

On the evolution of language: A unified view*

PHILIP LIEBERMAN

University of Connecticut, Storrs

Abstract

Language can be operationally defined as a communications system that permits the exchange of new, unanticipated information. Different forms of language appear to have been present in earlier stages of hominid evolution. Human language is unique, at the present time, since it makes use of 'encoded' speech to achieve a rapid transfer of information. The supralaryngeal vocal tract of modern Homo sapiens is a useful factor in this encoding process which also involves special neural mechanisms. Other factors like cognitive ability and 'automatization' are also necessary for language. Those factors are, however, important for many aspects of human and non-human behavior besides language.

The evolution of language appears to have been a gradual process that first led to systems that relied on mixed gestural and vocal communication. Some hominids appear to have retained this system until comparatively recent times. Other hominids appear to have placed a greater reliance on vocal communication. Reconstructions of fossil supralaryngeal vocal tracts show that some forms, Australopithecines and 'classic' Neanderthal, lacked the supralaryngeal vocal tract that is necessary for the production of fully encoded human speech. Other fossil forms, Steinheim and Es-Skhül V, had functionally modern vocal tracts. Others, like Broken Hill, represent intermediate forms. The evolution of human language can be viewed as a three stage process that involved (a) increased reliance on vocal communication in activities like hunting, (b) the enhancement of the vocal repertoire with the evolution of the human supralaryngeal vocal tract which produces acoustic signals that are both more distinct and more resistant to articulatory errors, and (c) the evolution of neural mechanisms that made use of the preadapted properties of the supralaryngeal vocal tract for rapid encoded speech communication.

* Paper prepared for the Ninth International Congress of Anthropological and Ethnological

Sciences, Chicago, U.S.A., September 1973.

I shall attempt to develop a unified theory for the evolution of human language in this paper. Though this theory centrally involves the comparative, ontogenetic, and evolutionary studies of speech production with which I and my colleagues are closely identified, it also crucially involves the consideration of other recent, and not so recent, studies of cognitive ability in non-human primates: Hunting, bipedal posture, the neural correlates of auditory perception, visual perception in adult and infant humans, speech perception in humans, play activity, gesture, etc. In short, I shall attempt to synthesize a great deal of data into what I hope is a coherent theory. I somewhat redundantly stress that this will hopefully result in a theory that is testable. Like all theories it cannot account for everything. This theory does, however, appear to 'explain' and relate a number of phenomena that otherwise appear to be quite unrelated. It, moreover, appears to point to a coherent evolutionary process that relates the communications systems of other animals to human language. It most importantly points out a number of questions that can be resolved through controlled experiments and careful observations.

I have drawn on a number of seemingly disparate ethological, anatomical, psychological and anthropological sources because I think that it is obvious that there is no single factor that is, in itself, responsible for the evolution of human language. Evolution is a complex process that inherently involves all aspects of the life cycle and environment of a species and its relationships to other species. Though particular factors like, for example, gestural communication (Hewes, 1971), undoubtedly had an important role in the evolution of human language, no single factor can, in itself, provide, as it were, the 'central key' to the puzzle. Everything depends on everything else and the interaction is the 'crucial' factor if anything is. Gestural communication, for example, depends on the prior existence of visual pattern recognition, analysis, cognitive ability, and bipedal posture. Visual pattern identification probably depends, in turn, on natural selection for visual ability in an arboreal environment. Bipedal posture, in turn, probably again depends on prior selection for brachiation in an arboreal environment (Campbell, 1966).

Note that I am not saying that we cannot analyze the factors that underly the evolution of human language. I am proposing that the process involved many factors. One of these factors appears to be the process of 'preadaptation', that is, natural selection channeled development in particular directions because of previous modifications selected for some other role. Darwin's (1859) comments concerning the evolution of the lung from the swim bladder perhaps constitute one of the first and most convincing examples of preadaptation.

Let me begin by listing the evolutionary factors that I will discuss in this present paper. There probably are more factors, but I propose that these are the central factors in the evolution of human language. I shall order the factors in terms of their

probable role in differentiating the language of modern man from progressively earlier hominids and other animals. In other words, I shall first list the factors that I think were most important in the late stages of human evolution and proceed to factors that probably were more important in earlier stages. It is important to note that I am not categorically differentiating human language, i.e., the language of present day *Homo sapiens*, from other languages, e.g., the possible language of present-day chimpanzees.

Linguists have been somewhat anthropocentric in defining language to be necessarily human language. I will define a language to be a communications system that is capable of transmitting new information. In other words, I am operationally defining language as a communications system that places no inherent restriction on the nature or quality of the information transferred. It is obvious that this definition does not require that all languages have all of the properties of human language.

1. Factor 1 – Specialized speech encoding and decoding

Modern man's communications achieve a high rate of transmission speed by means of a process of speech encoding and decoding. The rate at which meaningful sound distinctions are transmitted in human speech is about twenty to thirty segments per second. That is, phonetic distinctions that differentiate meaningful words, e.g., the sounds symbolized by the symbols [b], [æ], and [t] in the word *bat*, are transmitted, identified, and sorted at a rate of twenty to thirty segments per second. It is obvious that human listeners *cannot* simply transmit and identify these sound distinctions as separate entities. The fastest rate at which sounds can be identified is about seven to nine segments per second. Sounds transmitted at a rate of twenty per second indeed merge into an undifferentiable 'tone'. The linguist's traditional conception of phonetic elements comprising a set of 'beads on a string' clearly is not correct at the acoustic level. How, then, is speech transmitted and perceived?

The results of the past twenty years of research on the perception of speech by humans demonstrated that the individual sounds like [b], [æ], and [t] are encoded, that is, 'squashed together', into the syllable-sized unit [b æ t] (Lieberman *et al.*, 1967). A human speaker in producing this syllable starts with his supralaryngeal vocal tract, i.e., his tongue, lips, velum, etc., in the positions characteristic of [b]. He, however, does not maintain this articulatory configuration but instead moves his articulators towards the positions that would be attained if he were instructed to maintain an isolated, steady [æ]. He never reaches these positions, however, because he starts towards the articulatory configuration characteristic of [t] before he ever reaches the 'steady state' (isolated and sustained) vowel [æ]. The articulatory gestures

that would be characteristic of each isolated 'sound' are never attained. Instead the articulatory gestures are melded together into a composite, characteristic of the syllable.

The sound pattern that results from this encoding process is itself an indivisible composite. Just as there is no way of separating with absolute certainty the [b] articulatory gestures from the [æ] gestures (you can't tell exactly when the [b] ends and the [æ] begins), there is no way of separating the acoustic cues that are generated by these articulatory maneuvers. The isolated sounds have a psychological status as motor control or 'programming' instructions for the speech production apparatus. The sound pattern that results is a composite, and the acoustic cues for the initial and final consonants are largely transmitted as modulations imposed on the vowel. The process is, in effect, a time-compressing system. The acoustic cues that characterize the initial and final consonants are transmitted in the time slot that would have been necessary to transmit a single isolated [æ] vowel.

The human brain decodes, that is, 'unscrambles', the acoustic signal in terms of the articulatory maneuvers that were put together to generate the syllable. The individual consonants [b] and [t], though they have no independent acoustic status, are perceived as discrete entities. The process of human speech perception inherently requires 'knowledge' of the acoustic consequences of the possible range of human supralaryngeal vocal tract speech articulation (Lieberman *et al.*, 1967; Lieberman, 1970, 1972). The special speech processing involved appears to crucially involve the dominant hemisphere of the human brain (Kimura, 1964; Shankweiler and Kennedy, 1967; Lieberman *et al.*, 1967). We will discuss the process of human speech perception in more detail with respect to its interrelation with the anatomy of the human vocal tract. For the moment, we will note that the special neural devices necessary for the 'decoding' of human speech may be comparatively recent evolutionary acquisitions.

2. Factor 2 – Special supralaryngeal vocal tract anatomy

Modern man's speech-producing apparatus is quite different from the comparable systems of living non-human primates (Lieberman, 1968; Lieberman *et al.*, 1969; Lieberman *et al.*, 1972b). Non-human primates have supralaryngeal vocal tracts in which the larynx exits directly into the oral cavity (Negus, 1949). In the adult human the larynx exits into the pharynx. The only function for which the adult human supralaryngeal vocal tract appears to be better adapted is speech production. Understanding the anatomical basis of human speech requires that we briefly review the source-filter theory of speech production (Chiba and Kajiyama, 1958; Fant, 1960). Human speech is the result of a source, or sources, of acoustic energy being filtered

by the supralaryngeal vocal tract. For voiced sounds, that is, sounds like the English vowels, the source of energy is the periodic sequence of puffs of air that pass through the larynx as the vocal cords (folds) rapidly open and shut. The rate at which the vocal cords open and close determines the fundamental frequency of phonation. Acoustic energy is present at the fundamental frequency and at higher harmonics. The fundamental frequency of phonation can vary from about 80 Hz for adult males to about 500 Hz for children and some adult females. Significant acoustic energy is present in the harmonics of fundamental frequency to at least 3000 Hz. The fundamental frequency of phonation is, within wide limits, under the control of the speaker who can produce controlled variations by varying either pulmonary air pressure or the tension of the laryngeal muscles (Lieberman, 1967). Linguistically significant information can be transmitted by means of these variations in fundamental frequency as, for example, in Chinese where these variations are used to differentiate different words.

The main source of phonetic differentiation in human languages, however, arises from the dynamic properties of the supralaryngeal vocal tract acting as an acoustic filter. The length and shape of the supralaryngeal vocal tract determines the frequencies at which maximum energy will be transmitted from the laryngeal source to the air adjacent to the speaker's lips. These frequencies, at which maximum acoustic energy will be transmitted, are known as formant frequencies. A speaker can vary the formant frequencies by changing the length and shape of his supralaryngeal vocal tract. He can, for example, drastically alter the shape of the airway formed by the posterior margin of his tongue body in his pharynx. He can raise or lower the upper boundary of his tongue in his oral cavity. He can raise or lower his larynx and retract or extend his lips. He can open or close his nasal cavity to the rest of the supralaryngeal vocal tract by lowering or raising his velum. The speaker can, in short, continually vary the formant frequencies generated by his supralaryngeal vocal tract. The acoustic properties that, for example, differentiate the vowels [a] and [i] are determined solely by the shape and length differences that the speaker's supralaryngeal vocal tract assumes in articulating these vowels. The situation is analagous to the musical properties of a pipe organ where the length and type (open or closed end) of pipe determines the musical quality of each note. The damped resonances of the human supralaryngeal vocal tract are, in effect, the formant frequencies. The length and shape (more precisely the cross-sectional area as a function of distance from the laryngeal source) determine the formant frequencies.

The situation is similar for unvoiced sounds where the vocal cords do not open and close at a rapid rate releasing quasiperiodic puffs of air. The source of acoustic energy in these instances is the turbulence generated by air rushing through a constriction in the vocal tract. The vocal tract still acts as an acoustic filter but the acoustic source

may not be at the level of the larynx as, for example, in the sound [s] where the source is the turbulence generated near the speaker's teeth.

The anatomy of the adult human supralaryngeal vocal tract permits modern man to generate supralaryngeal vocal tract configurations that involve abrupt discontinuities at its midpoint. These particular vocal tract shapes produce vowels like [a], [i], and [u] which have unique acoustic properties as well as consonants like [g] and [k]. The acoustic properties of these particular sounds will be discussed in detail, but for the moment I will simply note that they are sounds that minimize the problems of precise articulatory control. A speaker can produce about the same formant frequencies for an [i], for example, while he varies the position of the midpoint area function discontinuity by one or two centimeters (Stevens, forthcoming). They are also sounds that are maximally distinct acoustically. They, moreover, are sounds that a human listener can efficiently use to establish the size of the supralaryngeal vocal tract that he is listening to. This last property relates to Factor 1, the specialized speech encoding and decoding that characterizes human language. The reconstructions of the supralaryngeal vocal tracts of various fossil hominids that my colleague Edmund S. Crelin has made indicate that some extinct hominids lacked the anatomical basis for producing these sounds while other hominids appear to have the requisite anatomical specializations for human speech. I will, of course, return to this topic.

3. Factor 3 - Cognitive ability and automatization

There are two interrelated aspects to the cognitive abilities that underly language. One is the process that I will term automatization. Human language involves rapidly executing complex sequences of articulatory maneuvers or making equally complex perceptual decisions regarding the identity of particular sound segments. At a higher level, complex phonologic and morphophonemic relationships must be determined. None of these processes is, however, what the speaker or listener is directly concerned with. The semantic content of the message is the primary concern of the speaker or listener. The sending and receiving processes are essentially automatic. No conscious thought is expended in the process of speech production, speech perception, or any of the syntactic or morphophonemic stages that may intervene between the semantic content of the message and the acoustic signal. It is clear that 'automatized' skills are not unique to human language. Other aspects of human activity, dance, for example, involve similar phenomena. The novice dancer must learn the particular steps and movements that characterize a particular dance form. Once the steps have been learned they must become automatized. The dance itself involves the complex sequences. Playing the piano or violin, skiing, or driving a car all involve automatized behavior.

