XX. SPEECH COMMUNICATION*

Academic and Research Staff

Prof. K. N. Stevens
Prof. M. Halle
Prof. W. L. Henke
Prof. A. V. Oppenheim
Dr. Mary C. Bateson
Dr. Margaret Bullowa
Dr. A. W. F. Huggins
Dr. R. D. Kent
Dr. D. H. Klatt
Dr. Paula Menyuk
A. R. Kessler

Graduate Students

R. W. Boberg
D. E. Dudgeon
R. W. Hankins
Emily F. Kirstein
R. M. Mersereau
B. Mezrich
J. S. Perkell
M. R. Sambur
J. S. Siegel
H. A. Sunkenberg
A. Weinreb
M. L. Wood, Jr.
V. W. Zue

RESEARCH OBJECTIVES AND SUMMARY OF RESEARCH

The broad aim of our research in speech communication is to gain an understanding of the nature of the processes of human speech production and perception and to learn how these processes are acquired. A practical aim is to utilize knowledge gained through study of speech communication to devise procedures that will permit limited communication between men and machines by means of speech. Several projects directed toward these goals are active at present.

Studies of the mechanism of speech production have included examination of turbulence noise generation and of glottal activity during consonants, and these studies are leading toward a revised specification of the distinctive features related to laryngeal configurations. Projects on the modelling of the speech production process continue, with work on improved rules for the synthesis of certain consonants in stressed syllables, using a terminal analog synthesizer, and with the development of a transmission-line simulation of speech production using the speech group computer facility. In addition to the modelling of speech production our interests have broadened to include the development of techniques for on-line computer-aided music composition and synthesis.

Modern digital signal processing techniques are being applied to speech in several projects. These include development of a method of frequency warping, for computation of spectra with unequal resolution using the fast Fourier transform. This technique is being applied to the problem of helium speech translation. Work on procedures for computer generation of speech spectrograms is also in progress.

Studies of the perception of voiced sounds characterized by time-varying fundamental frequency (F_o) are in progress, with an initial set of experiments on the discrimination of shifts in the slope of linearly changing fundamental frequency, and projected experiments on F_o contours of the type that occur in tone languages.

Research on language acquisition includes studies of the sounds that are used by infants in the first few months of life in interactive "conversational" situations and in other situations that can be reasonably well specified from the context, using previously acquired synchronized tape and film of children in their natural environment. Attempts are being made to classify these sounds according to the situations in which they occur and according to their acoustic characteristics. Experimental studies attempting to

*This work was supported in part by the U.S. Air Force Cambridge Research Laboratories under Contract F19628-69-C-0044; and in part by the National Institutes of Health (Grant 2 RO1 NB-04332-08), and the Joint Services Electronics Program (U.S. Army, U.S. Navy, and U.S. Air Force) under Contract DA 28-043-AMC-02536(E).

delineate the capabilities of children to produce and perceive sounds that are distinguished on the basis of certain prosodic features continue, and these experiments will be broadened to include investigation of particular segmental features as they occur in different phonetic contexts.

Most of the research projects described above make use of a digital computer facility comprising a PDP-9 computer and various peripheral items. The hardware and software capabilities of this facility are undergoing continual evolution. Current and recent activities include the addition of an in-house designed inexpensive but flexible display processor with a stroke-type character generator, and the specification and implementation of semantic extensions to a FORTRAN programming system oriented toward programming for on-line graphics and sonics. These extensions include machinery for the manipulation of "data structures" that greatly facilitate the design of graphics-using systems. Some general-purpose systems implemented using this extended FORTRAN system include DYNAMO – a general continuous system simulator, and MITSYN – a music synthesis-oriented system. MITSYN includes independently useful subsystems for on-line graphical specification of signal-processing networks built up by the interconnection of signal-processing primitives, and a parameter notation and editor used to create files of parameter values vs time for control of systems (often sound synthesis) which include temporally varying parameters as input.

A facility for presenting arbitrary sequences of prerecorded audio stimuli is in the final stage of development. The stimuli can be monotic or dichotic, and can be synthetic, natural, or edited-natural waveforms. The addition of disk-storage to our PDP-9 computer has made it possible to present the sequences in real time, and thus run subjects on-line, with the result that adaptive procedures can be used, in addition to the more mundane application of producing experimental tapes.

