Coarticulation as incomplete interpolation

Björn Lindblom[1]*, Diana Krull*[1] *& Harvey M Sussman*[2]
[1]*Dept of Linguistics, Stockholm University, Stockholm Se 10691 Sweden.*
[2]*Dept of Linguistics, Austin Texas 78712 USA.*

Abstract

The main result of this project is that we successfully replicated Öhman´s (1966) study and were able to describe both his and our own data using 2-D locus equations. Restating this finding we can say that, for V_1CV_2 with C = [b], [d] or [g], F2 transition offsets and onsets could be accurately predicted from information on the F2 values of V1 and V2 and the identity of the consonant. These results are fully compatible with an account that describes speech movements as unfolding phoneme-by-phoneme and with de-activation of articulatory structures not recruited by the current phoneme. According to this view vowel-dependent coarticulation of consonants in V_1CV_2 utterances can be characterized as a process of incomplete interpolation between articulatorily specified goals.

Background and goals

Studies of coarticulation usually take the classic work of the sixties as a point of departure. The present report is no exception.

Öhman (1966) investigated formant transitions for [b], [d] and [g] in V_1CV_2 sequences. He found patterns incompatible with a context-independent ***'locus'*** value for each place of articulation as proposed by Delattre et al (1955). Öhman's work shaped how phoneticians subsequently came to look at the relationship between phonetic segments and their acoustic correlates. Liberman & Mattingly (1985) captured that view in an often quoted remark:

"…… there is simply no way to define a phonetic category in purely acoustic terms".

Later Sussman and colleagues redefined 'locus' as the observable onset of a F2 transition (Sussman et al 1991). Analyzing CV syllables they used 'locus equations' (LE:s) - F2 transition onset plotted against F2 midpoint of the following vowel - to demonstrate robust linear relationships.

$$F2_{onset}(V2)=a+b*F2_{mid}(V2) \qquad (1)$$

In a number of publications, Sussman and colleagues have shown that the slopes and the intercepts of the LE metric provide distinct representations of place of articulation in stops and, as suggested by Krull (1987), can serve as useful indices of degree of coarticulation (Sussman (in prep) for a summary).

In recent work (Agwuele et al 2008, Lindblom et al 2007) the metric has been modified to handle the effect of both preceding and following context on F2 onsets:

$$F2_{onset}(V2)=a+b*F2_{mid}(V2)+c*F2_{mid}(V1) \qquad (2)$$

The formula says that, in V_1CV_2 sequences, the $F2_{onset}$ of the second vowel is a linear weighted combination of the $F2_{mid}$ values of V1 and V2. As in Eq (1) *'a'* and *'b'* are coefficients representing intercept and slope; *'c'* is also a slope measure.

One goal of the present project is to evaluate these formulas with the aid of Öhman's (1966) measurements and some new data of our own.
A second aim is to show how the new results throw new light on the process of coarticulation.

Öhman's data revisited

The (1966) speech samples consisted of all combinations of /b/ /d/ and /g/ with /yː/ /øː/ /ɑː/ /oː/ and /uː/ in initial and final position. Each word occurred five times. Accordingly there were 125 (5*5*5) items for each place. The subject was a male Swedish speaker who was instructed to pronounce the speech samples on a monotone pitch and with equal stress on the initial and final vowels.

Öhman used data of the kind shown in Figure 1 as counter-evidence against the 'locus'. Originally this notion was defined as an invariant underlying frequency value representing the starting point of all CV transitions in practice found by extrapolating the transitions backwards in time for different vowels to a common point of intersection. Figure 1 illustrates a case for which such a common point cannot be identified.

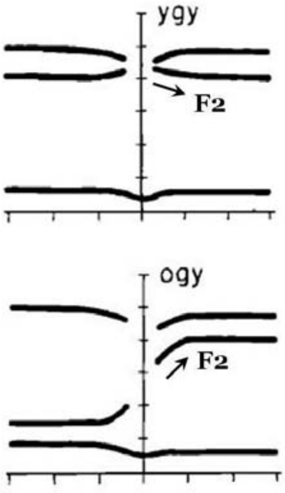


Figure 1. Stylized formant patterns for two test words (from Öhman's Fig 9). In /ygy/ $F2_{onset}$ is higher than the following $F2_{mid}$ and the transition is falling. In /ogy/ $F2_{onset}$ is lower than $F2_{mid}$ and the F2 transition is rising.