The bases for the automatized behavior that is a necessary condition for human language may reside in cross-modal transfers from other systems of hominid and hominoid activity. Tool use, for example, requires a high degree of automatization if it is to be effective. You can't stop to think how to use a hammer every time you drive a nail in. Hunting is perhaps a still stronger case. A successful hunter must be able to accurately thrust his spear without pausing to think about the mechanics of spear thrusting. Natural selection would quickly favor the retention of superior automatization. Automatized behavior pervades all aspects of culture. Indeed a cultural response is, to a degree, a special case of automatized behavior. In simpler animals cultural responses are perhaps less subject to environmental pressures. In humans they may be more subject to external forces rather than innate mechanisms, but they are no less automatized once learned.

A special factor that may be germane to automatized behavior is that a 'plastic' period appears to be involved. It is comparatively easy to shape behavior during the 'plastic' period. It afterwards is either impossible or relatively difficult to modify automatized behavior. Puppies thus can be trained more readily than adult dogs. All humans can readily learn different languages in their youth. Most humans can learn a foreign, sic unfamiliar, language only with great difficulty (or not at all) during adult life. The same comments probably apply to learning to play the violin, tight-rope walking, etc., though no definitive studies have yet been made.

4. Cognitive ability

Cognitive ability is a necessary factor in human language. Linguists often tend to assume that cognitive ability *is* linguistic ability. Indeed, since the time of Descartes the absence of human language in other animals has been cited as a 'proof' of man's special status and the lack of cognitive ability in all other species. Human language has been assumed to be a necessary condition for human thought. The absence of human language has been, conversely, assumed to be evidence of the lack of all cognitive ability.

It is clear that cognitive, i.e., logical, abilities can be demonstrated or observed in many animals. Behavioral conditioning, for example, which can be applied with great success to pigeons and rats, itself can be viewed as a demonstration of logical ability on the part of the 'conditioned' animal. Pavlov's dogs had to make a logical association between the bell and food. Calling the animal's response a 'conditioned reflex' obscures the fact that the animal had to be able to *logically* connect the sound of the bell with food. The same 'conditioned' response often can be observed as a human gourmet regards the menu. In both cases cognitive ability must interpose between the token

of the food that is anticipated and the observed physiologic response. The human gourmet is hopefully more flexible, adaptive and discriminating than Pavlov's dogs; however, the basic process is similar. In *Homo sapiens* the cognitive abilities that underly this particular aspect of behavior are simply more complex than is the case for *Canus familiarus*. The difference is, however, quantitative rather than qualitative.

The particular cognitive abilities that are associated with presumably 'unique' human behavioral patterns like tool use have been observed in chimpanzees (Goodall, 1971) and sea otter (Kenyon, 1969). Some of the cognitive abilities that have been traditionally associated with human language have likewise been demonstrated by Gardner and Gardner (1969) and by Premack (1972). Premack's experiments, in particular, clearly demonstrate that cognitive ability and human language *cannot* be regarded as the same biologic ability.

Chimpanzees do not possess the phonetic apparatus of human language. They have available a subset of the phonetic distinctions that are available to modern man. Chimpanzees *could*, using the phonetic distinctions that are available to them, establish a language. This language's phonetic system might not be as efficient as modern man's, but it could form the basis of a language where we have operationally defined a language to be a communications system capable of transmitting unanticipated, new knowledge. The difference, at the phonetic level, between human language and this hypothetical chimpanzee language would be quantitative rather than qualitative. Premack's experiments demonstrate that the cognitive abilities of chimpanzees are, at worst, restricted to some subset of the cognitive abilities available to humans. The difference at the cognitive level is thus also probably quantitative rather than qualitative.

It is important to note, at this point, that quantitative functional abilities can be the bases of behavioral patterns that are qualitatively different. I think that this fact is sometimes not appreciated in discussions of gradual versus abrupt change. A modern electronic desk calculator and a large general purpose digital computer, for example, may be constructed using similar electronic logical devices and similar magnetic memories. The large general purpose machine will, however, have 1,000 to 10,000,000 times as many logical and memory devices. The structural differences between the desk calculator and general purpose machine may thus simply be quantitative rather than qualitative. The 'behavioral' consequence of this quantitative difference can, however, be qualitative. The types of problems that one can solve on the general purpose machine will differ in kind, as well as in size, from those suited to the desk calculator. The inherent cognitive abilities of humans and chimpanzees thus could be quantitative and still have qualitative behavioral consequences.

The cognitive abilities that are typically associated with human language may have their immediate origins in the complex patterns of hominid behavior associated with

tool use, tool making and hunting. Hewes (1971) makes a convincing case for the role of gestural communication in the earliest forms of hominid language and associates language with the transference of cognitive ability from these complex behavioral patterns. I would agree with Hewes, but I would not limit the earlier hominid languages to gestures nor would I restrict the cognitive abilities that underly language to hominids. Tool use and hunting certainly are not exclusively hominid patterns of behavior.

We can get some insights on the neural abilities that non-human primates possess by taking note of the phylogenetic evolution of the peripheral systems involved in information gathering and communication. The acute color vision of primates, for example, would have had no selective advantage if it were not coupled with matching cognitive processes. Gestural communication is consistent with the evolution and retention of increasingly complex facial musculature in the phylogenetic order of primates. It is likewise unlikely that gestural communication was at any stage of hominid evolution the *sole* 'phonetic' medium. Negus (1949), by the methods of comparative anatomy, demonstrates that the larynges of non-human primates are adapted for phonation at the expense of respiratory efficiency. The far simpler larynx of the lung fish is better adapted for respiration and protecting the lungs. Clearly mutations that decreased respiratory efficiency would not have been retained over a phylogenetic order unless they had some selectional advantage. The cognitive skills that underly linguistic ability in hominids thus probably evolved from cognitive facilities that have a functional roles in the social behavior and communications of other animals. Like automatization, these skills would appear to be part of the biologic endowment of many species, and their continued development in 'higher' species is concomitant with behavioral complexity. The transference of these cognitive skills to human language thus could be viewed as yet another instance of 'preadaptation', the use of cognitive processes for language that originally evolved because of the selective advantages conferred on activities like hunting, evading natural enemies, food gathering, etc.

5. The speech abilities of Neanderthal and other fossil hominids

As I noted before, it is apparent that no single factor can be in any reasonable way identified as the 'key' to language. The two factors that appear to be most recent in shaping the particular form of *human* language are, however, Speech Encoding and Speech Producing Anatomy. Certain neural mechanisms must be present for the perception of speech (Lenneburg, 1967). It is difficult to make any substantive inferences about the presence or absence of particular neural mechanisms in the

brains of extinct fossil hominids since we can deduce only the external size and shape of the brain from a fossil skull. Also, we lack a detailed knowledge of how the human brain functions. We could not really assess the linguistic abilities of a modern man simply by examining his brain. Fortunately, we can derive some insights on the nature of speech perception in various fossil hominids by studying their speech-producing anatomy. The relationship between speech anatomy and speech perception is very much like that which obtains between bipedalism and the detailed anatomy of the pelvic region. The anatomy is a necessary condition, though neural ability is also necessary.

The methodology that has enabled us, and I must emphasize that this research has been a joint enterprise, to reconstruct the speech-producing anatomy of extinct hominids is that proposed by Charles Darwin. Darwin in chapters 10 and 13 of *On the origin of species* (1859) discussed both the 'affinities of extinct Species to each other, and to living forms', and 'Embryology'. We have applied the methods of comparative and functional anatomy to the speech-producing anatomy of present day apes and monkeys and to normal human newborn. We first assessed the speech-producing abilities of these living animals in terms of their speech-producing anatomy. We found that their supralaryngeal vocal tracts inherently restricted their speech-producing abilities. We then noted that certain functional aspects of the morphology of the skulls of these living animals resembled similar features of extinct fossil hominids.

The reconstructions of the supralaryngeal vocal tracts of the La Chapelle-aux-Saints, Es-Skhül V, Broken Hill, Steinheim, and Sterkfontein 5 fossils were made by my colleague Edmund S. Crelin by means of the homologues that exist between these skulls and living forms, the marks of the muscles on the fossil skulls, and the general methods of comparative anatomy. Crelin's (1969) previous experience with the anatomy of the newborn was especially relevant since we can see in the human newborn many of the relevant skeletal features associated with the soft tissue structures that must have occurred in certain of these now extinct hominid forms. In most cases we made use of casts of the fossil material made available by the Wenner-Gren foundation. For the La Chapelle-aux-Saints and Steinheim fossils, casts made available by the University Museum, Philadelphia, Pennsylvania were employed. The original La Chapelle-aux-Saints fossil as well as the La Ferrassie, and La Quina child's fossil were also examined with the cooperation of the Musée de l'Homme in Paris and the Musée des Antiquités Nationales in St. Germain-en-Laye. We attempted to examine the original Steinheim fossil but were not successful. The details of the reconstructions are discussed in our published and forthcoming papers (Lieberman and Crelin, 1971; Lieberman *et al.*, 1972b; Crelin *et al.*, forthcoming). I will, however, note some of the salient points in the discussion of particular fossils. I will first discuss the computer modelling technique that we employed to arrive at a functional assessment of these

supralaryngeal vocal tracts. I think that it makes sense to approach the discussion of the reconstructions by first discussing the modelling technique because one of the points that I hope will emerge from the discussion of the modelling technique is the question of how much of the details of the supralaryngeal vocal tract's morphology we need to know in order to make meaningful statements about speech ability. The answer to this question is that we really need to know only a few, fairly gross aspects of the morphology of the supralaryngeal vocal tract. The reason that this is so is itself one of the functional characteristics of human speech.

I'll begin the discussion of the modelling technique by returning to our studies of the speech capabilities of living non-human primates. This is a useful way to start since we can compare the results of our modelling with the actual phenomena. Figure 1 shows the left half of the head and neck of a young adult male chimpanzee sectioned in the midsagittal plane. Silicone rubber casts were made of the air passages, including the nasal cavity, by filling each side of the split air passages separately in the sectioned head and neck to insure perfect filling of the cavities. The casts from each side of a head and neck were then fused together to make a complete cast of the air passages. In Figure 2 the cast of the chimpanzee airways is shown together with casts made, following the same procedures, for newborn human and adult human. A cast of the reconstructed supralaryngeal airways of the La Chapelle-aux-Saints fossil also appears in this figure. In Figure 3 equal sized outlines of the air passages for these four vocal tracts are sketched.

Note the high position of the larynx in the newborn human, and adult chimpanzee vocal tracts where the soft palate and epiglottis can be approximated. In the adult human vocal tract the soft palate and epiglottis are widely separated and cannot be approximated. The tongue is likewise at rest in newborn human and chimpanzee completely within the oral cavity, whereas in adult man the posterior third of the tongue is in a vertical position forming the anterior wall of the supralaryngeal pharyngeal cavity. Note, in particular, that there is practically no supralaryngeal portion of the pharynx present in the direct airway out from the larynx when the soft palate shuts off the nasal cavity in newborn human and in chimpanzee. In adult man half of the supralaryngeal vocal tract is formed by the pharyngeal cavity.

This difference between the chimpanzee and newborn supralaryngeal vocal tracts and that of adult *Homo sapiens* is a consequence of the opening of the larynx into the pharynx directly behind the oral cavity. In other words, the larynx opens almost directly into the oral cavity. This is the case for all living animals (Negus, 1949) with the exception of adult *Homo sapiens*. We really should use the term 'adult-like' rather than 'adult' since these differences appear to be fully developed by two years of age and are probably largely differentiated by six months of age (Lieberman *et al.*, 1972a).

The functional distinctions that these anatomical differences confer on adult humans have been determined for respiration, swallowing, and the sense of smell. Kirchner (1970) notes that the respiratory efficiency of the adult human supralaryngeal airways is about half that of the newborn. The right angle bend in the adult human supralaryngeal airway increases the flow resistance. The non-human supralaryngeal anatomy allows the oral cavity to be sealed from the rest of the airway during inspiration. This aids the sense of smell (Negus, 1949) and also allows an animal to breathe while its mouth contains a liquid (e.g. when a dog laps water). The adult human supralaryngeal airways also increase the possibility of asphyxiation. Food lodged in the pharynx can block the entrance to the larynx. This is not possible in non-humans since the supralaryngeal pharynx serves as a pathway for both food and liquids and as an airway only in adult *Homo sapiens*.¹

The functional distinctions that the differences in the anatomy of the supralaryngeal airways confer on speech production can be determined by modelling techniques. The source-filter theory of speech production, as I have noted before, states that speech is the result of the filtering action of the supralaryngeal vocal tract on the acoustic sources that excite it. Since the filtering properties of the filter are uniquely determined by the shape and length (the cross-sectional area function) of the supralaryngeal vocal tract it is possible to assess the properties of a particular vocal tract once we know the range of shapes that it can assume.

