ARTICULATORY COORDINATION IN CORONAL STOPS: IMPLICATIONS FOR THEORIES OF COARTICULATION

Christine Ericsdotter, Björn Lindblom & Johan Stark
Stockholm University

ABSTRACT

Articulatory data obtained from X-ray films were analyzed to investigate articulatory coordination in Swedish vowels and dental and retroflex stops. Several quantitative frameworks for predicting observed tongue contours were evaluated. Performance was assessed on a set of articulations traced at the V_1:C boundary in V_1:CV_2 sequences. The best model, as determined by an rms-criterion, had 3 parameters: length of blade (dental/retroflex), degree of coarticulation, pharynx width. According to this account the coarticulatory (vowel-dependent) variations in [d] and [ɖ] tongue contours are seen as arising from the overlap of a gradual deactivation of the V_1: gesture and the movement towards a vowel-independent consonant target, a process successfully captured by interpolation in the model. This result is discussed in terms of co-production (Öhman, [1]) and the overlapping innervation wave theory (Joos [2]).

1. COARTICULATION AND 'LOCUS EQUATIONS'

There are probably few phonetic topics that have generated so much experimentation and theoretical speculation as coarticulation. Nonetheless, it is fair to say that, despite massive research efforts, our understanding of the physiological origins and perceptual motivation for this phenomenon still remains rather incomplete.

One approach to coarticulation is the use of 'locus equations' as promoted by Sussman and colleagues [3]. This format arises from the fact that the F2 onset of a consonant tends to be a linear function of the F2 of the following vowel. Numerous studies have shown that LE slopes & intercepts vary systematically and robustly with consonant place. Descriptively LE's can be seen as restating the fact that adjacent phonemes interact physically. In capturing that interaction, LE's offer a way of quantifying 'coarticulation'.

What are constraints that determine possible 'locus' patterns? Are LE's always linear? When linear, what is the origin of their slopes and intercepts? According to Sussman et al [4] LE's have a strong perceptual motivation, whereas Fowler [5] maintains that LE properties (linearity, slopes, intercepts) are inevitable consequences of production constraints. On this view a consonant produced at a certain place is supposed to offer a fixed degree of articulatory 'resistance' to adjacent vowels. Therefore, within a given stop place category, coarticulation tends to be uniform across vowel contexts. LE's simply reflect uniform coarticulation. We shall return to this issue at the end of the paper.

2. AN X-RAY DATABASE

Recent collaboration with Danderyd Hospital in Stockholm on an X-ray project provides a unique opportunity to re-examine some of the traditional issues raised by coarticulation. So far the database consists of X-ray films of 11 subjects (Swedish, Tamil, Hindi) recorded for 20 seconds at 50 images per second. Speech samples were selected to elucidate how movements are coordinated in the coarticulation of vowels and labial, dental, retroflex and velar stops. For background and methodology of this research see Stark et al, this conference [6].

3. QUESTION AND SPEECH SAMPLE

Here we summarize the results of an in-depth analysis of the dental-retroflex contrast for a single Swedish subject [7]. The relevant speech samples are listed in Table 1 below.

Table 1. The speech samples. The first vowel carries main stress and accent 2. The second vowel is unstressed.

Consonant place	**Vowel (V_1)** *anterior*	*posterior*
dental [d] *anterior*	y:dɛ e:dɛ	u:dɛ o:dɛ ɑ:dɛ
retroflex [ɖ] *posterior*	y:ɖɛ e:ɖɛ	u:ɖɛ o:ɖɛ ɑ:ɖɛ

The question addressed is: How is the movement of the tongue modified when the tongue blade gesture for [d] and [ɖ] interact with tongue body motion in various V_1_V_2 contexts?

4. TWO VIEWS OF COARTICULATORY OVERLAP: JOOS AND ÖHMAN

A classic attempt to account for VCV coarticulation as observed in X-ray data, is the numerical articulatory model proposed by Öhman [1]. It is summarized in equation 1:

$$s(x,t) = v(x) + k(t)[c(x)-v(x)]w(x) \qquad (Eq\ 1)$$

Here x represents position along the vocal tract, and t is time. Eq (1) states that, at any given moment in time, the shape of the tongue, $s(x)$, is a linear combination of a vowel shape, $v(x)$, and a consonant shape: $c(x)$. As the interpolation term, $k(t)$, goes from 0 to 1, a movement is generated that begins with a pure vowel, $v(x)$, and then changes into a consonant configuration which will to some extent retain aspects of the vowel contour owing to the weighting function, $w(x)$.

