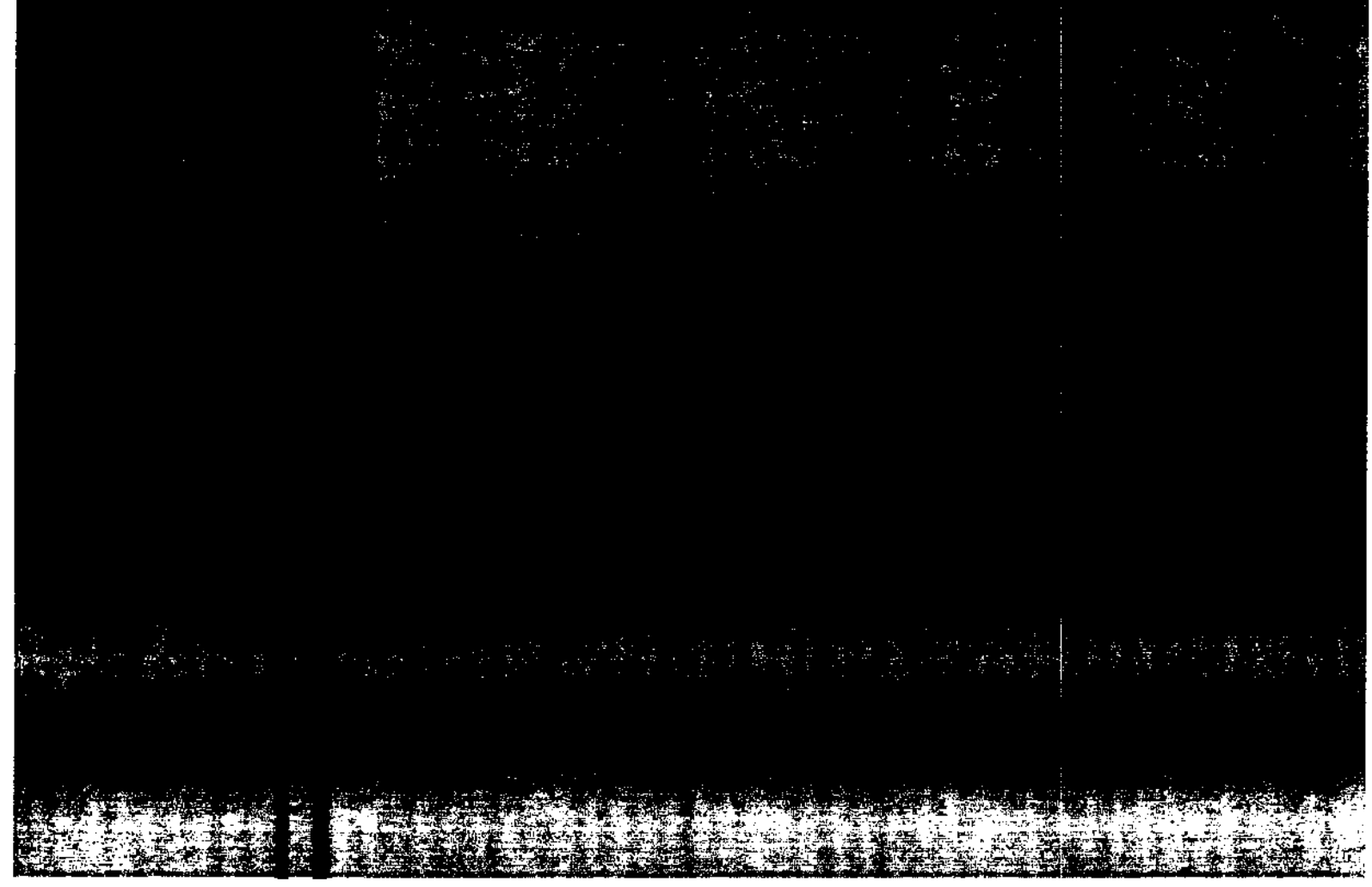
PAPERS IN LABORATORY PHONOLOGY I

BETWEEN THE GRAMMAR AND PHYSICS OF SPEECH

EDITED BY
JOHN KINGSTON AND
MARY E. BECKMAN



Löfqvist, A. and H. Yoshioka. 1980a. Laryngeal activity in Swedish obstruent clusters. *Journal of the Acoustical Society of America* 68: 792–801.
Löfqvist, A. and H. Yoshioka. 1980b. Laryngeal activity in Icelandic obstruent production. *Haskins Laboratories. Status Report on Speech Research* SR-63/64: 272–292.
Löfqvist, A. and H. Yoshioka. 1984. Intrasegmental timing: Laryngeal–oral
Munhall, K. & A. Löfqvist. (forthcoming). Gestural aggregation in sequential segments
Petursson, M. 1977. Timing of glottal events in the production of aspiration after [s]. *Journal of Phonetics* 5: 205–212.
Westbury, J. R. 1983. Enlargement of the supraglottal cavity and its relation to stop consonant voicing. *Journal of the Acoustical Society of America* 73: 1322–1326.
Yoshioka, H., A. Löfqvist, and H. Hirose. 1981. Laryngeal adjustments in the production of consonant clusters and gemminates in American English *Journal of the Acoustical Society of America* 70: 1615–1623.

26

The window model of coarticulation: articulatory evidence

PATRICIA A. KEATING

26.1 Introduction

26.1.1 *Phonetics and phonology*

Much recent work in phonetics aims to provide rules, in the framework of generative phonology, that will characterize aspects of speech previously thought to be outside the province of grammatical theory. These phonetic rules operate on a symbolic representation from the phonology to derive a physical representation which, like speech, exists in continuous time and space. The precise nature of phonological representations depends on the theory of phonology, but certain general distinctions between phonological and phonetic representations can be expected. Only in the phonology are there discrete and timeless segments characterized by static binary features, though phonological representations are not limited to such segments. Even in the phonology, segments may become less discrete as features spread from one segment to another, and less categorical if features assume non-binary values. However, only in the phonetics are temporal structure made explicit and features interpreted along physical dimensions; the relations between phonological features and physical dimensions may be somewhat complex.

Thus phonological representations involve two idealizations. They idealize in time with *segmentation*, by positing individual segments which have no duration or internal temporal structure. Temporal information is limited to the linear order of segments and their component features. Phonological representations idealize in space with *labeling*, by categorizing segments according to the physically abstract features. These idealizations are motivated by the many phonological generalizations that make no reference to quantitative properties of segments, but do make reference to categorial properties. Such generalizations are best stated on representations without the quantitative information, from which more specific and detailed representations can then be derived.