Note that this type of analysis will not tell us anything about the total range of phonetic variation. We would have to know the properties of the laryngeal source as well as the degree of motor control that a particular organism possessed. We, however, can assess the restraints that the supralaryngeal vocal tract itself imposes on the possible phonetic repertoire. The situation is similar to that which would occur if we found an ancient woodwind instrument made of brass. We would probably not be able to say very much about the reed, which would have decayed, but we would be able to determine some of the constraints that the instrument imposed on a performance. These constraints obviously would inherently structure the musical forms of the period. You can't write music that cannot be performed. We would not know all of the constraints, we could not say very much about the manual dexterity of the players or the general musical theory, but we would know more than would be the case if we had not found the ancient instrument.

1. The human vocal tract is also inferior to the vocal tracts of hominids like La Chapelle-aux-Saints with respect to chewing. The reduction in the body of the mandible in modern *Homo sapiens* has reduced the tooth area. Dental studies have determined (Manly and Braley, 1950; Manly and Shiere, 1950; Manly and

Vinton, 1951) that chewing efficiency in primates is solely a function of swept tooth area. Hominid forms that have smaller tooth areas have less efficient chewing. The reduction of the mandible in modern man therefore cannot be ascribed to enhancing chewing efficiency.

We are in a somewhat better position when we study the reconstructed supralaryngeal vocal apparatus of an extinct hominid. We can tell something about the constraints on the phonetic repertoire. The interconnections that exist between the vocal apparatus and the perception of speech in *Homo sapiens*, moreover, allow us to make some more general inferences than would otherwise be the case.

The technique that we have employed to assess the constraints imposed by the supralaryngeal vocal apparatus of an animal makes use of a computer model of the vocal tract. We really don't have to make use of this model. It would be possible, though somewhat tedious, to make actual models of possible supralaryngeal vocal tract configurations. If these models, made of plastic or metal, were excited by means of a rapid quasiperiodic series of puffs of air (i.e., an artificial larynx) we would be able to hear the actual vowel-like sounds that a particular vocal tract configuration produced. If we systematically made models that covered the range of possible vocal tract configurations we could determine the constraints that the supralaryngeal vocal tract morphology imposed, independent of the possible constraints determined by limitations on motor control, etc. We would be, of course, restricted to steady-state vowels since we could not rapidly change the shape of the vocal tract but we could generalize our results to consonants since we could model the articulatory configurations that occur at the start and end of typical consonant-vowel sequences. Note that these modelling techniques allow us to assess the limits on the phonetic repertoire that follow from the anatomy of the supralaryngeal vocal tract, independent of muscular or neural control and independent of the dialect, habits, etc., of the animal whose vocal tract we would be modelling. The technology for making these mechanical models existed at the end of the eighteenth century. Von Kempelen's (1791) famous talking machine modelled the human vocal tract by mechanical means. The method that we have employed simply makes use of the technology of the third quarter of the twentieth century.

6. Chimpanzee newborn and adult *Homo sapiens*

In Figure 4 three area functions are shown for the chimpanzee vocal tract derived from the sectioned head and neck shown in Figure 1. The silicone rubber casting and schematic drawing of this vocal tract are shown in Figures 2 and 3 respectively. The area functions shown in Figure 1 represent the best approximations that we could get to the human vowels [a], [i], and [u]. We systematically drew area functions on an oscilloscope input to a PDP 9 computer with a light pen. The computer had been programmed to calculate the formant frequencies that corresponded to these area functions. The details of the computer program are discussed by Henke (1966). The

computer allowed us to conveniently and rapidly make hundreds of possible supralaryngeal vocal tract models. We thus could explore the acoustic consequences of all possible chimpanzee supralaryngeal vocal tract configurations without waiting for a chimpanzee to actually produce these shapes. We used the same procedure to explore the possible range of supralaryngeal vocal tract shapes for the newborn human supralaryngeal vocal tract shown in Figures 2 and 3. We were guided in these simulations by the morphology of the head and neck, i.e., the relative thickness and position of the tongue, the lips, the velum and the position of the pharynx relative to the larynx and oral cavity. We were also able to make use of cineradiographic pictures of newborn infants during cry and swallowing (Truby *et al.*, 1965). The results of these simulations are shown in Figure 5. In Figure 5 the formant frequencies of the three area functions of Figure 4 are plotted, together with an additional data point (X) for human newborn. The loops labelled with phonetic symbols represent the data points for a sample of real utterances derived from 76 adult men, women and adolescent children producing American-English vowels (Peterson and Barney, 1952). In Figure 6 we have reproduced the actual data points for this sample of real human vowels. Note that the chimpanzee and newborn human utterances only cover a small portion of the adult human 'vowel space'. In other words, the chimpanzee and newborn vocal tracts according to this modelling technique inherently do not appear to be able to produce vowels like [a], [i], and [u].

All normal human speakers can inherently produce these vowels. Any human, if he is raised in an American-English environment will be able to produce these vowels. The modelling of the chimpanzee and newborn vocal tracts indicates that they could not, even if they had the requisite motor and neural abilities. The question that we are addressing is thus not whether chimpanzees and newborns can speak American-English. It is whether they have the anatomical apparatus that would allow them to speak.

The results of the modelling technique can, of course, be checked against the actual utterances of chimpanzees and newborn *Homo sapiens*. When this is done it is evident that the actual vowels of newborn *Homo sapiens* agree with the computer simulation (Irwin, 1948; Lynip, 1951; Lieberman *et al.*, 1972a). The chimpanzee simulation appears to encompass a greater range than has been observed so far in the acoustic analysis of chimpanzee vocalizations (Lieberman, 1968). This may merely indicate that the acoustic analyses so far derived from chimpanzee do not represent the total chimpanzee repertoire. It is, however, apparent that the computer simulation does not appear to be showing a *smaller* vowel space than is actually the case. The computer simulation for adult *Homo sapiens* corresponds with that observed (Chiba and Kajiyama, 1958; Fant, 1960; Peterson and Barney, 1952) and is not plotted here.

The vowel diagrams in Figures 5 and 6 are really an indirect way of showing that



Figure 1. *Left half of the head and neck of a young adult male chimpanzee sectioned in the midsagittal plane (after Lieberman et al., 1972)*

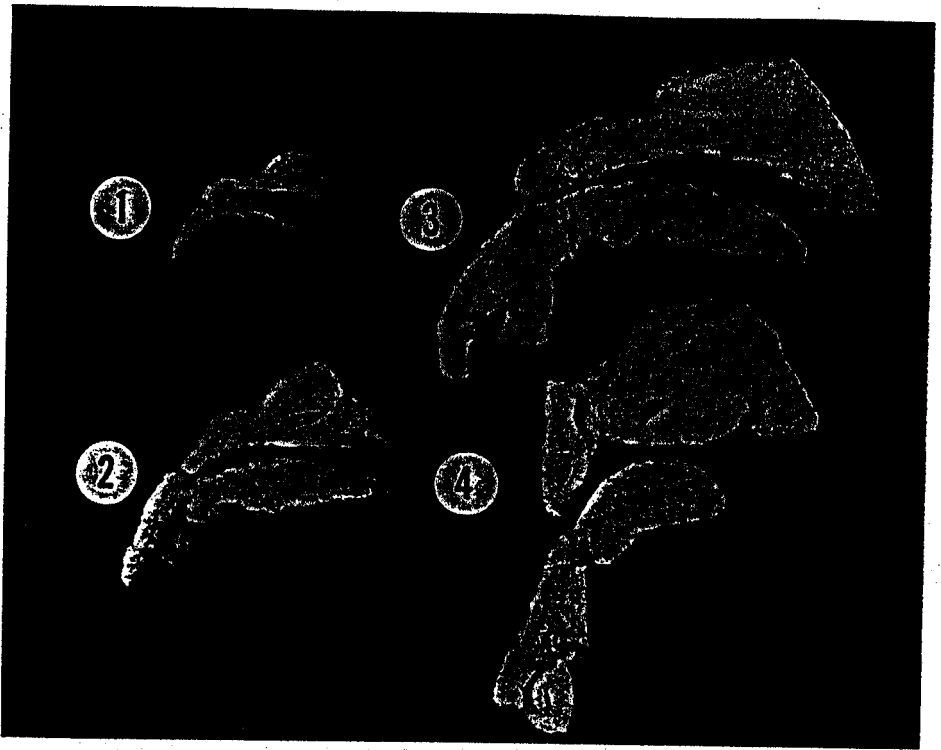


Figure 2. Casts of the nasal, oral, pharyngeal and laryngeal cavities of (1) newborn *Homo sapiens*, (2) adult chimpanzee, (3) *La Chapelle-aux-Saints* reconstruction, and (4) adult *Homo sapiens* (after Lieberman *et al.*, 1972)

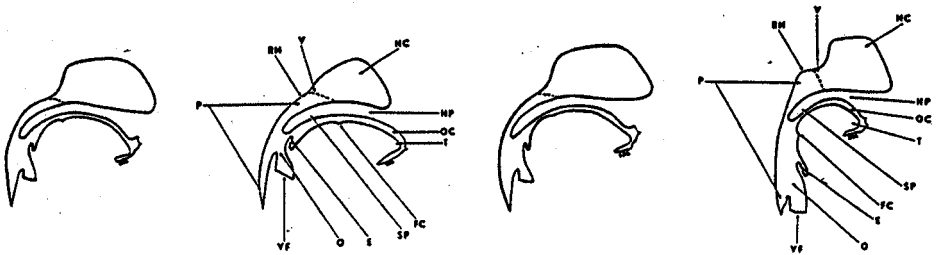
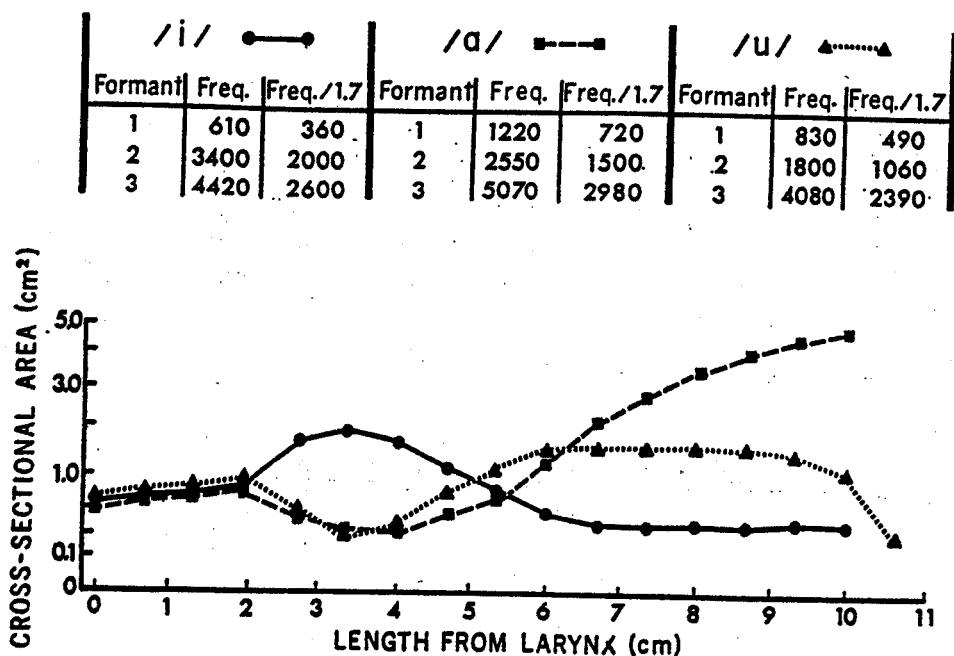


Figure 3. Diagrams of the air passages of (a) newborn human, (b) adult chimpanzee, (c) Neanderthal man, and (d) adult human. The anatomical details that are keyed on the chimpanzee and adult man are as follows: P = Pharynx, RN = Roof of Nasopharynx, V = Vomer Bone, NC = Nasal Cavity, HP = Hard Palate, OC = Oral Cavity, T = Tongue, FC = Foramen Cecum, SP = Soft Palate, E = Epiglottis, O = Opening of Larynx into Pharynx, VF = Level of Vocal Folds (after Lieberman *et al.*, 1972)

Figure 4. Chimpanzee supralaryngeal vocal tract area functions modeled on computer. These functions were the 'best' approximations that could be produced, given the anatomic limitations of the chimpanzee, to the human vowels [i], [a], and [u]. The formant frequencies calculated by the computer program for each vowel are tabulated and scaled to the average dimensions of the adult human vocal tract (after Lieberman et al., 1972)



the chimpanzee and newborn cannot generate supralaryngeal vocal tract area functions like those shown in Figure 7. These three configurations are the limiting articulations of a vowel triangle that is language universal (Troubetzkoy, 1939). It is not a question of the chimpanzee and newborn not being able to produce American-English vowels. They could not produce the vowel range that is necessary for any other language of *Homo sapiens*. Particular modern languages may lack one of these articulations, but they always include at least one of these vowels and/or the glides [y] and [w] which are functionally equivalent to [i] and [u]. It is important to remember that we are discussing the *phonetic* level rather than the *phonemic*. Claims that a particular language, e.g. Kabardian (Kuipers, 1960) has only 'one' centralized vowel generally concern the phonemic level, i.e., the claim is that a particular language does not

