, B. Merker and S. Brown (eds.) *The Origins of Music*. MIT Press, London, p. 103-123.

Geissman, T. & Orgeldinger, M. (1998) Duet or divorce!, *Folia Primatologica*, Vol. 69, p. 283.

Geschwind, N., Quadfasel, F., & Segarra, J. (1965) Isolation of the speech area, *Neuropsychologia* Vol. 6, p. 327-340.

Glory, A. (1964) La Grotte de Roucador, *Bulletin de la Société Préhistorique Française* Vol. 61, p. clxvi-clxix.

Glory, A. (1965) Nouvelles découvertes de dessins rupestres sur le causse de Gramat, *Bulletin de la Société Préhistorique Française* Vol. 62, p. 528-536.

Glory, A., Vaultier, M. & Farinha Dos Santos, M. (1965) La grotte ornée d'Escoural (Portugal), *Bulletin de la Société Préhistorique Française* Vol. 62, p. 110-117.

Goldbeck, C. (2001) Der Steinzeitflöte, *http://www.quarks.de/musik/002.htm.*

González Morales, M. R. (1986) Inventory of the bone and antler industry from the 1976-79 excavations at La Riera cave, in L. G. Straus and G. A. Clark (eds.) *La Riera Cave: Stone Age Hunter-Gatherer Adaptations in Northern Spain*, Arizona State University Anthropological Research Papers No. 36.

Gotfredsen, A. B. (1997) Sea bird exploitation on coastal Inuit sites, West and Southeast Greenland, *International Journal of Osteoarchaeology* Vol. 7, p. 271-286.

Gould, R. A. (1969) *Yiwara: Foragers of the Australian Western Desert*. Scribner's, New York.

Gould, R. A. (1980) *Living Archaeology*. Cambridge University Press, Cambridge.

Gould, R. A. & Watson P. J. (1982) A dialogue on the meaning and use of analogy in etnoarchaeological reasoning, *Journal of Anthropological Archaeology* Vol. 1, p. 355-381.

Gowlett, J. A. J. (1984) Mental abilities of early man: a look at some hard evidence, in R. Foley (ed) *Hominid Evolution and Community Ecology: Prehistoric Human Adaptation in Biological Perspective*. Academic Press, London, p. 167-192.

Greiser, D. & Kuhl, P. (1988) Maternal speech to infants in a tonal language: support for universal prosodic features in motherese, *Developmental Psychology* Vol. 24, p. 14-20.

Gros-Louis, J. (2002) Contexts and behavioural correlates of trill vocalisations in wild white-faced Capuchin monkeys (*Cebus capucinus*), *American Journal of Primatology* Vol. 57, p. 189-202.

Hagan, E. H. & Bryant, G. A. (*in press*) Music and dance as a coalition signalling system, *Human Nature*.

Hahn , J. & Münzel, S. (1995) Knochenflöten aus den Aurignacien des Geissenklösterle bei Blaubeuren, Alb-Donau-Kreis, *Fundberichte aus Baden-Würtemberg* Vol. 20, p. 1-12.

Harrison, R. A. (1978) A pierced reindeer phalanx from Banwell Bone Cave and some experimental work on phalanges, *Proceedings of the University of Bristol Spelaeological Society* Vol. 15, p. 7-22.

Hatfield, E., Cacioppo, J. T. & Rapson, R. L. (1994) *Emotional Contagion*. Cambridge University Press, Cambridge.

Hawkes, C. (1954) Archaeological theory and method: some suggestions from the Old World, *American Anthropologist* Vol. 56, p. 155-168.

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

Henschen-Nyman, O. (1988) Cup-marked standing stones in Sweden, in E. Hickman & D. W. Hughes (eds.) *The Archaeology Of Early Music Cultures; Third International Meeting Of The ICTM Study Group On Music Archaeology*. Verlag für systematische Musikwissenschaft GmbH, Bonn, p. 11-16.

Henshilwood C. S., Sealy, J. C., Yates, R., Cruz-Uribe, K., Goldberg, P., Grine, F. E., Klein, R. G., Poggenpoel, C., Van Niekerk, K., & Watts, I. (2001) Blombos Cave, Southern Cape, South Africa: preliminary report on the 1992-1999 excavations of the Middle Stone Age levels, *Journal of Archaeological Science* Vol. 28, p. 421-448.

Hewes, G. W. (1973) Primate communication and the gestural origin of language, *Current Anthropology* Vol. 14, p. 5-24.

Hewes, G. W. (1992) Primate communication and the gestural origin of language, *Current Anthropology* Vol. 33, p. 65-84.

Hewes, G. W. (1994) Comment on Armstrong, Stokoe and Wilson, *Current Anthropology* Vol. 35, p. 360-361.

Hitchcock, R. K. (1999) Introduction: Africa, in R. B. Lee & R. Daly (eds.) *The Cambridge Encyclopedia of Hunters and Gatherers*. Cambridge University Press, Cambridge, p. 175-184.

Holderman C.-S. & Serangeli, J. (1998a) Flöten an Höhlenbärenknochen: Spekulationen oder Beweise?, *Mitteilungsblatt der Gesellschaft für Urgeschichte* Vol. 6, p. 7-19.

Holderman C.-S. & Serangeli, J. (1998b) Einige Bemerkungen zur Flöte von Divje babe I (Slowenien) und deren Vergleichsfunde aus dem Östereichischen Raum und angenzenden Gebeiten, *Archäologie Österreichs* Vol. 9, p. 31-38.

Holloway, R. (1981) Volumetric and asymmetry determinations on recent hominid endocasts: Spy I and II, Djebel Ihroud I, and the Salè *Homo erectus* specimens, with some notes on Neanderthal brain size, *American Journal Of Physical Anthropology* Vol. 55, p. 385-393.

Holloway, R. (1983) Human palaeontological evidence relevant to language behavior, *Human Neurobiology* Vol. 2, p. 105-114.

