Primate Vocalizations and Human Linguistic Ability

PHILIP LIEBERMAN

University of Connecticut, Storrs, Connecticut 06268
and
Haskins Laboratories, New York, New York 10017

Some representative vocalizations of captive rhesus monkey, chimpanzee, and gorilla were recorded and analyzed by means of sound spectrograms and oscillograms. It was found that these animals' vocal mechanisms do not appear capable of producing human speech. The laryngeal output was breathy and irregular. A uniform cross-section, tracheal configuration appeared to underlie all the vocalizations. These animals did not modify the shape of their supralaryngeal vocal tracts by means of tongue maneuvers during a vocalization. Formant transitions occurred in some vocalizations, but they appeared to have been generated by means of laryngeal and possibly velar or lip movements. The nonhuman primates lack a pharyngeal region like man's, where the cross-sectional area continually changes during speech. The data suggest that speech cannot be viewed as an overt act function that makes use of a vocal tract that has evolved solely for respiratory and deglutitious purposes; the skeletal evidence of human evolution shows a series of changes from the primate vocal tract that may have been, in part, for the purpose of generating speech. Articulate speech may not have been fully developed in some of man's ancestors. The study of the peripheral speech-production apparatus of a fossil is thus may be useful in the assessment of its phylogenetic grade.

INTRODUCTION

We are accustomed to speak about the "vocal tract" when we refer to the articulatory apparatus that is used to produce human speech. One of the most common statements about speech production, however, is that it is an "overlaid" function insofar as it involves the manipulation of muscles and structures that have evolved for the purposes of eating and breathing. According to this view, there is, strictly speaking, no "vocal tract." Man has a set of devices that have evolved so that he can eat and breathe. He has happily been able to make use of this set of breathing and eating devices to communicate. This view, which is rather pervasive, stems from the anatomical and philosophical studies of the seventeenth and eighteenth centuries.

In the first half of the seventeenth century, Descartes developed his concept of the "bête machine," i.e.,

2 Other factors in the evolution of man's vocal tract are also sometimes cited, e.g., erect posture and man's visual acuity, which reduced the importance of the sense of smell, is cited by Negus as the reason for the degeneration of the ability of the epiglottis to seal the mouth off from the rest of the respiratory system.

animals are machines or mechanisms. In contrast to all other animals, only man possesses abstract thought and language. The basis for this distinction between man and all other animals appeared to rest solely on man's mental abilities, since animals like the apes appeared to have all the output mechanisms that are necessary for speech. Studies like Ferrault's and Tyson's comparative anatomies of the chimpanzee showed that the larynx, teeth, lips, and jaws of the nonhuman primates were similar to those of man. However, the monkeys and apes lacked speech and language and they could not think in abstract terms. They lacked the mechanism for abstract thought, that is, they lacked language.

Since Descartes, many people have tried to show that there is no fundamental difference between man and the animals. La Mettrie, in L'Homme Machine, stated

5 C. Ferrault, Mémoires Pour Servir à l'Histoire Naturelle des Animaux (L’Imprimerie Royale, Paris, 1676).
6 E. Tyson, Orang-outang. Sieve Homo Sapiens?: or, the Anatomy of a Pygmy Compared With That of a Monkey, an Ape, and a Man (Thomas Bennet and Daniel Brown, London, 1699). (The 1730 edition is available on microfilm from the Library of Congress.)
that man also was a machine. La Mettrie says that the
apes are, in effect, retarded people. Since they have the
necessary mechanism for speech production, La Mettrie
believed that with a little effort it should be possible to
teach an ape to talk. If an ape were carefully tutored as
though he were, for example, a deaf child, it would be
possible to teach him to speak. The ape would then, in
La Mettrie’s terms, “...be a perfect little gentleman.”
The belief that apes have a speech output mechanism
that would be adequate for speech production has
persisted to the present time. Osgood, for example, states
that, “the chimpanzee is capable of vocalizations almost
as elaborate as man’s.” Yerkes and Learned identify
more than 32 speech sounds for the chimpanzee.
Attempts to teach chimpanzees to talk continue.
A recent study by Hayes, for example, centered about
an attempt to teach a chimpanzee to talk by raising
it as though it were a retarded child. No one, however,
had ever been able to teach an ape to talk.

