

NEWBORN INFANT CRY AND NONHUMAN PRIMATE VOCALIZATION

PHILIP LIEBERMAN, KATHERINE S. HARRIS,
PETER WOLFF, and LORRAINE H. RUSSELL

Haskins Laboratories, New Haven, Connecticut

Cries were recorded from 20 normal newborn infants from birth to the fourth day of life. Sound spectrograms showed that these cries were similar to the vocalizations of nonhuman primates insofar as the infants seemed to produce sounds by means of a uniform cross-section, schwalike, vocal tract configuration. Under certain conditions the laryngeal excitation was breathy and formant frequencies corresponding to an open boundary condition at the glottis were generated. The infants did not produce the range of sounds typical of adult human speech. This inability appears to reflect, in part, limitations imposed by the neonatal vocal apparatus, which, like the non-human primate vocal tract, appears to be inherently incapable of producing the full range of human speech. The initial restrictions on the sound-making repertoire of human infants are also evident in previous perceptually based transcriptions of the utterances of infants as well as in spectrographic and cineradiographic studies.

It is possible to differentiate at least three stages in the acquisition of speech by children: cry, babble, and word acquisition. The object of this study is to examine the earliest stage of infant cry, that is, neonatal cry. We shall attempt to relate our results to previous cineradiographic, acoustic, and perceptual studies of infant cry and to the latter stages of language acquisition. We shall also compare human infant cry with the vocalizations of nonhuman primates.

METHOD

Cries were recorded from 20 newborn infants from birth to the fourth day of life. An Ampex type 601 tape recorder was used with an Electro-Voice 633A microphone. The recordings were made in the hospital delivery room and in a room adjoining the hospital nursery. The tape recordings were edited and spectrograms were made using a Kay Electric Sound Spectrograph and a Voice-print Sound Spectrograph. The sample analyzed included "birth cries," "fussing cries," "angry cries," "gurgles," "hunger cries," "shrieks," and inspiratory "whistles." The descriptive terms for these cries are consistent with clinical observations and impressions formed through an extensive study of infant behavior by one of the authors, Peter Wolff. Most of the cries were spontaneous, some were elicited by pinches. The vocalizations encompassed the range that is normally produced by infants in good health.

RESULTS

In Figure 1 a spectrogram of a cry produced during the first five minutes of life by a male infant is presented. The band width of the spectrograph's analyzing filter was 300 Hz. The fundamental frequency of phonation was about 400 Hz. The glottal excitation apparently was breathy since the effects of noise excitation are evident in the spectrogram. The noisy excitation indeed made it possible to clearly resolve the energy concentrations that appear at approximately 1.1, 3.3, and 5.8 kHz. These energy concentrations must mirror the transfer function of the supralaryngeal vocal tract since they are spaced farther

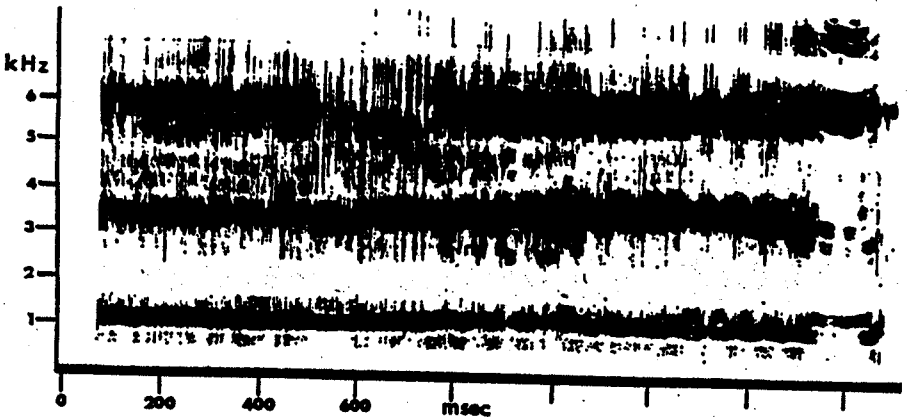


FIGURE 1. Spectrogram of cry produced during first five minutes of life. The fundamental frequency is about 400 Hz. Note the formant frequencies at 1.1, 3.3, and 5.8 kHz.