Hudson, T. (Ed.) (1977) The Eye of the Flute: Chumash Traditional History and Ritual, by F. Librado. Santa Barbara Museum of Natural History, Santa Barbara.

Hull, D. L. (1980) Individuality and selection, *Annual Review of Ecology and Systematics* Vol. 11, p. 311-332.

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

Huyge, D. (1990) Mousterian skiffle? Note on a Middle Palaeolithic engraved bone from Schulen, Belgium, *Rock Art Research* Vol. 7, p. 125-132.

Huyge, D. (1991) The "Venus" of Laussel in the light of ethnomusicology, *Archeologie in Vlaanderen* Vol. 1, p. 11-18.

Ichikawa, M. (1999) The Mbuti of Northern Congo, in R. B. Lee & R. Daly (eds.) *The Cambridge Encyclopedia of Hunters and Gatherers*. Cambridge University Press, Cambridge, p. 210-214.

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

Ito, S. & Hinoki, M. (1991) Human and animal semicircular canal functionduring circular walking, *Acta Oto-Laryngologica* Supplemental 481, p. 339-342.

Jablonski, N. J. & Aiello, L. C. (eds.) (1998) *The Origin and Diversification of Language*, Wattis Symposium Series in Archaeology: Memoirs of the California Academy of Sciences 24. California Academy of Sciences, San Francisco.

Joffe, T. H. (1997) Social pressures have selected for an extended juvenile period in primates, *Journal of Human Evolution* Vol. 32, p. 593-605.

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

Johnston, T. F. (1989) Song categories and musical style of the Yupik Eskimo, *Anthropos* Vol. 84, p. 423-431.

Jones, T. A. (1983) Australia, in D. Arnold (ed.) *The New Oxford Companion To Music* Vol. 1. Oxford University Press, Oxford, p. 117-119.

Jurgens, U. (1992) On the neurobiology of vocal communication, in H. Papousek, U Jurgens & M. Papousek (eds.) *Nonverbal Vocal Communication*. Cambridge University Press, Cambridge.

Jurgens, U. (1998) Neuronal control of mammalian vocalisation, with special reference to the squirrel monkey, *Naturwissenschaften* Vol. 85, p. 376-388.

Jurgens, U. & Alipour, M. (2002) A comparative study of the cortico-hypoglossal connections in primates, using biotin dextranamine, *Neuroscience Letters* Vol. 328, p. 245-248.

Jurgens, U. & Von Cramon, D. (1982) On the role of the anterior cingulate cortex in phonation - a case-report, *Brain and Language* Vol. 15, p. 234-248.

Jurgens, U. & Zwirner, P. (1996) The role of the periaqueductal grey in limbic and neocortical vocal fold control, *Neuroreport* Vol. 7, p. 2921-2923.

Juslin, P. N. & Sloboda, J. A. (eds.) (2001) *Music and Emotion: Theory and Research*. Oxford University Press, Oxford.

Kappas, A., U. Hess, and K. R. Scherer (1991) Voice and emotion, in R. S. Feldman & B. Rime (eds.), *Fundamentals of Nonverbal Behavior: Studies in Emotion and Social Interaction*. Cambridge University Press, Cambridge.

Karow, C. M., Marquardt, T. P., & Marshall, R. C. (2001) Affective processing in left and right hemisphere brain-damaged subjects with and without subcortical involvement. *Aphasiology* Vol. 15, p. 715-729.

Kay, R. F., Cartmill, M. & Balow M. (1998) The hypoglossal canal and the origin of human vocal behaviour, *Proceedings of the National Academy of Science of the USA*, p. 5417-5419.

Kehoe, A. B. (1999) Blackfoot and other hunters of the North American Plains, in R. B. Lee & R. Daly (eds.) *The Cambridge Encyclopedia of Hunters and Gatherers*. Cambridge University Press, Cambridge, p. 36-40.

Kirzinger, A. & Jurgens U. (1982) Cortical lesion effects and vocalization in the squirrel-monkey, *Brain Research* Vol. 233, p. 299-315.

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

Kitamura, C., Thanavisthuth, C., Burnham, D. & Luksaneeyanawin, S. (2002) Universality and specificity in infant-directed speech: pitch modifications as a function of infant age and sex in a tonal and non-tonal language, *Infant Behaviour and Development* Vol. 24, p. 372-392.

Kivy, P. (1989) *Sound Sentiment*. Temple University Press, Philadelphia.

Klein, R. G. (1999) *The Human Career (2nd Edition)*. University of Chicago Press, Chicago.

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

Kogan, N. (1997) Reflections on aesthetics and evolution, *Critical Review* Vol. 11, p. 193-210.

Kordos, L. & Begun, D. R. (1997) A new reconstruction of RUD 77, a partial cranium of *Dryopithecus brancoi* from Rudabanya, Hungary, *American Journal of Physical Anthropology* Vol. 103, p. 277-294.

Kraut, R. E. & Johnston, R. (1979) Social and emotional messages of smiling: an ethological approach, *Journal of Personality and Social Psychology* Vol. 37, p. 1539-1553.

Krumhansl, C. L. (1997) An exploratory study of musical emotions and psychophysiology, *Canadian Journal of Experimental Psychology* Vol. 51 , p. 336–352.

Kuhl, P. K. (1988) Auditory perception and the evolution of speech, *Human Evolution* Vol. 3, p. 19-43.

Kunej, D. & Turk, I. (2000) New perspectives on the beginnings of music: archaeological and musicological analysis of a Middle Palaeolithic bone “flute”, in N. L. Wallin, B. Merker & S. Brown (eds.) *The Origins of Music*. MIT Press, London, p. 235-268.

Laitman, J. T. (1984) The anatomy of human speech, *Natural History* Vol. 93, p. 20-27.

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

Laitman, J. T. & Heimbuch, R. C. (1982) The basicranium of Plio-Pleistocene hominids as an indicator of their upper respiratory systems, *American Journal of Physical Anthropology* Vol. 59, p. 323-344.

