Proceedings of The Institute of Acoustics

THE TARGET THEORY OF SPEECH PRODUCTION IN THE LIGHT OF MANDIBULAR DYNAMICS

BJÖRN LINDBLOM AND RICHARD SCHULMAN

DEPT OF LINGUISTICS, STOCKHOLM UNIVERSITY

1. The concept of vowel target.

In a previous study of vowel reduction (Lindblom 1963) formant frequencies and durations were measured for Swedish vowels produced in summetrical consonantal environments under varying temporal conditions. An exponential curve was fitted to the data to describe the extent to which vowel formant frequencies tended to reach their target values as a function of vowel duration. A target was specified by the formant frequency asymptotes and was found to be independent of consonantal context and duration. It was thus an invariant attribute of the vowel.

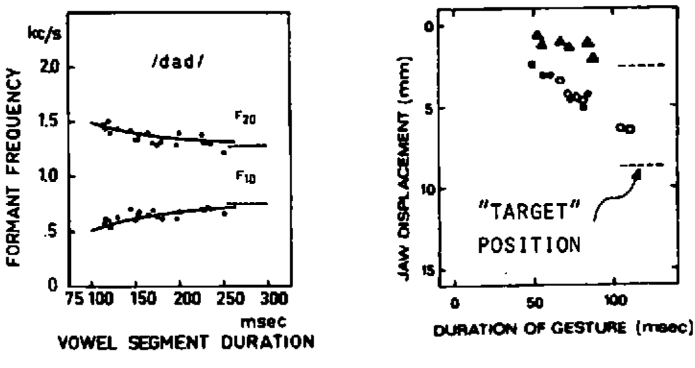

FIG 1 Two aspects of the vowel [a] in [d-d] contexts: Acoustic (left) and articulatory (right) "undershoot" associated with a decrease in vowel duration. Hypothetical targets indicated by dashed lines.

In FIG 1 we present results from two separate investigations. In both halves data on [a] in the environment of [d-d]. To the left acoustic data from Lindblom (1963) is shown. To the right the maximal mandibular excursion during [a] is plotted against the duration of this vowel for four speakers. We used the number of syllables per word to control vowel duration. In other words, rather than asking our subjects explicitly to speak at a particular rate (cf. Gay 1978, Harris 1978, Tuller, Harris and Kelso 1982) we introduced the polysyllabic shortening effect which has been demonstrated for many languages and operates on the main-stress vowel of a word (e.g. rob, robber, robbery). By this procedure vowel duration could be varied over a substantial range while keeping stress constant. It could be argued that this indirect method is a more satisfactory way of controlling "tempo" since it avoids the subjective element involved in interpreting instructions about speaking rate.

2. How reconcile observations of some effort-dependent phonetic variations with the notion of a unique, vowel-specific articulatory target?

FIG 2 shows more data on jaw movement as a function of duration.

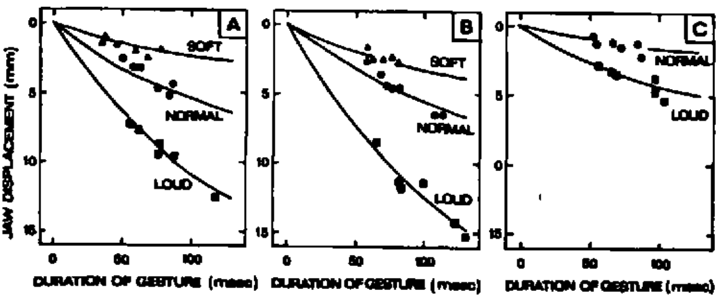

FIG 2 Average measurements of maximal mandibular excursion in [a] as a function of gesture duration (≃ half the acoustic vowel duration). Test words: dadd, daddal, ladadd, daddalal, ladaddal, laladadd. Three degrees of vocal effort. Three subjects.

