Phonetics Lab Copy

UCLA Working Papers in Phonetics 69

February 1988

koʷɑɹtɪkjuleɪʃən

COARTICULATION

Coarticulation

UCLA Working Papers in Phonetics 69

February 1987

Coarticulation and timing*

Patricia A. Keating

Coarticulation refers to overlap of articulatory gestures associated with separate speech segments, and by extension to its acoustic effects. As a result of coarticulation, segments vary according to their contexts. Thus the /t/ in *tea* may have spread lips while the /t/ in *two* may have more rounded lips, anticipating the rounding of the following vowel. Similarly, the /u/ in *choose* may have a fronted tongue position, due to the surrounding consonants, than the /u/ in *poof*; and the vowel in *man* may be nasalized, unlike that in *bad*. The term coarticulation was contributed in the 1930s by P. Menzerath, who (like others before him, especially E. W. Scripture) argued against the view that successive speech sounds consist of discrete steady states and transitions. Hardcastle (1981) provides a history of early work on coarticulation.

Though coarticulation primarily refers to timing of articulations, it can also involve contextual effects on the spatial extent of articulations, such as reductions in articulatory gestures at faster rates of speech. Such effects eliminate extreme articulatory movements, and thus coarticulation has been linked to ease of articulation.

Coarticulation is also sometimes used interchangeably with assimilation. However, Menzerath apparently meant to call attention to the physiological basis of assimilation: coarticulation the cause, assimilation the effect. Nonetheless, there are many types of assimilation, not all articulatory, and today there are many different opinions on the relation between coarticulation and assimilation.

Many models of coarticulation have been proposed. One (by Menzerath, but also independently influential later) is that articulations begin as

---

*To appear as an entry in the *Oxford International Encyclopedia of Linguistics*, edited by William Bright for the Oxford University Press.

early as possible; another is that vowel articulations begin during consonants and vice versa. Alternatively, certain articulations begin a fixed amount of time before other articulations. It has also been proposed that higher levels of structure, such as syllables, words, or phrases, may be involved in coarticulation.

Other studies of coarticulation (dating back to the earliest work) have been concerned with possible cross-language differences, and more recently, with general principles that would account for such differences. A related question concerns the acquisition of coarticulation: if languages differ, then speakers must learn the language-specific pattern.

The study of timing in speech can also involve measuring durations of particular events or units, in either physiological or acoustic records (Lehiste 1970). Due to coarticulation, measurements of segment durations are necessarily arbitrary, but have been shown to vary across contexts. Study of segment durations has not been well integrated with the study of coarticulation, though both are concerned with speech timing. However, Carol Fowler (1980) has proposed to unify the two, claiming that measured durations of segments reflect the extent to which they overlap. While this proposal remains controversial, the effort is an important one.

Most of the work on coarticulation has been published in the major phonetics and speech science journals, rather than in books. The *Journal of Phonetics* has published several review and debate papers since 1977.

References

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

Harcastle, W. J. (1981). Experimental studies in lingual coarticulation. In R. E. Asher & J. A. Henderson, eds., *Towards a history of phonetics.* Edinburgh University Press.

Lehiste, Ilse. (1970). *Suprasegmentals.* MIT Press.

The window model of coarticulation: articulatory evidence*

Patricia A. Keating

1. Introduction

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 categorial 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

---

*This paper is a revised version of one presented at the First Conference on Laboratory Phonology, held at the Ohio State University in June 1987, and will appear in the proceedings, edited by J. Kingston & M. Beckman.

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 property, or assign an amount that changes 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? In this paper, a new account of the spatial aspect of continuous representations is proposed. 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 in with a segmental feature analysis, traditional or autosegmental. In both of 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, categorial 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 1960's and early 1970's, 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 & Daniloff (1971) and Benguerel & Cowan (1974), for example, were completely phonological in character. Not surprisingly, binary spreading analyses generally proved inadequate (see, for example, Kent & Minifie 1977). The data that were being analyzed were continuous 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.

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 process of connecting, target 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 F0 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 F0 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 Ohman (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, & 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). Ohman (1966) used this mechanism, interpolation between sparse values, to produce tongue body coarticulatory effects on consonants, and Fowler (1980) draws on Ohman'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.

2. The window model of contour construction

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 of 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. 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

---

[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.