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

Lavy, M. M. (2001) *Emotion and the Experience of Listening to Music: A Framework for Empirical Research*, Unpublished Ph.D. thesis, University of Cambridge.

Lawson, G. & d'Errico, F. (2002) Microscopic, experimental and theoretical re-assessment of Upper Palaeolithic bird-bone pipes from Isturitz, France: ergonomics of design, systems of notation and the origins of musical traditions, in E. Hickman, A. D. Kilmer & R. Eichman (eds.), *Studien zur Musikachäologie III*. Verlag Marie Leidorf, Rahden.

Lawson, G., Scarre, C., Cross, I. & Hills, C. (1998) Mounds, megaliths, music and mind: some acoustical properties and purposes of archaeological spaces, *Archaeological Review From Cambridge* Vol. 15, p. 111-134.

Leakey, R. (1994) *The Origin of Humankind*. Weidenfeld and Nicolson, London.

Lefevre, C. (1997) Sea bird following in southern Patagonia: a contribution BQ understanding the nomadic round of the Canoeros Indians, *International Journal of Osteoarchaeology* Vol. 7, p. 260-270.

Leocota, F. (2001) Osservazioni sui "flauti" paleolitici, *Revista di Science Preistoriche* LI, p. 177-200.

Levenson, R. W., Ekman, P. & Friesen, V. W. (1990) Voluntary facial action generates emotion-specific autonomic nervous system activity, *Psychophysiology* Vol. 27, p. 363-384.

Levi-Strauss, C. (1970) *The Raw and the Cooked* (Translated by J. and D. Weightman). Cape, London.

Levitin, D. L. & Bellugi, U. (1997) Musical abilities in individuals with Williams Syndrome. Paper presented at the 1997 Society for Music Perception and Cognition. Massachusetts Institute of Technology, Cambridge, MA.

Lewis-Williams, D. (2002) *The Mind in the Cave*. Thames and Hudson, London.

Lewkowicz, D. J. (1998) Infants' response to the audible and visible properties of the human face: II. Discrimination of differences between singing and adult-directed speech, *Developmental Psychobiology* Vol. 32, p. 261-274.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P. & Studdert-Kennedy, M. (1967) Perception of the speech code, *Psychological Review* Vol. 74, p. 431-461.

Librado, F. (1977) The Eye of the Flute: Chumash Traditional History and Ritual (Ed. T. Hudson). Santa Barbara Museum of Natural History, Santa Barbara.

Lieberman, P. (1984) *The Biology and Evolution of Language.* Harvard, Cambridge, Massachusetts.

Lieberman, P. (1989) The origins of some aspects of human language and cognition, in P. Mellars & C. Stringer (eds.) *The Human Revolution: Behavioural and Biological Perspectives in the Origins of Modern Humans*. Edinburgh University Press, Edinburgh, p. 391-414.

Lieberman, P. (1991) *Uniquely Human: The Evolution of Speech, Thought and Selfless Behaviour*. Harvard University Press, Cambridge, Mass.

Lieberman, P. (1994) Comment on Armstrong, Stokoe and Wilson, *Current Anthropology* Vol. 35, p. 362-363.

Lieberman, P. & Crelin, E. S. (1971) On the speech of Neanderthal man, *Linguistic Enquiry* Vol. 11, p. 203-222.

Lieberman, P., Crelin, E. S. & Klatt, D. H. (1972) Phonetic ability and related anatomy of the newborn and adult human, Neanderthal man and the chimpanzee, *American Anthropology* 74, p. 287-307.

Lieberman, P., Klatt, D. H. & Wilson, W. H. (1969) Vocal tract limitations on the vowel repertoires of rhesus monkeys and other nonhuman primates, *Science* 164, p. 1185-1187.

Livingstone, F. B. (1973) Did The Australopithecines Sing?, *Current Anthropology* Vol. 14, p. 25-29.

Locke, D. (1996) Africa: Ewe, Mande, Dagbamba, Shona and BaAka, in J. T. Titon (ed.) *Worlds of Music: An Introduction to the Music of the World's People* 3rd Edition, Schirmer, New York, p. 71-143.

Locke, J. L. (1998) Social sound-making as a precursor to modern language, in J. Hurford, M. Studdert-Kennedy & C. Knight (eds.), *Approaches to the Evolution of Language*. Cambridge University Press, Cambridge.

Locke, J. L., Bekken, K. E., McMinnLarson, L. & Wein, D. (1995) Emergent control of manual and vocal motor activity in relation to the development of speech, *Brain and Language* Vol. 51, p. 498-508.

MacLarnon, A. M. (1993) The vertebral canal, in A. Walker & R. Leakey (eds.) *The Nariokotome* Homo erectus *Skeleton*. Harvard University Press, Cambridge MA, p. 359-390.

MacLarnon, A. M. & Hewitt, G. P. (1999) The evolution of human speech: the role of enhanced breathing control, *American Journal of Physical Anthropology* Vol. 109, p. 341-363.

McAllester, D. P. (1996) North America/Native America, in J.T. Titon (ed.) *Worlds of Music: An Introduction to the Music of the World's People* 3rd Edition. Schirmer, New York, p. 17-70.

McBrearty, S. & Brooks, A. S. (2000) The revolution that wasn't: a new interpretation of the origin of modern human behavior, *Journal of Human Evolution* Vol. 39, p. 453-563.

McBurney, C. B. M. (1967) *The Haua Fteah (Cyrenaica) and the Stone Age of the South East Mediterranean*. Cambridge University Press, Cambridge.

McClave, E. (1994) Gestural beats: the rhythm hypothesis, *Journal of Psycholinguistic Research* 23, p. 45-66.

McNeill, D. (1992) *Hand and Mind: What Gestures Reveal About Thought*, University of Chicago Press, London.

McNeill, D. (ed.) (2000) *Language and Gesture*. Cambridge University Press, Cambridge.