The object of this study is thus to examine the cries
of nonhuman primates in order to determine what as-
pects of their vocalizations are similar to human speech
and what aspects are different. In particular, we would
like to determine the articulatory and anatomical bases
of the differences so that we can tell with greater cer-
tainty the direction in which human speech-producing
capability has evolved from these related animals, the
apes and monkeys. In so doing, we may be able to gain
some insights into the evolution of man’s linguistic
abilities by comparing these animals with the skeletal
remains of man’s ancestors. These questions are, of
course, relevant to whether speech is an overlay func-
tion, and we, of course, should be able to determine
whether it is inherently possible to teach an ape to
produce human speech.

I. PROCEDURE

Vocalizations of captive 2- and 3-year-old gorillas
(Gorilla gorilla), 2-year old chimpanzees (Pan), and 1-
to 6-year old rhesus monkeys (Macaca mulatta) were
recorded. The range of vocalizations recorded for the
rhesus monkeys was judged by their keepers to be char-
acteristic of the animals’ range. The ape vocalizations
were judged by their keepers to be characteristic of a
good part of these animals’ “public” range. The vocali-
zations furthermore are consistent with those reported
by Rowell and Hinde for captive rhesus monkey and
by Andrew for captive rhesus monkey and chimpanzee.

These two studies made use of tape recordings and spec-
trographic analysis, so we have a reasonable basis for
comparison. The cries also appear to be consistent with
subjective transcriptions of ape vocalizations in their
natural environment.

Tape recordings were made in the monkey colony of
the University of Connecticut at Storrs, at the Central
Park and Prospect Park Zoos in New York City, and
at the Fairmount Park Zoo in Philadelphia. Sony type
TC 800 tape recorders were used with Sony type F85
and General Radio type 1560 P-5 microphones at a tape
speed of 7.5 in./sec. The microphones were placed 5-25
cm from the monkeys. The microphone-to-mouth dis-
tances for the apes ranged from 5 cm to 8 m. The upper
limit of the system’s frequency response was 12 kHz.

Sound spectrograms of these recordings were made,
using a Voiceprint sound spectrograph. Some of the
tape recordings were analyzed at half-speed and one-
quarter speed to increase the effective bandwidth of the
spectrograph’s analyzing filter. The effective band-
widths of the analyzing filters thus ranged from 50 to
1200 Hz. Oscillograms were also made, using a Honey-
well Visicorder.

II. RESULTS

In Fig. 1, a wide-band spectrogram is presented of a
vocalization of Gorilla Kathy, who is 3 years old and
lives in Philadelphia. The gorilla was producing a signal
at moderate intensity when food was withheld. The
bandwidth of the analyzing filter was 300 Hz, and the
spectrogram was made using the FLAT position of the
spectrograph, since there is more high-frequency energy

The Journal of the Acoustical Society of America 1575
Fig. 2. Spectrogram of cry produced at high level of intensity by 2-year-old chimpanzee. The bandwidth of the analyzing filter was 300 Hz. Note the presence of voicing "striations" during the transitions in the initial part of the cry. The fundamental frequency of phonation is 140 Hz. The transitions thus must reflect changes in the length of the supralaryngeal vocal tract. (After reduction of Figure, scale is now ~1.03 in. = 1 sec.)

The chimpanzee's vocalizations are also the case for the chimpanzee and monkey vocalizations that are described next.

The fundamental frequency of phonation was, however, rather unstable. Large pitch perturbations occurred. The laryngeal output appears to be very noisy and turbulent. Energy concentrations can be noted in Fig. 1 at 500, 1500, and 2400 Hz. Measurements of the skull and mandible of an adult gorilla yield an estimated vocal tract length of 17.8 cm. If a gorilla thus uttered the schwa vowel, that is, a vowel having a vocal tract shape that approximates a uniform tube open at one end, the vocal tract frequency would be at 470, 1414, and 2353 Hz. We can therefore infer that the energy concentrations in the spectrogram of gorilla Kathy's vocalization reflect the transfer function of her supralaryngeal vocal tract in the schwa configuration.

The last third of this spectrogram shows some effects that may resemble interaction of the supralaryngeal vocal tract with the activity of the vocal cords. We will return to these effects later. The main characteristic of this utterance is that the output of the gorilla's larynx is being modified by the resonances of the supralaryngeal vocal tract, as is the case for human speech. Note that this is in sharp contrast to the calls of birds, where the fundamental frequency and harmonics of the syrinx's output completely characterize the acoustic nature of the cry.