Particularly noteworthy is the fact that Öhman succeeded in deriving the observed context-dependent shape variations for each phoneme from a single, invariant description of the underlying vocal tract shape. That is, for a given [V_1dV_2] sequence, each vowel had its unique *v(x)*. The consonant [d] was specified by a single context-independent *c(x)* and its associated similarly unique "coarticulation function" *w(x)*. By many, this finding has been taken to suggest that phonetic invariance is present at an articulatory level but absent in the acoustics owing to the *'co-production'* of phonetic gestures.

Another point made by this research concerns how coarticulation is assumed to be motorically implemented. VCV coarticulation is seen as a vowel-to-vowel movement on which an intervocalic consonant is superimposed. Let this organization be termed the *dual-channel* mode.

Justifiably, this work has had a strong impact on the field. However, its influence appears to have been that a competing scenario, the *overlapping innervation wave theory* proposed earlier by Joos [2], has more or less taken the back seat. This theory was also presented to account for coarticulation effects, but, significantly, it did not postulate two separate channels but more parsimoniously assumed that speech is produced *phoneme by phoneme*.

The key difference between these two scenarios is linked to whether vowel production can be *deactivated* or not during the consonant. In the dual-channel mode, there is no de-activation of vowel gestures, V_1 always changing smoothly into V_2 in a *legato* fashion. In the overlap pictured by Joos, on the other hand, there is indeed a possibility for V_1 to die out before V_2 begins, since no dual grouping or V_1-to-V_2 transition is imposed.

Clearly, these alternatives would on occasion be expected to show marked acoustic differences.

5. FITTING ÖHMAN'S NUMERICAL MODEL TO THE X-RAY DATA

Attempts were first made to fit Eq (1) to [d] and [ɖ] across vowel contexts. The data consisted of tracings made at the V_1C boundary. First note that, for each test word, *s(x,t)* and *v(x)* can be measured from the X-ray tracings. The interpolation factor *k(t)* becomes unity during the occlusion. The remaining terms, *c(x)* and *w(x)*, are unknown and have to be computed. That is done by setting up the equations for *s(x,t)* in two vowel contexts, plugging in the known and solving for the unknown terms. There are five vowel contexts per consonant in this speech sample which means that ten pairs of equations were available for deriving ten paired estimates of *c(x)* and *w(x)*.

Using this procedure, we found that these estimates were too different to be meaningfully averaged or collapsed into a single pair of 'context-independent' *c(x)* and *w(x)* terms. Such strong vowel dependence was not reported by Öhman. Unfortunately, deducing *c(x)* and *w(x)* from Equation 1 is not feasible under such conditions.

The preceding point can be made more simply. As shown also by Dixit [8] and Krull et al [9], we found that the place of articulation in [ɖ] varied markedly as a function of the preceding vowel. To produce such vowel-dependent variation in place of articulation with the Öhman model, separate target shapes *c(x)* (and their associated *w(x)* functions) must be used for each vowel context.

6. MAPPING ARTICULATORY OBSERVATIONS ONTO APEX PARAMETERS

In a second attempt to quantify the articulatory data, we compared the X-ray tracings with contours generated by APEX as calibrated for the speaker under analysis. The APEX model is described in Stark et al, this conference [6]. An example of this matching procedure is given in Figure 3 of that paper. In this model, the tongue blade, a parabolic function attached to the tongue body, is controlled by two parameters, protrusion and elevation. The tongue body is specified by two parameters: anterior-posterior position and deviation from neutral. The position dimension ranges from palatal to pharyngeal via velar. Displacement regulates the size of the tongue 'hump', zero displacement corresponding the neutral contour. The class of 'possible tongue bodies' is generated by interpolating between four reference configurations. They correspond to contours observed for [i], [u], [ɑ] and 'rest'. The subject's geometry of fixed VT landmarks as well as the reference shapes were measured from the X-ray films. 'Rest' was defined as the schwa occurring in [eːdɛ] and [eːɖɛ] in which [eː] was produced as [eᵊː].

In its current implementation this model showed certain limitations. Essentially derived from vowels its repertoire of tongue body shapes was found to be too limited to adequately describe the contours for [d] and [ɖ]. A major shortcoming was that APEX tongues were not 'flat' enough. This observation is illustrated in Figure 1 which compares the tongue contour for a schwa (the flattest tongue in the APEX space) with [d] and [ɖ] contours sampled after schwa.