McNeill, W. H. (1995) *Keeping Together inTtime*. Harvard University Press, London.

Magriples, U. & Laitman, J. T. (1987) Developmental change in the position of the fetal human larynx*, American Journal of Physical Anthropology* Vol. 72, p. 463-472.

Mang, E. (2000) Intermediate vocalizations: an investigation of the boundary between speech and songs in young children's vocalisations, *Bulletin of the Council for Research in Music Education* Issue 147, p. 116-121.

Marin, O. S. M. & Perry, D. W. (1999) Neurological aspects of music perception and performance, in D. Deutsch (ed.) *The Psychology of Music* (2nd ed.). Academic Press, New York, p. 653-724.

Marshack, A. (1972) *The Roots Of Civilisation*. Weidenfeld and Nicholson, London.

Martin, D. F. (1999) Cape York Peoples, North Queensland, Australia, in R. B. Lee & R. Daly (eds.) *The Cambridge Encyclopedia of Hunters and Gatherers*. Cambridge University Press, Cambridge, p. 335-338.

Masataka, N. (2000) The role of modality and input in the earliest stages of language acquisition: studies of Japanese sign language, in C. Chamberlain, J. Morford & R. I. Mayberry (eds.) *Language Acquisition by Eye*. Erlbaum, Mahwah, NJ.

Matravers, D. (1998) *Art and Emotion*. Clarendon Press, Oxford.

Mayberry, R. I. & Jaques, J. (2000) Gesture production during stuttered speech: insights into the nature of gesture-speech integration, in D. McNeill (ed.) *Language and Gesture*. Cambridge University Press, Cambridge, p. 199-214.

Mazzucchi, A., Marchini, C., Budai, R. & Parma, M. (1982) A case of receptive amusia with prominent timbre perception defect, *Journal of Neurology, Neurosurgery, and Psychiatry* Vol. 45, p. 644-647.

Megaw, J. V. S. (1960) Penny Whistles and Prehistory, *Antiquity* Vol. 34, p. 6-13.

Mellars, P. (1989) Major issues in the emergence of modern humans, *Current Anthropology* Vol. 30, p. 349-385.

Mellars, P. (1994) The Upper Palaeolithic revolution, in B. Cunliffe (ed.) *Prehistoric Europe: An Illustrated History*. Oxford University Press, Oxford.

Mellars, P. (2000) The emergence of modern cognitive abilities, paper presented at the One-day conference on *The Speciation of Modern Homo sapiens,* at the British Academy, March 28th.

Mellars, P. & Gibson, K. (1996) *Modelling the Early Human Mind*. McDonald Institute Monographs, Cambridge.

Mellars, P. & Stringer, C. (1989) *The Human Revolution: Behavioural and Biological Perspectives in the Origins of Modern Humans*. Edinburgh University Press, Edinburgh.

Menuhin, Y. & Davis, C. W. (1979) *The Music of Man*. Methuen, London.

Merker, B. (1999) Synchronous chorusing and the origins of music, *Musicae Scientiae* Special Issue, p. 59-73.

Messinger D. S. & Fogel A. (1998) Give and take: the development of conventional infant gestures, *Merrill-Palmer Quarterly-Journal of Developmental Psychology* Vol. 44, p. 566-590.

Meyer, L. B. (1956) *Emotion and Meaning in Music*. University of Chicago Press, Chicago.

Michell, R. W. & Gallaher, M. C. (2001) Embodying music: matching music and dance in memory, *Music Perception* Vol. 19, p. 65-85.

Miller, G. F. (2000a) Evolution of human music through sexual selection, in N. L. Wallin, B. Merker & S. Brown (eds.) *The Origins of Music*. MIT Press, London, p. 329-360.

Miller, G. F. (2000b) *The Mating Mind: How Sexual Choice Shaped The Evolution Of Human Nature*. Heinemann, London.

Mithen, S. J. (1996) *The Prehistory of the Mind*. Pheonix, London.

Mithen, S. J. (1998) Was there a creative explosion?, in S. J. Mithen (ed.) *Creativity in Human Evolution and Prehistory*. Routledge, London. p. 193-210.

Mondragon-Ceballos, R. (2002) Machiavellian intelligence in primates and the evolution of social brain, *Salud Mental* 25, p. 29-39.

Morley, I. (2000) *The Origins and Evolution of the Human Capacity for Music*. MA Thesis, University of Reading, Reading, Berkshire.

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

Morley, I. (2002b) A cross-continental chorus: commonalities in the music of hunters and gatherers, in *Proceedings of the 9th International Conference On Hunting and Gathering Societies*, in press. Currently available online at http://www.abdn.ac.uk/chags9/1morley.htm

Morton, E. S. (1977) On the occurrence and significance of motivation-structural rules in some bird and mammal sounds, *American Naturalist* Vol. 111, p. 855-869.

Morton, E. S. (1994) Sound symbolism and its roe in non-human vertebrate communication, in L. Hinton, J. Nichols & J. J. Ohala (eds.) *Sound Symbolism*. Cambridge University Press, Cambridge, p. 348-365.

Morton, J. (1999) The Arrernte of Central Australia, in R. B. Lee & R. Daly (eds.) *The Cambridge Encyclopedia of Hunters and Gatherers*. Cambridge University Press, Cambridge, p. 329-338.

Mulvaney, J. (1999) The chain of connection: the material evidence, in N. Peterson (ed.) *Tribes and Boundaries In Australia*. Australian Institute of Aboriginal Studies, Canberra, p. 72-94.

Myers, F. R. (1999) Pintupi-speaking Aboriginals of the Western Desert, in R. B. Lee & R. Daly (eds.) *The Cambridge Encyclopedia of Hunters and Gatherers*. Cambridge University Press, Cambridge, p. 348-357.

Nattiez, J. J. (1983) Some aspects of Inuit vocal games, *Ethnomusicology* Vol. 27, p. 457-475.

