Principal components analysis of tongue shapes in symmetrical VCV utterances

Björn Lindblom & Harvey M Sussman
Department of Linguistics, Stockholm University & the University of Texas at Austin

Abstract

Current accounts of speech motor control do not predict tongue movement during labial stops in symmetrical VCV sequences. Nevertheless, there is clear evidence that such movement does occur. A Principal Components analysis was performed on X-ray data from a Swedish speaker. For all vowel contexts, it revealed that, as articulation approached the VCVs intervocalic consonant, the tongue contour tended to change in the direction of a more neutral shape. On the basis of this result (and previous findings) we suggest that the so far unexplained tongue movement (the 'trough' effect) should be seen as a de-activation of V_1 activity during the consonant closure and before the onset of V_2. The theoretical implication of that view is that VCV sequences should be reinterpreted as produced in a phoneme-by-phoneme fashion rather than as combining a diphthongal movement of the tongue body articulator from V_1-to-V_2 with a superimposed movement associated with the intervocalic consonant. Theories of coarticulation must be re-written accordingly.

Introduction

A small, but theoretically significant effect

In this paper we focus on the 'trough'. This phenomenon has been described as an "apparent discontinuity in anticipatory coarticulation" (Perkell 1986) and can be reliably demonstrated in [ibi] and [ipi] sequences; that is, in VCV utterances that have identical vowels surrounding a labial stop. For V_1 there appears to be a 'turning off' of underlying muscle activation with an associated downward movement of the tongue which continues into the stop closure interval. In approaching the second vowel the tongue returns to the target of the first vowel.

A classical demonstration of the trough effect is found in Houde's 1967 monograph. It reports X-ray data on tongue lowering during the stop in VbV utterances. Look-ahead models of anticipatory coarticulation (Henke 1966) expect maximum V-to-V coarticulation to occur across a labial stop, especially for /i/-to-/i/ trajectories, so as to allow maintaining a high fronted tongue throughout the stop. At the same time, Öhman (1967) published his numerical model which is compatible with Henke's notion but contrary to Houde's data. It embodied the idea that in VCV utterances two independent movements take place: (1) a diphthongal movement of the tongue body articulator from V_1-to-V_2; and (2) a superimposed movement associated with the intervocalic consonant.

Results similar to Houde's were reported by others a few years later. Recently we undertook a review of the experimental history of the trough (Lindblom et al, submitted). Referring to a sizeable number of EMG and EPG studies we show that the phenomenon is subtle, both articulatorily and acoustically, but at the same time: It is definitely there.

What do we make of it? It clearly violates well accepted notions of anticipatory, right-to-left coarticulation. It is at variance with Öhman's conception of VCV as a diphthong with a consonant superimposed.

Is the trough sending us a theoretically important message? We believe it is. "Exactly where is the tongue going during labial stops? And why is it going there? To shed some more light on those questions, we undertook the following study.

Experimental procedures

X-ray data analysis

Our data comes from the X-ray database created in collboration with the Department of Radiology at Danderyd Hospital, Stockholm (Stark et al 1999). The procedure produced digital films showing the subject in lateral profile. They were recorded at 50 images/second with synchronized sound.

For this study we examined a male Swedish speaker producing sequences containing voiceless aspirated and voiced stops in symmetrical vowel contetxts: *i bipil, e bepek, a bapar, å båpål, o bopol.* Second test word was pronounced with the grave accent.

The images were displayed using Osiris software (see ref). After the application of sharpening, contours were traced and saved with each image. The tracings were extracted by opening the images in Papex 1.0, a program for producing an xy table for each traced contour (Bresin). Contours included all the information needed to construct a VT area function from the profile, i.e., the outlines of teeth and hard palate, lips, jaw, tongue (root-to-tip), epiglottis and larynx, back wall of pharynx, uvula and soft palate.

In Excel the following steps were then taken for each image: (i) Conversion of contour tables (in image pixel coordinates) into millimeters, a step that involves use of a calibration grid recorded in each speaker's midsagittal plane. (ii) Compensation for head movements. This correction was based on a thin 5 cm wire attached from upper incisors and running posteriorly along the hard palate.

The analyses of the tongue contours were performed in a jaw-based coordinate system. The tracing method produces a number of sample points that tends to vary from frame to frame and that are rarely equidistant. For the purpose of the Principal Components analysis (see below) every contour was re-sampled as follows: a. Total contour length is first measured by summing linear distances between adjacent points on the contour. b. The x- and y-values of each contour is next plotted separately against cumulative contour length. c. High-order polynomials are fitted to these plots to provide separate equations for the x- and y-coordinates as a function of cumulative length. d. Step size is determined by dividing total length by the sampling rate. The present analyses used 25 points. e. New x and y equidistant values are then computed using the polynomial expressions. This procedure was applied to all the selected frames of all the test words (85 frames).

Principal Components Analysis

The 85 frames so obtained were subjected to Principal Components analysis (Maeda 1990). The output of this method is (i) a small set of Principal Components; (ii) a table of scale factors describing the weight that each Principal Component carries in describing any given contour. (iii) the variance accounted for by each PC.