apart than the harmonics of the laryngeal excitation and at inharmonic intervals. These energy concentrations may not exactly specify the formant frequencies since harmonics of the laryngeal excitation are spaced at almost 400-Hz intervals. However, taking this uncertainty into account, we can approximate the formant frequencies and thereby infer the configuration of the infant's supralaryngeal vocal tract for this vocalization by making use of the acoustic theory of speech production (Chiba and Kajiyama, 1958; Fant, 1960).

This theory allows us to infer that the supralaryngeal vocal tract configuration of this infant approximated a 7.5-cm long uniform tube open at one end. The formants of such a tube will occur at 1.1, 3.3, and 5.5 kHz since it will have resonances at intervals of

$$\frac{(2k + 1) C}{4L}$$

where C = velocity of sound, L = length of tube, and k is an integer ≥ 0 .

There is a surprising scarcity of information on the expected length of the neonatal vocal tract. Our estimate of 7.5 cm for the length of the infant's supralaryngeal vocal tract is consistent with comparative studies. Hopkin (1967), for example, notes that the neonatal tongue is approximately half the length of the adult tongue. Since the larynx is positioned higher in the neonate's vocal tract than is the case in an adult (Noback, 1923), this estimate of 7.5 cm, which is slightly less than half the length of the adult vocal tracts measured by Chiba and Kajiyama (1958) and Fant (1960), is quite reasonable.

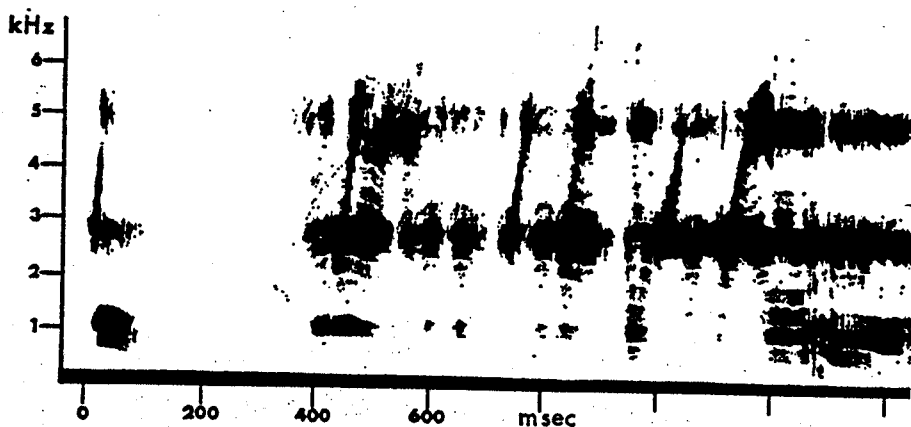


FIGURE 2. Spectrogram of infant cry. Note the formants at 1, 3, and 5 kHz. These energy concentrations again reflect a supralaryngeal vocal tract configuration that approximates a uniform cross-section tube terminated at one end.

In Figure 2 another cry recorded at the birth of this same infant is presented. The analyzing filter's band width was 300 Hz. Note that this cry consists of a short vocalization followed by a longer vocalization. The infant produced a short "gurgle" followed by a cry. Energy concentrations are again apparent at 1, 3, and 5 kHz for both episodes of vocalization, while harmonics of the laryngeal excitation are apparent at intervals of approximately 400 Hz. We can again infer that the supralaryngeal vocal tract configuration of this infant approximated the schwa vowel (i.e., a uniform cross-section tube) since the formants again occur at odd integral multiples.