Nelson, D. E. (1997) Radiocarbon dating of bone and charcoal from Divje babe I cave, in I. Turk (ed.), *Mousterian "Bone Flute" and other finds from Divje babe I cave site in Slovenia*. Institut za archaeologijo, Znanstvenoraziskovalni Center Sazu., Ljubljana, p. 51-64.

Nelson, E. W. (1899/1983) *The Eskimo About Bering Strait.* Smithsonian Institution, Washington.

Nettiez, J. J. (1983) Some aspects of Inuit vocal games, *Ethnomusicology* Vol. 27, p. 457-475.

Nettl, B. (1956) *Music in Primitive Culture*. Harvard University Press, Cambridge, MA.

Nettl, B. (1989) *Blackfoot Musical Thought: Comparative Perspectives*. Kent State University Press, Kent, Ohio.

Nettl, B. (1992) North American Indian music, in B. Nettl, C. Capwell, P. Bohlman, I. Wong, & T. Turino (eds.) *Excursions in World Music*. Prentice Hall, Englewood Cliffs, NJ, p. 260-277.

Nicolson, N. A. (1977) A comparison of early behavioral development in wild and captive chimpanzees, in S. Chavalier-Skolnikoff & F. E. Poirier (eds.) *Primate bio-social development*. Garland, New York.

Nobe, S. (1996) Cognitive rhythms, gestures, and acoustic aspects of speech. *Unpublished Ph.D. thesis*, University of Chicago. Cited in Mayberry and Jaques, 2000.

Noffsinger, D. (1985) Dichotic-listening techniques in the study of hemispheric asymmetries, in D. F. Benson & E. Zaidel (eds.) *The Dual Brain*. Guildford Press, New York, p. 127-141

O'Connell, J. F. (1995) Ethnoarchaeology needs a general theory of behaviour, *Journal of Archaeological Research* Vol. 3, p. 205-255.

Otte, M. (2000) On the suggested bone flute from Slovenia, *Current Anthropology* Vol. 41, p. 271-272.

Palastanga, N. P., Field, D. G. & Soames, R. (2002) *Anatomy and Human Movement*. Butterworth Heinemann, London.

Panksepp, J. (1995) The emotional sources of "chills" induced by music, *Music Perception* Vol. 13, p. 171-208.

Panksepp, J. & Bernatsky, G. (2002) Emotional sounds and the brain: the neuro-affective foundations of musical appreciation, *Behavioural Processes* Vol. 60, p. 133-155.

Papaeliou, C. & C. Trevarthen (1998). The infancy of music, *Musical Praxis* Vol. 1, p. 19-33.

Papousek, H. (1996). Musicality in infant research, in I. Deliege & J. A. Sloboda (eds.), *Musical Beginnings*. Oxford University Press, Oxford.

Papousek, M. (1996) Intuitive parenting: a hidden source of musical stimulation in infancy, in I. Deliége & J. Sloboda (eds.) *Musical Beginnings: Origins and Development of Musical Competence*. Oxford University Press, Oxford.

Papousek, M., Papousek, H. & Symmes, D. (1991) The meanings of melodies in motherese in tone and stress languages, *Infant Behavior and Development*, Vol. 14, p. 415-440.

Passemard, E. (1923) Une flûte Aurignacienne d'Isturitz, *Congrès de l'Association Français pour l'Avancement des Sciences* 46th Session, Montpelier, p. 474-476.

Patel, A. D., Peretz, I., Tramo, M., & Labreque, R. (1998) Processing prosodic and musical patterns: a neuropsychological investigation, *Brain and Language* Vol. 61, p. 123-144.

Pequart, M. & Pequart, S. J. (1960) *Grotte du Mas d'Azil (Ariege)*. Annales de Paleontologie, Paris.

Peretz, I. (1993) Auditory atonalia for melodies, *Cognitive Neuropsychology* Vol. 10, p. 21-56.

Peretz, I. & Morais, J. (1980) Modes of processing melodies and ear asymmetry in non-musicians, *Neuropsychologia* Vol. 18, p. 477-489.

Peretz, I. & Morais, J. (1983) Task determinants of ear differences in melody processing, *Brain and Cognition* Vol. 2, p. 288-292.

Peretz, I. & Kolinsky, R. (1993) Boundaries of seperability between melody and rhythm in music discrimination: a neurological perspective, *The Quarterly Journal of Experimental Psychology* Vol. 46A, p. 301-325.

Peretz, I., Kolinsky, R., Tramo, M., Labrecque, R., Hublet, C., DeMeurisse, G., & Belleville, S. (1994) Functional dissociations following bilateral lesions of auditory-cortex, *Brain* Vol. 117, p. 1283-1301.

Peterson, N. (1999) Introduction: Australia, in R. B. Lee & R. Daly (eds.) *The Cambridge Encyclopedia of Hunters and Gatherers*. Cambridge University Press, Cambridge, p. 317-313.

Peterson, S. E., Fox, P. T., Posner, M. I., Mintum, M., & Raichle, M. E. (1988) Positron Emission Tomographic studies of the cortical anatomy of single-word processing, *Nature* Vol. 331, p. 585-589.

Petitto, L. A. & Marentette, P. F. (1991) Babbling in the manual mode – evidence for the ontogeny of language, *Science* Vol. 251, p. 1493-1496.

Petsche, H., Rappelsberger, P., Filz, O., & Gruber, G. H. (1991) EEG studies in the perception of simple and complex rhythms, in J. Sundberg & L. Nord (eds.) *Music, Language, Speech and Brain*. MacMillan Academic and Professional, Basingstoke.

Piette, E. (1874) La Flûte Composée a l'âge du Renne, *Comptes Rendus de l'Academie des Sciences* LXXIX, p. 1277.

Piette, E. (1875) Les Vestiges de la Période Neolithique Comparés à ceux des âges Anterioers, *Congrès de l'Association Français pour l'Avancement des Sciences*, Congres de Nantes, 1875, p. 20.

