# XV. SPEECH COMMUNICATION*


Academic and Research Staff

Prof. K. N. Stevens
Prof. M. Halle
Prof. W. L. Henke
Prof. D. H. Klatt
Prof. A. V. Oppenheim
Dr. Margaret Bullowa
Dr. Paula Menyuk
Dr. J. S. Perkell

Graduate Students

M. E. Barron
A. J. Goldberg
A. R. Kent
M. F. Medress
R. M. Sachs
G. L. Wickelgren
J. J. Wolf


## A. CHILD'S PRODUCTION OF INITIAL CONSONANT CLUSTERS

### 1. Problem

It has been observed in inventories of the speech sound acquisition of children that the articulation of consonant clusters is mastered comparatively late in the sequence of speech sound mastery.[1] The singleton initial consonants L, R, and S are also late acquisitions. In normal acquisition, the initial consonants L, R, and S are frequently produced with substitutions and distortions.[2] Consonantal clusters are also produced with omissions, as well as substitutions and distortions. One might hypothesize that this is due to the fact that English initial consonant clusters are composed of C + L, (W) or R, or S + C (C), and each element is subject to systematic production errors. The manner in which aspects of initial clusters may be stored in the child's lexicon, however, and the changes that occur in the perception and production of these clusters during the developmental period have yet to be described. Indeed, even the facts of production are largely unknown. No spectrographic, X-ray film, or electromyographic studies of children's articulation of clusters (or other speech sounds) during the period when speech sounds are being mapped into words have been carried out.

A recent linguistic discussion proposed that initial clusters may best be described as single elements in the lexicon. Sequences containing clusters plus a vowel might be described as ($\overset{C}{+}$ features of L, W or R) (V) rather than (C) ( $\begin{Bmatrix} L \\ W \\ R \end{Bmatrix}$ ) (V). The sequence S + L or W are described as strident L and W and the sequence S + C as (S) (C). Segmentalization rules spell out the complete set of features for L, W, and R, and thus separate the elements.[3] The features that distinguish a speech sound from a cluster would be part of the matrix of features for the sound and the cluster in the following manner:


*This work was supported principally by the U. S. Air Force (Electronics Systems Division) under Contract AF 19(628)-5661; and in part by the National Institutes of Health (Grant 2 RO1 NB-04332-06), the Joint Services Electronics Programs (U. S. Army, U. S. Navy, and U. S. Air Force) under Contract DA 28-043-AMC-02536(E).

| | K | Kr | Kl | Kw |
|---|---|---|---|---|
| Obstruant | + | + | + | + |
| Consonantal | + | + | + | + |
| Flat | - | + | - | + |
| Coronal | - | + | + | - |

There is some historical evidence to support the hypothesis that clusters should be considered single elements in the lexicon with some distinctive feature differences indicating differences in point and type of release from those of singleton consonants.[4] An analysis of the process of acquisition of consonantal clusters may give us further evidence of how initial clusters may be stored in the lexicon for productive purposes.

2. Procedure

The utterances of a child, previously recorded at weekly intervals,[5] were sampled at 6-month periods from age 18 months to age 30 months, and again at 34 months. Utterances containing initial clusters were identified and isolated. At the last age at which utterances were sampled, initial clusters were still not being accurately produced in the identified morphemes in the estimation of 4 listeners, and spectrographic analysis indicated that this was indeed the case.

During the first observation, the child produced the following utterances containing initial clusters several times: Brian, truck, plane, and sky. The word Brian was heard as CVC (Bian) and CWVC (Bwian). The word truck was heard as CVC (tuck), CWVC (twuck), and CVC (guck). The word plane was heard only as CVC (bane) and the word sky only as CV (guy). Throughout the age range alternations of CVC and CWVC were heard in some instances of C + L or R; however, instances of CVC decreased and instances of CWVC increased at older age periods. There were instances in which this was not the case. The words <u>driveway</u> and <u>truck</u> were heard as 'jiveway' and 'chuck' at 30 months. During the first two observations (18 and 24 months), S + C clusters were heard as $\emptyset$ CVC (bank for spank). During later observations, however, the rules that the child might be observing to generate these clusters became less obvious. The word <u>sneaker</u> was heard as 'neager' and the word <u>spoon</u> as 'boon,' but the word <u>swimming</u> was heard as 'shimming' and the word <u>swim</u> as 'sim.' Obviously the C in the C + L or R clusters and the C in S + C clusters affected production.

