Vocal Tract Limitations on the Vowel Repertoires
of Rhesus Monkey and other Nonhuman Primates*

Philip Lieberman†
Haskins Laboratories, New York City
and
Dennis L. Klatt‡‡
and
William A. Wilson‡‡‡

Abstract. The vowel repertoire of a rhesus monkey (Macaca Mulatta) was explored by means of a computer program that calculated formant frequencies from the area function of the animal's supralaryngeal vocal tract, which was systematically varied within the limits imposed by anatomical constraints. The resulting vowels were compared with human vowels and recorded vocalizations of nonhuman primates. The computer model indicates that the acoustic "vowel space" of a rhesus monkey is quite restricted compared to the human range. This limitation is due to an anatomical feature of nonhuman primates, the lack of a pharyngeal region that can change its cross sectional area. These animals thus lack the output mechanism that is necessary for the production of human speech. Man's speech output mechanism is apparently species specific.

*Scheduled for publication in Science, 164, 1185-1187.

†Also, University of Connecticut
‡‡ Research Laboratory of Electronics, Massachusetts Institute of Technology
‡‡‡University of Connecticut
In a recent study (1) vocalizations of captive rhesus monkey, chimpanzee, and gorilla were recorded and analyzed by means of sound spectrograms and oscillograms. The acoustic analysis suggested that these animals lack the ability to produce the articulatory maneuvers that are necessary to produce the full range of human speech. The general assumption that the vocal mechanisms of these animals are sufficiently well developed to permit the articulation of words (2) would thus be wrong.

Human speech is essentially the product of a source (the larynx for vowels) and a supralaryngeal vocal tract transfer function. The supralaryngeal vocal tract, in effect, filters the source (3). The activity of the larynx determines the fundamental frequency of the vowel while its formant frequencies are the resonant modes of the supralaryngeal vocal tract transfer function. The formant frequencies are determined by the area function of the supralaryngeal tract (3). The vowels /a/ and /i/, for example, have different formant frequencies though they may have the same fundamental frequency. The object of this study is to extend the acoustic analysis (1) that indicated that the nonhuman primates' vocalizations are restricted to schwa-like cries produced by means of a supralaryngeal vocal tract with a cross section that is uniform along its length. (An example of the schwa is the first vowel in the word about.) Our acoustic analysis was perforce limited to the sounds that the animals actually uttered. Our present method makes use of a computer implemented model of the supralaryngeal vocal tract of a rhesus monkey (Macaca Mulatta) that we systematically manipulated. We thus were able to explore the full range of vowels that a rhesus monkey could produce if he exploited all the degrees of freedom of his supralaryngeal vocal tract. Our analysis of the possible range of monkey vocal-
izations thus can be independent of the restrictions inherent in the analysis of a limited set of actual utterances. There is, of course, no guarantee that a monkey will in fact use all of the articulatory maneuvers that we simulate. Itani (4) for example, reports that wild Japanese monkeys seldom use their lips during cries though they are physically able to move their lips. However, we can explore the inherent limits of the output device.

**Procedure.** A plaster casting was made of the oral cavity of a monkey soon after death. The monkey's tongue and lips were positioned in an approximation of an aggressive "bark" (5). The plaster casting was then sectioned at intervals of 0.5 cm and the cross sectional areas were determined by weighing paper tracings of the sections on an analytical balance. This area function is presented as the solid line in Fig. 1.

The acoustic waveform corresponding to a vowel can be regarded as the output of a vocal tract filter system which is excited by vocal cord vibrations. It is the transfer function of the vocal tract that determines the vowel uttered because the volume velocity waveform at the vocal cords (the source) and the radiation impedance at the lips (the load) are relatively constant during vowel production, and independent of the particular vowel.

The (frequency domain) transfer function of the vocal tract is determined by an area function which gives the cross-sectional area of the vocal tract as a function of position along the tract. For frequencies of interest, the vocal tract behaves as a linear system satisfying the one-dimensional wave equation (6). A closed form solution to the wave equation for arbitrary area functions is not known so it is necessary to use an algorithm to find an approximate solution for individual
sample area functions.