Piette, E. (1907) *L'Art Pendant l'Âge du Renne*, Paris. Plate 1, figs 6-7.

Pinker, S. (1997) *How The Mind Works*. Allen Lane, London.

Pinker, S. (1994) *The Language Instinct: How The Mind Creates Language*. William Morrow, New York.

The PoinTIS Spinal Cord Occupational Therapy site of the SCI Manuals for Providers, http://calder.med.miami.edu/providers/OCCUPATIONAL/over2.html.

Renfrew, A. C. & Bahn, P. G. (2000) *Archaeology: Theories, Methods and Practice* (3rd Edition). Thames and Hudson, London.

Repp, B. H. (1991) Some Cognitive and Perceptual Aspects of Speech and Music, in J. Sundberg, L. Nord & R. Carlson (eds.) *Music, Language, Speech and Brain*. MacMillan Press, Basingstoke, p. 257-268.

Reznikoff, I. & Dauvois, M. (1988) La dimension sonore des grottes ornées, *Bulletin de la Société Préhistorique Française* Vol. 85, p. 238-246.

Riches, D. (1984) Hunters, herders and potlatchers: towards a sociological theory of prestige, *Man* Vol. 19, p. 234-251.

Riches, D. (1995) Hunter-gatherer structural transformations, *Journal of the Royal Anthropological Institute* Vol. 1, p. 679-701.

Richter D., Waiblinger J., Rink W. J. & Wagner G. A. (2000) Thermoluminescence, electron spin resonance and C-14-dating of the Late Middle and Early Upper Palaeolithic site of Geissenklosterle Cave in southern Germany, *Journal Of Archaeological Science* Vol. 27, p. 71-89.

Roederer, J. G. (1984) The search for a survival value of music, *Music Perception* Vol. 1, p. 350-356.

Ross, C. F. & Ravosa, M. J. (1993) Basicranial flexion, relative brain size, and facial kyphosis in nonhuman primates, *American Journal of Physical Anthropology* Vol. 91, p. 305-324.

Sacks, O. (1996) *An Anthropologist on Mars*. Random House, New York.

Sadie, S. (ed.) (2001) *The New Grove Dictionary of Music and Musicians* 2nd edition. Grove, London.

Saint-Périer, R. de (1947) Les derniers objêts Magdaleniens d'Isturitz, *L'Anthropologie* Vol. 51, p. 393-415.

Saint-Périer, R. de, & Saint-Périer, S. de, (1952) La Grotte d'Isturitz III. Les Solutreens, Les Aurignaciens et Les Mousteriens, *Archives de l'Institute de Paléontologie Humaine* Vol. 25.

Scarre, C. (1989) Painting By Resonance, *Nature* Vol. 338, p. 382.

Schacter, S. & Singer, J. E. (1962) Cognitive, social, and physiological determinants of emotional states, *Psychological Review* Vol. 69, p. 379-399.

Scherer, K. R. (1985) Vocal affect signalling: a comparative approach, *Advances in the Study of Bahaviour* Vol. 15, p. 189-244.

Scherer, K. R. (1986) Vocal affect expression: a review and model for future research, *Psychological Bulletin* Vol. 99, p. 143-165.

Scherer, K. R. (1991) Emotion expression in speech and music, in J. Sundberg, L. Nord & R. Carlson (eds.) *Music, Language, Speech and Brain*, MacMillan Press, Basingstoke, p. 146-156.

Schmidt, K. L. & Cohn, J. F. (2001) Human facial expressions as adaptations: evolutionary questions in facial expression research, *Yearbook of Physical Anthropology* Vol. 44, p. 3-24.

Schrive, C. (1984) *Past and Present in Hunter-Gatherer Studies*. Academic Press, London.

Schweiger, A. (1985) Harmony of the spheres and the hemispheres: the arts and hemispheric specialisation in D. F. Benson & E. Zaidel (eds.) *The Dual Brain*. Guildford Press. New York, 359-373.

Schweiger, A & Maltzman, I. (1985) Behavioural and electrodermal measures of lateralisation for music perception in musicians and nonmusicians, *Biological Psychology* Vol. 20, p. 129-145.

Scothern, P. M. T. (1992) *The Music-Archaeology of the Palaeolithic Within its Cultural Setting*. Unpublished Ph.D. thesis, University of Cambridge.

Seewald, O. (1934) *Beitrage zur Kenntnis der Steinzeitlichen Muzikinstrumente Europas*. Schroll, Vienna.

Serjeantson, D. (1997) Subsistence and symbol: the interpretation of bird remains in archaeology, *International Journal of Osteoarchaeology* Vol. 7, p. 255-259.

Seyfarth, R. M. & Cheyney, D. L. (1992) Meaning and mind in monkeys, *Scientific American* Vol. 267, p. 78-84.

Seyfarth, R. M. & Cheyney, D. L. (1997) Behavioural mechanisms underlying vocal communication in nonhuman primates, *Animal Learning and Behaviour* Vol. 25, p.249-267.

Shanon, B. (1982) Lateralisation effects in music decision tasks, *Neuropsychologia* Vol. 18, p. 21-31.

Shennan, S. (2002) *Genes, Memes and Human History*. Thames and Hudson, London.

Simmons, T. & Nadel, D. (1998) The Avifauna of the early Epipalaeolithic site of Ohalo II (19,400 years BP), Israel: species diversity, habitat and seasonality, *International Journal of Osteoarchaeology* Vol. 8, p. 79-96.

Slater, P. J. B. (2000) Birdsong repertoires: their origin and use, in N. L. Wallin, B. Merker and S. Brown (eds.) *The Origins of Music*. MIT Press, London, p. 49-63.

Sloboda, J. A. (1985) *The Musical Mind: The Cognitive Psychology of Music*. Oxford University Press, Oxford.

Sloboda, J. A. (1991) Music structure and emotional response: some empirical findings, *Psychology Of Music* Vol. 19, p. 110-120.