In Fig. 2, a cry uttered by a 2-year-old chimpanzee is presented. The bandwidth of the spectrograph's analyzing filter was 300 Hz. The two energy concentrations at 1500 and 2800 Hz occur after the initial part of the vocalization where transitions to and from 1300 to 1700 to 1300 Hz and from 2800 to 3200 to 2500 Hz occur.

Note the presence of clearly defined "voicing striations" during these transitions. The fundamental frequency of phonation as determined from the oscillogram is 240 Hz. The energy concentrations that can be seen in Fig. 2 thus must reflect the transfer function of the chimpanzee's supralaryngeal vocal tract. The transitions in the initial part of Fig. 2 must reflect changes in the overall length of the chimpanzee's vocal tract, since both F1 and F2 rise and fall together.

In American English, changes in the overall length of the supralaryngeal vocal tract usually are the result of lip rounding. It is possible, however, to change the length of the vocal tract by moving the larynx up or down. Infants do this in their birth cries and in some languages, e.g., certain dialects of French, laryngeal motion is a normal distinctive articulatory gesture. Since the chimpanzee's lips were retracted while he uttered this cry, he probably moved his larynx upwards and downwards to change the length of his supralaryngeal vocal tract during the early part of the utterance in Fig. 2.

Energy concentrations occurred at 1500, 2800, and 4500 Hz during the steady-state portion of this cry. The length of a chimpanzee vocal tract was estimated at 12 cm from an adult skull and mandible. The resonances of a uniform 12-cm tube open at both ends are 1400, 2800, and 4200 Hz. If the chimpanzee's vocal tract looked like a uniform tube, open at both ends, we would expect to find the energy concentrations that are apparent in Fig. 2. The boundary condition looking back at the subglottal system from the chimpanzee's larynx would thus have to be similar to the boundary condition at the chimpanzee's lips for this open tube approximation to hold. The chimpanzee's glottal opening would therefore have to be large during the cry for this to be true. This may be what is happening. Kelemen, in his

---

the shape of the supralaryngeal vocal tract seems to approximate a uniform tube, that is, the schwa vowel. In all of the analyses of the ape cries that we recorded, the acoustic signal indicated that the supralaryngeal vocal tract configuration approximated either a tube of uniform cross section or a slightly flared tube. The data sample is admittedly small, as it involves only six captive apes, but the cries recorded are consistent with Andrew's spectrographic investigation. 32

In several subjective studies of ape cries, 3,12 transcriptions like /aw/ are used for certain cries. This, of course, implies that the ape is moving its tongue during the cry since this is what human speakers do when they produce the diphthong /aw/. In Fig. 4, a spectrogram is presented of a chimpanzee cry that sounds like /aw/. The cry was uttered at a high degree of vocal effort. The fundamental frequency was unstable. The oscillogram showed that it varied about 200 Hz, but the excitation was, in general, very breathy. The sound spectrogram in Fig. 4 was made with an analyzing filter bandwidth of

---

**Fig. 4.** Spectrogram of loud chimpanzee cry that sounds like diphthong /aw/. The bandwidth of the analyzing filter was 600 Hz. Note the decrease in the high-frequency energy of the glottal excitation that is apparent in the change in density in the second, third, and fourth "hairs" towards the end of the cry. Note that there are no transitions involving the first and second formants. The ape's tongue thus did not change the shape of this supralaryngeal vocal tract. (After reduction of Figure, scale is now ~1.03 in. = 1 sec.)

**Fig. 5.** Spectrogram of aggressive "bark" of rhesus monkey. The bandwidth of the analyzing filter was 1200 Hz. Formant frequencies occurred at 1.5, 3, and 6-8 kHz. (After reduction of Figure, scale is now ~1.03 in. = 1 sec.)