K. N. Stevens, M. Halle, W. L. Henke,
A. V. Oppenheim, D. H. Klatt

A. THE INTERPERSONAL CONTEXT OF INFANT VOCALIZATION

Early in infancy, interactional sequences appear between mother and child which have the appearance of conversation: constant or nearby constant communication in one modality (visual) and intermittent, alternating communication in another (vocal). Working with the longitudinal corpus of films and tapes collected by Dr. Margaret Bullowa,[1] a first sample of 5 such sequences, termed "proto-conversations," collected between the ages of 49 and 105 days from one mother-infant pair, has been isolated and subjected to initial statistical analysis. (The total number of utterances is 284.) Several comparable sequences have been isolated for two other mother-infant pairs, and nonsystematic interviewing suggests that, while not all mothers are aware of participating in such interactions, they are a part of the experience of many infants in this culture.[2] Work by other researchers has established the importance of eye-to-eye contact in early social development[3] and the possibility of using operant conditioning techniques to increase the frequency of non-cry infant vocalizations[4]: a mixed social stimulus serves to reinforce vocalization, but the influence of maternal vocalization (as contrasted with touch, visible presence, smiling, and so forth) is preponderant. Such research, however, which sums the total number of vocalizations over a time period and makes the maternal response automatic, ignores the interplay between mother and infant in which each is

affected by the behavior of the other and the two are coparticipants in an on-going event.

There is a general theoretical problem in describing mother-infant interactions, namely that the two participants act on the basis of different codes, and yet achieve mutual calibration. Some of this calibration is based on movement synchrony,[5] probably depending on an orchestration of physiological rhythms,[6] while in other cases there is alternation, most frequently described with regard to games such as peek-a-boo and pat-a-cake. The thrust of the present study is to describe the structure of the interactions themselves, with special attention to the development of patterns of alternation in which contexts for speech may be constructed. Such an analysis depends on a concept of conversation that focuses primarily on the importance of vocal exchange in affirming contact (which Malinowski referred to as "phatic communion"[7]), rather than on content or the exchange of information. This orientation has been followed in a number of studies of adult communication.[8] From this point of view, the "proto-conversations" of preverbal infants and their mothers can be treated as equivalent to the conversations of adults. The development of the capacity for participation in complex sequenced behavior must lay the groundwork for participation in games[9] and for the development of linguistic performance; an ability to manipulate and recombine sequences may be significantly related to the development of linguistic competence. An understanding of interactional sequences may also be expected to lay important groundwork for better descriptions of imitative behavior.

The corpus provides half-hour stretches of behavior, filmed and taped in the home, during which mother and baby followed their normal routines as nearly as possible and the camera was trained on the infant. Sequences selected for detailed study occurred amidst stretches of infant sleeping, crying, or playing alone, or being taken care of (bathed, fed, dressed) by the mother. The following criteria were used in selecting sequences: Frequent vocalization by both mother and infant (excluding crying); sustained eye contact; and absence of caretaking activities that might have distorted the temporal pattern of vocalization. Sequences selected by these criteria showed several regularities. They tended to follow periods of active caretaking; and typically mother and infant were less than a yard apart, with the mother frequently squatting at the baby's level. On the other hand, there are occasions in the sample on which the mother apparently tried to start a "proto-conversation," using the same postures and utterances, but without engaging the infant's attention, so that eye contact and responsive vocalization are absent.

That the mother's participation is patterned on conversation, and that her participation is constructed around an expected participation on the infant's part can be seen from the text in Fig. XX-1, which shows the recording technique used in preparing the data for analysis. In order to obtain an accurate record of the vocalizations of mother and

infant, a number of graphic-level recorder tracings were made of each sequence. Each time a tracing was made, the vocalizations of one participant were marked by hand with a separate stylus. A coded subsonic signal makes it possible to associate sounds on the tape (and marks on the graphic-level recordings) with specific frames of the film.