Sloboda, J. A. (1998) Does music mean anything?, *Musicae Scientiae* Vol. 2, p. 21-31.

Sloboda, J. A. (2001) The psychology of music: affect, in L. Macy (ed.) *The New Grove Dictionary of Music Online*. <http://www.grovemusic.com> (Accessed 1st July 2003).

Sloboda, J. A. & Juslin, P. N. (2001) Psychological perspectives on music and emotion, in P. N. Juslin & J. A. Sloboda (eds.) *Music and Emotion: Theory and Research*. Oxford University Press, Oxford, p. 71-104.

Sloboda, J. A., Hermelin, B. & O'Connor, N. (1985) On the anatomy of the ritard: a study of timing in music, *Journal of the Acoustical Society of America* Vol. 97, p. 53-67.

Snow, D. (2000) The emotional basis of linguistic and nonlinguistic intonation: implications for hemispheric specialization. *Developmental Neuropsychology* Vol. 17, p. 1-28.

Solow, B. (1966) The pattern of craniofacial associations, *Acta Oolontologica Scandinavia* Vol. 24, Supplementum 46, p. 125.

Sperber, D. (1996) *Explaining Culture*. Blackwell, Oxford.

Spoor, F. (1996) The ancestral morphology of the hominid bony labyrinth: the evidence from *Dryopithicus*, *American Journal of Physical Anthropology* Vol. 22 (Supplement), p. 219.

Spoor, F., Wood, B. & Zonneveld, F. (1994) Implications of hominid labyrinthine morphology for evolution of human bipedal locomotion, *Nature* Vol. 369, p. 645-648.

Spoor, F. & Zonneveld, F. (1998) Comparative review of the human bony labyrinth, *Yearbook of Physical Anthropology* Vol. 41, p. 211-251.

Stern, D. N. (1977) *The First Relationship: Infant and Mother*. Harvard University Press, Cambridge MA.

Stern, D. N., Speiker, S., Barnett, R. K. & MacKain, K. (1983) The prosody of maternal speech: infant age and context related changes, *Journal of Child Language* Vol. 10, p. 1-15.

Steward, J. H. (1942) The direct historical approach to archaeology, *American Antiquity* Vol. 7, p. 337-343.

Stiner M. C., Munro, N. D. & Surovell, T. A. (2000) The tortoise and the hare - Small-game use, the broad-spectrum revolution, and Paleolithic demography, *Current Anthropology* Vol. 41, p. 39-73.

Stockman, D. (1986) On the early history of drums and drumming in europe and the mediterranean, in C. S. Lund, (ed.) *The Second Conference Of the ICTM Study Group On Music Archaeology* Volume I: General Studies. Royal Swedish Academy of Music, Stockholm, p. 11-28.

Stokoe, W. C. (2000) Gesture to sign (language), in D. McNeill (ed.) *Language and Gesture*. Cambridge Unversity Press, Cambridge, p. 388-399.

Stringer, C. & Gamble, C. (1993) *In Search of the Neanderthals*. Thames and Hudson, London.

Strong, W. D. (1936) Anthropological theory and archaeological fact, in R. H. Lowie (ed.) *Essays in Anthropology*. University of California Press, Berkeley, p. 359-369.

Strong, W. D. (1942) Historical approach to anthropology, in A. L. Kroeber (ed.) *Anthropology Today*. University of Chicago Press, Chicago, p. 386-397.

Tartter, V. C. (1980) Happy talk: perceptual and acoustic effects of smiling on speech, *Perception and Psychophysics* Vol. 27, p. 24-27.

Tartter, V. C. & Braun, D. (1994) Hearing smiles and frowns in normal and whisper registers, *Journal of the Acoustical Society of America* Vol. 96, p. 2101-2107.

Taylor, C. (ed.) (1991) *The Native Americans: The Indigeous People of North America*. Salamander Books, London.

Taylor, C. (1991) The Plains, in C. F. Taylor (ed.) *The Native Americans: The Indigeous People of North America*. Salamander Books, London.

Thaut, M. H., McIntosh, G. C. & Rice, R. R. (1997) Rhythmic facilitation of gait training in hemiparetic stroke rehabilitation, *Journal of Neurological Sciences* Vol. 151, p. 207-212.

Thompson, R. H. (1956) The subjective element in archaeological inference, *Southwestern Journal of Anthropology* Vol. 12, p. 327-332.

Tonkinson, R. (1999) The Ngarrindjeri of Southeastern Australia, in R. B. Lee & R. Daly (eds.) *The Cambridge Encyclopedia of Hunters and Gatherers*, Cambridge University Press, Cambridge, p. 343-347.

Trainor, L. J., Clark, E. D., Huntley, A. & Adams, B. (1997) The acoustic basis of infant preferences for infant-directed singing, *Infant Behavior and Developmend* Vol. 60, p. 383-396.

Trainor, L. J., Austin, C. M. & Desjardins, R. N. (2000) Is infant-directed speech prosody a result of the vocal expression of emotion?, *Psychological Science* Vol. 11, p. 188-195.

Trehub, S. E., Trainor, L. J. & Unyk, A. M. (1993) Music and speech processing in the first year of life, in H. W. Reese (ed.) *Advances in Child Development and Behaviour* Vol. 24, p. 1-35. Academic Press, New York.

Trevarthen, C. (1999) Musicality and the intrinsic motive pulse: evidence from human psychobiology and infant communication, *Musicae Scientiae* Special Issue, p. 155-215.

Turino, T. (1992) The music of Sub-Saharan Africa, in B. Nettl, C. Capwell, P. Bohlman, I. Wong & T. Turino (eds.) *Excursions in World Music*. Prentice Hall, Englewood Cliffs, NJ, p. 165-195.

Turk, I. (ed.) (1997) *Mousterian "Bone Flute" and other finds from Divje babe I cave site in Slovenia*. Institut za archaeologijo, Znanstvenoraziskovalni Center Sazu., Ljubljana.