Because phonological and phonetic representations are different, the rules that can operate on each must be different. Phonological representations, which are essentially static and categorial, can be acted on by phonological rules, which change or rearrange the features which comprise segments. The output phonological representations can be acted on by phonetic rules, which interpret these features in time and space. Phonetic rules can thus, for example, assign a segment only a slight amount of some parameter [illegible] over time during the segment. The result is a representation which provides continuous time functions along articulatory or acoustic dimensions. In this paper, the representations discussed will be articulatory; they depict articulatory movements in space as a function of time, simply because of the interest of some available articulatory data. No special status for articulatory representations is intended.

In considering the relation between segments and the speech signal, phoneticians have always seen coarticulation as a key phenomenon to be explained. Coarticulation refers to articulatory overlap between neighboring segments, which results in segments generally appearing assimilated to their contexts. In the spatial domain, transitions must be made from one segment to the next, so that there are no clear boundaries between segments. Whatever the feature values of adjacent segments, some relatively smooth spatial trajectory between their corresponding physical values must be provided. What mechanisms are available to provide such enriched representations? A new account of the spatial aspect of continuous representations is proposed here. In this account, continuous spatial representations are derived using information about the contextual variability, or coarticulation, of each segment.

Though our concern here will be with coarticulation that is phonetic, i.e. quantitative, in character, it must be noted that rules of coarticulation, like other rules, could be either phonological or phonetic in the sense described above. A consequence of work in autosegmental and CV phonology (e.g. Goldsmith 1976; Steriade 1982) is that some segmental overlap can now be represented phonologically. Phonological rules of feature-spreading will produce partial or complete overlapping of segments, including assimilation over long spans. Phonological rules nonetheless have only limited access to segment-internal events. Phonetic rules, on the other hand, can (and more typically will) affect portions of segments, or affect them only slightly, or cause them to vary continuously in quality.

The distinction between phonological and phonetic coarticulation is brought out in, for example, discussions of Arabic tongue backing ("emphasis") by Ghazeli (1977) and Card (1979). In previous work, it was largely assumed that this phenomenon was phonological in nature, because the effects of backing could extend over a span of several segments. The phenomenon then appears to be a

prime candidate for an autosegmental account in which one or more feature values (i.e. [+back]; possibly also [+low]) spread from certain contrasting segments to other segments in a word. However, Ghazeli and Card, in studies of different dialects and different types of data, both find difficulties with a segmental feature analysis, traditional or autosegmental. In these studies, facts about the gradient nature of contextual tongue backing are presented. The phenomena discussed include partial backing of front segments by back segments and vice versa; weakening (as opposed to blocking) of the spread of backness by front segments; dependence of the amount of backness on distance from the trigger to the target segment. Clearly, categorical phonological rules cannot describe such effects. The difficulties are discussed explicitly by Card. For example, she notes that underlyingly backed segments are "more emphatic" than derived segments are, apparently requiring that phonological and phonetic levels of representation be kept distinct in output. However, neither Card nor Ghazeli actually provides a phonetic analysis of any of these phenomena.

Much of the coarticulation literature is confusing on this issue of levels, in that phenomena that are clearly phonetic are often given (unsatisfactory) phonological treatments. In the late 1960s and early 1970s, studies of coarticulation were extended to include effects over relatively long spans. These effects were modeled in terms of spreading of binary feature values; analyses of phonetic nasalization and lip rounding proposed by Moll and Daniloff (1971) and Benguerel and Cowan (1974), for example, were completely phonological in character. Not surprisingly, binary spreading analyses generally proved inadequate (see, for example, Kent and Minifie 1977). The data that were being analyzed were continous physical records, and the analyses were intended to account for such things as details of timing. The analyses failed because such phonological accounts, which make no reference to time beyond linear sequencing, in principle cannot refer to particular moments during segments. The point, though, is not to make the opposite category error by assuming that all coarticulation and assimilation must be phonetic in character. Rather, the point is to determine the nature of each case.

What we want, then, is a way of describing those coarticulatory effects which do not involve phonological manipulation of segmental feature values, but instead involve quantitative interactions in continuous time and space. To simplify matters, we will consider only one sub-type of coarticulation; coarticulation involving a single articulator used for successive segments. Coarticulation involving the coordination of two different articulators will not be considered, as further principles are then required for inter-articulator alignment. In single articulator coarticulation, the given articulator must accommodate the spatial requirements of successive segments. If two such requirements are in conflict, they could be moved apart in time (temporal variation), or one of them could be modified (spatial variation). The question, then, is how phonetic rules deal with such situations.

26.1.2 *Target models*

The traditional, and still common, view of what phonetic rules do is that segmental features are converted into spatio-temporal targets (e.g. MacNeilage 1970), which are then connected up. Segmental speech synthesis by rule typically uses some kind of targets-and-connections model. Targets were formerly seen as invariant, the defining characteristic of a phoneme class. In the [illegible] values may not always be reached; e.g. targets may be undershot or overshot due to constraints on speed of movement, thus resulting in surface allophonic variation. The approach in Pierrehumbert (1980) and related work on intonation is in a similar vein, though with an important difference. In this work, target F_0 values are assigned in time and space by a context-sensitive process called "evaluation." A tone is evaluated with reference to various factors, such as the speaker's current overall pitch range, the phonological identity of the previous tone, the phonetic value assigned to the previous tone, and the particular tonal configurations involved. The use of context-sensitive evaluation, instead of invariant targets, minimizes the need for processes of undershoot and overshoot to deal with systematic deviation of observed contours from targets. While Pierrehumbert believes that crowded tones can give rise to overshoot and undershoot, in cases where tones are sparse their targets are always reached. Targets are connected by rules of "interpolation," which build contours. Interpolation functions are usually monotonic, with target values usually providing the turning points in a contour, and in general the intention is a theory in which speech production constrains interpolations. However, when tones are sparse, interpolations may vary; for example, in the 1980 work on English, "sagging" and "spreading" functions are used to sharpen and highlight F_0 peaks.