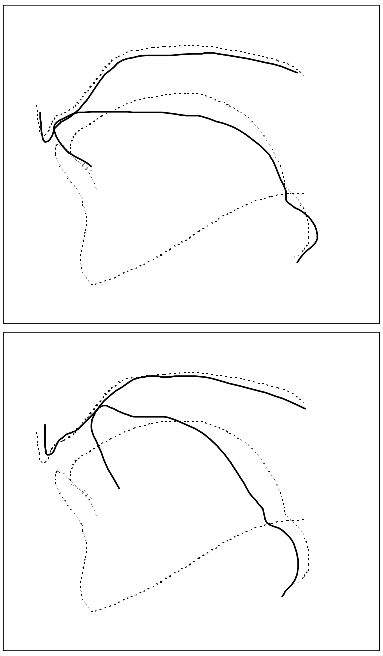

Figure 1. Comparing the contours for the consonants after ə with the vowel ə. Top: dental articulation, bottom: retroflex articulation.

7. COARTICULATION AS GESTURAL OVERLAP & DEACTIVATION: A PHONEME-BY-PHONEME ACCOUNT

The information in Figure 1 proved crucial in suggesting a third quantitative approach. A satisfactory solution to the problems reviewed above was obtained after the following two assumptions had been adopted (cf the Joos scenario).

> (a) Articulators that are not active default to rest position. In other words, muscles active only for the vowels shut down during the consonant;
>
> (b) Raising the tongue blade for dental [d] and retroflex [ɖ] affects not only the anterior portions of the tongue (as currently assumed in APEX), but reshapes the entire tongue contour all the way to the root.

In keeping with these assumptions, we stipulated that the tongue contours observed after schwa be defined as *target contours* for [d] and [ɖ], (analogous to *c(x)*). The schwa was assumed to approximate a state of 'deactivated' vowel musculature. The tongue contours predicted for the V_1C boundary were then derived by interpolating by a *single* coefficient between the V_1 contour and the target shapes for [d] and [ɖ] respectively. This rule together with a parameter of '*pharynx width*' and a two-valued parameter of '*tongue blade length*' (dental or retroflex) was found to produce highly accurate matches to the empirical contours (cf rms-values in mm in Table 2 below).

In Figure 2 a representative example is shown. The diagram shows four tongue contours plotted in a mandible-based coordinate system: (i) the contour for [y:], (ii) the consonant target; (iii) the observed consonant contour; and (iv) the predicted consonant contour.

The table below contains the results of fitting this three-parameter model to the data. The blade parameter refers to the length of the blade contour (in %). The columns labeled *degree of coarticulation* contains the interpolation coefficient used in the predictions. By definition this number varies between 0 and 1. At each point along the contour it was applied to the distances between the vowel - consonant target contours. In the pharynx a correction was made by superimposing a curve of fixed shape but whose amplitude varied and is given in mm in the table. Note especially the large vowel-dependent variations in degree of coarticulation which indicate that, for the present speech samples, coarticulation was strongly *non-uniform*.

Table 2. Parameter values and matching results.

speech sample	**degree of coart.**	**pharynx corr.** (mm)	**blade length** (%)	**r m s** (mm)
y:dɛ	0.48	-8	100	1.4
y:ɖɛ	0.81	-2	70	1.5
u:dɛ	0.47	-9	100	2.8
u:ɖɛ	0.61	-7	70	1.6
o:dɛ	0.32	-3	100	1.1
o:ɖɛ	0.15	-4	70	1.5
ɑ:dɛ	0.54	2	100	0.5
ɑ:ɖɛ	0.35	-2	70	2.3