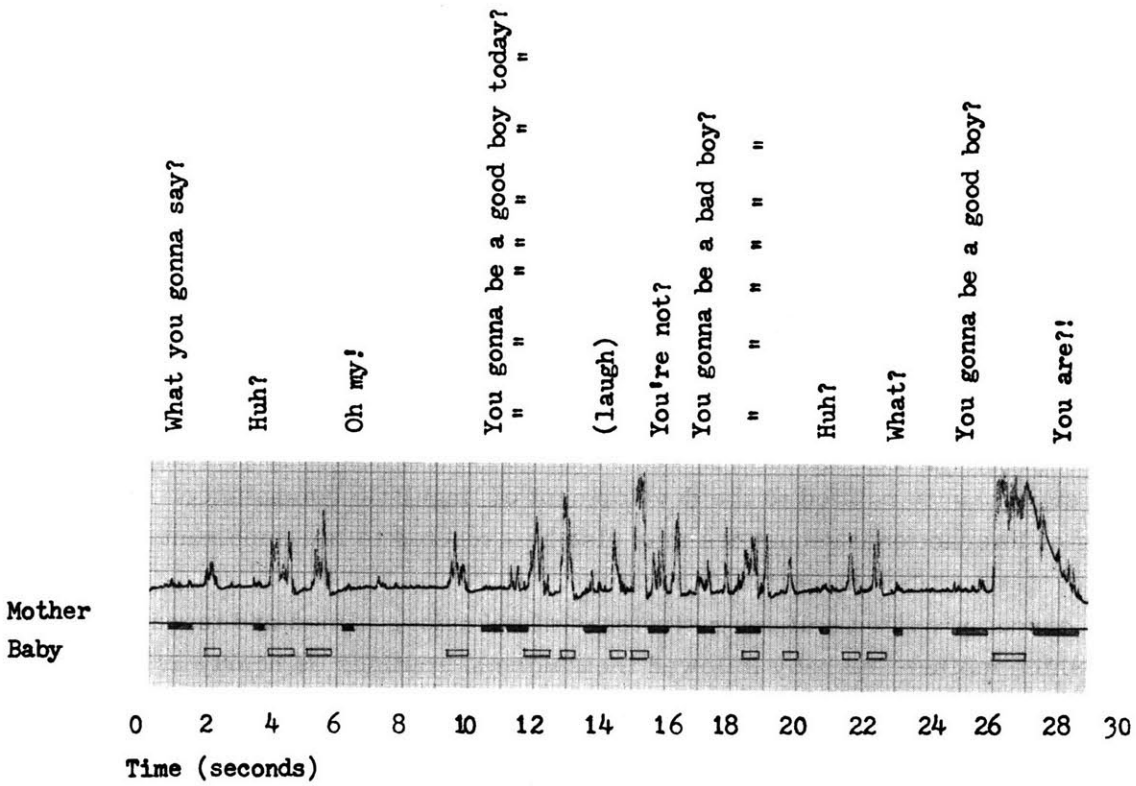

Fig. XX-1. Graphic level recording of a 30-s portion of an interaction between Mackie (age 98 days) and his mother.

The text represents interpretation, since the mother's vocalizations were typically murmured and elliptical. Although the sound was recorded on a single boom-mounted microphone in a naturalistic setting, with considerable extraneous noise, onset figures for vocalizations are estimated as reliable ±0.2 s.

The infant's vocalizations, while obviously not structured in the same sense as the mother's, are apparently of the type referred to as cooing and grunting, or more generally as "happy vocalization," although for statistical purposes all vocalizations falling within the analyzed passages were included. Figure XX-2 has been included to provide examples of typical "conversational" vocalizations.