Figure 5. Plot of formant frequencies for chimpanzee vowels of Figure 4, data points (1), (2), and (3), scaled to correspond to the size of the adult human vocal tract. Data point (X) represents an additional point for human newborn. The closed loops enclose 90 percent of the data points derived from a sample of 76 adult men, women, and children producing American-English vowels (Peterson and Barney 1952). Note that the chimpanzee and newborn vocal tracts cannot produce the vowels [i], [u], and [a] (after Lieberman et al., 1972)

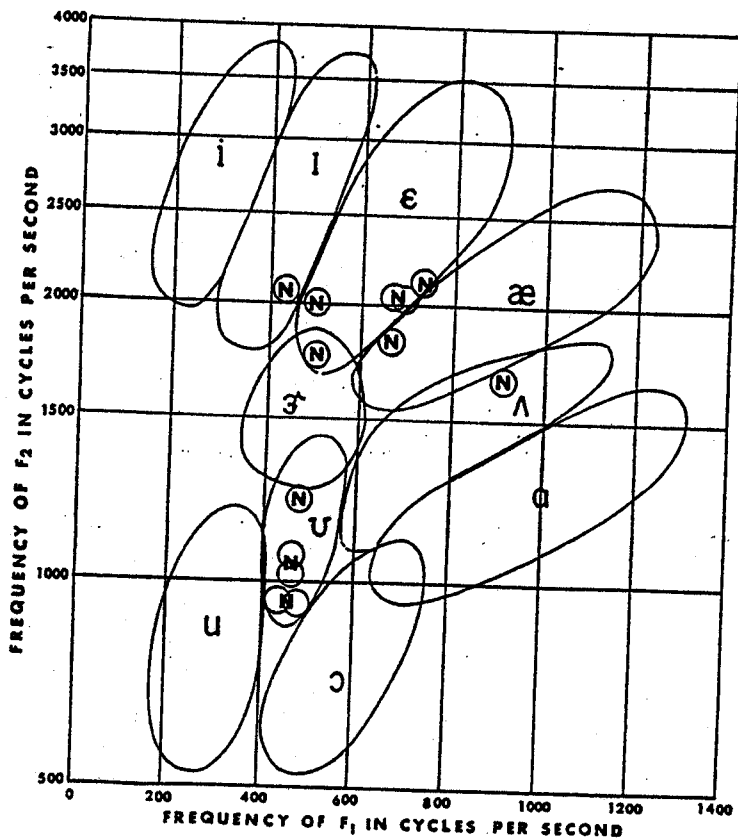
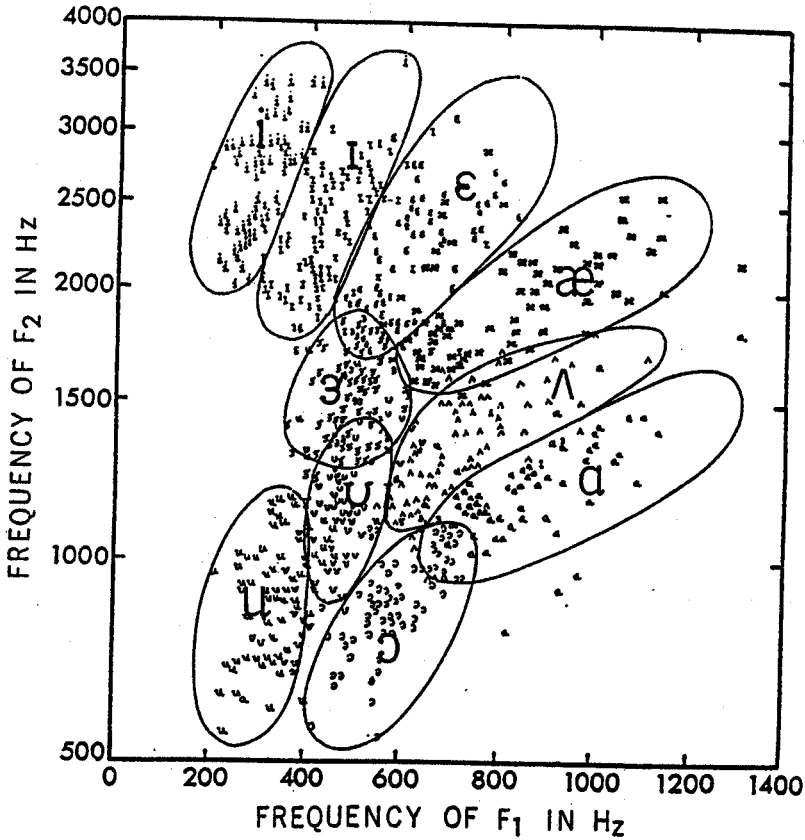
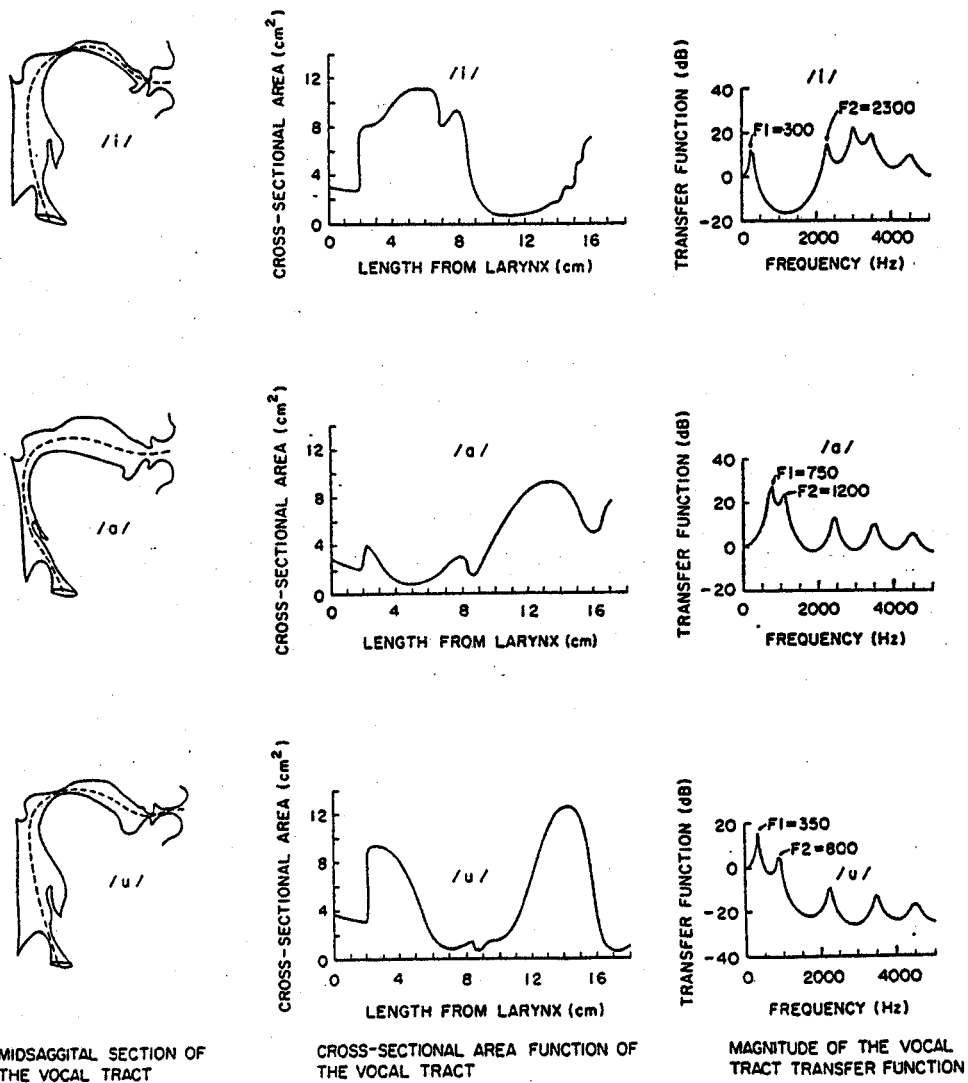


Figure 6. *Formant frequencies of American English vowels for a sample of 76 adult men, adult women and children. The closed loops enclose 90 percent of the data points in each vowel category (after Peterson and Barney, 1952)*



differentiate words at the phonemic level through vowel contrasts. At the phonetic level these languages make use of vowels like [i], [u], and [a] though these vowels' occurrences are conditioned by other segments. It is also important to note that a vocal tract that cannot produce the area functions necessary for [i], [a], and [u] also cannot produce velar consonants like [g] and [k]. These consonants also involve discontinuities at the midpoint of the supralaryngeal vocal tract. Dental and bilabial consonants like [d], [t], [b] and [p] are, however, possible.

Figure 7. Illustrations of approximate (a) midsagittal sections, (b) cross-sectional area functions, and (c) acoustic transfer functions of the vocal tract for the vowels [i], [a], and [u] (after Lieberman et al., 1972)



MIDSAGGITAL SECTION OF THE VOCAL TRACT

CROSS-SECTIONAL AREA FUNCTION OF THE VOCAL TRACT

MAGNITUDE OF THE VOCAL TRACT TRANSFER FUNCTION

Figure 7 shows a midsagittal outline of the vocal tract for the vowels [i], [a], and [u], as well as the cross-sectional areas of the vocal tract (Fant, 1960) and the frequency domain transfer functions for these vowels (Gold and Rabiner, 1968). Ten to one discontinuities in the area function at the vocal tract's midpoint are necessary to produce these vowels. It is possible to generate these discontinuities with the 'bent' adult human supralaryngeal vocal tract since the cross-sectional areas of the oral and pharyngeal cavities can be independently manipulated in adult humans while a midpoint constriction is maintained. The supralaryngeal vocal tract in adult humans thus can, in effect, function as a 'two tube' system. The lack of a supralaryngeal pharyngeal cavity in the direct airway from the larynx, at a right angle to the oral cavity, in chimpanzee and newborn humans restricts these forms to 'single tube' resonant systems. In adult humans, muscles like the genioglossus can pull the pharyngeal portion of the tongue in an anterior direction, enlarging the pharyngeal cavity while the oral cavity is constricted, as in the production of [i]. In the production of [a], in adult humans, the pharyngeal constrictors reduce its cross sectional area while the oral cavity is opened by lowering the mandible. It is impossible to articulate these extreme discontinuities in the chimpanzee and newborn supralaryngeal vocal tracts. They can only attempt to distort the tongue body in the oral cavity (see Figures 2, 3, and 4) to obtain changes in cross-sectional area. The intrinsic musculature and elastic properties of the tongue severely limit the range of deformations that the tongue body can be expected to employ. This is evident in cineradiographic observations of newborn cry and swallowing (Truby *et al.*, 1965), and baboon cries (Zhinkin, 1963), and the deformations of the oral and pharyngeal portions of the tongue in adult humans (Perkell, 1969).

Note that Figure 7 shows that the discontinuities in the [a], [i], and [u] are functions which occur at or near the midpoint of the supralaryngeal vocal tract. Stevens (in press) has shown that the midpoint area discontinuity has an important functional value. It allows human speakers to produce signals that are acoustically distinct with relatively sloppy articulatory maneuvers. The first and second formant frequencies are maximally separated for [i], maximally centered for [a], and maximally lowered for [u]. When a human speaker wants to produce one of these vowels it is not necessary for him (or her) to be very precise about the position of the tongue. All that is necessary is an area function discontinuity within one cm or so from the midpoint. The formant frequencies will not perceptibly vary² (Flanagan, 1955) when the discontinuity shifts plus or minus one cm from the midpoint. This would not be the case for similar articulations if they were generated at any point other than the midpoint of the vocal

2. Flanagan (1955) shows that human listeners are not able to discriminate stimuli that differ

solely with respect to a single formant frequency unless the difference exceeds 60 Hz.

tract. The vowels [a], [i], and [u] are thus optimal acoustic signals for communication. The speaker can produce maximally differentiated sounds without having to be terribly precise. All other vowels are both less distinct and less 'stable'. The speaker must be more precise to produce acoustic signals that are not as distinct and separable. This factor is germane to one of the points that I raised earlier: How precise does the reconstruction of the supralaryngeal vocal tract of an extinct hominid have to be to yield meaningful data? The answer is that we can derive useful information without having to reconstruct fine detail since the crucial factor is essentially the ability to generate area discontinuities at or near the midpoint.