The algorithm that has been used in the computer program represents the vocal tract by a series of contiguous cylindrical sections, each of fixed area (7). Each section can be described by a characteristic impedance and a complex propagation constant, both of which are well-known quantities for uniform cylindrical tubes. Junctions between individual sections satisfy the constraints of continuity of pressure and conservation of volume velocity. The transfer function (magnitude and phase) is calculated directly as a function of frequency. Natural frequencies (formants) are determined from the phase spectrum.

In this fashion the computer program calculated the three lowest formant frequencies. (The lowest resonances are the perceptually most important aspects of a vocalization produced with a given supralaryngeal vocal tract configuration.) These formant frequencies are presented in Fig. 1.

We then systematically explored the possible range of supralaryngeal vocal tract area functions that a rhesus monkey could make by moving his tongue, lips and jaw. The computer program was used to determine the formant frequencies of the first three formants for each configuration. We estimated the range of articulatory maneuvers by manipulating the supralaryngeal vocal tract of an anesthetized monkey and by taking into consideration the continuity constraints imposed by the monkey's tongue as well as the effects of different jaw angles and lip rounding. In doubtful cases we allowed greater deviations from the "unperturbed" area function derived from the casting.

Discussion. In Fig. 1 the dashed lines represent vocal tract configurations (for two different degrees of tongue height) that would be most likely to lead to the production of an unrounded high vowel. These result from changes in the tract toward that
shape of a human vocal tract that is characteristic of the production of /i/. Note that $F_1$ decreases relative to the unperturbed vocal tract while $F_2$ increases. In Fig. 2 the solid line again represents the unperturbed vocal tract while the three dashed lines represent vocal tract configurations analogous to low back vowels. The terminology, low back vowel, is shorthand for a vowel produced with a tongue constriction towards the back of the mouth while the jaw is open or low, for example /a/. In Fig. 3 the dashed lines represent configurations for rounded back vowels, that is vowels in which the lips are rounded, for example, /u/. In configuration 8 we have tried to account for the lengthening of the vocal tract that can occur with lip rounding, as, for example, in the human vowel /u/.

The ratio of the maximum constrictions for these vocal tract configurations relative to the dimensions of the unperturbed vocal tract is similar to that measured by Fant (3) for human vowels. We have plotted the calculated first and second formant frequencies that correspond to our unperturbed and perturbed monkey vocal tract configurations in Fig. 4. We have also plotted the formant frequencies measured by Fant for a male human speaker for the vowels /a/, /u/, and /i/. These three vowels delimit the human "vowel space". The length of this speaker's supralaryngeal vocal tract was 17 cm for /a/. We have therefore multiplied the formant frequencies measured by Fant by the ratio 2.6 to take account of the fact that the tract length of the monkey is 6.5 cm while that of the adult human measured by Fant was 17 cm. This procedure is valid for $F_1$ and $F_2$ of /u/ and /i/ where the behavior of the vocal tract can, as a first approximation, be represented by means of two or three tube models (3, 8).

We have also plotted the formant frequencies of the ape and monkey cries previously analyzed that were produced with
the vocal tract terminated by a high laryngeal impedance (1, 9). We have also scaled these formant frequencies up in frequency to take account of the longer vocal tracts of these animals relative to rhesus monkey. The formant frequencies that would correspond to a uniform 6.5 cm length tube terminated at one end are also plotted on Fig. 4. Note that the actual monkey and ape cries, which are noted by the letters C (Chimpanzee), G (Gorilla), and R (Rhesus monkey), occupy only part of the vowel space of our computer generated vowels. The only natural cry that is a significant deviation from this schwa vowel is the chimpanzee cry which was produced by the animal with its lips rounded (1); the formant frequencies of this cry correspond most closely to configuration 7 of Fig. 3, which represents the least rounded of our simulated rounded back vowels. Our computer-modeled perturbed monkey vocal tract configurations thus encompass and extend beyond the "acoustic vowel space" that was measured for actual utterances of non-human primates. The non-human primates previously recorded did not, in fact, use all of the articulatory maneuvers that we simulated for the rhesus monkey by means of the computer model.

The computer model further indicates that the possible acoustic vowel space of a monkey is quite restricted compared to the human range. In other words, even if a rhesus monkey were able to manipulate his supralaryngeal vocal tract to make use of all of the possibilities that we considered in our computer model, he would not be able to produce the full range of human vowels. We can thus conclude that the vocal apparatus of the rhesus monkey is inherently incapable of producing the range of human speech.