---

600 Hz by analyzing the tape recording at half-speed. The time scale of the spectrogram is thus stretched. Note that the energy in the higher formants decreases towards the end of the cry. Note, however, that the formants continue to be evenly spaced at the end of the cry. There is no transition in either the first- or the second-formant frequencies. The apparent change in vowel quality in this cry is due to a change in the energy content of the glottal excitation, rather than to a change in the configuration of the supralaryngeal vocal tract. In other words, the cry sounds like /aw/ because the energy balance between the higher and the lower formant frequencies shifts to the lower formant frequencies as the spectral content of the glottal excitation shifts towards the end of the cry. Similar though less marked changes in the spectral content of the glottal excitation are also at the end of a human cry. If the animal is in the inspiratory configuration while the subglottal air pressure is maintained, vocal sound is produced at a moderate degree of vocal effort while the monkey, closed his teeth. We recorded two young monkeys over a period of 6 months in the monkey colony of the University of Connecticut at Storrs. This particular recording was made with the Sony type F85 microphone. In Fig. 6, part of the oscillogram of this cry is presented. The first two "bursts" are presented in the oscillogram that was made as the tape recording was played back at one-quarter speed. The fundamental frequency of phonation is approximately 400 Hz. Note that the fundamental periodicity is very unstable at best. Parts of the waveform appear to be very turbulent. The waveform, in all, looks very much like those associated with pathologic human larynges where a hoarse vocal output results. The rhesus monkeys, like the gorillas and chimpanzees, are unable to produce sustained vocalizations that have a steady fundamental periodicity.

The spectrogram in Fig. 5 was also made from a tape recording played back at one-quarter speed. The effective bandwidth of the spectrograph was thus 1200 Hz. The fundamental frequency of phonation was about 400 Hz. There were approximately 25 msec between each burst and glottal activity seems to have been sustained between bursts. Thus, the cry is similar to a sequence of voiced stops in intervocalic position.

Unlike voiced stops in human speech, the closure of the vocal tract seems to have been effected by the animal's epiglottis and velum. The monkey's lips were retracted, exposing his teeth throughout the cry, so he could not have used his lips to close his vocal tract. There is also evidence that the vocal tract of the monkey is in a narrow configuration. If the epiglottis is considered to be closed and the monkey is quiet, it is in contrast to the position of the human vocal tract, and the epiglottis can seal the mouth open at the end phase of a typical phonation.

Note that this cry is quite different from that of the macaque monkey. The energy concentration at 2.5 and 3.5 kHz was much higher, and the resonant frequency of the vocal tract was open at one end. We anesthetized a 5-year-old male monkey and measured the length of his supralaryngeal vocal tract. With his lips rounded, the length of his supralaryngeal vocal tract was 7.6 cm. The resonances of a 7.6-cm-long tube open at one end are 1100, 3300, and 5500 Hz. We recorded a number of cries that this monkey made with his lips rounded at a low level of vocal effort. The recordings were made in a quiet room using the General Radio 1560-FS microphone. The average values of F1, F2, and F3 were 1300, 3000, and 4200 Hz, respectively. Thus, the monkey was producing these cries with a slightly flared supralaryngeal vocal tract.

In Fig. 7, photographs of a casting of the oral cavity of a rhesus monkey are presented. The monkey's tongue...
and lips were positioned in an approximation of an aggressive "bark"\textsuperscript{21} and a plaster-of-Paris casting was made shortly after an experiment in which the monkey was sacrificed (for other purposes). Note that the vocal tract of the monkey approximates a uniform cross section passage with a flared portion at the laryngeal end. Also note the shallowness of the pharyngeal "bend" and the flatness of the monkey's tongue, which is apparent in the side view. (The monkey's tongue fills up the shallow section delimited by the depth of the "bend" at the laryngeal end of the oral cavity.)

In Fig. 8, a distress cry of a rhesus monkey is presented. This cry was produced at an extremely high level of vocal effort. The monkey had its lips retracted. It was clinging to the back of another monkey at the rear of its cage. The bandwidth of the spectrograph filter was 1200 Hz, since the tape recording was analyzed at one-quarter speed. Note the energy concentrations in the initial part of this vocalization. The supraglottal vocal tract length of this monkey is about 4 cm when his lips are fully retracted and he is anesthetized. The lowest energy concentration ranges from 4 to 4.3 kHz during the initial part of the cry. This frequency range is consistent with a 4-cm vocal tract tract length resonating as a half-wave resonator. The next highest energy concentration ranges from 8.5 to 9 kHz during the initial part of the cry. This too is consistent with the second resonance of a uniform tube that has similar boundary conditions at both ends. There is no low-frequency fundamental frequency apparent in the initial part of this cry. Examination of the oscillogram shows energy present only at the two resonances of the supralaryngeal vocal tract. We have here a case where the resonances of the supralaryngeal vocal tract apparently control the excitation function. In other words, the resonances of the vocal tract determine the energy components of the laryngeal excitation. The system is behaving like a trumpet where the resonances of the trumpet determine the rate at which the musician's lips vibrate. Similar though smaller effects have been noted during normal human speech where the vocal cords can be seen to vibrate at the first formant frequency.\textsuperscript{27} Flanagan has observed similar effects\textsuperscript{28} in a model of the human larynx. In the gorilla cry in Fig. 1, energy can be seen at multiples of the first formant during the last third of the spectrogram. The abrupt "bars" may be caused by the vocal cords of the gorilla vibrating at the first formant frequency. Similar effects also seem to occur in human speech from dysarthric subjects.\textsuperscript{19}