Turk, I. & Kavur, B. (1997) Palaeolithic bone flutes – comparable material, in I. Turk (ed.), *Mousterian "Bone Flute" and other finds from Divje babe I cave site in Slovenia*. Institut za archaeologijo, Znanstvenoraziskovalni Center Sazu., Ljubljana, p. 179-184.

Turk, I., Dirjec, J. & Kavur, B. (1997) Description and explanation of the origin of the suspected bone flute, in I. Turk (ed.), *Mousterian "Bone Flute"*, Znanstvenoraziskovalni Center Sazu, Ljubljana, p. 157-178.

Turk, I., Dirjec, J., Bastiani, G., Pflaum, M., Lauko, T., Cimerman, F., Kosel, F., Grum, J. & Cevc, P. (2001) Nove analize "piscali" iz Divjih bab I (Slovenija), *Arheoloski vestnik* Vol. 52, p. 25-79.

Turnbull, C. (1962) *The Forest People*. Simon and Schuster, New York.

Vanderhorst, V. G. J. M., Teresawa, E., & Ralston, H. J. (2001) Monosynaptic projections from the nucleus retroambiguus region to laryngeal motoneurons in the rhesus monkey, *Neuroscience* Vol. 107, p. 117-125.

Van Lancker, D. R. & Canter, G. J. (1982) Impairment of voice and face recognition in patients with hemispheric damage. *Brain and Cognition* Vol. 1, p. 185-95.

Vaneechoutte, M. & Skoyles, J. R. (1998) The memetic origin of language: modern humans as musical primates, *Journal of Memetics - Evolutionary Models of Information Transmission* Vol. 2. Also available online at http://www.mmu.ac.uk/jom-emit/1998/vol2/vameechoutte_m&skoyles_jr.html

VanWijngaarden-Bakker, L. H. (1997) The selection of bird bones for artefact production at Dutch Neolithic sites, *International Journal of Osteoarchaeology* Vol. 7, p. 339-345.

Vekua, A., Lordkipanidze, D., Rightmire, G. P., Agusti, J., Ferring, R., Maisuradze, G., Mouskhelishvili, A., Nioradze, M., de Leon, M. P., Tappen, M., Tvalchrelidze, M. & Zollikofer, C. (2002) A new skull of early homo from Dmanisi, Georgia, *Science* Vol. 297, p. 85-89.

Vincent, A. (1988) L'os commune artefact au Paléolithic moyen: princpes d'étude et premieres resultants, in *L'Homme de Néanderthal* Vol. 4: La Technique. Études et Recherches Archéologiques de l'Université de Liège 31, Liège, p. 185-196.

Wagner, H. (1989) The peripheral physiological differentiation of emotions, in H. Wagner & A. Manstead (eds.), *Handbook of psychophysiology: emotion and social behavior*. Wiley and Sons, Chichester.

Walker, A. (1993) Perspective on the Nariokotome discovery, in A. Walker & R. Leakey, *The Nariokotome* Homo erectus *Skeleton*. Harvard University Press, Cambridge, MA, p. 411-432.

Wallin, N. (1991) *Biomusicology: Neurophysiological, Neuropsychological and Evolutionary Perspectives on the Origins and Purpose of Music.* Pendragon Press, Stuyvesant, NY.

Wallin, N. L., Merker, B., & Brown, S. (2000) *The Origins of Music*. MIT Press, London.

Watt, R. J. & Ash, R. L. (1998) A Psychological Investigation Of Meaning In Music, *Musicae Scientiae* Vol. 2, p. 33-54.

Werker, J. F., Pegg, J. E. & McLeod, P. J. (1994) A cross-language investigation of infant preference for infant-directed communication, *Infant Behaviour and Development* Vol. 17, p. 323-333.

Wild, B., Erb, M. & Bartels, M. (2001) Are emotions contagious? Evoked emotions while viewing emotionally expressive faces: quality, quantity, time course and gender differences, *Psychiatry Research* Vol. 102, p. 109-124.

Williams, L. (1980) *The Dancing Chimpanzee* 2nd Edition. Allison and Busby, London.

Wilson, D. S. (1997) Incorporating group-selection into the adaptationist program: a case study involving human decision-making, in J. A. Simpson & D. T. Kendrick, (eds.) *Evolutionary Social Psychology*. Lawrence Erlbaum Assoc., New Jersey.

Wilson, D. S. & Sober, E. (1994) Reintroducing group selection to the behavioural sciences, *Behavioral and Brain Sciences* Vol. 17, p. 585-654.

Wind, J. (1990) The evolutionary history of the human speech organs, in J. Wind *et al*. (eds.) *Studies In Language Origins* Vol. 1. Benjamins, Amsterdam, p. 173-197.

Wylie, A. (1985) The reaction against analogy, in M. Schiffer (ed.) *Advances in Archaeological Method and Theory* Vol. 8. Academic Press, New York, p. 63-111.

Wynn, T. (1993) Two developments in the mind of early Homo, *Journal of Anthropological Archaeology* Vol. 12, p. 299-322.

Yellen, J. E. (1977) *Archaeological Approaches to the Present*. Academic Press, New York.

Zatorre, R. J. (1984) Musical perception and cerebral function: a critical review, *Musical Perception* Vol. 2, p. 196-221.

Zubrow, E., Cross, I. & Cowan, F. (2001) Musical behaviour and the archaeology of the mind, *Archaeologia Polona* Vol. 39, p. 111-126.

APPENDIX

Table 1: Inventory of Palaeolithic reputed pipes and flutes

Table 2: Inventory of Palaeolithic reputed phalangeal whistles

Table 3: Palaeolithic Pipes and flutes* by *fauna used

The appendix of the thesis, which constitutes a database of reputed Palaeolithic musical artefacts, is unavailable here pending further analysis and subsequent publication in printed form, at the recommendation of the examiners of the thesis, Prof. Paul Mellars and Prof. Leslie Aiello.

Many apologies for any inconvenience this may cause.