In Fig. 5 we have presented schematized area functions for the human vowels /a/, /u/, and /i/ where we have approximated the vocal tract by means of uniform tubes for illustrative
purposes. We have based these approximations on Fant's data (3). Note that the supralaryngeal vocal tract can essentially be divided into an anterior and a posterior cavity. The cross sectional area of the pharyngeal region in man can be constricted while the front of the mouth is open as in /a/. A large cross sectional area can also be produced in the pharyngeal region with either a constricted anterior passage as in /i/ or a large cavity as in /u/. The nonhuman primates cannot produce vocal tract area functions like man's because both the apes and monkeys lack a pharyngeal region like man's (1, 10) where the body of the tongue forms a movable anterior wall. In Fig. 6 we have reproduced an illustration from Negus (10). Note the relative position of the palate and larynx in the nonhuman primates and in man. The nonhuman primates lack a pharyngeal region like man's where the cross-sectional area continually changes during speech. The inability of apes to mimic human speech (2) is thus an inherent limitation of their vocal mechanisms. Some of man's recent ancestors also may have been unable to produce the full range of human speech; the skeletal evidence of human evolution shows a series of changes from the primate vocal tract that may have been, in part, necessary for the generation of speech (1). The human speech output mechanism thus should be viewed as part of man's species specific linguistic endowment.
Fig. 1

Area functions of supralaryngeal vocal tract modeled by computer and corresponding vowel formant frequencies. Curve 0 is the unperturbed vocal tract of the rhesus monkey. Curves 1 and 2 are perturbed area functions analogous to human high front vowels.
Fig. 2

Area functions of supralaryngeal vocal tract modeled by computer and corresponding vowel formant frequencies. Curve 0 is the unperturbed vocal tract of the rhesus monkey. Curves 3, 4 and 5 are perturbed area functions analagous to human low, back vowels.
Fig. 3

Area functions of supralaryngeal vocal tract modeled by computer and corresponding vowel formant frequencies. Curve 0 is the unperturbed vocal tract of the rhesus monkey. Curves 6, 7, and 8 are perturbed area functions analogous to human rounded, back vowels.
Vowel "space" of simulated monkey vocal tract relative to human vowel space and natural ape and monkey cries (1). Chimpanzee cry noted by letter C, Gorilla by G and Rhesus monkey by R. The formant frequencies that would correspond to a uniform 6.5 cm length tube terminated at one end are also plotted. All of the formant frequencies have been frequency scaled towards those of the rhesus monkey to compensate for differences in overall vocal tract length.
Fig. 5

Schematized area functions for the human vowels /a/, /u/, and /i/. Note that the area of the pharyngeal region is independent of the area of the front part of the supralaryngeal vocal tract.
Fig. 5
Vocal tracts of monkey, ape and man after Negus (10). Note that the nonhuman primates lack a pharyngeal region where the root of the tongue forms a movable anterior wall. The nonhuman primates thus lack a speech production mechanism where the area of the "back" pharyngeal region is variable and independent of the area of the "front" of the vocal tract.
by Negus
References


Some of the ape and monkey cries were apparently produced while the animals' larynges were open wide (1). Those cries are therefore not plotted in Fig. 4 since the vocal tract boundary conditions do not correspond to the computer model. However the acoustic analysis indicated that the shape of the animals' supralaryngeal vocal tract when they produced these cries still appeared to approximate a uniform tube. These cries therefore would not change our conclusions concerning the range of supralaryngeal vocal tract configurations that underlie these animals' vocalizations.

11. The authors wish to thank Professor William Henke for providing the computer program to calculate formant frequencies from area functions. This work was supported in part by grants from National Institute of Dental Research (Grant DE 01774), National Institute of Child Health and Human Development (Grant HD 01994), National Institutes of Health (Grant NB 04332-06), U. S. Public Health Service (Grant MH 10972). The computation facilities were provided by Project MAC, a Massachusetts Institute of Technology program sponsored by the Advanced Projects Research Agency, Department of Defense, under contract with the Office of Naval Research.