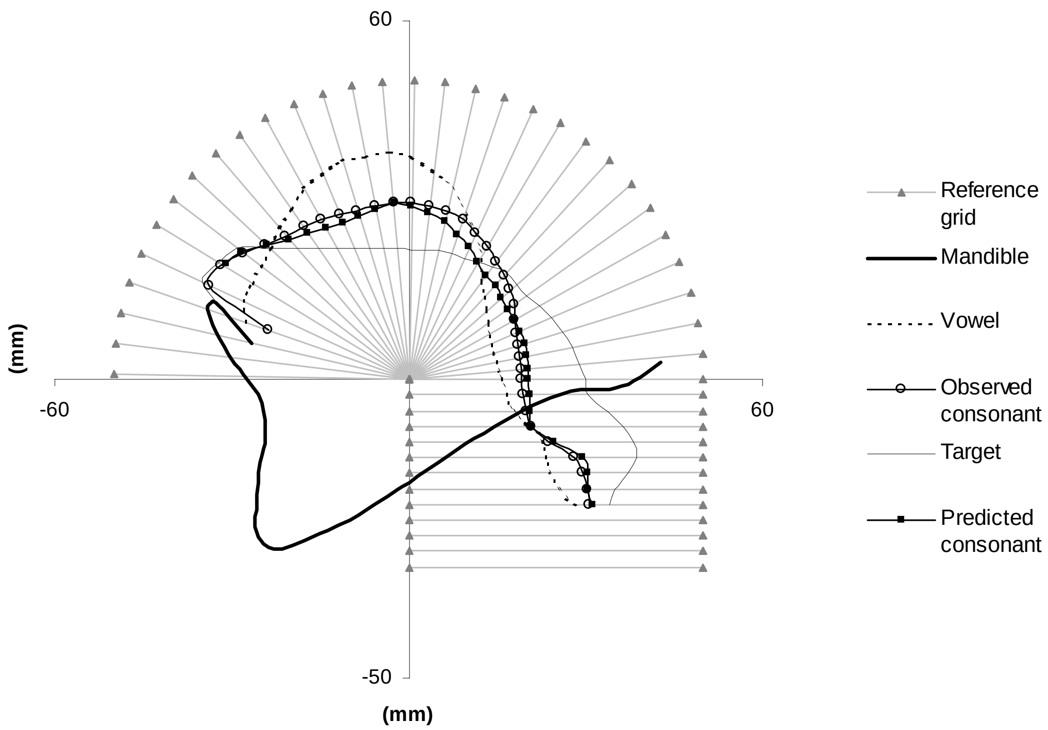

Figure 2.

Returning to Figure 2 we can examine how these parameters interact. Note that, in the anterior part of the diagram, the observed consonant contour is about half-way between the vowel and the consonant target. Here the coarticulation coefficient is 0.48. In the tongue root region the consonant curve stays close to that for the vowel. The predictions reflect that effect as a result of the -8 mm pharynx correction.

8. CONCLUDING REMARKS

How is the movement of the tongue modified when the tongue blade gesture for [d] and [ɖ] interact with tongue body motion in V_1_ε contexts? Despite the limited number of samples analyzed so far, a few preliminary remarks seem possible. Raising the tongue blade for [d] and [ɖ] affects not only the anterior portions of the tongue, but modifies the entire tongue contour all the way to the root. We showed that this effect could be satisfactorily captured by letting the [d] and [ɖ] tongue contours after schwa be defined as *target contours*. In addition, the consonant contours exhibited significant non-uniform (!) vowel-dependent changes. For the present V_1C samples those effects were accurately modeled by interpolating between V_1 and the consonant target and making the numerical degree of coarticulation vowel-dependent.

The present findings differ from Öhman's results in two important ways. Our definition of consonant target is more in the spirit of the Joos theory than the co-production scenario. Unlike the latter, the former does allow for vowel deactivation during the consonant.

The second point concerns the non-uniform coarticulation effects. Conceivably, they might originate from the fact that different vowels deactivate with different time constants.

If the non-uniformity effect holds up in future articulatory analyses, the regularities captured by locus equations can not be dismissed as inevitable consequences of production constraints. Clearly more articulatory data and further quantitative modeling are needed to resolve those issues.

ACKNOWLEDGMENTS

This work is supported by the Swedish Council for Research in the Humanities and Social Sciences (HSFR F0707/97) and the Bank of Sweden Tercentenary Foundation (RJ 95-5173:03).

REFERENCES

[1] Öhman, S. 1967. Numerical Model of Coarticulation. *J. Acoust. Soc. Am.* 41, 310-320.

[2] Joos M. 1948. Acoustic Phonetics. *Language* 24, 1-136.

[3] Sussman H M, McCaffrey H & Matthews S. l991. An investigation of locus equations as a source of relational invariance for stop place categorization. *J Acoust Soc Am*, 90:1309-1325.

[4] Sussman H M, Fruchter D D, Hilbert J & Sirosh J. 1998. Linear corrrelates in the speech signal: The Orderly Output Constraint. *Behavioral & Brain Sciences*, 21, 241-299.

[5] Fowler C. 1994. Invariants, specifiers cues: An investigation of locus equations as information for place of articulation. *Percept Psychophys* 55, 597-610.

[6] 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. *Proceedings from the XIVth ICPhS, San Francisco*, August 1-7 1999.

[7] Ericsdotter, C., Lindblom, B. Stark, J., & Sundberg, J. 1998. The coarticulation of coronal stops: A modeling study. *J. Acoust. Soc. Am.* 104, 1804.

[8] Dixit, R. P. 1990. Linguotectal contact patterns in the dental and retroflex stops of Hindi. *J. of Phonetics* 19: 189-201.

[9] Krull, D., Lindblom, B., Shia, B.-E. & Fruchter, D. 1995. Cross-linguistic aspects of coarticulation: an acoustic and electropalatographic study of dental and retroflex consonants. *Proceeding of the XIIIth ICPhS, Stockholm, Sweden, on 13-19 August*, 1995, 436-439.