

## Computer Synthesis of Infant Vocalizations†

Spectrograms of stops produced by infants in the developmental studies currently underway show the necessary features for measuring voice onset time (release of air pressure and voicing). The formants, however, are often poorly defined. This is due, in part, to the fact that the resonances of the vocal tract of an infant are excited by only a few harmonics. In addition, the formant bandwidths of young children's vocal tracts probably differ from those of adults. On the other hand, the stops analyzed in the developmental studies generally show consistent listener agreement. It appears, then, that while stops produced by children are less well structured in an acoustic sense, sufficient information is contained in the speech sound to permit unambiguous perceptual differentiation.

Relatively few studies have been done on the acoustic cues underlying the perceptual differentiation of children's speech. To investigate some of these cues, a project has been initiated to study the acoustic structure of stops in initial position occurring in the vocalizations of children about one year of age. The perception of these stops is being investigated as a function of systematic variation of acoustic dimensions in synthetically produced replicas.

An utterance [bæ] produced by a 13 month old female was chosen as a model for the initial work since it was clearly heard as [bæ] by a set of judges and also showed well defined acoustic features such as fundamental frequency (about 380 cps), voice onset time, apparent vowel formant structure, and easily visible formant transitions. The use of the synthesizer at Haskins has

required a reduction of the measured values of F1, F2 and F3 center frequencies by a factor of 1/2. A synthetic speech sound having the appropriate frequency and time characteristics can then be obtained by tape recording the output of the synthesizer and playing it back at two times the recording speed. This procedure is necessary since computer control of fundamental frequencies has an upper limit of about two hundred cycles (with the synthesizer adjusted to the adult male range) and, in addition, the range of formant frequencies of the synthesizer then lies below those required for children's speech.

To date, satisfactory progress has been made in synthesizing the vowel and transition portions of the model in terms of formant center frequencies and relative formant amplitudes. Present synthesizer settings, however, produce formant bandwidths which appear greater than those observed in the model. A set of judges (16 out of 18) have been able to correctly identify the synthetic production as [bæ]. Much remains, however, to be done in improving the naturalness of the speech sound. Once this problem has been overcome, perceptual studies employing systematic variation of second formant transition and voice onset time will be undertaken.

Malcolm S. Preston

John K. Cullen, Jr.

† The research reported here is part of a cooperative effort with the Neurocommunications Laboratory, the Johns Hopkins University Medical School.