ARTICULATORY COORDINATION IN SELECTED VCV UTTERANCES: ACOUSTIC-AUDITORY CONSIDERATIONS

Olle Engstrand
Institute of Linguistics, Uppsala

Introduction

The basic argument of this paper is that the predictive power of articulatory models largely depends on the extent to which they are supported by an explicit theory of the acoustic-auditory properties of speech sounds. More specifically, the purpose is to demonstrate a case where the choice of motor strategy is determined by acoustic-auditory criteria rather than by constraints on the motor system.

The speech sample consists of cineradiographic recordings of a few utterances of the vowel-consonant-vowel (VCV) type, namely ipi, ipu, upi, upu, and ipa. The experimental task is to study the trajectories of the tongue body movements from the first vowel (V_1) to the second (V_2) across the intervening consonant. For the sake of the discussion, we adopt as our null-hypothesis that these trajectories are approximately linear. From an articulatory point of view, this is a reasonable assumption since there are no obvious anatomical restrictions on the mutual independence between tongue and lip movements.

As our alternative hypothesis, we simply take the contrary, i.e., the tongue body trajectories from V_1 to V_2 will deviate from a straight course. If this turns out to be true, further qualifications as to the nature of the deviation will have to be made.

We assume, as a matter of fact, that the null-hypothesis will have to be rejected. This assumption is based on the following two grounds:

1. An optimal articulatory strategy will be such that the crucial acoustic-auditory effects intended by the speaker are kept (a) essentially invariant and (b) free from conflicting noise.

Strong evidence in favor of acoustically invariant properties of stop sounds has recently been offered by Blumstein and Stevens (1979). They showed that properties such as 'diffuse-rising', 'diffuse-falling', and 'compact', residing in the short-time release spectra, are not appreciably affected by changes in the immediate vowel context.

The characteristic (manner) feature of Swedish stops is taken to be a short burst of fricative noise with an abrupt onset. As is well known, however, the production of a voiceless stop is often accompanied by numerous acoustic events. This immediately necessitates a justification of the choice of one particular property, or event, as the 'object of the speaker's intention' in the sense of von Wright (1971, p. 89). To this end, we can try to show that all acoustic events that tend to materialize in connection with the noise burst can be plausibly regarded as causally related to the burst in a certain way (cf. Öhman et al. 1979). Then, in terms of production, the closing of the mouth in combination with an egressive airflow generated in the lungs can be looked upon as adjustments made in preparation for the 'executive' opening gesture through which the intended burst is brought about. Those adjustments make the burst possible, so to speak, not the other way around.

Consequently, the fricative burst is taken to be the primary, intended effect of the articulatory movements in question. The acoustic consequences of preparatory (and postcipatory) movements, e.g. silent interval and formant transitions, will then be regarded as side-effects, secondary to the burst.

2. In order to bring about the primary acoustic properties of a p sound, i.e. a short noise burst characterized by a diffuse-falling spectrum and an abrupt onset, some aero-dynamic conditions have to be met with.

According to Stevens (1971), the pressure drop P_d at a constriction is proportional to the density of the air ρ, the volume velocity U, and the cross-sectional area A of the constriction by

$$P_d = k(\rho U^2/2A^2),$$

where k is a constant.

Now consider a pi or pu sequence. We assume (Stevens 1971) that the cross-sectional area A at the tongue-palate constriction is in the order of 0.3 cm^2 for the high vowels i and u, and, further, that the volume velocity U rises to about 1500 $cm^3 sec^{-1}$ as an immediate consequence of the p release. Then, if both these conditions are present simultaneously, i.e. if the tongue fully anticipates the vowel at the release, the results of Stevens' calculations show that a considerable turbulence noise source will be generated at the tongue-palate constriction. Consequently, the (primary) p burst will be (a) perceptually masked by the secondary noise, and, moreover, (b) physically weakened, since, by the above equation, the lowered volume velocity U which results from the presence of a secondary constriction will reduce the pressure drop P_d across the mouth orifice. Also, (c) the secondary noise will contain frequency components not compatible with an ideal p spectrum.