3. Method of Analysis

In tracings of the spectrograms obtained from the utterances heard as Bian and guck and Bwian and twuck, it was observed that there seemed to be a more gradual transition into the vowel in producing the last two words than in producing the first two, although

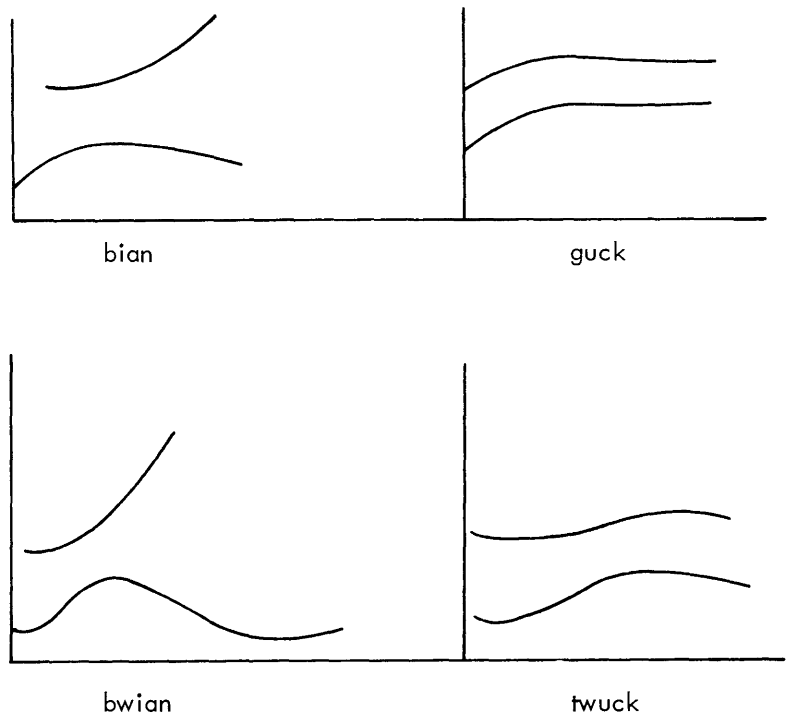


Fig. XV-1. Tracings of spectrograms.

a well-formed W was not being produced. This result is indicated in Fig. XV-1.

To pursue this finding, utterances that contained initial clusters and singleton initial consonants that were the same as those in first position in a cluster plus the same vowel were isolated to see if this distinction could be observed consistently. The utterances Brian and bike; broke, blow, and boat; and truck and tug were the only ones that could be found in the language samples examined. Tracings of spectrograms obtained from these utterances indicated that in the case of clusters there was a more gradual transition into the vowel than in the case of singleton consonants. These tracings are shown in Fig. XV-2. The age at which the utterance was produced is in parentheses.

Spectrograms of children's utterances, however, are often far from clear. A broadband spectrogram of a child's utterance, intended to be the word "Brian," appears in Fig. XV-3. The child has a fundamental frequency of the order of 300 Hz. Therefore the broadband spectrographic display is resolving individual harmonics of the glottal source. The formant frequencies, and thereby the articulatory pattern, must be estimated from the changing pattern of harmonic levels seen in the figure.

A spectrogram of the same word as spoken by an adult male is shown in Fig. XV-4. In this case the formant frequencies are clearly defined. The consonant cluster [br] is manifested as a rapid rise in all formant frequencies immediately following the release (characteristic of a labial stop), and this gesture is followed by a period of time during