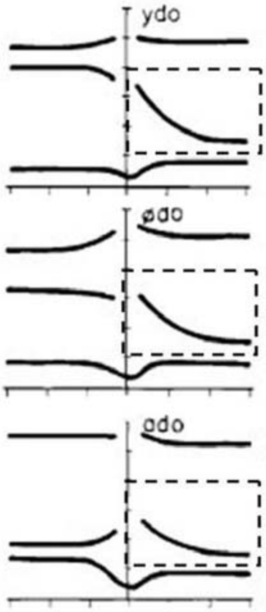
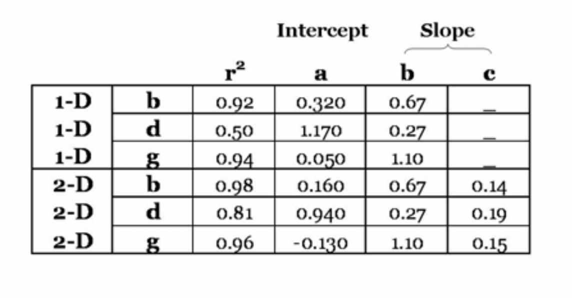


Figure 2. Stylized spectrograms of /y:do:/ /ø:do:/ and /ɑ:do:/. The dashed rectangles draw attention to the effect of V_1 on the onset of F2 in /do:/.

Looking at the stylized formant patterns (Öhman's Figs 7-9), one detects certain regularities in the data. Although, in Figure 2, the second syllable is kept the same, the F2 onsets are seen to depend on F2 in the first vowel. It was observations of this type that suggested that the 1966 data might be described in a simpler quantitative manner by taking both V_1 and V_2 into account.

To explore this possibility question we used the numbers of Tables II and IV (Öhman 1966) and made F2onset(V2)-versus-F2mid(V2) diagrams. We fitted both Eq (1) and Eq(2) to the data using a multiple regression method. Our aim was to establish standard LE slopes and intercepts using Eq (1) and to examine a possible effect of V1 on the F2 onsets of V2 with the aid of Eq (2). The result of the analyses is presented in Table I.



			Intercept	Slope	
		r^2	a	b	c
1-D	**b**	0.92	0.320	0.67	_
1-D	**d**	0.50	1.170	0.27	_
1-D	**g**	0.94	0.050	1.10	_
2-D	**b**	0.98	0.160	0.67	0.14
2-D	**d**	0.81	0.940	0.27	0.19
2-D	**g**	0.96	-0.130	1.10	0.15

Table 1. Top half: Results of fitting Eq(1) to the data in Öhman (1966). Lower part: Fitting Eq(2) to the same data. The intercept "a" is specified in kHz; "b" quantifies the effect of V; "c" reflects the V_1 dependence. All slopes and intercepts reached at least the 1% significance level.

Both equations provide adequate descriptions but it is clear that the 2-D model significantly improves the fit to the data. Comparing *c* slopes indicates that as expected V_1 plays a significant role in determining the F2onset(V2) values.

An indication of the numerical accuracy of the 2-D predictions is obtained by looking at the average absolute differences between of the model and the measurements. In Hz they were found to be 40 ([b]), 65 ([d]) and 60 Hz ([g]) which amounts to 4-5%.

A replication

To examine the generality of these findings we recorded 4 male speakers of Swedish producing the (1966) test words in accordance with Öhman's experimental procedures. The speech samples consisted of all combinations of /b/ /d/ and /g/ with /y:/ /ø:/ /ɑ:/ /o:/ and /u:/ in initial and final position. Each word occurred five times. The test items were pronounced on a monotone pitch and with equal stress on the initial and final vowels. Formant frequency measurements were made using Swell software (Ternström 2000).

The top panel of Figure 3 shows the [b]-words, the middle the [d]-words and the bottom the [g]-words. The data form quasi-linear patterns but the points cover a rather broader range along the ordinate than what is normally seen. Our statistical analyses indicate that this is mainly due to the effect of V_1.