These diagrams raise the following problems:

(i) If the "undershoot" model of Lindblom 1963 (= that of FIG 1) is applied to these data on [a], it becomes impossible to hypothesize, for this vowel, a single, unique target position for the jaw.

(ii) Does a speaker lower his jaw more when speaking louder irrespective of the vowel produced? Or will he do so only for those vowels for which a mandibular perturbation has less drastic consequences, e.g. for [a] but not for say [i] (cf. Lindblom and Sundberg 1971)?

3. An experiment using effort as a natural, dynamic "bite-block": Is the greater jaw opening of loud speech compensated for?

To address these issues we did an experiment in which five subjects read randomized lists of /hVt/-syllables at normal and loud vocal efforts. High-quality acoustic recordings were made for the purpose of measuring formants by means of an inverse-filtering program written by Liljencrants and Branderud (1981). Vertical lip separation and jaw movement were also recorded using a magnetometer system developed by Branderud.

Informal analyses of the articulatory and perceptual properties of the recorded speech indicate qualitatively similar results for all the subjects. For lack of space and time the following report will present detailed, quantitative data on one subject only.

3.1 Perceptual observations.

A listening test was administered to spot items not perceived as intended by the talker. For the subject of this report all test words were heard correctly in both conditions (cf. Pickett 1956).

3.2 Articulatory observations.

FIG 3 compares average measurements of vertical lip separation and jaw position for the two conditions.

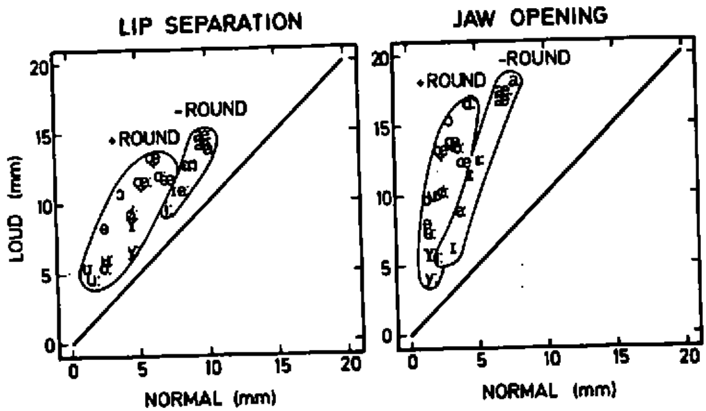

FIG 3 Average data on vertical lip separation and jaw position for Swedish vowels in /hVt/-context. Zero mm's refers to rest and clench respectively.

We note two major points:

(i) Speaking louder is associated with a proportional increase in jaw position for all vowels;

(ii) Vertical lip separation increases similarly for loud speech - but interestingly, not as much. We take this to indicate partial articulatory compensation.

3.3 Acoustic observations.

The formant frequency estimations of FIG 4 were made with the aid of an inverse-filtering program. This procedure bases the determination of formant locations not only on spectral peaks but also on displays of the inverse-filtered waveform and spectrum. The experimenter thus makes his measurements according to criteria such as shapes of residual waveform and of residual spectrum and aims at adjusting the zeros cancelling the vocal tract poles in such a way that certain predetermined features (glottal as well as subglottal) appear in these residues.

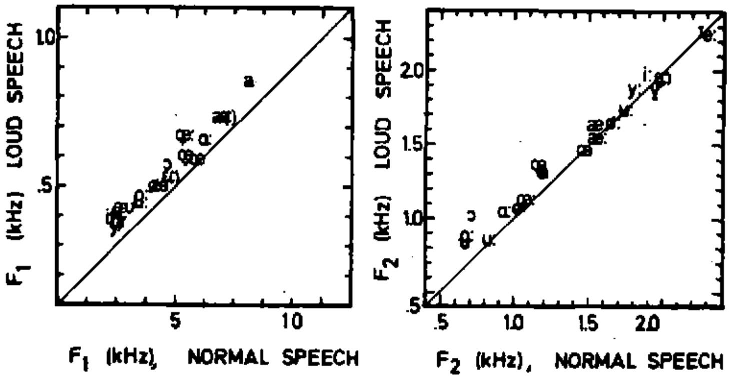

FIG 4 F_1 and F_2 values for normal and loud speech plotted against each other. Average data based on at least four repetitions per condition. Same speaker and test words as analyzed in FIG 3.