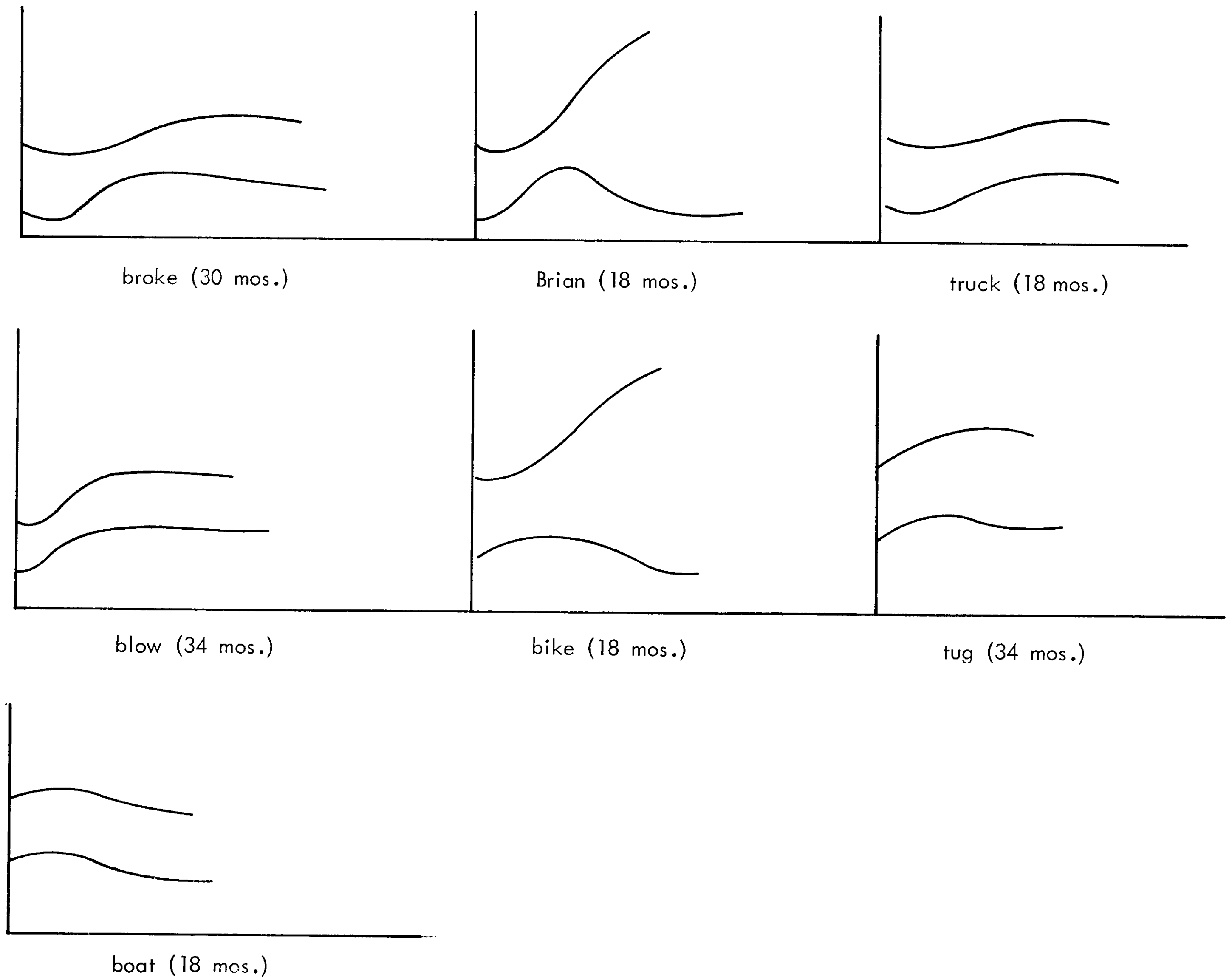


Fig. XV-2. Tracings of spectrograms.