Use of Principal Components allows the investigator to recreate any individual observed contour as a weighted sum of the PC's extracted. The accuracy of the description is determined by how many PC's are chosen. The more PC's, the higher the numerical accuracy. However, for most applications, 3 to 4 PC's tend to yield sufficient accuracy. In the present project we found that 2 PC's accounted for 98.6% of the total variance.

A common way of displaying the results of a PC analysis is to plot the the weights of the first two PC's for all the data. This format provides a 2D representation of the articulatory space of the data analyzed, see e.g., Figures 3 and 4.

Results

The jaw-based movement of the tongue.

The duration of vowel and consonant segments were measured from wide-band spectrograms. The degree of jaw opening was determined for each image. It was defined as zero for clenched teeth and elsewhere as the Euclidean distance between the upper and lower incisors. The resampling of the tongue contours yielded 25 'fleshpoints'. For a selected subset, their location along the vertical and horizontal dimensions were plotted as a function of time. Fig 1 shows a typical example of such a plot.

At the top: the vertical displacement of a given fleshpoint on the tongue dorsum during [iːbiːp^{h}iːl]. The lower curve is the jaw opening which remains stable for this token. Note the distinct minima, i.e., 'troughs', in the top curve. They occur during the [b] and the [p^h] segments respectively. Similar results were obtained for the other test words.

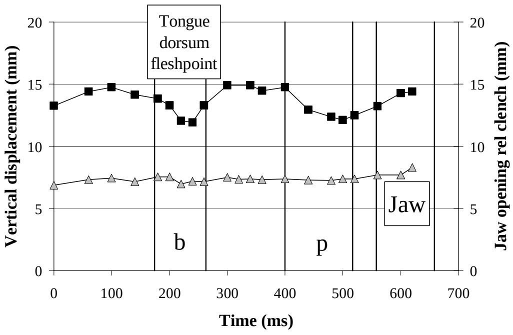

Figure 1. Top: Vertical displacement of point on tongue dorsum during [iːbiːp^{h}iːl]. Bottom: Jaw opening relative to clench.

Results for [ɑːbɑːp^{h}ɑːr] are shown in Figure 2 below. In this case there is more jaw movement. That component has been factored out in the diagram below.

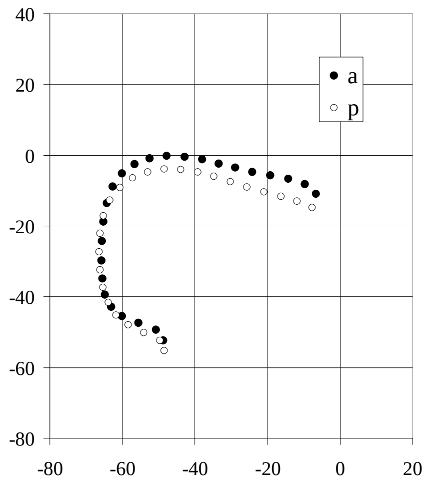

Figure 2. Comparison of tongue shapes for [aː] and [p] in [aːbaːp^{h}aːl]. Scales in mm.

For all the tokens analysed here it is clear that the tongue does not remain immobile during the labial stop closures. The question is: What is the nature of the movement?

In order to further examine the trough effect we decided to subject the present set of tongue contours to a Principal Components Analysis.

PCA of tongue movements.

The tongue contour data formed a matrix with 25 columns (fleshpoints) and 170 rows (x- and y-coordinates of 85 images, an average of 17 frames per test word).

When the weights for the first Principal Component were plotted for x values on the abscissa (Horizontal Component) and the y values on the ordinate (Vertical Component), the patterns in Figure 3 appear.

We begin by presenting the [i, a, u] data. Note that the numerical ordinate values of the PCA have been reversed to produce greater similarity to conventional vowel charts. Roughly speaking, Vertical dimension = open-close; Horizontal = front-back.

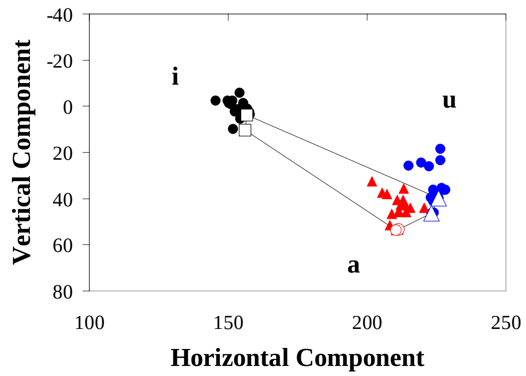

Figure 3. Corners of vowel space portrayed in terms of the weights for x and y values of first PC.

Figure 3 shows all the data points for *i bipil*, *a bapar*, *o bopol*. The minimum points (read: 'troughs') of the tongue movement in Figure 1 are plotted with unfilled symbols. The same was done for [a] and [u]. Note that these points - connected by lines for clarity - tend to be placed at the 'lower' ends of the distributions.