A targets-and-connections account of phonetic implementation is only part of a complete phonological and phonetic system. Such a system allows several types of phonological or phonetic contextual influences. Indeed, the system may well be so rich that it is difficult to determine the nature of any single observed effect. First there are the phonological rules which affect (i.e. change, insert, delete) binary feature values; these rules ultimately give rise to gross spatial changes when the feature values are interpreted quantitatively by later rules. Next there is the possibility of context-sensitive evaluation; in Pierrehumbert's scheme for intonation, all evaluation is context-sensitive. An example with segmental features would be that the precise spatial place of articulation of one segment could depend on that of an adjacent segment. This situation differs from a phonological feature change in that the spatial shift would presumably be small. Such context-sensitive "target selection" was used, for example, by Öhman (1967) to account for variation in velar consonants as a function of details of the vowel context; similar rules are often used in speech synthesis (e.g. Allen, Hunnicutt, and Klatt 1987).

Another locus of contextual effects is the temporal location of targets. Because spatial values must be assigned to particular points in time, contextual effects could arise from shifts in such time points, rather than the spatial values themselves. For example, a value for one segment could be assigned to a relatively early point in time, far from the value of the following segment. Subtle variation in the timing of targets will produce subtle phonetic effects, e.g. on-glides and off-glides to vowels due to consonants.

Interpolation between targets results in time-varying context effects. Pierrehumbert (1980) showed how this mechanism could be used to determine much of an intonation contour. When two points are connected up, both of them influence the entire transition between them. This mechanism becomes especially important when the targets to be connected up are located far apart in time. Since English tones, from which intonation contours can be generated, are sparse relative to the syllables of an utterance, the parts of the contour interpolated between the phonetic values of tones play an important role. The same would be true for segmental features if segments may be underspecified throughout the phonetics (Keating 1985). Öhman (1966) used this mechanism, interpolation between sparse values, to produce tongue body coarticulatory effects on consonants, and Fowler (1980) draws on Öhman's model in her own account of vowel and consonant coproduction.

In this paper I propose a somewhat different way of viewing the process of building a contour between segmental features. In this new model, variability, both systematic and random, plays a more central role, while targets, and turning points in contours, play a much lessened role.

26.2 The window model of contour construction

26.2.1 *Outline of model*

I propose that for a given physical articulatory dimension, such as jaw position or tongue backness, each feature value of a segment has associated with it a range of possible spatial values, i.e. a minimum and maximum value that the observed values must fall within. I will call this range of values a *window*. As will be seen below, this window is not a mean value with a range around that mean, or any other representation of a basic value and variation around that value. It is an undifferentiated range representing the contextual variability of a feature value. For some segments this window is very narrow, reflecting little contextual variation; for others it is very wide, reflecting extreme contextual variation. Window width thus gives a metric variability. There is no other "target" associated with a segment; the target is no more than this entire contextual range.

To determine the window for a segment or for a particular feature value, quantitative values are collected across different contexts. Since an overall range of values is sought, maximum and minimum values are the most important.

Therefore contexts which provide extreme values are crucial, and must be found for each segment or class. A window determined in this way is then used to characterize the overall contextual variability of a segment. Windows are determined empirically on the basis of context, but once determined are not themselves contextually varied. That is, a feature value or segment class has one and only one window that characterizes all contexts taken together; it does not have different windows for different contexts. Information about the possibilities for contextual variation is already built into that one window. Note, however, that the phonological feature values that are the basis for window selection need not be the same as the underlying values: phonological rules to change or spread feature values still apply before the phonetics. Thus, in terms of segments, windows are selected for extrinsic allophones rather than phonemes.

Windows are given for physical dimensions rather than for phonological features.[1] In some cases, the relation between a feature and a physical dimension is fairly direct; the relation between [nasal] and velum position is a standard example. In other cases, the relation may be less direct. Dimensions of tongue and jaw position relate to more than one phonological feature. Thus, phonetic implementation involves interpreting features as physical dimensions in a potentially complex way, though conceivably with the right set of physical dimensions and features this task would be more straightforward than it now seems. Furthermore, the physical interpretation may depend to some extent on other feature values for a given segment. For example, place of articulation may depend somewhat on manner. Thus, in what follows, attributing one window to all instances of a phonological feature value is probably an oversimplification.

On a given dimension, then, a sequence of segments' feature values can be translated into a sequence of windows. The process of interpolation consists of finding a path through these windows. Although the relevant modeling remains to be carried out, I assume that this path is constrained by requirements of contour continuity and smoothness, and of minimal articulatory effort, along the lines of minimal displacements or minimal peak velocities. Thus the process of interpolation can be viewed as an optimization procedure which finds smooth functions that fall within the windows. Most of the path must fit into the window, but some part of it will fall within narrow "transition" zones between windows; in the case of adjacent narrow windows, the entire transition will take place quickly between the windows. On this view, the job of "evaluation" (for example, determining turning points in curves) is divided between a mechanism which provides the windows, and the interpolation mechanism. The individual values associated with segments do not exist before an actual curve is built; there are no "targets" or assigned values. Thus whether there is a turning point associated with a given segment depends on the window for that segment and the windows of the context.

Windows are ranges within which values forming a path are allowed to fall.

The window model of coarticulation

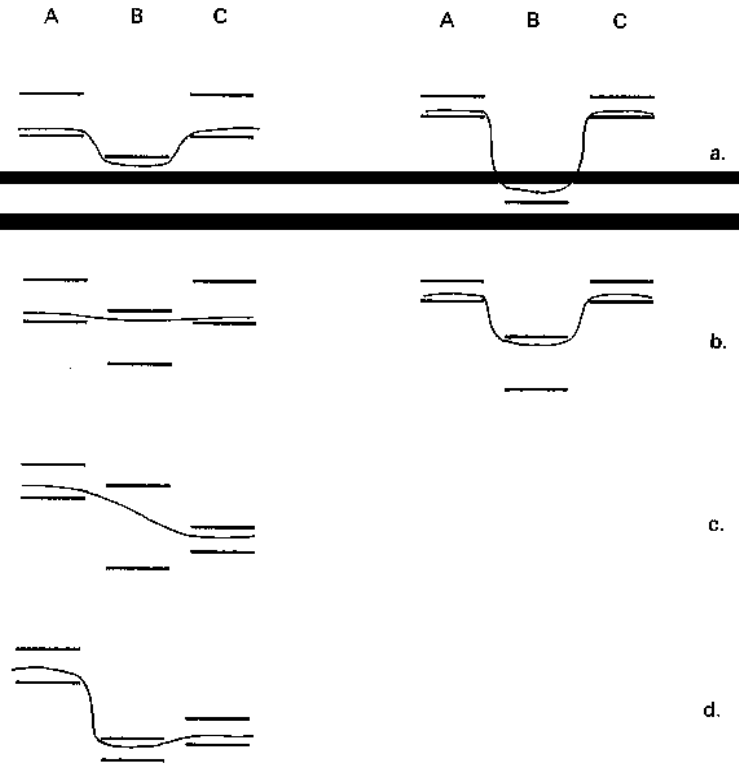


Figure 26.1 Illustrations of sequences of windows of various widths. See text for description of each sequence.