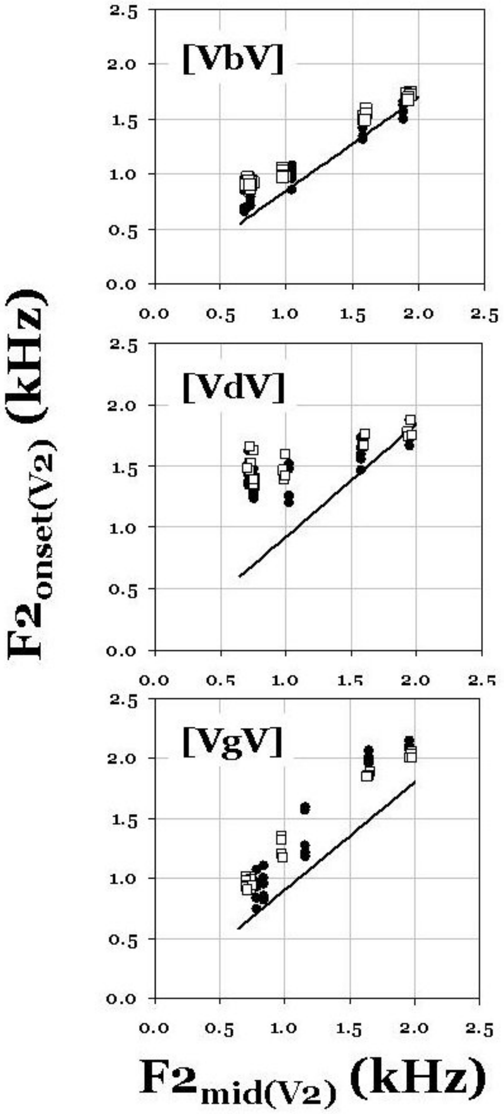
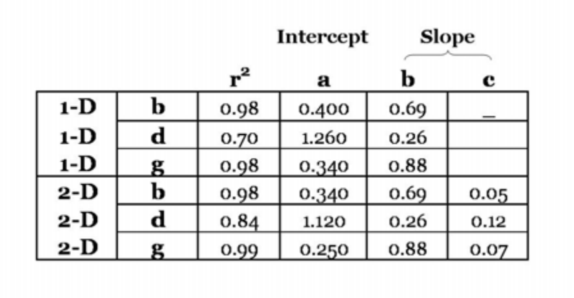


Figure 3. The data from Öhman (solid circles) compared with results from new measurements (unfilled squares). The straight lines are simulated 'locus lines' to be discussed later in the interpretation of the results.

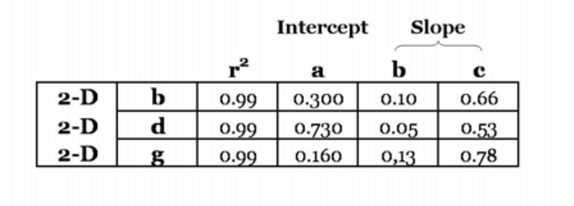
The statistical results for the new data are presented in Table 2. As we examine Figure 3 and compare Tables 1 and 2, we observe strong parallels. Slopes, intercepts and r^2 scores, pattern in a similar fashion. As expected, V2 has a stronger effect than V1 on the onsets of the F2 transitions.

Predictive accuracy - again as measured in terms of the average absolute difference between observed and calculated values comes out to be - 35 ([b]), 50 ([d]) and 45 Hz ([g]) which corresponds to an average error of 3%.



			Intercept	Slope	
		r^2	a	b	c
1-D	b	0.98	0.400	0.69	_
1-D	d	0.70	1.260	0.26	
1-D	g	0.98	0.340	0.88	
2-D	b	0.98	0.340	0.69	0.05
2-D	d	0.84	1.120	0.26	0.12
2-D	g	0.99	0.250	0.88	0.07

Table 2. Results for speech samples analyzed to replicate Öhman (1966). The organization of the table is identical with that of Table 1. "b slopes" are linked to V_2 and "c slopes" to V_1.



			Intercept	Slope	
		r^2	a	b	c
2-D	b	0.99	0.300	0.10	0.66
2-D	d	0.99	0.730	0.05	0.53
2-D	g	0.99	0.160	0,13	0.78

Table 3. 2-D regression analyses for $F2_{offset(V1)}$. The numbers should be compared with those in the lower part of Table 2.

2-D analyses were also made for the F2 transitions in V1 (Table 3). Comparing the results with the correspoinding numbers in Table 2 we note that they indicate strong linearity for all three places. As can be expected the dependence on V2 is reduced whereas the effect of V1 is strengthened.

Discussion

The 2-D model summarizes the two sets of observations with a fair degree of accuracy. The findings imply that, if we know the place of the C and the F2 values of V1 and V2, we can recreate the formant transitions in a meaningful quantitative manner. Conceptually the predictions would seem to represent a process of 'assimilation' in the sense that transition offsets and onsets are displaced in the direction of the surrounding vowels. Degree of context-dependence is captured by the slopes. The numbers indicate stronger anticipatory than regressive effects. V2 has a larger weight than V1 in modifying F2onset(V2). Conversely V1 dominates F2offset(V1).