These being negative consequences of full vowel anticipation by the tongue at the stop release in pi and pu, we assume that some articulatory measure is taken to avoid them. For instance, according to the above formula, a great enough increase in A will, all other things being equal, lead to a reduction of the pressure drop across the tongue-palate constriction. This means that the tongue should not be in a high vowel position at the moment of release. In other words, the tongue body movement from a high V_1 to a high V_2 across p has to deviate from its straight course so that a wide enough air outlet at the release moment is provided for. In the special case of the symmetrical ipi and upu sequences we would, consequently, expect a tongue lowering gesture to coincide with the p, whereafter the tongue body would resume its high vowel position.

Concerning the timing of the initiation of the tongue lowering gesture, we draw on a few pilot experiments with speech synthesis recently carried out in our laboratory[1)]. Those experiments indicate that, if the voice source is maintained slightly too long at a p implosion, listeners tend to hear the sequence bp, for instance ibpi for ipi. We therefore expect the glottis to begin its opening gesture before the labial closure in order to avoid the undesired impression of a voiced consonant. The concomitant rise in volume velocity will then create a turbulence noise if the high vowel constriction is still present. We therefore assume that the tongue lowering gesture will begin before the implosion so that turbulence noise is avoided.

These considerations lead us to reject the null-hypothesis - i.e. that the tongue body movement trajectory follows a straight line - and to accept the alternative hypothesis with the above temporal specifications. We will now compare these predictions with experimental evidence.

Experimental methods[2)]

The results presented in this report are drawn from a larger body of data including simultaneous high-speed cineradiographic and electromyographic recordings made in an attempt to cover various aspects of speech production. Here, the relevant cineradiographic observations will be singled out for inspection. No EMG data will be reported on in this paper.

The subjects were two Central Standard Swedish speakers, one female (LE) and one male (BG), both in the beginning of their thirties. During the recording session the subjects read various VCV utterances for two minutes each using systematically varying rate and stress on each utterance. The utterances studied here were produced at a fairly slow rate with the sentence accent on the VCV part which was embedded in the carrier phrase 'säja p___p igen' ('say p___p again'). The grave word accent characteristic of Swedish compounds was used. In combination with the sentence accent, this gives roughly equal prominence to both syllables.

The observed X-ray frames were sampled from a film of the mid-sagittal aspect of the subjects' vocal tracts taken with a 16 mm cine camera at a speed of 60 frames/sec. Small radio-opaque pellets were attached to three positions along the midline of the tongue dorsum and to one position each on the lower teeth and the upper and lower lips. A reference pellet was attached to the upper teeth. A contrast medium was applied to the lips of both subjects and to the upper tongue surface of subject LE. For the purpose of simultaneous and immediately subsequent EMG-recording, both subjects had needle electrodes implanted in a few muscles (lips, jaw, and tongue). Both subjects reported, after a few minutes practice, that the presence of the pellets and the needles was not felt to disturb their speech.

The subjects were seated comfortably with their heads positioned towards a hollow head-support. No other head-holder was used. The subjects were asked not to move during the run. Their steady position was continuously monitored, and no appreciable movement was detected.

The film analysis was carried out in different ways. Several frames were traced in order to obtain the outlines of the mandible, tongue, lips, and palate at certain points in time. For all test utterances, except those rejected because of subject errors, the x and y pellet coordinates relative to the reference pellet were fed into a computer, and their movements were plotted as a function of time. Those diagrams will be published in a forthcoming report.

Results[3)]

The predictions made in the introductory section are fully corroborated by the data. A consistent cross-subject and cross-utterance feature of the data is the striking non-linearity of the trajectories of tongue body movement from V_1 to V_2 across the intervening consonant (see Fig. 1). The tracings evidence a slight movement away from the extreme V_1 tongue position before the p implosion. This movement goes in the direction of the approximately neutral vocal tract which, except for the lips, characterizes the moment of p release. Inspection of consecutive frames indicate that a movement towards V_2 has started at the moment of release but that most part of the trajectory still remains to be completed at that time. In Fig. 2, the strikingly similar vocal tract shapes at the p release of ipi, ipu, and ipa are shown. A comparison with the fully developed shapes associated with i, a, and u, which can be seen in Fig. 3, reveals that the second vowel anticipation at the release is a very weak one.