Depending on the particular context, a path through a segment might pass through the entire range of values in the window, or span only a more limited range within the window. The paths depend on the context. This is why the window is not taken to be a mean plus a range around the mean. It is not clear how information about a mean value (across all contexts) could be useful in constructing a path for any one context; one could not, for example, constrain a path to pass through the mean, or show the mean as a turning point.

In this paper I will offer no explicit procedure for constructing paths through windows, e.g. what functions are possible, whether construction proceeds directionally, how large a span is dealt with at a time. I also leave aside the question of how timing fits into this scheme, e.g. the time interval over which paths are constructed, and whether windows have variable durations, or are purely notional. These are, of course, crucial points in actually implementing the model, but the guiding ideas should be clear.

To show how the model is to work, various combinations of wide and narrow windows, and schematic interpolations through them, are given in figure 26.1. The contours were drawn by eye-and-hand. In this figure, imagine a single articulatory dimension, with each example showing a range of values that spans some subset of the total possible physical range for this articulator. First, in (a), consider cases of a segment with a narrow window between two identical segments. [illegible] the middle segment. This middle segment imposes strong constraints on the interpolation, shows little variation across contexts, and affects the interpolation in the adjacent segments. Next, in (b), consider cases of a segment with a wide window between the same identical segments. The wide-window segment assimilates its turning point to the context, and in some contexts will show no turning point at all. Finally, in (c) and (d), consider wide and narrow windows between unlike segments. The wide window allows straight interpolations between many different segment types; the narrow window more often makes its own contribution to the curve.

Note that the contours shown are not the only possible interpolations through these sequences, since even minimal curves can be moved higher or lower in sequences of wider windows. When only wide windows are found in a sequence, or when a segment is in isolation, multiple interpolations will also be possible. Indeed, the prediction of this model is that speakers, and repetitions of a single speaker, should vary in their trajectories through window sequences which underdetermine the interpolation. Whether such variability is in fact found remains to be seen. It may be that instead there are speaker-specific strategies for limiting variation in such cases, requiring the model to be revised. For example, windows as currently viewed are essentially flat distributions of observed values; instead, empirical distributions could be associated with windows, and used to calculate preferred paths.

26.2.2 *An example: English velum position*

An example will illustrate how this model is derived from and applied to data. A case from the literature that receives a natural analysis under this model involves velum height in English. It is well-known that the degree of velum opening for [+nasal] segments, and the degree of velum closing for [−nasal] segments (both consonants and vowels), varies across segment types and contexts. Thus figure 26.2, after Vaissière (1983), shows the ranges of values covered by nasal consonants as opposed to oral consonants. These values group together all places of articulation, though Vaissière goes on to show that velum height varies with consonant place. Thus the windows for individual consonants should be somewhat narrower than these ranges of values.

For vowels, velum position is more variable than during either oral or nasal consonants. Again, vowels are discussed here as a group, though again differen

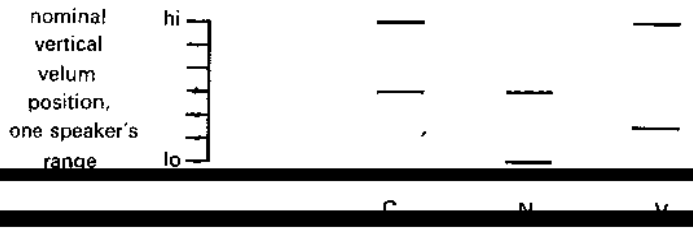


Figure 26.2 Ranges of values for vertical position of a point on the velum in sentences, based on figures in Vaissière (1983).

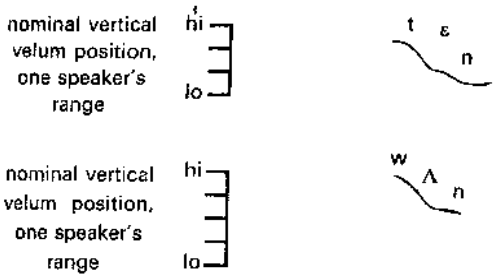


Figure 26.3 Vertical position of a point on the velum in CVN sequences from sentences, after Kent et al. (1974).

Vowel phonemes will be expected to have different windows for velum height. The difference between consonants and vowels is traditionally treated as being due to the fact that English consonants, but not vowels, contrast in nasality. Velum position for vowels can vary more because the vowels carry no contrastive value for nasality; velum positions for vowels are more affected by context than are consonant positions. When a vowel precedes a nasal, velum lowering begins at vowel onset, and velum height is interpolated through the vowel, as shown in tracings of velum position in fig. 26.3, after Kent et al. (1974).

A window analysis of these facts of English runs as follow. Velum height windows are suggested for oral consonants, nasal consonants, and vowels in figure 26.4. Nasal and oral consonants nearly divide up the range of velum positions, with nasal consonants having lower positions. Vowels have a very wide window, from low to high velum positions, but excluding maximal lowering. Sequences of segmental [nasal] values are represented as sequences of these windows, as in figure 26.4. Contours are derived by tracing smooth paths through these sequences. Vowels, with their wide windows, easily accommodate most interpolations between consonants; any values that would be encountered in interpolating between two oral consonants, or an oral and a nasal consonant, will satisfy the vowel window.

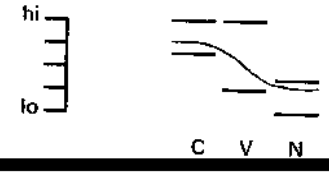


Figure 26.4 Schematic velum height windows, based on data in figure 26.2 and other [illegible]

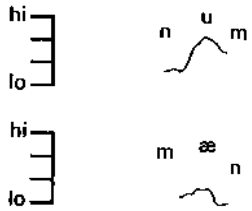


Figure 26.5 Observed raising of velum for vowel between two nasal consonants, after Kent et al. (1974).

The vowel window will, however, exclude a straight-line interpolation between two nasal consonants with maximally open values; to satisfy the vowel window, a slight raising of the velum would be required. Such raising was noted and discussed by Kent et al. (1974); two of their examples are shown in figure 26.5. It is also apparent in Vaissière's (1983) data. On the window analysis, then, this slight velum raising is the minimally required satisfaction of the vowel's window.