6.1 *La Chapelle-aux-Saints*

In Figures 2 and 3 a silicone rubber model and a sketch of the supralaryngeal vocal tract of the La Chapelle-aux-Saints Neanderthal fossil are shown. It obviously was not possible to obtain this information directly from the soft tissue of this fossil hominid. The reconstruction of the supralaryngeal airways was effected by Edmund S. Crelin using the similarities that exist between this fossil and newborn human as a guide (Lieberman and Crelin, 1971; Lieberman *et al.*, 1972b). The possible arthritic condition (Straus and Cave, 1957) of the La Chapelle-aux-Saints fossil has been raised in some criticisms of Crelin's reconstruction. Arthritic changes could no more have affected his supralaryngeal vocal tract than is the case in modern man.

Figure 8 shows a lateral view of the skull, vertebral column, and larynx of newborn and adult *Homo sapiens* and the reconstructed La Chapelle-aux-Saints fossil. Note that the geniohyoid muscle in adult *Homo sapiens* runs down and back from the hyoid symphysis of the mandible. This is necessarily the case because the hyoid bone is positioned below the mandible in adult *Homo sapiens*. The two anterior portions of the digastric muscles, which are not shown in Figure 8, also run down and back from the mandible for the same reason. When the facets into which these muscles are inserted at the symphysis of the mandible are examined, it is evident that the facets are likewise inclined to minimize the sheer forces for these muscles. The human chin appears to be a consequence of the inclination of these facets. The outwards inclination of the chin reflects the inclination of the inferior plane of the mandible at the symphysis. Muscles are essentially 'glued' in place to their facets. Tubercles and fossae in this light may be simply regarded as adaptations that increase the strength of the muscle to bone bond by increasing the 'glued' surface area. The inclination of the digastric and geniohyoid facets likewise serves to increase the functional strength of the muscle to bone bond by minimizing sheer forces. As Bernard Campbell (1966, p. 2) succinctly notes, 'Muscles leave marks where they are attached to bones, and from such marks we assess the form and size of the muscles'. This is no less true for living than for

Figure 8. *Skull, vertebral column and larynx of Newborn (A), and adult Man (C), and reconstruction of Neanderthal (B). G-Geniohyoid Muscle, H-Hyoid Bone, S-Stylohyoid Ligament, M-Thyrohyoid Membrane, T-Thyroid Cartilage, CC-Cricoid Cartilage. Note that the inclination of the styloid process away from the vertical plane in Newborn and Neanderthal results in a corresponding inclination in the stylohyoid ligament. The intersection of the stylohyoid ligament and geniohyoid muscle with the hyoid bone of the larynx occurs at a higher position in Newborn and Neanderthal. The high position of the larynx in the Neanderthal reconstruction follows, in part, from this intersection (after Lieberman and Crelin, 1971)*

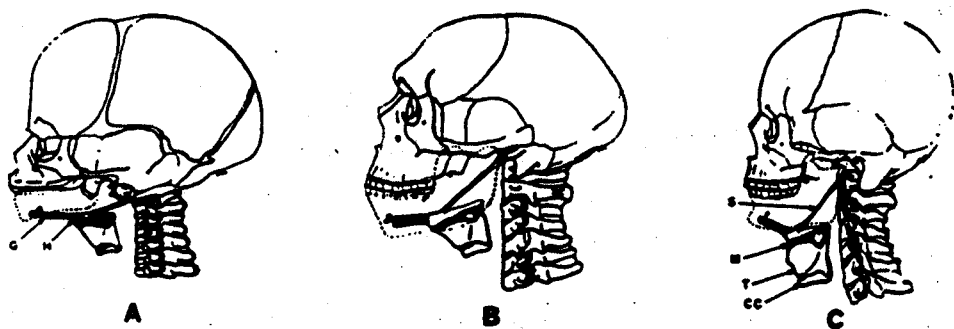


Figure 9. *Inferior views of base of skull of Newborn (A), Neanderthal (B), and adult Man (C). D-Dental Arch, P-Palate, S-Distance Between Palate and Foramen Magnum, V-Vomer Bone, BO-Basilar Part of Occipital, O-Occipital Condyle (after Lieberman and Crelin, 1971)*

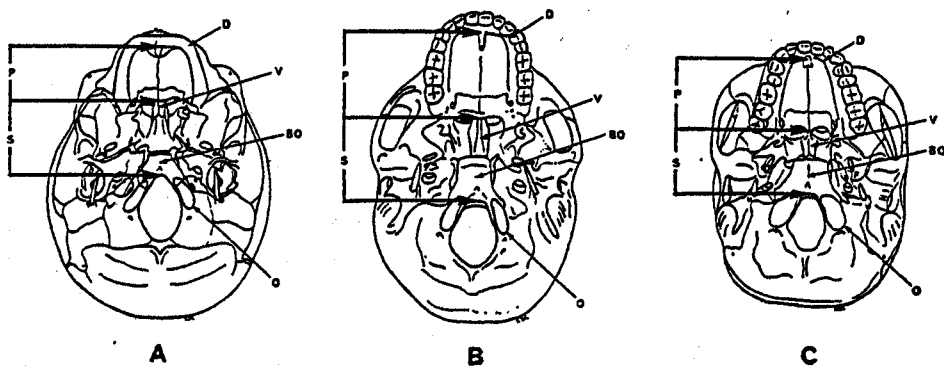
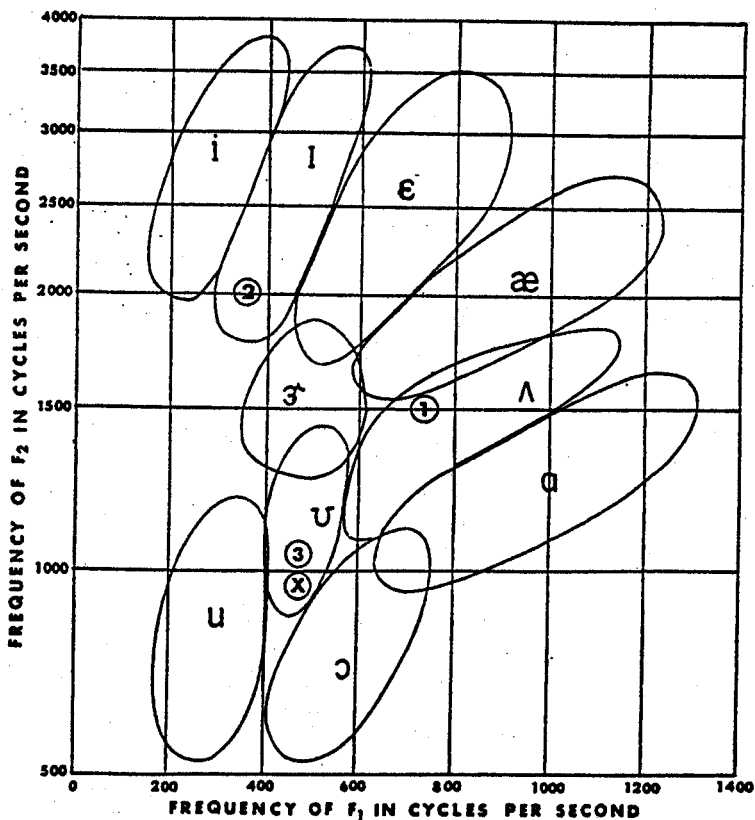


Figure 10. Plot of formant frequencies for reconstructed La Chapelle-aux-Saints supralaryngeal vocal tract in attempts to produce the vowels [i], [u], and [a]. Note that none of the data points (N) falls into the vowel loops that specify these vowels (after Lieberman and Crelin, 1971)



extinct forms. When the corresponding features are examined in newborn *Homo sapiens* (Figure 10) it is evident that the nearly horizontal inclination of the facets of the geniohyoid and digastric muscles is a concomitant feature of the high position of the hyoid bone (Crelin, 1969, pp. 107-110). These muscles are nearly horizontal with respect to the symphysis of the mandible in newborn *Homo sapiens*. The facets therefore are nearly horizontal to minimize shear forces. Newborn *Homo sapiens* thus lacks a chin.³ When the mandible of the La Chapelle-aux-Saints fossil is examined,

3. The human chin is sometimes stated to be a reinforcement for the mandible. This is

probably not the case. It more likely is a stress concentration point. It would be rather simple

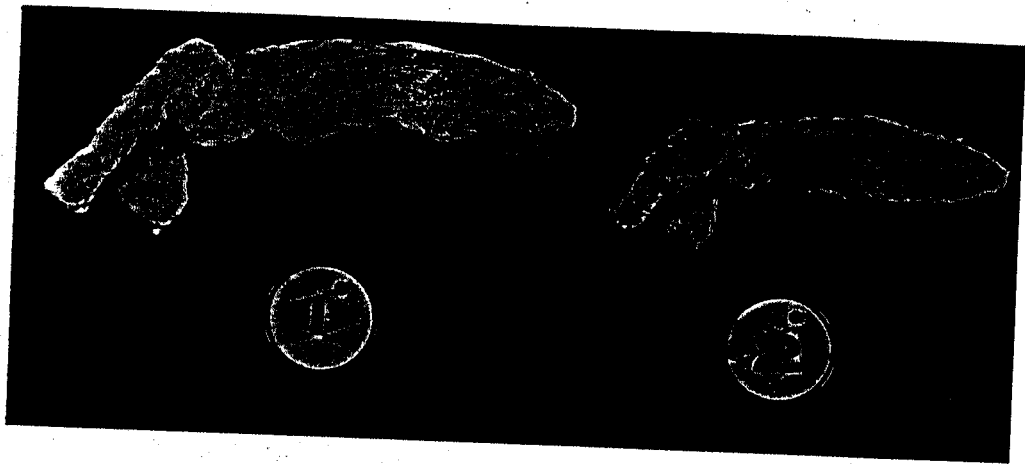


Figure 11. *Casts of the oral, pharyngeal, and laryngeal cavities of (a) Sterkfontein 5 reconstruction and (2) chimpanzee. Note that the supralaryngeal airways of the Australopithecine fossil and chimpanzee are almost identical except for their size. The nasal cavities have been omitted to make the similarities in these 'one tube' supralaryngeal vocal tracts more apparent.*

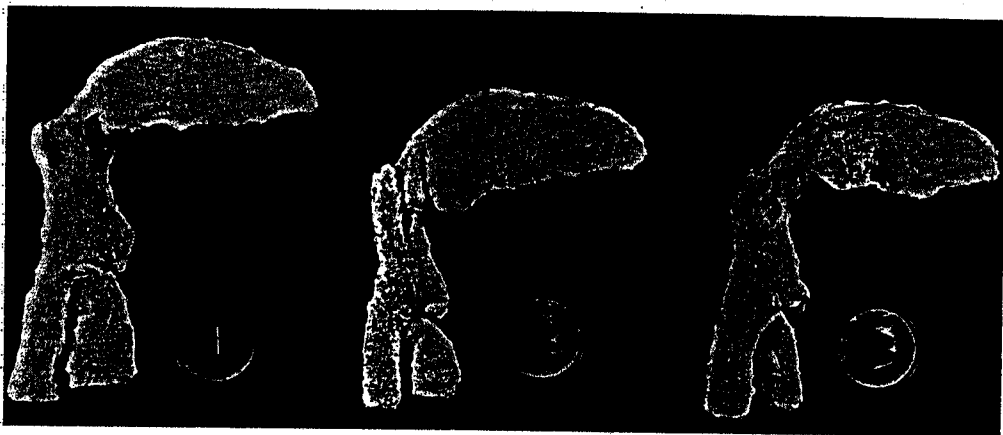


Figure 12. Casts of the oral, pharyngeal, and laryngeal cavities of (1) *Es-Skhul V* reconstruction, (2) *Steinheim* reconstruction, and (3) adult *Homo sapiens* (refer back to the key on Fig. 3 for anatomical details). Note that the two fossil reconstructions' supralaryngeal airways both have a right angle bend and a pharyngeal cavity similar to that of modern *Homo sapiens*.