Close examination of the set of spectrograms of the cries of 20 neonates revealed no formant patterns that were not consistent with a supralaryngeal vocal tract configuration that approximated either a uniform cross section or a slightly flared tube. In some instances the formants all moved higher or lower in frequency during the course of the cry. However, the intervals between the formants showed that the supralaryngeal vocal tract still approximated a tube (Lieberman, 1968). These formant transitions thus reflected changes in the overall length of the infants' supralaryngeal vocal tracts. Cineradiographic studies (Truby, Bosma, and Lind, 1965) show that changes in the overall

length of the neonatal supralaryngeal vocal tract occur during cries and are the result of laryngeal movements.

The formant pattern of the neonatal cries had energy present at intervals of $1F_0$, $3F_0$, and $5F_0$ (where F_0 = the first formant's center frequency) in approximately 80% of the cases. In the remaining 20%, energy instead was present at intervals of F_0 , $2F_0$, $3F_0$, etc., where $F_0 = 2F_0$. Under these conditions the neonates' supralaryngeal vocal tracts apparently resembled a uniform tube open at both ends. It was not possible to correlate this pattern with the descriptive terms that were used to characterize the cries except that it did not occur during any of the cries that were labelled "fussing." In Figure 3 an example of this formant pattern is presented.

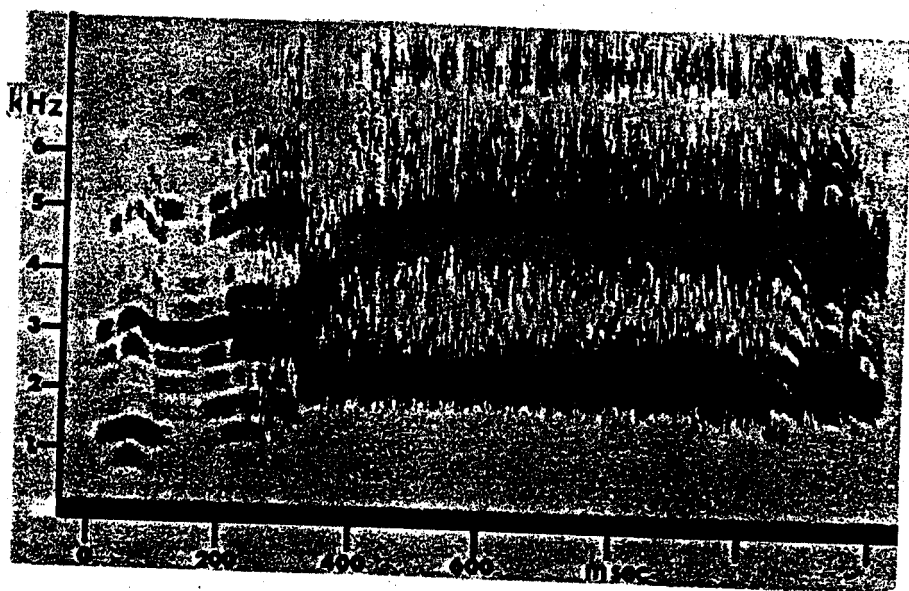


FIGURE 3. Example of cry that starts with periodic excitation shifting to aperiodic excitation. Note initial formants at 1.25, 3.0, and 5.0 kHz. Energy concentrations then shift to 2.25 and 4.8 kHz with aperiodic excitation. The supralaryngeal vocal tract configuration apparently approximates a tube with uniform boundary conditions when the glottal opening is large during the aperiodic breathy excitation.

This cry was produced seven minutes after birth by the same infant of Figures 1 and 2. The analyzing filter's band width was again 300 Hz. Note the presence of harmonics of the laryngeal excitation at the start of this utterance. Energy concentrations are present during the initial, voiced part of the utterance at 1.25, 3.0 and 5.0 kHz. Note that the cry loses its harmonic structure after 300 msec, where it becomes noisy. Note the abrupt discontinuity in the first energy concentration which shifts to 2.25 kHz. The second energy concentration occurs at 4.8 kHz during this noisy part of the cry. The supralaryngeal

vocal tract seems to be resonating as a half-wave rather than as a quarter-wave oscillator. The higher formants are multiples of the first formant which now occurs at twice the frequency of the first formant of the quarter-wave resonator. The resonances of a uniform tube open at both ends will occur at intervals of

$$\frac{(k) C}{2L}$$

where C = the velocity of sound, L = the length of the tube, and k is an integer ≥ 1 .