Inferring articulatory mechanisms

What do these results tell us about the underlying articulatory mechanisms? How do the patterns arise? For bilabial closure the tongue body is free to anticipate the following V. For [d] the tongue blade/tip is the primary articulator. The tongue body can be shaped by surrounding vowels although its mobility is somewhat limited by the anterior closure. Coarticulation in [g] differs from that in [b] and [d] in that the same articulator for C and V is used: the tongue body. [b] uses lips and tongue separately. [d] uses blade and body of tongue in a semi-independent manner.

Accordingly the three places present three different pictures of coarticulatory coordination. Thus each place will be considered separately.

APEX simulations

As a baseline for our interpretations we simulated LE's for [b] [d] and [g] using our articulatory model APEX. We used it to generate formant patterns and area functions for a representative sample of vowel articulations (top left panel Figure 4). As a method of simulating stop closures "coarticulated" with the 16 vowels selected we proceeded as follows. The jaw specifications of the vowels were changed to a value typical of the stops (=7 mm). New area functions were derived for the modfied vowels. Three additional series of area functions were then produced by introduing narrow constrictions (~ "stop closure") at three points. For [b] the final section (lips) was replaced; For [d] the constriction was placed just behind the teeth. For [g] the 'closure' occurred 3.5 cm posterior to the teeth. APEX calculated the F patterns for these 48 (3*16) peudo-stops and their F2's were plotted against the F2 of the corresponding vowels. The results are presented in Figure 4.

For all places data points form linear rather tight clusters. The three places produce similar slopes that all come close to the 45 degree line.

Although the simulations of "place" are physiologically crude they enable us to make comparisons with the real data and help us understand why observed LE's come out the way they do. For instance they tell us that introducing a local perturbation (narrow constriction 0.5 cm long) somewhere along the VT does not change the formants appreciably. Hence "F2onsets" plotted against "F2vowel" should approximate straight lines with slope=1.0. In other words, LE 'linearity' is to some extent a consequence of the articulation-to-acoustics mapping.

Coarticulation in [VbV] sequences

The least unrealistic APEX simulations are those for [b]. It is interesting that observed LE's for symmetrical [VbV] often show slopes significantly smaller than 1.0.

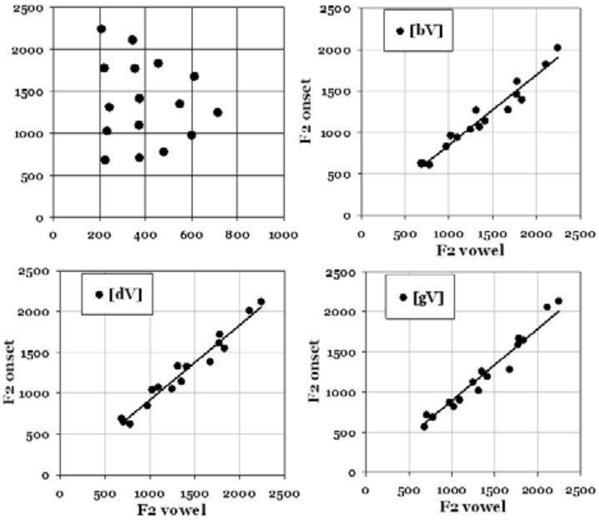


Figure 4 Top row: Vowel formant patterns (left); Simulated LE for "[bV]" (right); Bottom panels: Simulated LE:s for "[dV]" (left) and "[gV]" (right).

One hypothesis is that speech is produced in a phoneme-by-phoneme fashion (Joos 1948) and that the vowel activity during the [VbV] is *de-activated* during the [b] closure. Articulatory tongue traces of tongue height in [ibi] tend to show a trough effect (Lindblom et al 2002) which may be stronger in aspirated stops (Engstrand et al 1997). The latter observation suggest the possibility that the the trough shape of the tongue is also related to aerodynamic factors (Hoole et al 1998).

Producing [d]

The de-activation hypothesis can also be applied to the production of [d] as suggested by the following investigation.