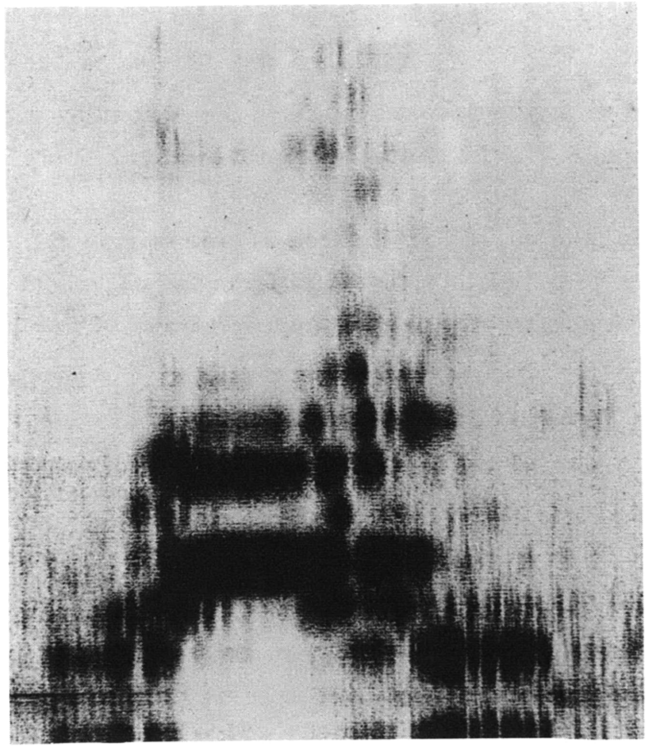

Fig. XV-3. Spectrogram of the word "Brian" spoken at age 18 months.

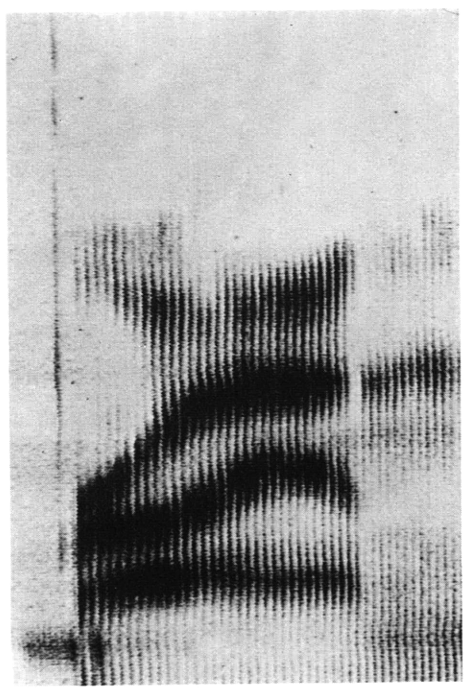

Fig. XV-4. Spectrogram of the word "Brian" spoken by an adult male.

which the third formant is very low and close to the second formant (characteristic of a retroflex articulation). The formants then move continuously through trajectories of the diphthong [aI].

The extent to which the child's production is an adequate replication of an adult production is not obvious. Fortunately, the data from the child includes a contrasting utterance, "bike." A spectrogram of this word appears in Fig. XV-5.

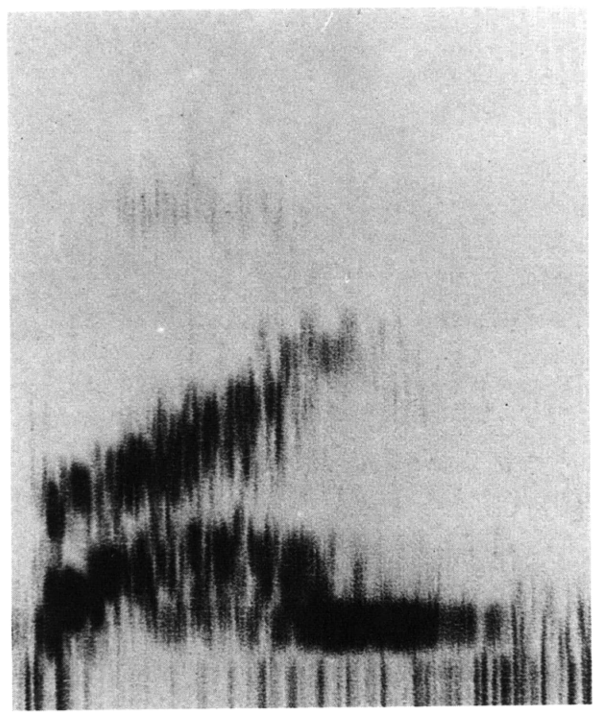

Fig. XV-5. Spectrogram of the word "bike" spoken at age 18 months.

Figures XV-3 and XV-5 do not permit a sufficiently accurate estimation of formant frequency locations to make a definite comparison. Therefore, a computer-controlled filter-bank spectral analyzer was used to supplement the analysis.[6] Figure XV-6 contains one frame from a computer-controlled "spectral movie" of the words "Brian" and "bike" as spoken by the child.[7] The spectral frame corresponds to a time approximately 50 msec after the release of the initial stop in each word.