The first formant of a supralaryngeal vocal tract that resembles a uniform tube will therefore abruptly double in frequency when the boundary condition at the larynx changes from a closed state to an open state. The formants of a uniform tube open at both ends will also occur at regular intervals. The boundary condition at the infant's larynx therefore probably approximated an open termination when the laryngeal excitation changed to aperiodic noise at 300 msec. The most likely explanation for the change in the laryngeal excitation's character at 300 msec is that the infant fails to increase the medial compression of his vocal cords as he increases his subglottal air pressure. He would thus blow his vocal cords apart preventing phonation and producing noise-like aperiodic excitation of his supralaryngeal vocal tract which would be terminated by the open glottis.

Truby, Bosma, and Lind (1965) in their study of neonatal vocalizations present spectrograms that show similar effects. They also present simultaneous plots of esophageal air pressure which indicate that these effects occur when the subglottal air pressure exceeds a critical level (about 6 cm H₂O above the mean subglottal pressure). The infant's vocal cords are then thrown into an open position because he apparently does not modify the medial compression of his larynx (Van den Berg, 1960) during the cry.

Note that the laryngeal excitation becomes periodic at the end of the utterance in Figure 3 (after 1000 msec) where subglottal air pressure typically falls at the end of the unmarked breath-group (Lieberman, 1967). The fundamental frequency also abruptly falls during the last 100 msec of the breath group since the fundamental frequency of phonation is a function of both laryngeal muscular maneuvers and the transglottal air pressure drop.

It seems reasonable to attribute the changes in formant pattern and laryngeal excitation associated with the utterance in Figure 3 to a state of minimum control or, indeed, of no control by the infant. The infant does not adjust his larynx so that phonation can continue as the subglottal air pressure rises. Perhaps the infant larynx is inherently incapable of withstanding high subglottal air pressures. However, whether the problem is a matter of laryngeal control or of laryngeal development, the changes in formant pattern are concomitant with the change in laryngeal source characteristics from periodic excitation (which involves an adducted state of the larynx) to aperiodic noise which involves an open glottis. Truby and Bosma, in the work cited, attribute similar changes in spectrographic displays to a special mode of phonation which they

term *hyperphonation*. We believe instead that these energy concentrations at widely spaced regular intervals reflect the formants of a supralaryngeal vocal tract shape that resembles a uniform tube that has similar, open boundary conditions at each end. Similar effects also appear to occur during the vocalizations of nonhuman primates.

Other phenomena also differentiate neonatal cry from adult speech. We found that the fundamental frequency of phonation, in general, was not stable during neonatal cry. In Figure 4 a spectrogram is presented of a cry from the