This velum raising is better accounted for by the window model than by two previous approaches. In one approach, the vowel would be assumed to have an "oral," raised-velum, target, but that target would be undershot because of constraints on quickly raising and then lowering the velum. The result of the undershoot would be only a slight raising of the velum, as observed. The problem with this account is that in other contexts, especially that of a nasal followed by an oral consonant, much faster velum movements are in fact observed (e.g. Vaissière 1983). An undershoot approach cannot easily explain these differences in observed speed of movement.

In the other approach, vowels are said to be completely unspecified for nasality and therefore impose no requirements on the velum. However, as Kent et al. (1974) noted, the behavior of a vowel between two nasals is a counterexample to this hypothesis. The vowel instead appears to be specified as "at least weakly oral." The window model states exactly this generalization, in a quantitative fashion.

26.2.3 *Consequences*

Wide windows, like the velum window for vowels, have an interesting effect in cases where the two sides of the context have very different values for window width and window placement within the dimension. If B is a segment with a wide window, in a sequence of segments ABC, an interpolated trajectory between [illegible] will satisfy B's window. Thus B's values, and in fact its entire trajectory, will usually depend completely on the context. Yet these are the sort of cases which have been described as resulting from surface phonetic underspecification (Keating 1985). That is, cases of apparent "underspecification" can be seen as cases of very wide windows. True surface underspecification would be equivalent to a window covering the entire range of possible values. In most cases, though, even a wide window will span something less than the entire range of values, and so in at least some context will reveal its own inherent specification. If a segment appears to be unspecified along some dimension, then it should be examined between identical segments with both extremes of values. One or both of these extremes should show the limits, if any, of the putatively unspecified segment's window. An example of this was seen in the case of English vowels and velum position, where vowels in the context of flanking nasals do not appear as unspecified as they do in other contexts.[2]

The implication of this point is that, in a window theory, *phonetic* underspecification is a continuous, not a categorial, notion. The widest windows that produce the apparent lack of any inherent phonetic value are simply one extreme; the other extreme is a window so narrow that contextual variation cannot occur.

In a window model, it is possible to assign a target that is equivalent to a single point in space – namely, a maximally narrow window – for segments where this is appropriate. However, not all targets need to be specified this narrowly, and indeed the assumption behind the model is that they should not. Instead the model assumes that most segments cannot be so uniformly characterized. In traditional models, each segment is viewed as having an idealized target or variant that is uncontaminated by contextual influences; context systematically distorts this ideal, and random noise further distorts it unsystematically. (The ideal may be identified with the isolation form, but even there it may be obscured because in practice speakers have difficulty performing in this unusual situation.) The window model stands this view on its head. Context, not idealized isolation, is the natural state of segments, and any single given context *reduces, not introduces*, variability in a segment. Variability is reduced because the windows provided by the context contribute to determining the path through any one window. However, in most cases there will still be more than one possible path through a given sequence of windows, especially at the edges of the utterance. This

indeterminacy is seen as an advantage of the model; it says that speakers truly have more than one way to say an utterance, especially in cases of minimal context.

26.2.4 *Another example: English jaw position*

Another example of how the window model works, which shows the effects of narrower windows, is provided by jaw position for consonants in some dual of Amerman, Daniloff, and Moll (1970). The fact that jaw position is not directly controlled by any single phonological feature suggests that contextual effects on jaw position are unlikely to be phonological in nature. Jaw positions for vowels depend on (tongue) height features, but for consonants are less directly if at all specified by such features. (For example, the jaw is generally high for coronals, especially alveolars, but is relatively low for velars, though these are usually described as [+high].) Given this difference between vowels and consonants, phoneticians expected effects of vowels on consonant jaw position to extend over long time spans, and looked at data to see just how far such effects could extend.

Amerman et al. (1970) looked at the influence of an open vowel /æ/on jaw position in various numbers of preceding consonants. Because they wanted to look at maximal sequences of consonants in English, the first consonant was almost always /s/. They used the distance between the incisors as observed in X-ray motion pictures as the measure of jaw opening. Data analysis consisted of noting during which consonant the jaw lowering gesture for /æ/ began. In over 90% of their cases, this consonant was the /s/. They concluded that jaw lowering for /æ/ extends over one or two preceding consonants up until an /s/:

> Results for jaw lowering show that coarticulation of this gesture definitely extends over two consonants [sic] phonemes preceding the vowel /æ/, probably irrespective of ordinary word/syllable positions. This result could have presumably been extended to four consonants, had the /s/ phoneme not shown itself unexpectedly to be contradictory to jaw lowering. [...] coarticulation of a gesture begins immediately after completion of a contradictory gesture.

Amerman et al. did not describe jaw lowering as a feature, or use the term "feature-spreading" to describe coarticulation of jaw position. However, some aspects of their discussion certainly suggest this kind of analysis. They share with feature-spreading analyses two central ideas: first, that coarticulation is the anticipation of an upcoming gesture, and second, that contradictory gestures block this anticipation. In the case at hand, it was expected that a jaw lowering gesture for a vowel would be anticipated during previous consonants, and it was found, surprisingly, that /s/ is contradictory to such a lowering gesture. This finding could easily be given a feature-spreading formulation: a low vowel had a feature for low jaw position which will spread to preceding unspecified segments; /s/ is

a consonant specified with a feature for a high jaw position which therefore blocks the spreading of the vowel's low position feature. While Amerman et al. did not give such a formulation, Sharf and Ohde (1981), in their review of coarticulation, do: they cite this study, plus Gay (1977), and Sussman, MacNeilage, and Hanson (1973), as showing "feature spreading and shifts in target position" for the jaw. In jaw position during consonants, do we find evidence of either a gesture or a feature value for jaw lowering? In the sample figures given by Amerman et al. (as well as in similar figures in Keating 1983) it seems clear that jaw position is continuously changing between the closest position, associated with the entire /s/, and the most open position, associated with the extreme for /æ/. An example is shown in figure 26.6. Any consonants between /s/ and /æ/ are affected equally by those two extreme positions; it might as well be said that there is left-to-right carryover assimilation of jaw raising from the /s/, to which /æ/ is contradictory. Even /æ/ shows some effect of the /s/, since much of the /æ/ is spent reaching the extreme open position. In fact, then, both extreme endpoints appear important to the intermediate segments, in that both determine the trajectory from high to low position. Intermediate segments "assimilate" in the sense that they lie along the (curved) interpolation. In these terms, it makes little sense to ask "how many segments" lowering can "coarticulate across." Instead, we want to know which segments provide which extreme values for an articulatory dimension such as jaw lowering.