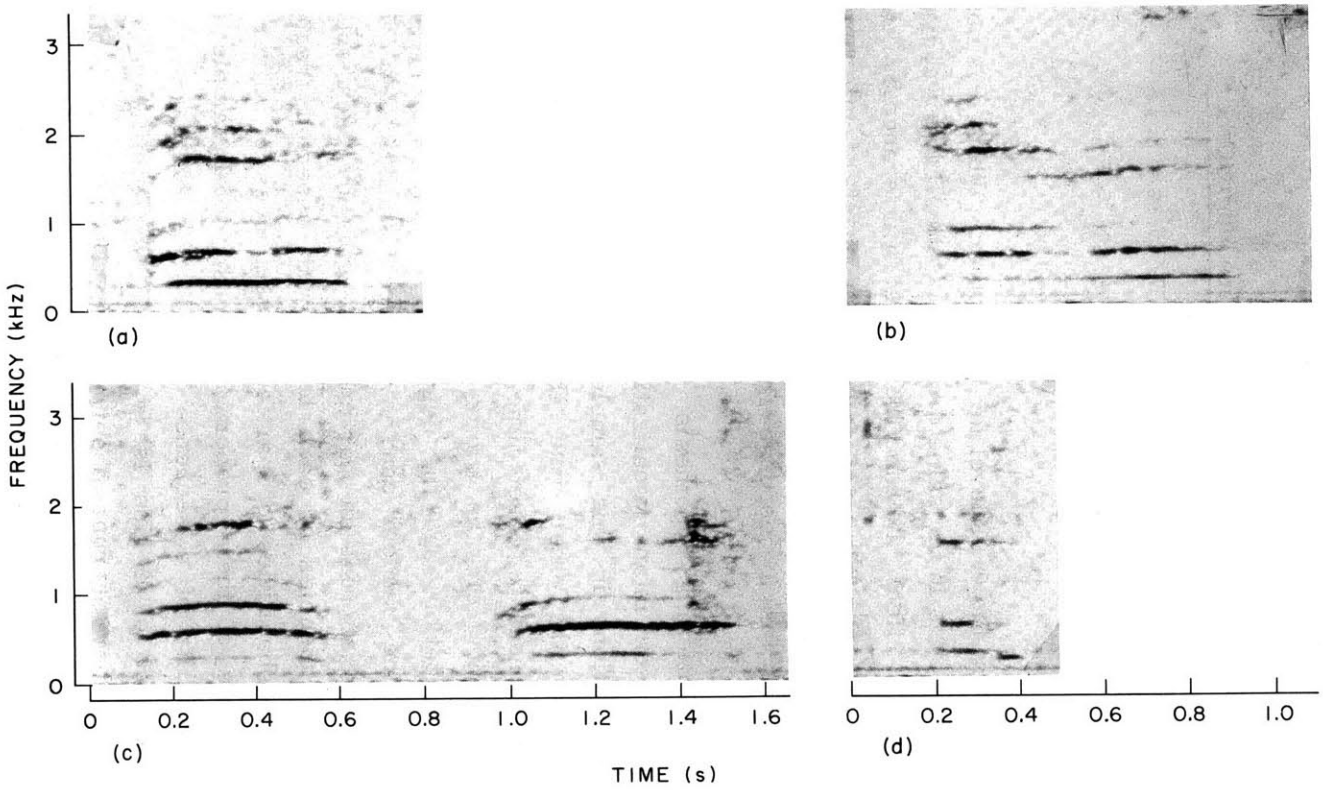

Fig. XX-2. Narrow-band spectrograms of several of Mackie's vocalizations in interaction with his fundamental-frequency contour and onset. (a) Vocalization beginning with a glottal stop. (b) Diphthong. (c) Series of two vocalizations without initial glottal stop, but with the second terminating with glottalization. (d) Brief "grunt."

For both mother and infant, the mean time of utterance onset from onset of previous utterance is longer when the previous utterance was by self than when the previous utterance was by other (times given in fifths of seconds).

	Preceded by Self	Preceded by Other
Mo	$\bar{x}$ = 10.75, s = 4.94	$\bar{x}$ = 7.15, s = 8.33
Ba	$\bar{x}$ = 7.64, s = 6.05	$\bar{x}$ = 6.87, s = 6.29

This difference is significant at the 0.01 level for the mother. While it is not significant for the infant, the tendency is in the same direction, with all interval types somewhat briefer for the infant. In effect, the time from onset of utterance by other may be characterized as response time, while the time from onset of utterance by self may be characterized as elicitation time: apparently the mother says something (often a question) and waits for a response from the infant before renewing her vocalization.

Complementing the conditioning hypothesis, is the hypothesis that the mother inserts her comments between infant vocalizations. Thus the mean time between onset of all infant vocalizations in this sample, regardless of maternal vocalizations, is 15.7 (s=21.33); on the other hand, the same thing might be said of the infant, with the mean onset to onset time for the mother at 14.79 (s=11.19), and not significantly different from the timing of the infant. The over-all effect is one of alternating vocalization, and the degree to which the alternation of the two parties deviates from the random can be explored statistically. This was done by pooling the number of runs of particular lengths by each speaker and comparing the observed frequencies of each length with expected frequencies, by a chi-square test. The probabilities of particular sequences of either Mo or Ba were computed on the basis of their proportions in the total sample (p_M=.54, p_B=.46). Thus, the probability of a sequence of only one Mo is $p_M \cdot p_B$, and the probability of a sequence of n Mos is $(p_M)^n p_B$, and similarly for sequences of Ba. Since the total number of runs observed was 183, the probabilities of particular run lengths were used to compute the number of runs of each length to be expected in a sample of 183 runs, given the constraint that runs by Mo and Ba must alternate and the total number of vocalizations must =284.[10] The test was significant at well over the 0.01 level (d.f.=7). In effect, mother and infant vocalize by turns.