Before proceeding, we need to add a note on the phonetics of the dialect spoken by the subject. He pronounces /e:/ and /o:/ as many speakers of his dialect and age do to-day, *viz* with a schwa off-glide: [eə] and [oə].

What do those results mean? Figure 4 attempts to address that question and includes data also for *e bepek* and *å båpål*. As we examine the complete diagram with all five vowel contexts and with the open symbols for the minima (troughs), a certain lawful pattern emerges. The minima of [eə] and [oə] seem somewhat deeper than those of the other contexts. *Again they occur at the lower ends of the distributions!* Recall though that in those two cases, the troughs are adjacent to schwa-like articulations. It is as if, for /e:/ and /o:/, the troughs have slipped out into the open and have clear acoustic consequences: For instance, in *e bepek*, F1 is clearly seen to increase towards end of the vowel segments.

The overall impression evoked by the points in Figure 4 is that of a wedge converging on a point in the lower right of the diagram (drawn by hand in Figure 4). That point, we would like to argue, represents a state of 'de-activation' of the tongue to which it returns irrespective of the identity of the preceding V_1 articulation. It is this 'back-to-baseline' movement that underlies the trough.

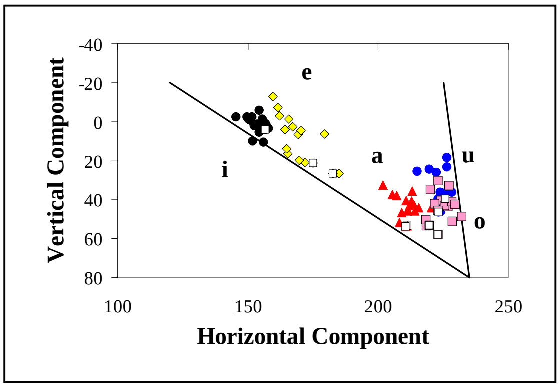

Figure 4. The complete set of data (five utterances, 85 film frames) represented in terms of the weights for x and y values of the first PC.

Discussion and conclusions

The present data are compatible with the large number of studies that have demonstrated the existence of the articulatory 'trough'. The tongue does *indeed* move during the closure of labial stops in symmetrical VCVs. It is significant that this effect is similar for the voiced and the voiceless aspirated labial stops, [b] and [p^h].

It might be assumed that a small dip in the articulatory records of VCVs should not force us to re-evaluate the way we are used to thinking about coarticulation. However we believe that the 'trough' has enormous significance in how we conceptualize the production of speech.

On the basis of the present X-ray findings and PC analyses, we feel justified in proposing that the most reasonable interpretation, for now, is that the 'trough' represents a *de-activation* of the muscle activity for V_1 during the consonant closure and before the onset of V_2. Moreover, the theoretical implication of that view is that VCV sequences should be analyzed as produced in a phoneme-by-phoneme fashion rather than as combining a 'diphthongal movement' of the tongue body articulator from V_1-to-V_2 with a superimposed movement associated with the intervocalic consonant (cf Öhman's co-production model). New frameworks must be developed that are in closer agreement with the '*overlapping neural waves*' scenario first discussed by Joos in 1948.

Acknowledgements

This research was supported by NIH (Grant No 5 R01 DC02014, NIDCD to UT Austin (Sussman) and Subcontract #UTA00-452 to Stockholm University (Lindblom). Special thanks are due to Hans-Jerker Lundberg and Jaroslava Lander for their roles in the acquisition of the X-ray data base.

References

Bresin R (1998). PapEx 1.0. Dept of Speech, Music and Hearing, KTH Stockholm, Sweden.

Henke W (1966). Dynamic articulatory model of speech production using computer simulation. *Ph D dissertation*, MIT.

Houde R A (1967). A study of tongue motion during selected speech sounds. *Ph D dissertation*, U. Michigan (1967).

Joos M (1948). Acoustic Phonetics. *Lg 24*, 1-136.

Lindblom B, Sussman H M, Ghavami G M & Burlingame E (2002). The trough effect: Implications for speech motor programming. submitted, *Phonetica*.

Maeda S (1990). Compensatory articulation during speech: Evidence from the analysis and synthesis of vocal-tract shapes using an articulatory model. In: Hardcastle W J & Marchal A, eds, *Speech Production and speech modeling*, 131-149, Netherlands: Kluwer.

Osiris Medical Imaging 3.5. University Hospital of Geneva, Digital Imaging Unit. Switzerland. http://www.expasy.ch/UIN/html1/UIN.html

Perkell J (1986). Coarticulation strategies: preliminary implications of a detailed analysis of lower lip protrusion movements. *Speech Comm. 5*: 47-68.

Stark J, Ericsdotter C, Branderud P., Sundberg J, Lundberg H.-J & Lander J (1999). The APEX model as a tool in the specification of speaker-specific articulatory behavior. *Proc XIVth ICPhS, San Francisco*.

Öhman S E G (1967). Numerical model of coarticulation. *J Acoust Soc Am 41*, 310-320.