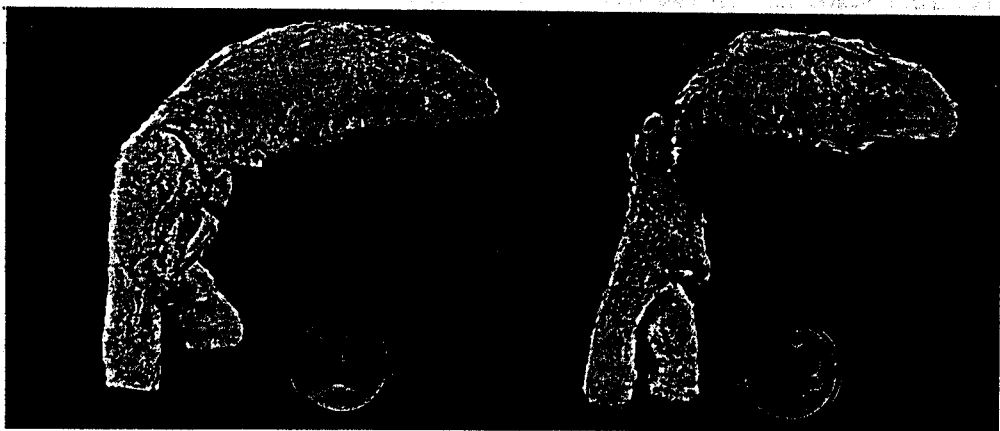


Figure 13. Casts of the oral, pharyngeal, and laryngeal cavities of (1) *Broken Hill* reconstruction, and (2) adult *Homo sapiens* (refer back to the key on Fig. 3 for anatomical details). Note that the fossil reconstruction's supralaryngeal airway is an intermediate form. It has a right angle bend but has a pharyngeal cavity smaller than that of adult, modern *Homo sapiens*.

it is evident that the facets of these muscles resemble those of newborn *Homo sapiens*. The inclination of the styloid process away from the vertical plane is also similar in newborn *Homo sapiens* and the La Chapelle-aux-Saints fossil. When the base of the skull is examined (Fig. 9) for newborn and adult *Homo sapiens* and the La Chapelle-aux-Saints fossil it is again apparent that the newborn *Homo sapiens* and fossil forms have many common features that differ from adult *Homo sapiens*. The sphenoid bone is, for example, exposed in newborn *Homo sapiens* and the La Chapelle-aux-Saints fossil between the vomer and the basilar part of the occipital. This is a skeletal feature that provides room for the larynx which is positioned high with respect to the mandible. There has to be room for the larynx behind the palate in newborn *Homo sapiens* and in the La Chapelle-aux-Saints fossil. The qualitative difference in the morphology of the base of the skull, i.e., the exposure of the sphenoid, is a skeletal consequence of this anatomical necessity.

We do not claim that all the features of the La Chapelle-aux-Saints fossil are found in newborn *Homo sapiens*. This is definitely *not* the case. We are claiming that certain features, particularly those relating to the base of the skull and mandible, are similar. These similarities make possible a reasonably accurate reconstruction of the supralaryngeal vocal tract of the La Chapelle-aux-Saints fossil. Our observations are in accord with the results of Viček's (1970) independent 'Onto-phylogenetic' study of the development of a number of fossil skulls of Neanderthal infants. Viček notes the presence of skeletal characteristics that are typical of both infant and adult Neanderthal fossils, that are manifested during particular phases of the ontogenetic development of contemporary man. Other features that characterize adult Neanderthal man never appear in the ontogenetic development of contemporary man, while still other features that characterize contemporary man never are manifested in the fossil skulls. I will return to this data when I discuss the status of classic Neanderthal man. For the moment, it is relevant as an independent replication of the similarities between newborn *Homo sapiens* and the La Chapelle-aux-Saints fossil. Crelin's reconstruction of the supralaryngeal vocal tract of this fossil is also in accord with earlier attempts like that of Keith, which is discussed by Negus (1949) as well as the inferences of Coon (1966).

In Figure 10 the vowel space of the reconstructed La Chapelle-aux-Saints supralaryngeal vocal tract is presented. Each of the data points (N) represents attempts to produce vowels like [a], [i], or [u]. The labelled loops again refer to the Peterson-Barney (1952) data for actual human vowels. Note that the vowel space of the fossil is a subset of the human vowel space and that it is impossible to produce the 'extreme' vowels [a], [i], and [u]. It is likewise impossible to produce the glides [y] or [w] or velar

consonants like [g] and [k]. The Neanderthal supralaryngeal vocal tract also probably is not capable of making nasal versus non-nasal contrasts. Everything will tend to be nasalized. The modelled Neanderthal vowel space is probably too large since we allowed articulatory maneuvers that would have been rather acrobatic in modern man (Lieberman and Crelin, 1971). We tried to err on the side of making this fossil's phonetic ability more humanlike whenever we were in doubt.

6.2 *Sterkfontein 5*

In Figure 11 a silicone rubber model of the airways of the reconstructed supralaryngeal vocal tract of the Sterkfontein 5 cranium (Mrs. Ples) is shown together with the chimpanzee airways that appeared in Figure 2. Note the similarities. Crelin's reconstruction follows from the similarities that exist between this fossil and present-day orangutan and to a lesser degree chimpanzee. The reconstructed vocal tract has the same phonetic limitations as present-day apes. The details of this reconstruction and the others that follow will be discussed in detail in a separate paper (Crelin *et al.*, forthcoming).

6.3 *Es-Skhül V and Steinheim*

In Figure 12 silicone rubber models of the reconstructed airways of the Es-Skhül V and Steinheim fossils are shown together with the supralaryngeal airways of adult *Homo sapiens*. Note that the reconstructed supralaryngeal airways both have right-angle bends, that the pharyngeal cavity is part of the direct airway out of the larynx, that both resemble the supralaryngeal airways of adult modern man. The reconstructed Es-Skhül V airway is completely modern. It would place no limits on its owner's phonetic repertoire if he attempted to produce the full range of human speech. The Steinheim supralaryngeal airway, though it has some pongid features, is also functionally equivalent to a modern supralaryngeal vocal tract. It would have placed no restrictions on its owner's phonetic repertoire if he attempted to produce the full range of human speech.

6.4 *Broken Hill (Rhodesian Man)*

In Figure 13 a silicone rubber model of the reconstructed supralaryngeal airways of Rhodesian man is shown together with a casting of the supralaryngeal airways of adult *Homo sapiens*. Note that despite the large oral cavity which follows from the large palate of this fossil, there is a right angle bend in the supralaryngeal airway. This vocal tract appears to be an intermediate form. When it is modelled it can

produce acoustic signals appropriate to the human vowels [a], [i], and [u] though the supralaryngeal vocal tract configurations that are needed are not as stabile, i.e. resistant to articulatory sloppiness, as equivalent human vocal tract configurations. Note that the large palate in this fossil form occurs with a bent supralaryngeal vocal tract. The reduction of the palate in forms like Steinheim, Es-Skhul V and modern *Homo sapiens* therefore cannot be the factor that caused the larynx to descend.

7. Significance of results

In Table 1 the results of the reconstructions and computer modelling so far discussed are presented together with the results that would be obtained for various fossils that are similar to the ones that we have examined. I have not attempted to list all the similar forms. Note that we have divided the table into two categories: Fossil hominids who had the anatomical specializations that are necessary for human speech and fossils who lacked these specializations.

Table 1.

- Human Supralaryngeal Vocal Tract		+ Human Supralaryngeal Vocal Tract
Australopithecines: africanus robustus bosei		
Neanderthal	Saccopastore I	Steinheim
	Monte Circeo	
	Teschik-Tasch (infant)	
	La Ferrassie I	
	La Chapelle-aux-Saints	
	La Quina (infant)	Broken Hill
	Pech-de-l'Azé	
	Solo II	Es-Skhul V Djebel Kafzeh
	Shanidar I	Cro-Magnon modern <i>Homo sapiens</i>

7.1 Neotony

The first point that I want to make is that the anatomy necessary for producing the full range of sounds necessary for human speech represents a particular specialization that, at the present time, occurs only in normal adult *Homo sapiens*. It is clear that

adult *Homo sapiens* does not particularly resemble newborn *Homo sapiens*.⁴ This is, in general, true of all primates (Schultz, 1968). The infantile forms of primates often do not resemble their adult forms. Schultz (1944, 1955), moreover, shows that the infantile forms of various non-human primates resemble newborn *Homo sapiens* whereas the adult forms of these non-human primates diverge markedly from adult *Homo sapiens*. This, however, does not mean that adult *Homo sapiens* has evolved by preserving neonatal features (Montagu, 1962), since it is apparent that modern man has his own unique specializations. The unique specializations of modern man include the anatomy necessary for the production of human speech. Table 1 shows that these specializations have evolved over at least the past 300,000 years and that until comparatively recent times, various types of hominids existed, some of whom lacked the anatomical mechanisms necessary for articulate human speech.

7.2 The 'Neanderthal problem'

Note that Table 1 places a number of fossil forms that lacked speech into a category labelled 'classic Neanderthal'. A view that has enjoyed some popularity in recent years is that Neanderthal fossils do not substantially differ from modern *Homo sapiens*, that they form a subset of hominids who have characteristics that grade imperceptibly with those typical of the modern population of *Homo sapiens*. An extreme formulation of this view is, for example, that '... no single measurement or even set of measurements can set Neanderthals apart from modern man' (Nett, 1973). In other words, that Neanderthal man can not be regarded as a separate species or even a separate variety distinct from *Homo sapiens*. This claim can be substantiated only if one includes fossils like Steinheim and Es-Skhül V in the same class as forms like La Chapelle-aux-Saints. Quantitative multivariate analysis like that of Howells (1968) demonstrates that fossils like La Chapelle and La Ferrassie form a class that is quite distinct from modern man. The measurements contained in Patte's (1955) comprehensive work as well as the observations of Vlček (1970) on the ontogenetic development of Neanderthal infants, indicate that this class of fossils, classic Neanderthal man, represents a specialization that diverged from the line (or lines) that are more direct ancestors of *Homo sapiens*. Fossils like Steinheim and Es-Skhül V, which are sometimes categorized as 'generalized' Neanderthal, are functionally distinct from classic Neanderthal. These fossils exhibit the anatomical specializations necessary for human speech.

4. Benda (1969) shows that Down's Syndrome (Mongolism) involves the retention of infantile morphology. Victims of this pathology, in some instances, retain the general proportions of the newborn skull. Their supralaryngeal vocal

tracts retain the morphology of the newborn and they are unable to speak. They strikingly demonstrate that *Homo sapiens* has not evolved by retaining infantile characteristics.

A general overlap between modern man and Neanderthal man is possible only if forms like Steinheim and Es-Skhül V are put into the same class as La Chapelle, La Ferrasie, Monte Circeo, etc. Hominids who could have produced human speech would have to be classified with hominids who could not have produced human speech. This would be equivalent to putting forms that had the anatomical prerequisites for bipedal posture into the same class as forms that lacked this ability.

The question immediately arises. Is this category, i.e., set of fossils, labelled 'classic Neanderthal' a separate species? It is useful to remember Darwin's definition of the term species. Darwin (1859, p. 52) viewed the term species, '... as one arbitrarily given for the sake of convenience to a set of individuals closely resembling each other, and that it does not essentially differ from the term variety, which is given to less distinct and more fluctuating forms'. Darwin later notes (1859, p. 485), '... the only distinction between species and well-marked varieties is that the latter are known, or believed, to be connected at the present day by intermediate gradations, whereas species were formerly thus connected'. It is evident that intermediate fossil forms like Broken Hill man bridge the gap between classic Neanderthal man and modern *Homo sapiens*. We do not know, and we probably never will be able to know, all the traits that may have differentiated various hominid populations that are now extinct. We do not, for example, know whether viable progeny would have resulted from the mating of forms like Cro Magnon and La Quina. Even if we did know that viable progeny would result from the mating of classic Neanderthal and early *Homo sapiens* populations, we would not necessarily conclude that these forms were members of the same species. The term species as Darwin noted is simply a labelling device. *Canus lupus* and *Canus familiaris* are considered to be separate species even though they may freely mate and have viable progeny. The behavioral attributes of wolves and dogs make it important for people, for example shepherds, to place these animals into different species, even though some dogs, e.g. Chihuahuas and St. Bernards, are more distinct morphologically, behaviorally and can't mate. The question of separate species labels for classic Neanderthal and other fossil hominid populations is thus probably not a crucial question. We simply can note that different types of hominids apparently coexisted until comparatively recent times and that some of these hominids do not appear to have contributed to the present human gene pool.

Table 1 does have some bearing on the apparent absence of the specializations typical of classic Neanderthal man (e.g. La Chapelle, La Ferrasie, etc.) in modern man. Animal studies (Capranica, 1965) have established the role of vocalizations in courtship and mating. The presence or absence of humanlike speech probably would have served as a powerful factor in assortative mating. In the present population of modern man it is evident that linguistic differences and affinities play a powerful role in mate selection. We would expect this phenomenon to be accentuated when different

hominid populations inherently were unable to produce the sounds of other groups. Sexual selection determined by speech patterns may thus have played a significant role in the divergence of groups like classic Neanderthal in Western Europe and the ancestral forms of modern *Homo sapiens*.

7.3 *The evolutionary sequence*

There is, unfortunately, a large gap in Table 1 since we have not yet been able to examine specimens of *Homo erectus* that have intact skull bases. It is, however, likely that the situation that typifies later hominid forms will also characterize *Homo erectus*. Most will probably not have the anatomy necessary for the production of the full range of human speech. Some forms, however, will undoubtedly be found that either had the necessary anatomy or that were intermediate forms. Evolution goes in small steps and forms intermediate between Steinheim and the Australopithecines must have existed. We still are, like Darwin, at the mercy of the 'Imperfection of the Geological Record'.