Returning to Fig. 8, note that the energy concentrations at the end of the cry are at 2.5 and 7 kHz. These frequencies are consistent with a flared 4-cm tube resonating as a quarter-wave resonator. Note that there is less energy in the high part of the spectrum at the end of the cry. The oscillogram also shows low-frequency energy and a general noisillike to quasiperiodic nature where the fundamental frequency is about 600 Hz. The monkey's vocal tract apparently resembles a half-wave resonator during the initial part of the cry where the higher subglottal air pressure produces a large glottal opening. During the final part of the cry, the lower subglottal air pressure probably results in a smaller average glottal opening that results in resonances in the quarter-wave mode.

The cry in Fig. 8, which we have been discussing, is consistent with the distress cries\textsuperscript{20,41} recorded by Andrew\textsuperscript{20} and by Rowell and Hinde.\textsuperscript{41} Andrew, for example, notes that these cries have no energy below 2 kHz in contrast to the other cries he recorded. This is probably

\textsuperscript{21} One last comment should be made concerning these distress cries. When one listens to these cries played back at their normal speed, they sound like birdlike chirps. There is no sense of fright or terror. However, when the signals are scaled down in frequency by playing the tapes at one-quarter speed, the cries convey extreme fright. Darwin's theory [C. Darwin, The Expression of Emotion in Man and Animals (J. Murray, London, 1872)] of the innate expression of emotion through the vocal mechanism would seem to be correct. We do not normally interpret the rhesus monkey's fright cries correctly when we hear them, because the output mechanism of the rhesus monkey produces an acoustic signal that is scaled up in frequency relative to our range. When we hear the cry scaled down in frequency we are able to interpret it. The motor controls to the rhesus monkey's vocal mechanism when it expresses fright are probably similar to the motor controls that we would use, but the acoustic signal has been scaled up in frequency because the rhesus monkey's much smaller. When we play back the tape at one-quarter speed, we match the acoustic signal to our own perceptual recognition routines, which appear to be structured in terms of the constraints of our own speech-production mechanism; see also Ref. 31.


\textsuperscript{19} H. I. Soron, Air Force Cambridge Research Laboratories, has noted these effects in high-speed motion pictures of the human vocal cords during phonation.


\textsuperscript{28} K. S. Harris, Haskins Laboratories, New York (private communication).
area of the pharynx varies, for example, over a ten-to-one range for the vowels /a/ and /i/. The vowel /a/ is produced with a small pharyngeal cross section, while the /i/ is produced with a large cross section. These variations in pharyngeal cross-sectional areas are characteristic for consonants as well as vowels and they are essential in the production of human speech.

In Fig. 10, a semidigrammatic representation of the nose, palate, tongue, pharynx, and larynx of a monkey, an ape, and man are reproduced from Negus’ Comparative Anatomy and Physiology of the Larynx. Note the relative positions of the palate and larynx. The basis for the nonhuman primates’ lack of tongue mobility appears to be anatomical. The pharyngeal region, which can vary its shape in man, has no real counterpart in these animals. Their larynges are positioned quite high compared to the human larynx, almost in line with the roof of the palate. The tongues of these animals are thin compared to man’s. The nonhuman primates do not have a pharynx where the root of a thick tongue forms a movable anterior wall. Zhinkin, for example, in a cineradiographic study of baboon cries, shows that the baboon can not vary the size of his pharynx. The tongues of the nonhuman primates are long and flat and their supralaryngeal vocal tracts cannot assume the range of shape changes characteristic of human speech.