The data presented are not sufficient for a window analysis. Determining the contribution of each segment to the overall contour requires information about the variability of each segment type, yet Amerman et al., with their different experimental goal, examine only one type of segment sequence. From this kind of data we cannot conclude anything about window widths.

As it happens, data that address this hypothesis with respect to jaw lowering are available in unpublished work done by me with Bjorn Lindblom and James Lubker. Our experiment recorded jaw position over time in one dimension in VCV tokens. The vowels were /i,e,ɑ/ and were the same in any one token; the consonants were /s,t,d,r,l,n,b,f,k,h/. Although we recorded both English and Swedish speakers, I will discuss only the five English speakers here; each speaker produced each item six times. The measurements made were maximum opening for the vowels, and maximum closing for the consonants, that is, the extreme positions in a VCV. Measuring such extreme positions is relatively straightforward, though it may not represent the full range of possible variation. These data allow us to ask whether a given consonant has a fixed value, or a variable value, between vowels of different degrees of openness. What we find is that in VCV's, both the vowel and consonant extreme jaw positions vary as a function of the other, but vowels vary more than consonants. This result can be stated more generally by saying that overall higher segments (segments whose average position is higher) vary less than overall lower segments.

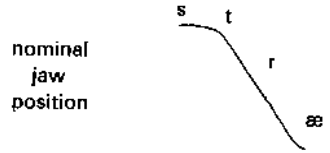


Figure 26.6 Vertical position of a point on the jaw in /stræ/, after Amerman et al. (1970).

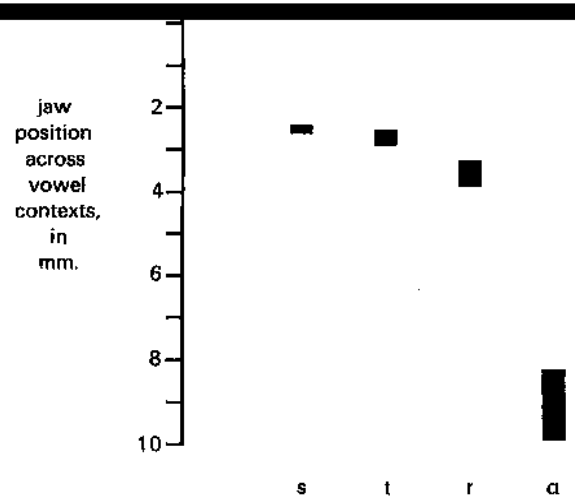


Figure 26.7 Range of mean extreme values for jaw position for four segments, from data of Keating et al. (1987).

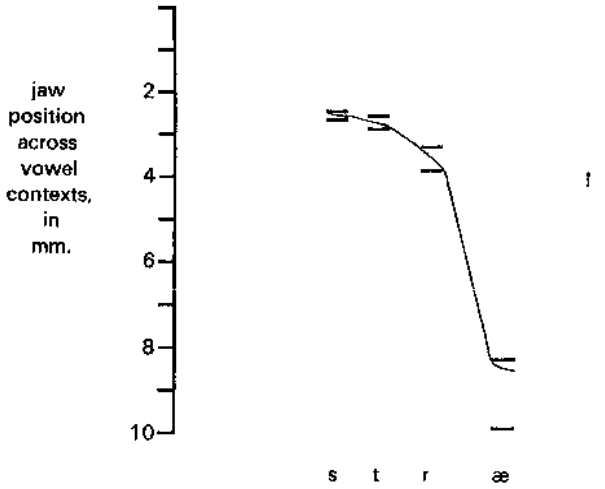


Figure 26.8 Sequence of schematic jaw windows for /stræ/, with contour.

Figure 26.7 shows mean variability of the consonants of interest (/s,t,r/) and the low vowel /ɑ/. For each of these segments in each of its contexts, an average extreme value was calculated across speakers and repetitions. The variability shown here indicates the minimum and maximum averages obtained in this way. From these data, windows for jaw position for each of these segments are proposed in figure 26.8: high and [illegible] for /s/, similar but slightly wider for /t/, middle and medium for /r/, and low and wide for /ɑ/. In the figure, windows are shown in sequence for /stræ/, together with a contour traced through these windows. The window for /s/ is the narrowest and thus exerts the most influence on the contour. The other consonants have windows that place them along a smooth trajectory from the /s/ to the vowel.

26.3 General discussion and conclusion

The window model is a proposal about how successive segments are accommodated in building a continuous contour along a single articulatory dimension. Two examples have been presented: one in which an articulatory dimension (velum position) is controlled by a single phonological feature ([nasal]), and another in which an articulatory dimension (jaw position) is less directly related to any single phonological feature (e.g. [high]). In this section, some general points about the proposal will be discussed.

26.3.1 *Underspecification*

The window model extends and refines an earlier proposal about phonetic underspecification. In Keating (1985) I claimed that formant trajectories during /h/ are determined by surrounding context, with the /h/ contributing no inherent specification of its own, other than the glottal state. I suggested that /h/ be analyzed as having no values for oral features even in phonetic representation, with interpolation alone providing the observed trajectories. Under this proposal, phonetic underspecification was viewed as a carryover into surface representation of phonological underspecification, and thus was an all-or-none possibility. Now, however, the degree of *phonetic* specification is differentiated from phonological underspecification. A wide window specifies relatively little about a segment, while a narrow window gives a precise specification, and all intermediate degrees are possible. Thus, for example, with respect to phonetic nasality, English vowels, with their wide but not maximal window, are "not quite unspecified."

Window width is to some extent an idiosyncratic aspect that languages specify about the phonetics of their sounds and features. The default rules and phonetic detail rules of a language will be reflected in window widths. However, it seems likely that, overall, phonetic variability will in part be a function of phonological contrast and specification. Thus it can be proposed that only phonologically unspecified features will result in very wide windows; if the two contrasting values

of a feature were each assigned wide-open windows, the "contrast" could hardly be maintained. Furthermore, window width may derive in part from the specification, or lack of it, of phonological features. It has already been noted that more than one feature may be reflected in a given physical dimension; if a segment is specified for all of the features relating to a particular dimension, then that segment could well have a narrower window for that dimension than a segment which was specified for only one of the relevant features. More generally, the more [illegible] for the various dimensions defining the overall phonetic domain.[3] Thus the more specified a segment, the smaller its total share of the phonetic domain, so that contrasting segments tend to occupy separate areas in that domain.