We can, despite this gap, draw several inferences from Table 1. I would like to propose the following evolutionary sequence. The first phase of the evolution of human language must have relied on a system of gestures, facial expressions and vocal signals like those of present-day apes to communicate the semantic, i.e. cognitive, aspects of language. The Australopithecines must have had cognitive abilities that surpassed those of present-day apes. Although early Australopithecines may have had cognitive abilities that were near the levels of present-day apes, late forms would have developed superior abilities as evolution continued step by step and mutations favoring larger relative brain sizes were retained. The retention of mutations leading to larger relative brain sizes is itself a sign that cognitive ability had a selective advantage. We can reasonably infer that activities like tool making and collective social enterprises like hunting were important attributes of Australopithecine culture.

Although the vocal apparatus of forms like *Australopithecus africanus* does not appear to differ significantly from those of present-day apes, vocal communications undoubtedly played a part in their linguistic system. Our reconstructions can tell us nothing about the larynx; however, it is almost certain that the laryngeal mechanisms of these forms was at least as developed as those of present-day apes. As Negus (1949) observed, there is a continual elaboration of the larynx as we ascend the phylogenetic scale in terrestrial animals. The larynges of animals like wolves are capable of producing a number of distinct calls that serve as vehicles of vocal communication. The same is true for the larynges of chimpanzees and gorillas. Studies like that of Kelemen (1948) which have attempted to show that chimpanzees cannot talk because of laryngeal deficiencies are not correct. Kelemen shows that the chimpanzee's larynx is

different from the larynx of a normal adult human male. The chimpanzee's larynx will not produce the range of fundamental frequencies typical of adult human males; however, it can produce a variety of sound contrasts. Many of these sound contrasts indeed occur in human languages. A present-day chimpanzee, if it made maximum use of its larynx and supralaryngeal vocal tract, could, for example, produce the following sound contrasts:

- a. *Voiced versus unvoiced*, i.e., excitation of the vocal tract by the quasi-periodic output of the larynx versus turbulent noise excitation generated by opening the larynx slightly and expelling air at a high flow rate.
- b. *High fundamental versus normal fundamental*, i.e., adjusting the larynx so phonation occurs in the falsetto register rather than the modal chest register (Van den Berg, 1960). The larynx has several modes of phonation which result in acoustic signals that are quite distinct. In falsetto the fundamental frequency is high and the glottal source's energy spectrum has comparatively little energy at its higher harmonics.
- c. *Low fundamental versus normal fundamental*, i.e., adjusting the larynx to a lower register. This lower register termed 'fry' produces very low fundamental frequencies (Hollien *et al.*, 1966) that are irregular (Lieberman, 1963).
- d. *Dynamic fundamental frequency variations*, e.g., low to high, high to low. Variations like these occur in many human tone languages.
- e. *Strident high-energy laryngeal excitation*, i.e., the high fundamental frequency, breathy output that can be observed in some chimpanzee vocalizations (Lieberman, 1968) as well as in the cries of human newborn (Lieberman *et al.*, 1972a).
- f. *Continuent versus interrupted*, i.e., the temporal pattern of laryngeal excitation can be varied. This can be observed in the calls of present day monkeys and apes (Lieberman, 1968).
- g. *Oral versus non-oral*, i.e., the animal can produce a call with his oral cavity sealed or with his oral cavity open. This can be observed in present-day gorilla where the low energy, low fundamental frequency sounds that sometimes accompany feeding appear to be produced with the oral cavity sealed by the epiglottis (Lieberman, 1968).
- h. *Lip rounding and laryngeal lowering*. Chimpanzees have the anatomic ability of rounding their lips and/or lowering their larynges while they produce a call. Both of these articulatory gestures could produce a formant frequency pattern that has falling transitions.
- i. *Flared lips and laryngeal raising*. Chimpanzees could either flare their lips and/or raise their larynges while they produced a call. This would generate a rising formant frequency pattern.
- j. *Bilabial closures and releases*. Sounds like [b] and [p] as well as prevoiced [b] (like

that occurring in Spanish, for example [Lisker and Abramson, 1964]) could be produced by controlling the timing between the opening and closing of the larynx and the lips.

- k. *Dental closures and releases.* Sounds like [d] and [t] (Lisker and Abramson, 1964) could be produced by varying the timing between a closure effected by the tongue blade against the alveolar ridge and the opening and closing of the larynx.

Australopithecines could have generated all of the above sound contrasts if they had the requisite motor control and the neural ability to perceive the differences in sound quality that are the consequences of these articulatory maneuvers. Most of these phonetic contrasts, i.e., 'features' (Jakobson *et al.*, 1952) have been observed in the vocal communications of present-day non-human primates. Present-day human languages make use of all of these sound contrasts. The combination of articulatory features like (j) and (k) and timing features like (f) could also generate sounds like [f], [v], [s], etc. It is quite probable that late Australopithecines and various forms of *Homo erectus* made use of these sound contrasts to communicate. The transference of patterns of 'automatized' behavior which was discussed early in this paper from activities like tool making and hunting would have facilitated the acquisition of the motor skills necessary to produce these sounds. The role of hunting would have placed a premium on communication out of the line of sight, communication that, furthermore, left the hunter's hands free.

The neural mechanisms necessary for the differentiation of these sounds appear to exist in present-day primates. Wollberg and Newman (1972), for example, have shown that squirrel monkeys (*Saimiri sciureus*) possess auditory receptors 'tuned' to one of the vocal calls that these monkeys make use of in their communications. Similar results have been demonstrated for frogs (Capranica, 1965). Although gestural communication (Hewes, 1971) undoubtedly played a more important role in the communications of these early hominids than is the case for modern man, I think that it is most unlikely that vocal communications also did not play an important role.

The crucial stage in the evolution of *human* language would appear to be the development of the 'bent' supralaryngeal vocal tract of modern man. Table 1 shows a divergence in the paths of evolution. Some hominids appear to have retained the communications system that was typical of the Australopithecines. A mixed system that relied on both gestural and vocal components. Other hominids appear to have followed an evolutionary path that resulted in almost total dependence on the vocal component for language, relegating the gestural component to a secondary 'para-linguistic' function. The process would have been gradual, following from the prior existence of vocal signals in linguistic communication.

As I have noted before, the bent supralaryngeal vocal tract that appears in forms like present-day *Homo sapiens*, and the Steinheim, Es-Skhül V, and Broken Hill

fossils, allows its possessors to generate acoustic signals that have very distinct acoustic properties and that are very easy to produce. These signals are, in a sense, optimal acoustic signals (Lieberman, 1970). If vocal communications were already part of the linguistic system of early hominids, then mutations that extended the range and the efficiency of the signalling process would have been retained in forms like Steinheim. The 'bent' supralaryngeal vocal tract is otherwise a burden for basic vegetative functions. It would not have been retained unless it had conferred an adaptive advantage. The initial adaptive value of the bent supralaryngeal vocal tract would have been its value in increasing the inventory of vocal signals and in, moreover, providing more efficient vocal signals.

The neural mechanisms necessary to perceive these new signals would, in all likelihood, have been available to hominids like Steinheim. Recent electrophysiological data (Miller *et al.*, 1972) shows that animals like Rhesus monkey, *Macaca mullata*, will develop neural detectors that identify signals important to the animal. Receptors in the auditory cortex responsive to an 200 Hz sinusoid were discovered after the animals were trained by the classic methods of conditioning to respond behaviorally to this acoustic signal. These neural detectors could not be found in the auditory cortex of untrained animals. The auditory system of these primates thus appears to be 'plastic'. Receptive neural devices can be formed to respond to acoustic signals that the animal finds useful. These results are in accord with behavioral experiments involving human subjects where 'categorical' responses to arbitrary auditory signals can be produced by means of operant conditioning techniques (Lane, 1965). They are also in accord with the results of classic conditioning experiments like those reported by Pavlov. The dogs learned to identify and to respond decisively to the sound of a bell, an 'unnatural' sound for a dog. The dog obviously had to 'learn' to identify the bell. Hominids like Steinheim who had the potential to make 'new' acoustic signals would also have had the ability to 'learn' to respond to these sounds in an automatized way. The plasticity of the primate auditory system would have provided the initial mechanism for 'learning' these new sounds.

Later stages in the evolution of human language probably involved the retention of mutations that had 'innately' determined neural mechanisms that were 'tuned' to these new sounds. By innately determined, I do *not* mean that the organism needs no interaction with the environment to 'learn' to perceive these sounds. The evidence instead suggests that humans are innately predisposed to 'learn' to respond to the sounds of speech. Experiments with 6-week old infants (Eimas *et al.*, 1971; Morse, 1971) show that they respond to the acoustic cues that differentiate sounds like [b] and [p] in the same manner as adults. These acoustic distinctions involve 10 msec differences in the timing of the delay between the start of the acoustic signal that occurs when a human speaker opens his lips and the start of phonation. It is most

improbable that 6-week old infants could 'learn' to respond to these signals unless there was some innate predisposition for this sound contrast to be perceived. This surely is not surprising. Human infants really do not 'learn' the complex physiologic maneuvers associated with normal respiration. They have built in 'knowledge'. The case for the neural mechanisms that are involved in the perception of human speech is not as simple as that for respiration. Some contact with a speech environment is necessary. Deaf children, for example, though they at first produce the vocalizations of normal children become quiet after six months of age (Lenneburg, 1967). Nottebaum (1970) shows similar effects in birds. Some aspects of the bird's vocal behavior are manifested even when the bird is raised in isolation. Other important aspects of the bird's vocal behavior develop only when the bird is exposed to a 'normal' communicative environment.

At some late stage, that is, late with respect to the initial evolution of the 'bent' supralaryngeal vocal tract, the neural mechanisms that are necessary for the process of speech encoding would have evolved. The human-like supralaryngeal vocal tract would have initially been retained for the acoustically distinct and articulatory facile signals that it could generate. The acoustic properties of sounds like the vowels [a], [i], and [u] and the glides [y] and [w], which allow a listener to determine the size of the speaker's supralaryngeal vocal tract, would have preadapted the communications system for speech encoding.

When a human listener hears a sound like the word *bat*, as it is produced by an intermediate sized supralaryngeal vocal tract, it is indeterminate. Ladefoged and Broadbent (1957), for example, show that a listener will perceive this sound as the word *bit* if he is led to believe that it was produced by a long vocal tract. The same listener will perceive the same sound as *but* if he is led to believe that it was produced by a small, i.e., short vocal tract. The listener, in effect, 'normalizes' the signal to take account of the acoustic properties of different sized vocal tracts. The listener responds as though he is interpreting the acoustic signal in terms of the articulatory gestures that a speaker would employ to generate the word. The perception of human speech is generally structured in terms of the articulatory gestures that underly the acoustic signal (Lieberman *et al.*, 1967). This process, as I noted earlier, is the basis of the encoding which allows human speech to transmit information at the rate of 20 to 30 segments per second. Signals like the vowels [a], [u], and [i] and the glides [y] and [w] are determinate in the sense that a particular formant pattern could have been generated by means of only one vocal tract using a particular articulatory maneuver (Stevens and House, 1955; Lindblom⁷ and Sundberg, 1969). A listener can use these vowels to instantly identify the size of the supralaryngeal vocal tract that he (or she) is listening to (Darwin, 1971; Rand, 1971). These vowels can indeed serve the same function in the recognition of human speech by computer. Gerstman (1967), for

example, derives the size of a particular speaker's vocal tract from these vowels to identify the speaker's other vowels. Without this information it is impossible to assign a particular acoustic signal into the correct vowel class. The computer, like a human, has to know the size of the speaker's supralaryngeal vocal tract. The process of speech encoding need not have followed the exact path that I have proposed. Other sounds, like [s] can provide a listener (or a computer) with information about the size of the speaker's vocal tract. As I noted before, the Australopithecines had the anatomical prerequisites for producing sounds like [s] so the process of speech encoding and the evolution of the human supralaryngeal vocal tract may have been coeval from the start. It is clear, however, that evolution goes by small steps and what we have in present-day man is a fully encoded speech system with a speech producing anatomy that is highly adapted to this function. Other, now extinct, hominids like classic Neanderthal man had speech producing anatomy that clearly was not as well adapted for speech encoding. It is, therefore, reasonable to conclude that speech encoding either was more rudimentary, or not present.

It is, however, important to conclude with the point that language does not necessarily have to involve the process of speech encoding and rapid information transfer. The remains of Neanderthal culture all point to the presence of linguistic ability. Conversely, birds may have the potential for rapid information transfer (Grenewalt, 1967); however, birds lack the cognitive ability that is also a necessary factor in language. It is most unlikely that birds could develop a complex language unless they also had larger brains.

Human language is the result of the convergence of many factors: Automatization, cognitive ability, and speech encoding. The particular form that human language has taken, however, appears to be the result of the evolution of the human supralaryngeal vocal apparatus. The supralaryngeal vocal apparatus that differentiates present-day *Homo sapiens* from all living animals thus is as important a factor in the late stage of hominid evolution as dentition and bipedal posture were in earlier stages.