The acoustic results are presented in FIG 4 and can be summarized as follows:

(a) F_1 shifts upward in frequency by about 100 Hz for all vowels.

(b) F_2 shifts upward among the back vowels whereby the front series exhibits little change.

(c) These modifications are in qualitative agreement with predictions derived from the mandible-based articulatory model of Lindblom and Sundberg (1971). A quantitative comparison will be undertaken to examine the possibility of vowel-specific effects and to assess the degree to which the loud speech involves compensatory articulatory behavior. Cf. however the lip/jaw interaction of FIG 3 which clearly indicates partial compensations.

4. Tentative conclusions.

Increasing vocal effort appears to be associated with a greater jaw opening for vowels. (This result seems to hold for all five subjects.) From work on articulatory modeling (Lindblom and Sundberg 1971) we would expect mandibular perturbations of this kind to have a rather marked effect on the F-pattern, in particular F_1. The perceptual results showed, however, that such effects were not marked enough to cause any misidentifications among the 21 major vowel qualities of Swedish. The subsequent acoustic analyses confirmed both the model predictions as well as the results of the listening tests in the following way: While it is true that F-pattern shifts in the expected directions did take place they were never so large as to give rise to acoustic ambiguity. The reason for this might be that our main speaker invoked a strategy of partial compensation for the loudness-induced articulatory changes (cf. FIG 3).

With respect to the concept of vowel target our (regrettably imcomplete) results appear so far to be compatible with a view that has emerged from work on compensatory articulation and so-called "bite-block" speech (Gay, Lindblom and Lubker 1981) which says that a vowel target is coded in terms of output-oriented, primarily perceptual goals and that a talker tends to invoke compensatory maneuvers to reach such goals only in so far as they are perceptually motivated (cf. also Perkell and Nelson 1982). We admit that a lot of research needs to be done before such a conclusion can be more than tentative and before observations of such behavior, if general, can be independently justified with reference to psychological and neurophysiological theory.

5. References.

(1) Lindblom, B. (1963): "Spectrographic study of vowel reduction", J. Acoust. Soc. Am. 35, 1773-1781.

(2) Gay, T. (1978): "Effect of speaking rate on vowel formant movements", J. Acoust. Soc. Am. 63(1), 223-230.

(3) Harris, K.C. (1978): "Vowel duration change and its underlying physiological mechanisms", Language and Speech 21(4), 354-361.

(4) Tuller, B., Harris, K.S. and Kelso, J.A.S. (1982): "Stress and rate: Differential transformations of articulation", J. Acoust. Soc. Am. 71(6), 1534-1543.

(5) Lindblom, B. and Sundberg, J. (1971): "Acoustical consequences of lip, tongue, jaw and larynx movement", J. Acoust. Soc. Am. 50, 1166-1179.

(6) Pickett, J.M. (1956): "Effects of Vocal Force on the Intelligibility of Speech Sounds", J. Acoust. Soc. Am. 28(5), 902-905.

(7) Gay, T., Lindblom, B. and Lubker, J. (1981): "Production of bite-block words: Acoustic equivalence by selective compensation", J. Acoust. Soc. Am. 69(3), 802-810.

(8) Perkell, J.S. and Nelson, W.L. (1982): "Articulatory targets and speech motor control: A study of vowel production", In Grillner, S., Lindblom, B., Lubker, J. and Person, A. (eds.): Speech Motor Control, Pergamon Press, London.