Another independent consideration that may influence window width is revealed in the data on jaw position variability. Jaw positions that were lower on average were also more variable. Possibly variability may be better measured on a different scale, one using percentages rather than absolute values for position. In any event, to the extent that the dimensions along which variation is measured may not be strictly linear, nonlinguistic constraints on variation may be found.

26.3.2 *Variability*

The window model expresses the observation that some segments vary more than others along various articulatory dimensions. Previous work has also addressed the issue of segment variability. Bladon and his colleagues (Bladon and Al-Bamerni 1976; Bladon and Nolan 1977) proposed an index of *coarticulatory resistance* to encode the fact that some segments are relatively insensitive to context, while others vary greatly as a function of context. However, Bladon's index was construed as a separate segmental "feature" with numerical values. That is, a given segment would have its values for the usual phonetic features, plus a value for the coarticulatory resistance feature. Thus the coarticulatory resistance indicated variability around some norm derived from the phonetic feature values. The basic insight of coarticulatory resistance is preserved in the window model: high coarticulatory resistance corresponds directly to a narrow window, and a lack of coarticulatory resistance corresponds directly to a wide window. But in the window model, the variability is not represented separately from observed, modal, or target values. Furthermore, it is values for features, rather than values for unanalyzed segments, which are related to numerical variability.

Lindblom (1983) employed a notion related to coarticulatory resistance and general segment *incompatibility*, called *coarticulatory propensity*. Using the idea that some segments are more prone to coarticulate with neighbors, Lindblom proposed that segment sequences are generally ordered so as to minimize conflicts due to incompatibility. In effect, then, Lindblom's incompatibility was related to incompatibility as used in feature-spreading models of coarticulation, where

features spread until blocked by segments having the opposite value for the spreading feature or specification for other features incompatible with it. The difference is that Lindblom's blocking was gradient in nature; segments would *more or less* block coarticulation. However, Lindblom did not quantify his dimension; indeed, it is not obvious how this could be done.

More recently, Manuel and Krakow (1984) and Manuel (1987) have discussed variability in terms of what they call *output constraints*. The idea here is that the [illegible] phoneme's variability. Phonemes are represented as areas or regions, not as points, in a phonetic domain such as a two-dimensional vowel graph (as also for Schouten and Pols 1979). The output constraints essentially say that no phoneme can intrude into another phoneme's area. Thus the size of the inventory strongly influences the size of each phoneme's area, and its possible contextual variability. In this work, the focus is on variation of an entire segment class within its physical domain, for example, all vowels in the vowel-formant domain. The constraints may be taken as properties of the class more than of the individual segments. Furthermore, no explicit mechanism is given for relating the constraints to the process of phonetic implementation. Manuel follows Tatham (1984) in saying simply that the phonetics "consults" the phonology to ascertain the constraints. The window model, in contrast, gives an account of how the phonetic values derive from the representation of variability. First, window width is related to (though not equated with) feature specification, making variability more a property of individual segments than of segment classes *per se*. Second, window widths are the basis of path construction, providing variable outputs for combinations of segments and contexts.

A hypothesis related to output constraints was discussed by Keating and Huffman (1984) and by Koopmans-van Beinum (1980, and later work), namely, that a language might fill the available vowel domain one way or another – if not with phonemes, then with allophonic variation. These authors place more emphasis than Manuel does on languages allowing extensive overlap among phonemes, that is, not constraining the output very severely. Thus, for example, a five-vowel language like Russian, with extensive vowel allophony and reduction, will show much more overlap among vowels than will five-vowel languages like Japanese. The window model allows this kind of arbitrary variation, since a language can have wider or narrower windows for all feature values. However, as noted above, there is a sense in which inventory will generally affect window widths: inventories affect the degree of phonological feature underspecification, and underspecification probably generally results in wide windows.

The output constraints model described by Manuel (1987) differs in another way from the window model. Output constraints are constraints on variability around a target or modal value for each phoneme, and phonemes are seen as having "canonical" variants. In addition, careful speech is hypothesized to involve

production of these canonical variants. The window model includes no such construct, and indeed in earlier discussion the notion of a canonical "isolation form," distorted in contexts, was rejected. Nonetheless, Manuel raises an important issue that future experimentation should address.

26.3.3 *Coarticulation*

Returning to our initial impetus, how does the window model differ from a targets-and-connections model in providing mechanisms for coarticulation? In the targets-and-connections model, coarticulation at the phonetic level could result from two mechanisms, determination of target values, and interpolation between them. Considering only spatial values, quantitative evaluation as a coarticulatory mechanism is trivialized in the window model. In this model, evaluation consists of looking up the appropriate window. Only narrow windows say anything about the values to be associated with a segment, and turning points in the contour have no special status in characterizing contextual effects on segments. Instead, contour building is done almost entirely by the equivalent of interpolation, which is nearly as powerful in this model as in the old. The hope is that replacing specific rules of context-sensitive evaluation with windows in effect tightens the model and eliminates some potential ambiguities in the targets-and-connections framework.

26.3.4 *Conclusion*

In this paper some aspects of continuous spatial movements have been examined for two articulatory dimensions, velum and jaw. When segments are combined in sequences, some are more variable, and thus more accommodating to context, than others along a given articulatory dimension. The window model presented here uses those differences in variability as the basis for deriving the continuous spatial movements. In effect, the feature values for segments are translated in the phonetics into more or less specified quantitative values along the articulatory dimensions. There are many aspects of this proposal that obviously require expanded and more precise formulations. More generally, there are many known cases of coarticulation and assimilation that the model must be tested against. What I have tried to do here is to provide a framework and terminology in which contextual variability and continuous representations can be discussed. The proposal is offered at this point with a view to such discussion and future development.

This work was supported by the NSF under grant BNS 8418580. I would like to thank Abigail Cohn, Carol Fowler, Bruce Hayes, Marie Huffman, Kenneth Stevens, and the editors for their comments.

1 In this paper, windows are discussed only in terms of certain articulatory measurements. However, it seems plausible to construct similar windows from variation in acoustic measurements, such as formant frequencies, since these are often used as physical representations. Nonetheless, the nonlinearity of articulatory–acoustic and acoustic–perceptual mappings means that variation in one domain will not translate directly into variation in another.

2 I have seen other examples in the acoustic domain, where a particular formant will reveal a "target" value only in extreme contexts. Thus the second formant for intervocalic [s] interpolates between the surrounding vowels, unless both vowels have very low, or very high, second formants. In such cases the [s] is shown to have its own value in the vicinity of 1,700 to 2,000 Hz, depending on the speaker.