REFERENCES

- Benda, C. E. (1969) *Down's syndrome, mongolism and its management*. New York, Grune and Stratton.
- Campbell, B. (1966) *Human evolution: An introduction to man's adaptations*. Chicago, Aldine.
- Capranica, R. R. (1965) *The evoked vocal response of the bullfrog*. Cambridge, Mass., M.I.T. Press.
- Chiba, T, and Kajiyama, M. (1958) *The vowel: Its nature and structure*. Tokyo, Phonetic Society of Japan.

- Coon, C. S. (1966) *The origin of races*. New York, Knopf.
- Crelin, E. W. (1969) *Anatomy of the newborn: An atlas*. Philadelphia, Lea and Febiger.
- Crelin, E. W., Lieberman, P., and Klatt, D. H. (In preparation) Anatomy and related phonetic ability of the Skhul V, Steinheim, and Rhodesian fossils and the Pleisianthropus reconstruction.
- Darwin, C. (1859) *On the origin of species* (Facsimile edition). New York, Atheneum.
- Darwin, C. J. (1971) Ear differences in the recall of fricatives and vowels. *Q. J. exper. Psychol.*, 23, 386-392.
- Eimas, P. D., Siqueland, E., R. Jusczyk, P., and Vigorito, J. (1971) Speech perception in infants. *Science*, 171, 303-306.
- Fant, G. (1960) *Acoustic theory of speech production*. The Hague, Mouton.
- Flanagan, J. L. (1955) A difference limen for vowel formant frequency. *J. acoust. Soc. Am.*, 27, 613-617.
- Gardner, R. A. and Gardner B. T., (1969) Teaching sign language to a chimpanzee. *Science*, 165, 664-672.
- Gerstman, L. (1967) Classification of self-normalized vowels. *Proceedings of IEEE conference on Speed Communication and Processing*, 97-100.
- Gold, B. and Rabiner, L.R. (1968) Analysis of digital and analog formant synthesizers. *IEEE-Trans. Audio Electroacoustics, AU-16*, 81-94.
- Goodall, J. v. L. (1971) *In the shadow of man*. New York, Dell.
- Grenewalt, C. A. (1967) *Bird song: Acoustics and physiology*. Washington, D.C., Smithsonian Institution.
- Henke, W. L. (1966) Dynamic articulatory model of speech production using computer simulation, Ph.D. dissertation, (appendix B). M.I.T.
- Hewes, G. W. (1971) *Language origins: A bibliography*. Dept. of Anthropology, Boulder, Colo., Univ. of Colorado.
- Hollien, H., Moore P., Wendahl, R. W., and Michel, J.F. (1966) On the nature of vocal fry. *J. Speech Hearing Res.*, 9, 245-247.
- Howells, W. W. (1968) Mount Carmel man: Morphological relationships. *Proceedings of the VIIIth International Congress of Anthropological and Ethnological Sciences*, Vol. I. Anthropol. Tokyo.
- Irwin, O. C. (1948) Infant speech: Development of vowel sounds. *J. Speech Hearing Dis.*, 13, 31-34.
- Jakobson, R., Fant, C. G. M., and Halle, M. (1952) *Preliminaries to speech analysis*. Cambridge, Mass., M.I.T. Press.
- Kelemen, G. (1948) The anatomical basis of phonation in the chimpanzee. *J. Morph.*, 82, 229-256.
- Kempelen, W. R. von (1791) *Mechanismus der menschlichen Sprache nebst der Beschreibung seiner sprechenden Maschine*. Vienna, J. B. Degen.
- Kenyon, K. W. (1969) *The sea otter in the eastern Pacific Ocean*. U.S. Gov't. Printing Office.
- Kimura, D. (1964) Left-right differences: The perception of melodies. *Q. J. exper. Psychol.*, 16, 355-358.
- Kirchner, J. A. (1970) *Pressman and Kelemen's physiology of the larynx* (Rev. Ed.). Rochester, Minn., Amer. Acad. Ophthal. and Otolaryngol.
- Kuipers, A. H. (1960) *Phoneme and morpheme in Kabardian*. The Hague, Mouton.
- Ladefoged, P. and Broadbent, D. E. (1957) Information conveyed by vowels. *J. acoust. Soc. Amer.* 29, 98-104.
- Lane, H. (1965) Motor theory of speech perception: A critical review. *Psychol. Rev.*, 72, 275-309.
- Lenneburg, E. H. (1967) *Biological foundations of language*. New York, Wiley.
- Lieberman, A. M., Cooper F.S., Shankweiler, D. P., and Studdert-Kennedy, M. (1967) Perception of the speech code. *Psychol. Rev.*, 74, 431-461.
- Lieberman, P. (1963) Some acoustic measures of the periodicity of normal and pathologic larynges. *J. acoust. Soc. Amer.* 35, 344-353.
- Lieberman, P. (1967) *Intonation, perception, and language*. Cambridge, Mass, M.I.T. Press.
- Lieberman, P. (1968) Primate vocalizations and human linguistic ability. *J. acoust. Soc. Amer.*, 44, 1574-1584.
- Lieberman, P. (1970) Towards a unified phonetic theory. *Ling. Inq.*, 1, 307-322.
- Lieberman, P. (1972) *The speech of primates*.

- The Hague, Mouton.
- Lieberman, P., Klatt, D. H. and Wilson, W. A. (1969) Vocal tract limitations on the vowel repertoires of rhesus monkey and other nonhuman primates. *Science*, 164, 1185-1187.
- Lieberman, P. and Crelin, E. S. (1971) On the speech of Neanderthal man. *Ling. Inq.*, 2, 203-222.
- Lieberman, P., Harris, K.S., Wolff, P. and Russell, L. H. (1972a) Newborn infant cry and nonhuman primate vocalizations, *J. Speech Hear. Res.*, 14, 718-727.
- Lieberman, P., Crelin, E. S. and Klatt, D. H. (1972b) Phonetic ability and related anatomy of the newborn, adult human, Neanderthal man and the chimpanzee. *Amer. Anthropol.*, 74, 287-307.
- Lindblom, B. and Sundberg, J. (1969) A quantitative model of vowel production and the distinctive features of Swedish vowels. *Speech transmission laboratory report 1*. Stockholm, Sweden, Royal Institute of Technology.
- Lisker, L. and Abramson, A. S. (1964) A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, 20, 384-422.
- Lynip, A. W. (1951) The uses of magnetic devices in the collection and analysis of the preverbal utterances of an infant. *Genet. Psychol. Mono.*, 44, 221-262.
- Manly, R. S. and Braley, L. C. (1950) Masticatory performance and efficiency. *J. Dent. Res.*, 29, 448-462.
- Manly, R. S. and Shiere, F. R. (1950) The effect of dental deficiency on mastication and food preference. *Oral Surg., Oral Med., Oral Path.*, 3, 674-685.
- Manly, R. S. and Vinton, P. (1951) A survey of the chewing ability of denture wearers. *J. dent. Res.*, 30, 314-321.
- Miller, J. M., Sutton, D., Pflingst, B., Ryan, A. and Beaton, R., (1972) Single cell activity in the auditory cortex of rhesus monkeys: Behavioral dependency. *Science*, 177, 449-451.
- Montagu, M. F. A. (1962) Time, morphology and neotny in the evolution of man. In M. F. A. Montagu (Ed.), *Culture and the evolution of man*. New York, Oxford University Press. p. 324-342.
- Morse, P. (1971) Speech perception in six-week old infants. Unpublished Ph.D dissertation, University of Connecticut Storrs, Conn.
- Negus, V. E. (1949) *The comparative anatomy and physiology of the larynx*. New York, Hafner.
- Nett, E. G. (In press) A note on phonetic ability. *Amer. Anth.*
- Nottebohm, F. (1970) Ontogeny of bird song. *Science*, 167, 950-956.
- Patte, E. (1955) *Les neanderthaliens, anatomie, physiologie, comparaisons*. Paris, Masson et Cie.
- Perkell, J. S. (1969) *Physiology of speech production: Results and implications of a quantitative cineradiographic study*. Cambridge, Mass., M.I.T. Press.
- Peterson, G. E. and Barney, H. L., (1952) Control methods used in a study of the vowels, *J. acoust. Soc. Amer.*, 24, 175-184.
- Premack, D. (1972) Language in chimpanzee?, *Science*, 172, 808-822.
- Rand, T. C. (1971) Vocal tract size normalization in the perception of stop consonants. *Haskins Laboratories status report on speech research*, SR-26/26, 141-146.
- Schultz, A. H. (1944) Age changes and variability in gibbons. *Amer. J. Phys. Anth.*, n.s. 2, 1-129.
- Schultz, A. H. (1955) The position of the occipital condyles and of the fact relative to the skull base in primates. *Amer. J. Phys. Anth.*, n.s. 13, 97-120.
- Schultz, A. H. (1968) The recent hominoid primates. In S. L. Washburn and P. C. Jay (Eds.) *Perspectives on human evolution I*. New York, Holt, Rinehart and Winston. Pp. 122-195.
- Shankweiler, D. and Studdert-Kennedy, M. (1967) Identification of consonants and vowels presented to left and right ears. *Q. J. exper. Psychol.*, 19, 59-63.
- Stevens, K. N. (In press) Quantal nature of speech. In E. E. David and P. B. Denes (Eds.) *Human communication: A unified view*. New York, McGraw Hill.
- Stevens, K. N. and House, A. S. (1955) Development of a quantitative description of vowel articulation. *J. acoust. Soc. Amer.*, 27, 484-493.
- Straus, W. L., Jr. and Cave, A. J. E. (1957)

- Pathology and posture of Neanderthal man. *Q. Rev. Biol.*, 32, 348-363.
- Troubetzkoj, N. S. (1939) *Principes de phonologie* (Trans. J. Cantineau, 1949) Paris, Klincksieck.
- Truby, H. M., Bosma, J. F., and Lind, J. (1965) *Newborn infant cry*. Uppsala, Almqvist and Wiksells.
- Van den Berg, J. W., (1960) Vocal ligaments versus registers. *Curr. Prob. Phon. Logoped.*, 1, 19-34.
- Vlček, E. (1970) Étude comparative ontogénétique de l'enfant du Pech-de-l'Azé par rapport à d'autres enfants néandertaliens. In D. Farembach *et. al.* (Eds.) *L'enfant du Pech-de-l'Azé*, Paris, Masson et cie. Pp. 149-186.
- Wollberg, Z. and Newman, J. D. (1972) Auditory cortex of squirrel monkey: Response patterns of single cells to species-specific vocalizations. *Science*, 175, 212-214.
- Zhinkin, N. I. (1963) An application of the theory of algorithms to the study of animal speech - Methods of vocal intercommunication between monkeys. In R. G. Bushel (Ed.) *Acoustic Behavior of Animals*. Amsterdam, Elsevier Publishing.

Résumé

Le langage peut être défini opérationnellement comme un système de communication permettant l'échange d'information nouvelle, non anticipée. Diverses sortes de langages semblent avoir existé à des stades antérieurs de l'évolution des hominiens. Le langage *humain* est unique, actuellement, dans la mesure où il utilise le discours vocal 'encodé' pour permettre une transmission rapide de l'information. Le canal vocal supralaryngal du moderne *Homo sapiens* est un facteur utile de ce processus d'encodage qui met en jeu également des mécanismes nerveux spéciaux. D'autres facteurs, tels que la capacité cognitive et l'automatisation, sont également nécessaires au langage. Ces facteurs sont, toutefois, importants pour divers aspects du comportement humain et non humain en dehors du langage.

L'évolution du langage semble avoir été un processus graduel qui mena d'abord à des systèmes qui reposent sur un mélange de communication gestuelle et vocale. Certains hominiens semblent avoir conservé ce système jusqu'à une époque relativement récente. D'autres semblent avoir donné beaucoup plus

d'importance à la communication verbale. Des reconstructions de canaux vocaux supralaryngaux fossiles montrent que certaines formes, les Australopithèques et l'homme de Neanderthal 'classique', ne possédaient pas le canal vocal supralaryngal qui est nécessaire à la production de la parole humaine complètement encodée. D'autres formes fossiles, celles de Steinheim et de Es-Skhu V, possèdent des canaux vocaux fonctionnellement modernes. D'autres, comme celle de Broken Hill, représentent des formes intermédiaires. L'évolution du langage humain peut être conçue comme un processus en trois stades qui impliqua (a) une importance croissante de la communication vocale dans des activités telles que la chasse, (b) un accroissement du répertoire vocal avec l'évolution du canal vocal supralaryngal humain qui produit des signaux acoustiques à la fois plus distincts et plus résistants aux erreurs articulatoires, et (c) l'évolution des mécanismes nerveux qui font usage des propriétés préadaptées du canal vocal supralaryngal en vue de la communication verbale encodée rapide.