In Fig. 11, we have reproduced a second illustration from Negus. Negus’ notes that there has been a continuing set of changes in the evolution of the upper respiratory system. He notes, for example, that Neanderthal man has “…no large pharyngeal resonator, as in modern man” and that “…the gap between the palate and the epiglottis has increased during evolutionary changes to that of modern man” (p. 195). If one examines the skulls of earlier hominoid fossils, like the one that Dart referred to as Australopithecus prometheus, the similarity between these phylogenetically primitive hominids and present-day nonhuman primates is quite striking.

—J. Bastian, “Primate Signaling Systems and Human Languages,” in Primate Behavior: Field Studies of Monkeys and Apes, I. Devore, Ed. (Holt, Rinehart and Winston, Inc., New York, 1965) notes that the auditory signals of humans and primates have certain similarities, but that the “…disposition of the parts of the auditory filtering system remains rather stable during species and period. Most of the departures from the relatively stable and open configurations occur at the beginnings of signals and appear to be most often due to the involvement of open parts at the very front (the lips) or the very back.”

Fig. 9. Schematicized view of the human oral and pharyngeal region. Note the relative thickness of the tongue. The anterior wall of the pharynx in man is formed by the tongue and the cross-sectional area of this back cavity can vary over a ten-to-one range. A variable pharyngeal region is essential for the production of back vowels and consonants.

apparent. A plaster cast of the reconstructed skull of *Australopithecus prometheus* was compared with a chimpanzee skull. The over-all lengths of both skulls were approximately 18 cm. The shape of the mandible and the palate, and the position of the foramen magnum relative to the palate, were very similar. The mandibles of both the chimpanzee and *Australopithecus* left room for only a relatively thin tongue. The length of the supralaryngeal vocal tract was estimated to be approximately 12 cm for both specimens. Insofar as both vocal tracts would reflect the gross skeletal similarities that exist between *Australopithecus* and a modern chimpanzee, they both would lack a variable pharyngeal area. *Australopithecus* is thus in line with the evolutionary changes in the pharyngeal region that Negus notes. The earlier the fossil, the smaller the pharyngeal region is. *Australopithecus prometheus*, in all likelihood, could not have produced human speech, since his vocal apparatus, insofar as we are able to make deductions from fossil remains, appears to be quite similar to those of present-day apes and monkeys. Man's closer ancestors may or may not have been able to produce the full range of human speech. Vallois, 24 in his survey of skeletal evidence, cites the difficulties that have beset past attempts to infer the presence or absence of speech from anatomical arguments. These past difficulties were due primarily to the lack of a quantitative acoustic theory of speech production, so that the interpretation of the acoustic consequences of anatomical structures was tenuous at best.

The evolution of the vocal tract thus reflects, in part, its role in speech production. Speech cannot be simply regarded as an overlaid function that makes use of a mechanism that has evolved solely for the purposes of eating and breathing. The apes and monkeys have no difficulty in either breathing or eating. In fact, they have better breathing systems than ours. The monkeys and most apes can, for example, seal their mouths off from the rest of the respiratory system because the high position of the larynx in these animals allows the epiglottis to close the mouth. The results of comparative anatomy indeed demonstrate that the role of the epiglottis is to close the mouth.1, 2 A dog or a monkey can breathe while its mouth is full of food or water. The low position of the larynx in man also leads to difficulties when food is lodged in the larynx. This often can have fatal consequences. In no sense is the human larynx optimal for the purposes of respiration. Negus, 3 for example, also shows that in contrast with the larynges of animals like the horse, the human larynx impedes the flow of air during respiration. Whereas the maximum opening of the larynx during respiration in a horse is greater than the area of the tracheas, in man the maximum laryngeal area is only half of the tracheal area.

The cries of the nonhuman primates also differ from human speech insofar as the output of the larynx is less periodic. The glottal period often varies from one period to the next. In other words, large pitch perturbations 4 often occur. At high levels of effort, the output of the larynx often appears to be quite noisy and the waveform of the acoustic signal recorded from the animal resembles a tuned circuit excited by bursts of wide-band noise. In some of the very loud cries, the laryngeal output seems to be coupled to the resonances of the supralaryngeal vocal tract and energy appears to be present only at the formant frequencies. Although similar interactions between the laryngeal output and the supralaryngeal vocal tract occur in human speech, they play a very small part in normal speech production.