Lindblom (2003) reports data from a 20-second X-ray film of a Swedish male speaker It contains information on tongue shapes and a Principal Components Analysis. The speech samples of relevance here are six test words: [ɛ:'di:], [ɛ:'de:], [ɛ:'da(s:)], [ɛ:'dɑ:(l)], [ɛ:'dɔ(lk)], [ɛ:'du:(s)]. The tracings for these words form a subset of a total of 400 frames analyzed. The images taken at 50 frames/sec portray a midsagittal articulatory profile. Tracings of acoustically relevant structures were made. Contours were converted into tables with x- and y-coordinates, calibrated in mm and corrected for head movements. The tongue contours were redefined in a jaw-based coordinate system and resampled at 25 equidistant ‘fleshpoints’ in preparation for a Principle Components Analysis (PCA).

The input to the PCA was a matrix whose rows specifed the indivual images and whose columns contained the x and y coordinates. The output consisted of a set of basic tongue shapes (the PC’s) and weights to be applied to the various PC’s to recreate observed contours.

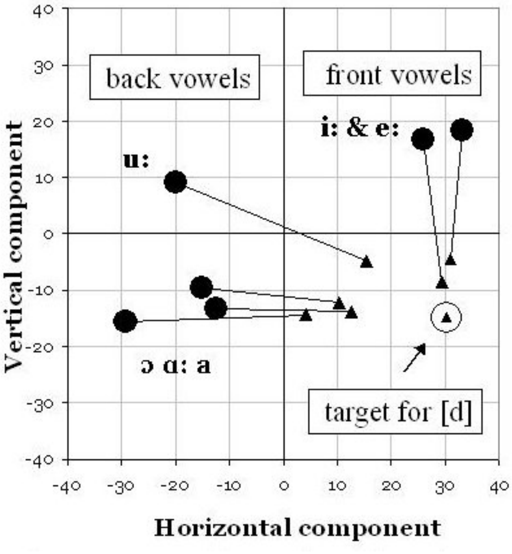


Figure 5. PCA representation of dV transitions. The vertical component of PC1 along the ordinate, and its horizontal component along the x-axis.

Using the weights for a given tongue shape we obtain the location of that shape in the articulatory space defined by the PCA. Figure 5 is such a diagram for the most important component (PC1). The vertical position (ordinate) is plotted against the horizontal position (abscissa) for tongue contours observed at the release of [d] (triangles) and vowel targets (solid circles). The straight lines stylize the [dV] transitions.

The lines show a systematic fanlike pattern. Interestingly it is possible to derive a mean point of intersection (encircled triangle). By definition it is context-free (cf ‘locus’). It has been labeled ‘[d] target’ but the reality of this point is unlikely to be a shape that the speaker “aims at”. More plausibly it should be interpreted as a shape for *complete de-activation* of vowel context. This account of [d] presents a neat parallel with evidence for de-activation in [b]. Since the de-activated tongue in [b] and [d] has a mid-range F2, the de-activation hypothesis has the further advantage of explaining where the LE slopes for [b] and [d] come from. The reason why they are < 1.0 ([d] more so than [b]) is that, in the limit (complete de-activation). they become horizontal lines.