The third formant is clearly visible and the positions of the first and second formants can be estimated from the relative amplitudes of the harmonics of the fundamental frequency. By watching a slow-motion replay of the spectral movie, it is possible to get an even better feeling of formant locations by the way in which harmonic amplitudes change as a function of time. The natural ability of the eye to estimate a spectral envelope from this moving pattern is thwarted on the spectrogram

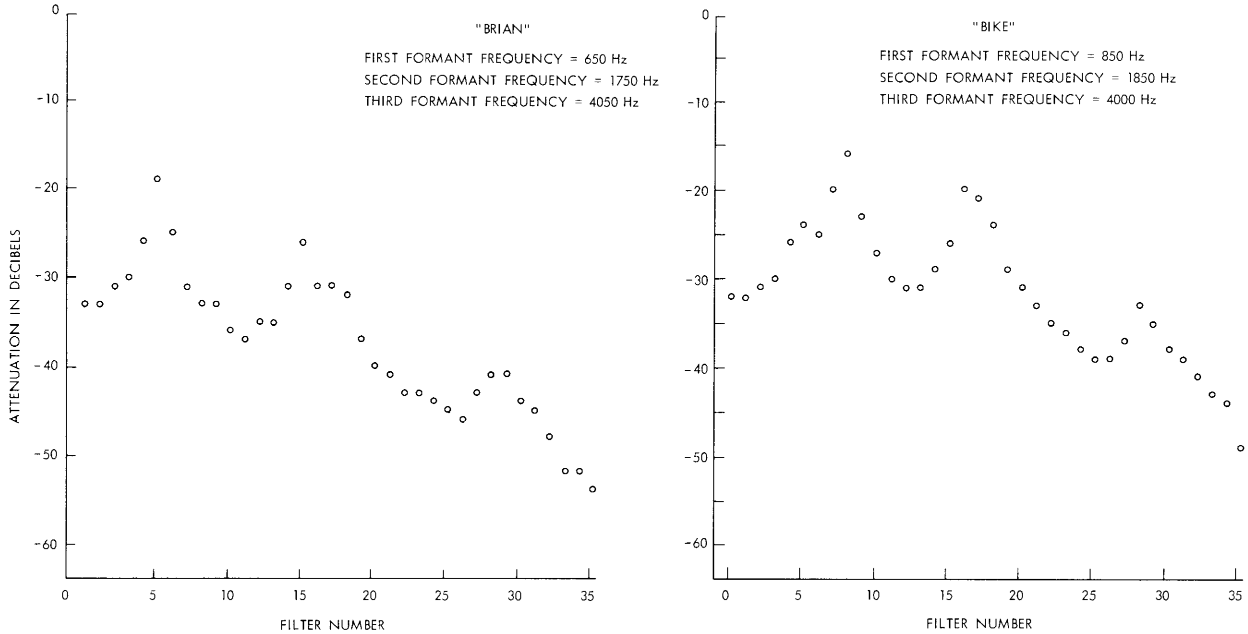


Fig. XV-6. Outputs from a filter bank 50 msec after the release of the initial stop in the words "Brian" and "bike" spoken by the child.

because of the poor amplitude resolution.

The words "Brian" and "bike" have been compared with the aid of the computer display. The results indicate that "Brian" does not have the required lowered third formant following release, but that there are some differences in the initial portions of the two words. If we assume that the peak in the first formant trajectory indicates the closest approach to the [a] articulation, it then follows that the word "Brian" takes longer to reach the [a], and that the first two formant frequencies are lower throughout the utterance than for the word "bike."

From these observations it might be concluded that an adult listener will not hear an [r] when presented with the word intended to be "Brian," but that he is likely to believe that some kind of phonetic segment is interposed between the [b] and the [aI]. This segment is acoustically most similar to a [w].[8]

In summary, then, "Brian" does not contain an acceptable [r], but does show some indication of an initial consonant cluster. The cluster is acoustically most similar to [bw].