3 Such phonetic domains are usually called *spaces*, e.g. the vowel space. This term is avoided here and later to minimize confusion with the notion of *spatial dimension* used through this paper.

References

Allen, Jonathan, Sheri Hunnicutt, and Dennis Klatt. 1987. *From Text to Speech.* Cambridge: Cambridge University Press.

Amerman, James, Raymond Daniloff and Kenneth Moll. 1970. Lip and jaw coarticulation for the phoneme /æ/. *Journal of Speech and Hearing Research* 13: 147–161.

Benguerel, A.-P. and H. Cowan. 1974. Coarticulation of upper lip protrusion in French. *Phonetica* 30: 41–55.

Baldon, R. A. W. and A. Al-Bamerni. 1976. Coarticulation resistance of English /l/. *Journal of Phonetics* 4: 135–150.

Bladon, R. A. W. and Francis Nolan. 1977. A videofluorographic investigation of tip and blade alveolars in English. *Journal of Phonetics* 5: 185–193.

Card, Elizabeth. 1979. A phonetic and phonological study of Arabic emphasis. Ph.D. dissertation, Cornell University.

Fowler, Carol. 1980. Coarticulation and theories of extrinsic timing. *Journal of Phonetics* 8: 113–133.

Gay, Thomas 1977. Articulatory movements in VCV sequences. *Journal of the Acoustical Society of America* 62: 183–193.

Ghazeli, Salem. 1977. Back consonants and backing coarticulation in Arabic. Ph.D. dissertation, University of Texas.

Goldsmith, John. 1976. Autosegmental phonology. Ph.D. dissertation, MIT. Distributed by Indiana University Linguistics Club.

Keating, Patricia. 1983. Comments on the jaw and syllable structure. *Journal of Phonetics* 11: 401–406.

Keating, Patricia. 1985. Phonological patterns in coarticulation. Paper presented LSA Annual Meeting, Seattle. MS in preparation under new title.

Keating, Patricia and Marie Huffman. 1984. Vowel variation in Japanese. *Phonetica* 41: 191–207.

Keating, Patricia, Bjorn Lindblom, James Lubker, and Jody Kreiman. 1987. Jaw position for vowels and consonants in VCVs. MS in preparation.

Kent, Raymond, Patrick Carney and Larry Severeid. 1974. Velar movement and timing: Evaluation of a model for binary control. *Journal of Speech and Hearing Research* 17: 470–488.

Kent, Raymond and Frederick Minifie. 1977. Coarticulation in recent speech production models. *Journal of Phonetics* 5: 115–133.

Koopmans-van Beinum, Floria. 1980. Vowel contrast reduction: An acoustic and

perceptual study of Dutch vowels in various speech conditions. Dissertation, University of Amsterdam.
Lindblom, Bjorn. 1983. Economy of speech gestures. In Peter MacNeilage (ed.) *The Production of Speech*. New York: Springer-Verlag.
MacNeilage, Peter. 1970. Motor control of serial ordering of speech. *Psychology Review* 77: 182–196.
Manuel, Sharon. 1987. Acoustic and perceptual consequences of vowel-to-vowel coarticulation in three Bantu languages. Ph.D. dissertation, Yale University.
Manuel, Sharon and Rena Krakow. 1984. Universal and language particular aspects of vowel-to-vowel coarticulation. *Haskins Laboratories. Status Report on Speech Research* SR77/78: 69–78.
Moll, Kenneth and R. Daniloff. 1971. Investigation of the timing of velar movements during speech. *Journal of the Acoustical Society of America* 50: 678–694.
Öhman, Sven. 1966. Coarticulation in VCV utternaces: spectrographic measurements. *Journal of the Acoustical Society of America* 39: 151–168.
Öhman, Sven. 1967. Numerical model of coarticulation. *Journal of the Acoustical Society of America* 41: 310–320.
Pierrehumbert, Janet. 1980. The phonology and phonetics of English intonation. Ph.D. dissertation, MIT.
Schouten, M. and L. Pols. 1979. Vowel segments in consonantal contexts: a spectral study of coarticulation – Part I. *Journal of Phonetics* 7: 1–23.
Sharf, Donald and Ralph Ohde. 1981. Physiologic, acoustic, and perceptual aspects of coarticulation: Implications for the remediation of articulatory disorders. In Norman Lass (ed.) *Speech and Language: Advances in Basic Research and Practice*. Vol. 5. New York: Academic Press.
Steriade, Donca. 1982. Greek prosodies and the nature of syllabification. Ph.D. dissertation, MIT.
Sussman, Harvey, Peter MacNeilage, and R. Hanson. 1973. Labial and mandibular movement dynamics during the production of bilabial stop consonants: Preliminary observations. *Journal of Speech and Hearing Research* 16: 397–420.
Tatham, Marcel. 1984. Towards a cognitive phonetics. *Journal of Phonetics* 12: 37–47.
Vaissière, Jacqueline. 1983. Prediction of articulatory movement of the velum from phonetic input. MS. Bell Laboratories.

27

Some factors influencing the precision required for articulatory targets: comments on Keating's paper

KENNETH N. STEVENS

1 Introduction

The framework that Keating has presented for describing articulatory movements has a number of desirable features. It incorporates the view that the positioning of articulatory structures to implement a particular feature or feature combination can occur within a certain range of values or "window," and that the movements that occur when there is a feature change need only follow trajectories that do not violate these windows. Whether, as Keating suggests, there is no preference for different target positions within a window might be open to debate, since there is at least some evidence that certain acoustic characteristics of clear speech show greater strength or enhancement than those for conversational speech (Picheny, Durlach, & Braida 1986). That is, there appears to be some justification for the notion that the strength of the acoustic correlate of a feature may vary, and one might conclude, then, that some regions within a window (as defined by Keating) are to be preferred over others.

In Keating's framework, a target region is characterized not only by a range of values for an articulator but also, through the specification of horizontal parallel lines, by a time span within which the articulatory dimension must remain as it follows a trajectory defined by a feature change. The implication is that an articulatory structure should remain within a target region over a time interval in order to implement adequately the required feature or feature combination. Thus, for example, a narrow window would require an almost stationary position for an articulator over a substantial time span. Examination of articulatory movements suggest, however, that during speech articulatory structures rarely remain in a fixed position for very long, except when two structures are approximated, as for a stop consonant. A possible modification of Keating's proposal would be to specify windows with particular widths, as she suggests, but to require that these windows be reached or passed through and not necessarily maintained over a time interval. Interpolation between targets would be determined through some