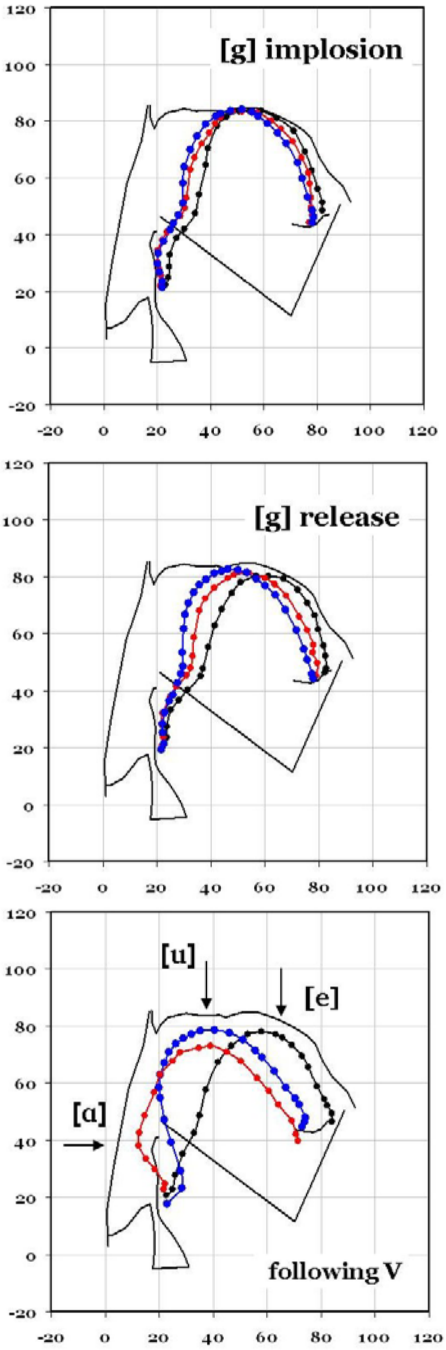


Figure 6. Tracings of articulatory profiles during [ɛ :'ge:] (black contour), [ɛ :'gɑ :(l)] (red), and [ɛ :'gu:(s)] (blue).

Articulation and acoustics of [g]

Since the same articulator (tongue body) is used both for [g] and the adjacent vowels the de-activation idea cannot explain coarticulation in [VgV]. What we need to account for is why data points for [g] with back vowels form a cluster whose slope often *exceeds* 1.0 (Figure 3). To think about that problem it is helpful to look at a few articulatory profiles. Figure 6 presents tracings from [ɛ:'ge:], [ɛ:'gɑ:(l)] and [ɛ:'gu:(s)].

The first diagram of the vertical series in Figure 6 compares the tongue contours at the moment of [g] implosion. Note how minimal the front-back differences are. The next picture shows the situation at the release. There is greater separation between the contours presumably in anticipation of the upcoming vowels. Nonetheless the three occlusions form a pretty tight group.

The third set presents profiles at the vowel targets. The significance of the arrows is that they indicate where these vowels have their maximal constrictions. If, using place of constriction as our criterion, we were to arrange standard vowels along a front-back continumm we would obtain the following series: [i] [u] [o] [ɔ] [ɑ] [a]. As shown by Figure 6 the profiles for [g] closure resemble [i] and [u] rather closely but their articulatory 'distance' to the other vowel increases as we move along the series. Interestingly so do the extents of F2 transitions: [gu] has a small transition whereas [ga] has a large one.

We can conclude with a refinement of a popular rule. Users of LE have often used LE slope as an index of degree of coarticulation (Krull 1987). For [b] and [d] this rule says that an *increase in slope* implies greater coarticulation (greater resemblance between F2onset and F2vowel). In the case of [g] data with slopes > 1.0 the rule must be changed into "a *decrease in slope* implies greater coarticulation (greater resemblance between F2onset and F2vowel).

References

Agwuele A, Sussman H M & Lindblom B (2008): "The effect of speaking rate on consonant vowel coarticulation", *Phonetica 65*:194–209

Delattre P C, Liberman A M & Cooper F S (1955): "Acoustic loci and transitional cues for consonants", *J Acoust Soc Am 27*:769-773.

Engstrand O & Lindblom B (1997): "The locus line: does aspiration affect its steepness?", Fonetik 97, *Phonum 4*:101-104, Umeå Universitet.

Hoole P, Munhall K & Mooshammer C (1998): "Do air-stream mechanisms influence tongue movement paths?", *Phonetica 55*:131-146.

Joos M (1948): "Acoustic phonetics", *Language 24*:2.

Krull D (1987): "Second formant locus patterns as a measure of consonant-vowel coarticulation", *PERILUS V, Stockholm University*, 43-61.

Liberman A & Mattingly I (1985): ‘‘The motor theory of speech perception revised’’, *Cognition 21*:1–36.

Lindblom B (2003): "A numerical model of coarticulation based on a Principal Components analysis of tongue shapes", *XVth International Congress of Phonetic Sciences*, Barcelona, Spain.

Lindblom B, Agwuele A, Sussman H M, Cortes E E (2007): The effect of emphatic stress on consonant vowel coarticulation", *J Acoust Soc Am 121 (6)*:3802–3813.

Sussman H M, McCaffrey H A & Matthews S A (1991): ‘‘An investigation of locus equations as a source of relational invariance for stop place categorization,’’ *J Acoust Soc Am 90*:1309–1325.

Sussman H M S (in press): "A Locus Equation perspective on coarticulation", in Embarki M. & Marchal A (eds): *Coarticulation: between physics and metaphysics*, Cambridge (UK): Cambridge University Press.

Ternström, S. (2000). Soundswell-signal workstation. Manual version 4.0 Hitech Development AB.

Öhman S (1966): "Coarticulation in VCV utterances: Spectrographic measurements", *J Acoust Soc Am 39*:151–168.