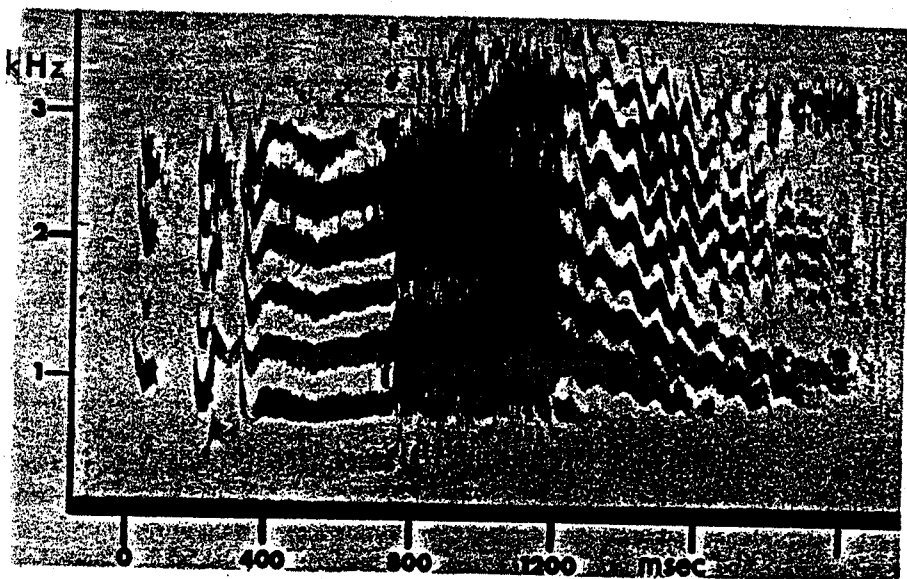


FIGURE 4. Spectrogram showing large periodic variations in fundamental frequency that occur at a rate of approximately 12 Hz. The effective band width of the spectrograph's analyzing filter was 150 Hz.

same infant in response to pain stimulation 35 minutes after birth. The band width of the analyzing filter was 150 Hz. The narrower band width was achieved by playing the tape recording back at twice its normal speed. The time scale of the spectrogram is thus compressed while the frequency scale of the spectrogram is expanded. Individual harmonics of the laryngeal excitation are clearly resolved except for the middle portion of the utterance where some turbulent noiselike energy also occurs. Note the large periodic variations in fundamental frequency that occur at a rate of approximately 12 Hz. Variations in fundamental frequency like these frequently occur during the vocalizations of newborn infants.

Correlations with Previous Studies of Infant Cry

The cineradiographic data reported by Bosma, Truby, and Lind (1965) confirm that the supralaryngeal vocal tract configuration for newborn infant cry

is almost rigid. They note that:

Direct inspection and radiographic observation reveal that the oral structures move little. The mandible is held tensely in open position, perhaps opening during cry and closing slightly during inspirations. The tongue tip becomes separated from its suckling or resting apposition to the lips and to oral surface of palate and maxillary alveolar ridge, and the tongue tip protrudes ventrad and cephalad from the tongue body. On inspection, the anterior portion of the tongue is seen to be tensed in concave contour . . . the mouth may be essentially immobilized in position. (p. 70)

The only movements that appear to occur during neonatal cry involve gross laryngeal maneuvers where the larynx moves upwards and downwards. Similar laryngeal maneuvers also occur during ape and monkey cries (Lieberman, 1968). Furthermore, the cineradiographic photographs reproduced by Bosma, Truby, and Lind showed that the velum was either open or closed throughout each utterance. The earliest vocalizations appeared to be nasalized; some of the later vocalizations appeared to be made while the velum was closed.

Lynip (1951), in a study of infant vocalizations, made spectrograms of utterances produced by one girl from birth to 60 weeks. He concluded that the "infant's pre-speech utterances are essentially incomparable to adult sounds." (p. 246) His conclusion is correct insofar as very young children produce schwalike utterances where all the formants are transposed to higher frequencies than those typical of adult vocalizations. Winitz (1960), in a study of infants whose ages ranged from 9-15 months (mean age, 11.5 months), found that they indeed produced the entire range of human vowels. The disagreement between Lynip's and Winitz's conclusions appears to reflect Lynip's paying most attention to utterances recorded between birth and one month of life while Winitz concentrates on much older infants. Our results are consistent with Lynip's data for cries recorded between birth and two to three months of life.

One of the oft-repeated statements about language acquisition is that children, when they babble, produce the sounds of all languages known to man (Miller, 1951). Our data indicate that infants certainly do not start with this capability at birth. Perceptually based studies of the utterances of infants tend to support our conclusions. Irwin (1948), for example, reports that infants do not use any back vowels during the first three months of life. He notes that 25% of their vocalizations are transcribed as /ɪ/, 45% as /e/, and 25% as /ʌ/. In all likelihood Irwin's observers categorized some of the infant schwalike sounds /ɪ/ or /e/ because the short length of the infant vocal tract produces higher formant frequencies than is the case for adult speech.