The differences between the nonhuman and human laryngeal output again appear to have an anatomical basis. Kelemen, 11, 16–18 in a series of detailed anatomical studies, has noted a number of differences between human and nonhuman larynges, although the larynges of all primates are superficially similar. One difference, which probably accounts, in part, for breathy excita-

tion, is the presence of what Kelemen terms the "hlatus intervocalis" in the primate larynx. The animal's larynx cannot be completely adducted, and a glottal shunt always exists. The differences in the outputs of the primate and human larynx are of course due to differences in the ability to control the larynx by making fine adjustments of the tensions of the laryngeal muscles. Nevertheless, it is clear that these differences are connected with vocalization rather than with respiration.

A. Linguistic Abilities of Apes and Monkeys

It is not always clear from the acoustic data of this paper what the monkeys and apes were doing when they produced their cries. The acoustic analysis indicates that some of the cries were made while the animal changed the over-all length of its supralaryngeal vocal tract by either lip rounding or by moving his larynx up and down. Some of the interrupted cries seem to have been made by the animal's closing his epiglottis and/or his velum, and in some of the cries, the vocal tract may have been behaving more like a half-wave, rather than a quarter-wave resonator. What is clear, however, from the acoustic and anatomical data is (1) that these animals do not move their tongues during a cry, and (2) that the laryngeal output tends to be aperiodic. These animals could not produce human speech even if they had the requisite mental ability. Their vocal apparatus is not adapted for the production of human speech; they lack laryngeal control and a pharyngeal region capable of varying its cross-sectional area.

B. Evolution of the Human Vocal Mechanism

Although we cannot say anything about the larynx of man's immediate ancestors, we can see the evolution of a variable "pharyngeal area" from the skeletal evidence. Insolar as the presence of an output mechanism is a necessary condition for human language, and insolar as the phonologic features have an abstract as well as a physical basis in language, we can say that the earliest hominid beings did not have language. The evolution of the vocal tract seems to move consistently towards the mechanism of modern-day man as we proceed from one phylogenetic grade to the next. We do not find any "puzzling regressions" like those that occur when one attempts to correlate "cosmetic" aspects of anatomy like brow ridges with phylogenetic grade.

C. Man's Acquisition of Language

It is not clear exactly when language came into being. If the arguments advanced by Dart concerning the

---

long static period in the development of human culture are valid, then the acquisition of speech may be comparatively recent. Dart claims that man's culture was static from the time of the *Australopithecines* to about 250,000 years ago. If the level of culture were an index of whether language was necessary or not, we could conclude that either all hominids from the *Australopithecines* onwards had speech, or that all these hominids lacked speech. Since *Australopithecus prometheus* probably did not have the ability to produce speech, we would therefore conclude that none of these hominids had speech. However, Dart's statement about the presence of a static culture until 250,000 years ago fails to take into account changes in toolmaking techniques that occurred in the lower paleolithic era, that is about 100,000 years ago. Dart also probably places too much importance on the presence of language in the development of culture. Although it is quite likely that an accelerated pace of cultural change at some period reflects the presence of language, which extends man's powers of abstract thought, other factors probably are always involved. Barnett,44 for example, notes the effects of intercultural contact on cultural change. The presence of language may therefore be only a necessary rather than a sufficient factor in man's cultural development. If we take the level of culture above the *Australopithecines* baseline as an index of the presence of language, it is clear that human speech was already present by the upper paleolithic era.

We cannot, on the basis of skeletal evidence, tell exactly when human speech first appeared. We cannot, for example, state with certainty whether Neanderthal man, who is a comparatively recent hominid, could or could not have articulated the full range of human speech. This is because we cannot determine the relationship between skeletal structure and soft tissue with the detail that would be necessary to justify a positive conclusion. We also do not even know the range of vocal tract dimensions that holds for the modern man, nor do we know to what extent small differences in the vocal tract are mirrored in the acoustic signal. Other motor skills that we cannot infer from skeletal evidence, like the ability to execute the rapid, controlled articulatory maneuvers that are typical for consonants, are also necessary for the production of speech. Thus, while we can say with reasonable certainty that older fossil hominids did not possess human speech, the nearer the vocal apparatus of a fossil is to that of modern man, the greater is our uncertainty regarding his ability to produce human speech. This uncertainty merely reflects the fact that, at the present time, we know some of the factors that are necessary for the production of human speech.

speech, but we do not know what peripheral mechanisms and central controls would be sufficient for the production of human speech.