4. Tentative Conclusions

Several hypotheses about the production and possible perception of clusters might be derived from the observation of the 4 listeners and the results of the analysis. The first is that CCV syllables are initially always produced as CV, and that the child has yet to observe any distinctions between morphemes composed of CV and CCV sequences. In that case, it would seem reasonable to suppose that either the first segment or the second segment of the cluster would be consistently omitted. In fact, it is always the L and R, the second segments in clusters, or the S, the first segment in clusters, which are consistently omitted. Another hypothesis, then, might be that the child has observed distinctions between CV and CCV syllables, but has yet to produce these distinctions, since it is always the predictable segment of these clusters which is omitted (that is, $C + \begin{matrix}\text{+consonantal}\\ \text{-obstruant}\end{matrix}$ is always C + L, or R; $C + \begin{matrix}\text{+consonantal}\\ \pm\text{obstruant}\end{matrix}$ is always S + stop or nasal). This in turn leads to the hypothesis that the child has acquired some of the morpheme structure rules of his language, and observed some segmental features so that these rules can be applied.

According to the analysis of the data, as well as observer estimation, the following factors may be occurring in the process of mastering the articulation of initial clusters. C + L, W, R, are produced initially with only the C of the cluster (bane for plane, Bian for Brian, guck for truck). Later, they are produced as $C + \left(\begin{matrix}\text{-obstruant}\\ \text{+gradual release}\end{matrix}\right)$ with the release possibly being homorganic (in the same place of articulation as C). Thus Bwian and bwoke are produced, but chuck and jiveway are also produced. (This obviously does not account for the twuck that was heard.) Clusters with S + C are produced as $\emptyset$ + C in the case of stops and nasals (guy for sky and neager for sneaker), but as +strident + $\emptyset$ in the case of W and L (shimming for swimming and sim for swim). Other factors may be operating as well. The voicing of stops in clusters (guck, bane, and guy) indicates, perhaps, that the child has observed the lack of aspiration in clusters.

From these results, one might hypothesize that children have very complicated rules for the production of clusters. On the other hand, one might hypothesize that clusters are indeed entered in the child's dictionary as single segments rather than as a sequence of segments for productive purposes. At age 34 months the rules for filling in the complete set of features (segmentalization rules) and for separating these elements for productive purposes have not yet been acquired. Thus, C + L, W, R may be initially entered as ($\underset{+}{C}$ gradual homorganic release) and additional features then added to the consonant (±flat, ±coronal). S + W or L may be entered as $\left(\begin{matrix}\text{+strident}\\ C\end{matrix}\right)$, and then additional features added to C (±coronal, ±anterior) and S + nasal or stop as $\left(\begin{matrix}\text{+obstruant}\\ C\end{matrix}\right)$, and additional features added to +obstruant (+strident, +coronal, +anterior).

As we have stated, the data are quite sparse and, therefore, a far from full account is possible. The validity of some of these assumptions will be tested in an experiment in which children are asked to identify morphemes with initial consonant clusters or singleton consonants. The various feature contrasts such as place (pr, tr, kr, etc.), voicing (pr, br, etc.) and cluster composition will be examined. The children will also be asked to reproduce sentences containing the same morphemes. The morphemes produced will then be analyzed according to the method discussed.

Paula Menyuk, D. H. Klatt

References

1. D. McCarthy, "Research in Language Development: Retrospect and Prospect," Monogr. Society for Research in Child Development 24, 3-24 (1959).
2. M. H. Powers, "Functional Disorders of Articulation – Symptomatology and Etiology," in L. E. Travis (ed.), Handbook of Speech Pathology (Appleton-Century-Crofts, New York, 1957).
3. T. R. Hoffman, "Initial Clusters in English," Quarterly Progress Report No. 84, Research Laboratory of Electronics, M.I.T., January 15, 1967, pp. 263-274.
4. T. B. Gamkrelidge, "A Typology of Common Kartvelian," Language 42, 69-83 (1966).
5. M. Bullova, L. Jones, and T. Bever, "The Development from Vocal to Verbal Behavior in Children," Monogr. Society for Research in Child Development 29, 101-107 (1964).
6. N. Benheim, "Real-Time Spectral Input System for Computer Analysis of Speech," Quarterly Progress Report No. 80, Research Laboratory of Electronics, M.I.T., January 15, 1966, p. 197.
7. The computer program was developed by J. J. Wolf of the Speech Communication Group of the Research Laboratory of Electronics.
8. An alternative [ℓ], can be ruled out, because of the absence of a stoplike discontinuous increase in the first-formant frequency typical of the release of an [ℓ] into the [aI].