DISCUSSION

Newborn human infants, like nonhuman primates, do not execute any maneuvers of their supralaryngeal vocal tracts during vocalizations except for gross laryngeal maneuvers. The shape of their supralaryngeal vocal tract appears to approximate a uniform cross-section, schwalike, configuration.

Human infants appear to start life equipped with a vocal tract that differs

from that of the adult vocal tract in configuration as well as in size. Indeed, the vocal tract of the newborn in some ways is more similar to the vocal tract of a nonhuman primate than it is to the adult human vocal tract. The position of the larynx in the human neonate is quite high. The epiglottis is at the level of the first cervical vertebra while the inferior border of the cricoid is at the level of the fourth cervical vertebra (Noback, 1928). These positions are similar in a gorilla (Negus, 1949). In an adult, these cartilages are, respectively, at the level of the third and sixth cervical vertebrae. The thyroid cartilage of the neonate lies contiguous to the hyoid bone (Eckenhoff, 1951) placing the epiglottis in fairly close proximity to the velum and keeping the root of the tongue within the oral cavity. Furthermore, the infant tongue is large, much closer to its adult size than is the oral cavity. The mandible will undergo dramatic downward, forward growth which, together with the downward growth of the upper alveolar process, the upward growth of the lower alveolar process, and the descent of the root of the tongue in conjunction with the descent of the larynx, will eventually enclose the tongue in an oral cavity such as is known in the adult (Brodie, 1949). However, in the neonate, the short, broad tongue fills the entire mouth and in its resting state is superiorly in contact with the entire length of the palate, laterally with the buccinator, and anteriorly with parts of the jaws so that the mouth is closed by the action of the lips (Scammon, 1923; Brodie, 1950). During suckle, a position is maintained between the tongue tip and the lower lip. Only during infant cry is the mouth wide open and the tongue separated from its apposition with the palate and lower lip (Bosma, 1967).

The upper pharynx also differs markedly in the neonate and as a result is much less mobile than in the adult. In the new-born, the upper pharynx is a narrow tube, the longest diameter of which runs anteroposteriorly rather than superoinferiorly as it does in the adult (Braislin, 1919). The roof of the infant's pharynx slopes gently downward from the chonae to the dorsal wall of the mesopharynx and, therefore, the epipharynx of the infant does not have a dorsal or posterior wall (Bosma and Fletcher, 1961). The newborn infant, like a nonhuman primate, thus lacks a pharyngeal region that can vary its cross-sectional area. In fully developed human speech, pharyngeal volume changes over a ten-to-one range as the root of the tongue, which forms the anterior pharyngeal wall, moves. These changes in pharyngeal volume are essential for producing phonetic contrasts, for example /a/ versus /ɪ/ (Chiba and Kajiyama, 1958; Fant 1960). Although some of the limitations of the vocal repertoire of human neonates may be due to deficiencies in the central control of the vocal apparatus, the newborn human infant, like the nonhuman primates, is restricted by the limitations of his vocal apparatus (Lieberman, 1968; Lieberman, Klatt, and Wilson, 1969).

Human infants, unlike monkeys and apes, eventually produce the full range of human speech. The question that confronts us now is, when and under what conditions do infants go beyond the nonhuman stage? Human infants start life with a speech production apparatus that in many ways resembles the nonhuman

primate vocal apparatus. We know that the nonhuman primate vocal apparatus is inherently incapable of producing the range of human speech. Human evolution involved, among many other factors, the development of the peripheral structures used in speech production. We need to study the development of cry and babble in human infants with respect to the development of both the output mechanism (i.e., the vocal tract) and the central control of the vocal tract. We think that the answers to these questions, as well as the overall sequence that is involved in the infant's acquisition of the phonetic level of language, will be relevant to the broader question of the nature of human linguistic ability.