D. Cries versus Language

The acquisition of language probably was an abrupt thing that came when the number of calls and cries that could be made with the available vocal mechanism increased to the point where it was more efficient to code features. We can speculate at a process in which the ability to make more and more cries gradually developed. The differentiated cries allowed the species to compete more successfully and mutations that led to the ability to make more cries were therefore retained. At some point, the number of different cries that could be made increased markedly; perhaps the mobility of the pharynx increased to the point where the phonologic features of back tongue position and high tongue position could be produced. If the computational abilities of the species were sufficiently advanced, it would have been efficient to recode the phonologic features leading to an arbitrary relationship between sound and meaning.

The difference between a system of cries, even though it may be highly developed, and a language, is that the relationship between meaning and sounds is fixed for cries. A high pitched /a/, for example, may be the cry for pain. It always "means" pain no matter what sounds precede or follow it. In contrast, the sound /a/ in a language may have no meaning in itself, nor might the sounds /ma/ or /an/ in isolation. The sound sequence /ma/ does have a particular semantic reference or meaning in English while the sound sequences /ma/ and /an/ have other meanings. Language essentially involves a two-level process where it is necessary to interpose the rules of a grammar and a dictionary between the sound sequence and its meaning.

E. Matching of Speech Production and Speech Perception

The two necessary conditions for the presence of speech and language, an output mechanism and central mental ability, may have developed together. Certainly everything that comes to mind about language seems to show this kind of optimization between output and internal computation. The simultaneous evolution of a mechanism for the production of speech and of man's mental ability would, for example, account for the close relationship that we find between speech production and speech perception. It would have been "natural" and "economical" for the constraints of speech production to be structured into the speech perception system if both of these abilities developed at the same pace. We would thus expect to find the speech recognition routines that involved a match with the constraints of speech production (the motor theory of speech perception) to be structured into a speech perception center that would be species specific, rather than in the peripheral or central auditory systems, which probably are similar for man and other animals.

V. Conclusion

Man has apparently developed special modifications of his vocal tract for the purposes of speech production. Just as an ability to use tools depends, in part, on having an opposable thumb and an upright posture, the ability to talk depends on our having a mouth, tongue, larynx, and pharynx that are adapted towards speech production. Speech production is thus not an overlaid function that makes use of a mechanism that has evolved solely for the purposes of feeding and breathing. The apes and monkeys lack the adaptations that are essential for the production of human speech and they obviously have no difficulties in either breathing or eating. Human infants, in a sense, begin at the same point as the nonhuman primates. They do not move their tongues during a cry for the first weeks of life. By the sixth week of life, however, they begin to change the configurations of their supralaryngeal vocal tracts during a vocalization.

The nonhuman primates never reach this point, though their general mental ability and physical dexterity are equivalent to, or better than, a human infant's at this age. Man's remote ancestors also lacked the output mechanism that is necessary for the production of speech and man may have acquired speech and speech adapted mechanisms at a comparatively recent time. We cannot say very much about the evolution of the central mechanisms that are necessary for speech and language, but looking at the "speech" abilities of present-day monkeys and apes gives us some insights into the nature, the evolution, and the acquisition of man's linguistic ability.

ACKNOWLEDGMENTS

Thanks are due to Ronald Ellis and Kenneth Welch of the Prospect Park Zoo, Brooklyn, New York, and John Fitzgerald of the Central Park Zoo, New York City, and Dr. W. A. Wilson of the University of Connecticut for their aid in recording their charges. Special thanks are also due to Miss Carolyn Ristau of the University of Pennsylvania who made her tape recordings of apes available. The acoustic analysis has also profited from the special insights of Dr. Franklin S. Cooper of Haskins Laboratories, while the interpretation of anthropological theories has benefited from the advice of Dr. D. J. Nash of the University of Connecticut and Dr. L. L. Eiseley of the University of Pennsylvania.

This research was supported in substantial part by grants from the National Institute of Child Health and Human Development and the National Institute of Dental Research.

---

48 Newborn human infants begin by making cries in which their tongues are immobile. They then start by making cries that are similar to those made by the nonhuman primates. See the forthcoming study by P. Lieberman, K. S. Harris, and P. Wolff, "Newborn Infant Cry in Relation to Nonhuman Primate Vocalizations," J. Acoust. Soc. Amer. 45, 369 (A) (1968).