Occurrences of Mo in a Run	Observed Frequencies	p	Expected Frequencies
1	68	.248	33.4
2	10	.134	18.0
3	6	.072	9.7
4	4	.039	5.3
5 or more	4	.047	6.3

Occurrences of Ba in a Run	Observed Frequencies	p	Expected Frequencies
1	62	.248	39.0
2	19	.114	18.0
3	7	.052	8.2
4	3	.024	3.8
5 or more	0	.022	3.5

Interestingly enough, runs of all lengths greater than one are low for the mother, while the high number of single-utterance runs for the baby is produced by a lower than expected number of runs of 3 or more, but the observed number of two-utterance runs corresponds to the prediction. In fact, pairs of acoustically similar vocalizations, spaced approximately 0.8 s apart (see Fig. XX-2c) were quite frequent for the infant, and further analysis of the spectrograms may provide a basis for treating these doublets as single utterances.

Continuing research on the interpersonal context of infant vocalization is being directed toward (a) classifying the types of vocalizations that occur in this kind of context, as contrasted with cries, etc. Bullowa's current effort to establish a typology of vocalizations from birth is expected to lead to a refinement of the current descriptions; (b) establishing the range of variation in the currently available longitudinal sample; (c) relating vocal and kinesic behavior; and (d) following the development of comparable interactions in the sample through the development of speech.

Mary C. Bateson

References

1. The methodology of data collection is described in Margaret Bullowa, L. G. Jones, and T. G. Bever, "Development from Vocal to Verbal Behavior in Children," Monographs of the Society for Research in Child Development No. 29, 1964.

2. William Caudill has shown that American mothers "chat" more with their infants and the infants produce more "happy vocalizations" than Japanese infants, which in turn have more direct body contact with the mother than U. S. babies. See W. Caudill, "Tiny Dramas: Vocal Communication between Mother and Infant in Japanese and American Families" (to be published in Proc. Second Conference on Culture and Mental Health, William Lebra (ed.) (Social Science Research Institute, University of Hawaii).

3. K. S. Robson, "The Role of Eye-to-Eye Contact in Maternal-Infant Attachment," J. Child Psychol. Psychiatr. 8, 13-25 (1967).

4. Harriet L. Rheingold, J. L. Gewirtz, and Helen W. Ross, "Social Conditioning of Vocalizations in the Infant," J. Comp. Physiol. Psychol. 52, 68-73 (1959); P. Weisberg, "Social and Nonsocial Conditioning of Infant Vocalization," Child Develop. 34, 377-388 (1963); G. A. Todd, A. Gibson, and B. Palmer, "Social Reinforcement of Infant Babbling," Child Develop. 39, 591-596 (1968).

5. W. S. Condon and W. D. Ogston, "Sound Film Analysis of Normal and Pathological Behavior Patterns," J. Nervous and Mental Disease 143, 338-347 (1966).

6. Margaret Bullowa, "The Onset of Speech," Presented at the Society for Research in Child Development, 1967.

7. B. Malinowski, "The Problem of Meaning in Primitive Languages," in C. K. Ogden and I. A. Richards, The Meaning of Meaning (Harcourt, Brace and World, New York, 9th edition, 1968), pp. 296-336.

8. For instance, N. McQuown (ed.), Natural History of an Interview (to be published by Grune and Stratton); A. E. Scheflen, Stream and Structure of Communicational Behavior, Behavioral Studies Monograph No. 1 (Philadelphia, Eastern Pennsylvania Psychiatric Institute, 1965); E. D. Chapple (with the collaboration of C. M. Arensberg), Measuring Human Relations: An Introduction to the Study of the Interaction of Individuals, Genetic Psychology Monographs No. 22, pp. 3-147, 1940.

9. J. D. Call, "Games Babies Play," Psychol. Today 3, 8, 34-37, 54 (1970).

10. The method for predicting expected numbers of runs was developed by Prof. D. H. Klatt.