ACKNOWLEDGMENT

Peter Wolff's address is Harvard Medical School, Boston, Massachusetts. In addition to their work at Haskins Laboratories, the other authors are associated with the following institutions: Philip Lieberman, University of Connecticut; Katherine S. Harris, City University of New York; and Lorraine H. Russell, Temple University, Philadelphia, Pennsylvania. Requests for reprints should be addressed to Philip Lieberman, Haskins Laboratories, 270 Crown Street, New Haven, Connecticut 06510.

REFERENCES

- BOSMA, J. F., Human infant oral function. Ch. 4 in J. F. Bosma (Ed.), *Symposium on Oral Sensation and Perception*. Springfield, Ill.: Thomas, 98-110 (1967).
- BOSMA, J. F., and FLETCHER, S. G., The upper pharynx, a review, part I, embryology and anatomy. *Ann. Otol. Rhinol. Laryng.*, 70, 953-973 (1961).
- BRAISLIN, W. C., A study of some casts of the infantile pharynx with special reference to the eustachian tubes. *Ann. Otol. Rhinol. Laryng.* 21, 36-74 (1919).
- BRODIE, A. G., Anatomy and physiology of the head and neck, musculature. *Amer. J. Orthop.*, 26, 831-837 (1950).
- BRODIE, A. G., On the growth pattern of the human head from three months to eight years of life. *Amer. J. Anat.*, 68, 209-259 (1949).
- CHIBA, T., and KAJIYAMA, M., *The Vowel, Its Nature and Structure*. Tokyo: Phonetic Society of Japan (1958).
- ECKENHOFF, J. E., Some anatomic considerations of the infant larynx influencing endotracheal anesthesia. *Anesthesiology*, 12, 401-410 (1951).
- FANT, C. G. M., *Acoustic Theory of Speech Production*. The Hague: Mouton (1960).
- HOPKIN, G. B., Neonatal and adult tongue dimensions. *Angle Orthodont.*, 27, 132-133 (1967).
- IRWIN, O. C., Infant speech: Development of vowel sounds. *J. Speech Hearing Dis.*, 13, 31-34 (1948).
- LIEBERMAN, P., *Intonation, Perception, and Language*. Cambridge, Mass.: MIT (1967).
- LIEBERMAN, P., Primate vocalizations and human linguistic ability. *J. acoust. Soc. Amer.*, 44, 1574-1584 (1968).
- LIEBERMAN, P., HARRIS, K. S., and WOLFF, P., Newborn infant cry in relation to nonhuman primate vocalizations. *J. acoust. Soc. Amer.*, 44, 365 (A) (1968).
- LIEBERMAN, P., KLATT, D. L., and WILSON, W. A., Vocal tract limitations on the vowel repertoires of rhesus monkey and other nonhuman primates. *Science*, 164, 1185-1187 (1969).
- LYNIP, A. W., The uses of magnetic devices in the collection and analysis of the preverbal utterances of an infant. *Genet. Psychol. Monogr.* 44, 221-262 (1961).
- MILLER, G. A., *Language and Communication*. New York: McGraw-Hill (1951).
- NEGUS, V. E., *The Comparative Anatomy and Physiology of the Larynx*. New York: Hafner (1949).
- NOBACK, G. J., The developmental topography of the larynx, trachea, and lungs in fetus, new-born, infant and child. *Amer. J. dis. Child.* 26, 515-533 (1923).

- SCAMMON, R., A summary of the anatomy of the infant and child. Chapter 3 in I. Apt (Ed.) *Pediatrics*, Vol. I. Philadelphia: W. B. Saunders, 296-297 (1923).
- TRUBY, H. M., BOSMA, J. F., and LIND, F., *Newborn Infant Cry*. Uppsala: Almqvist and Wiksells (1965).
- VAN DEN BERG, J., Vocal ligaments versus registers. *Curr. Probl. Phoniatr. Logoped.*, 1, 19-34 (1960).
- WINITZ, H., Spectrographic investigation of infant vowels. *J. Genet. Psychol.*, 96, 171-181 (1960).

Received June 4, 1969.