In ipi and upu, there is a tongue lowering gesture which coincides with the consonant, i.e., the tongue body does not remain in the i and u

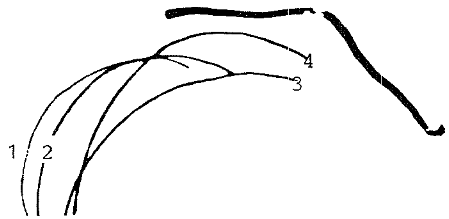

Fig. 1. Tongue contours for utterance upi. 1 = u, 2 = moment of lip closure for p, 3 = moment of p release, 4 = i. Subj. BG.

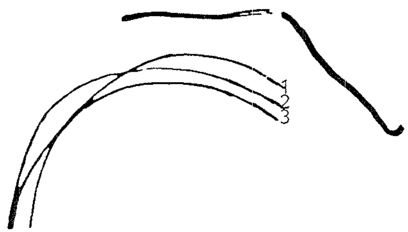

Fig. 2. Tongue contours at the moment of p release for (1) ipi, (2) ipu, and (3) ipa. Subj. BG.

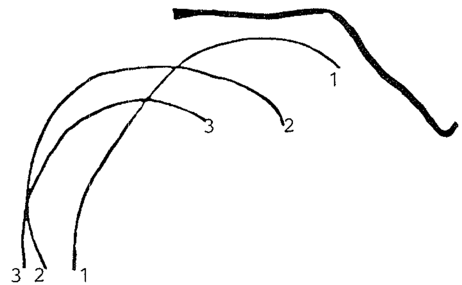

Fig. 3. Fully developed second vowel tongue contours for (1) ipi, (2) ipu, and (3) ipa. Subj. BG.

Fig. 4. Tongue contours for utterance upu. u_1 and u_2 have essentially similar contours, while the tongue body has been lowered at the moment of p release. Subj. LE.

positions even though the second vowel requires roughly the same position as the first one. The sequence upu is exemplified in Fig. 4. The 'trajectories' are, in other words, not straight in the symmetrical case either. Similar phenomena have been observed in American English subjects (Gay 1978).

Conclusion

The results referred to provide a case where one pattern of articulatory coordination (the 'alternative hypothesis') is preferred on its acoustic-auditory merits over another, different pattern (the 'null-hypothesis'). What makes this choice of motor strategy particularly interesting is that it cannot be readily explained on mere anatomical or physiological grounds. This shows the necessity of paying adequate attention to the real object of articulatory action, i.e. the bringing about of specific acoustic-auditory effects.

Notes

1. With Lennart Nordstrand.
2. Further details about experimental methods and procedures will be given in a forthcoming publication.
3. Due to space limitations, only a few representative examples of the data can be shown here.

References

Blumstein, S.E. and K.N. Stevens. 1979. Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants. J. Acoust. Soc. Am. 66(4), 1001-1017

Gay, T. 1978. Articulatory units: Segments or syllables? In Syllables and segments. Eds. Bell and Hooper. North-Holland, Amsterdam

McAllister, R. 1978. Temporal asymmetry in labial co-articulation. Papers from the Institute of Linguistics, University of Stockholm 35

Stevens, K.N. 1971. Airflow and turbulence noise for fricative and stop consonants: Static considerations. J. Acoust. Soc. Am. 50(4), 1180-1192

von Wright, G.H. 1971. Explanation and understanding. Routledge and Kegan Paul, London

Öhman, S., S. Zetterlund, L. Nordstrand and O. Engstrand. 1979. Predicting segment durations in terms of a gesture theory of speech production. In Proceedings of the 9th International Congress of Phonetic Sciences held in Copenhagen Aug. 6-11 1979. Vol. II