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 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 different from 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,

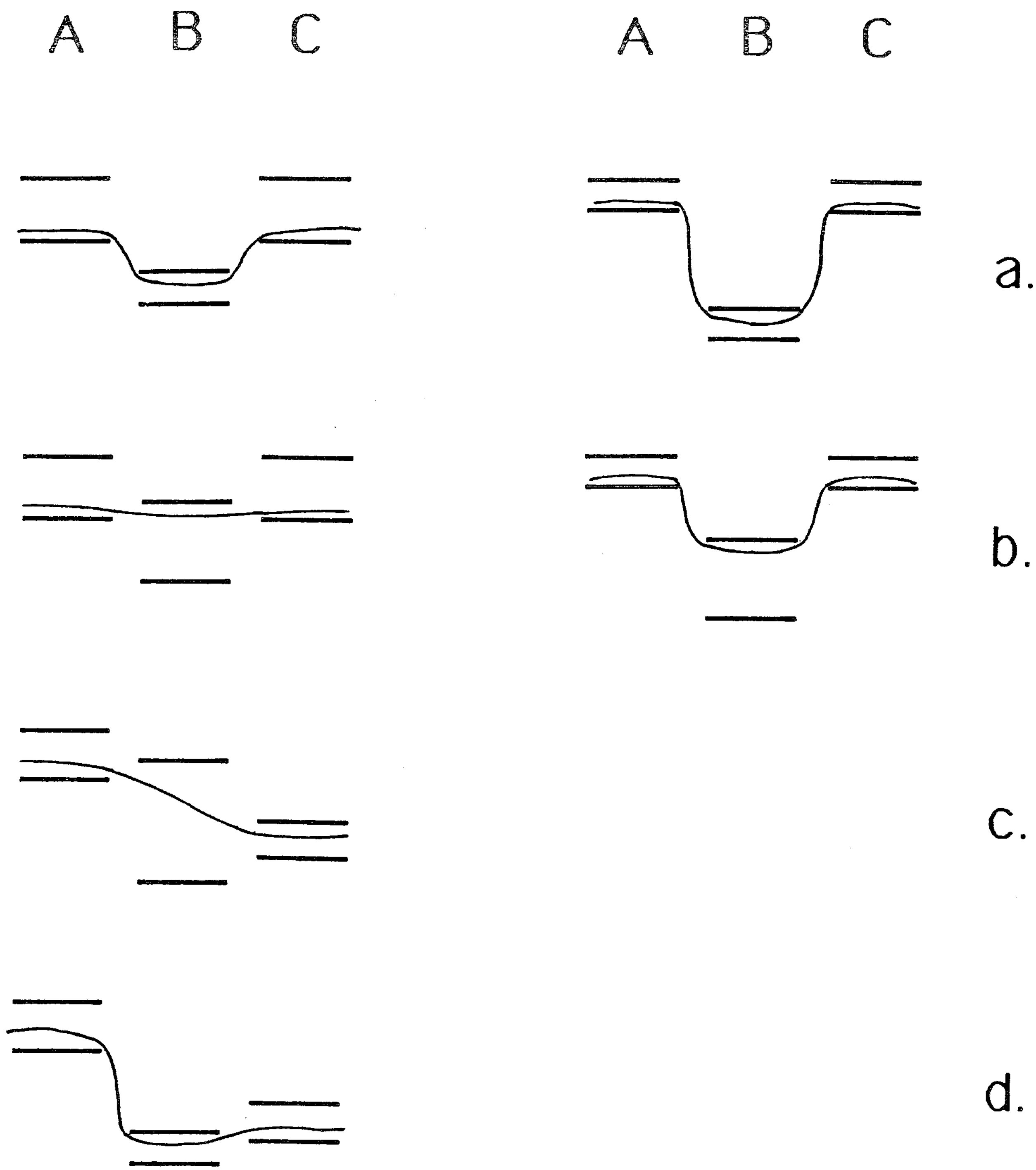


Figure 1. Illustrations of sequences of windows of various widths. See text for description of each sequence.

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.

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 2, after Vaissiere (1983), shows the ranges of values covered by nasal consonants as opposed to oral consonants. These values group together all places of articulation, though Vaissiere 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 different vowel phonemes will be expected to have different windows for velum height. The difference betweem 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 Figure 3, after Kent et al. (1974).

A window analysis of these facts of English runs as follows. Velum height windows are suggested for oral consonants, nasal consonants, and vowels in Figure 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 4. Contours are derived by tracing smooth

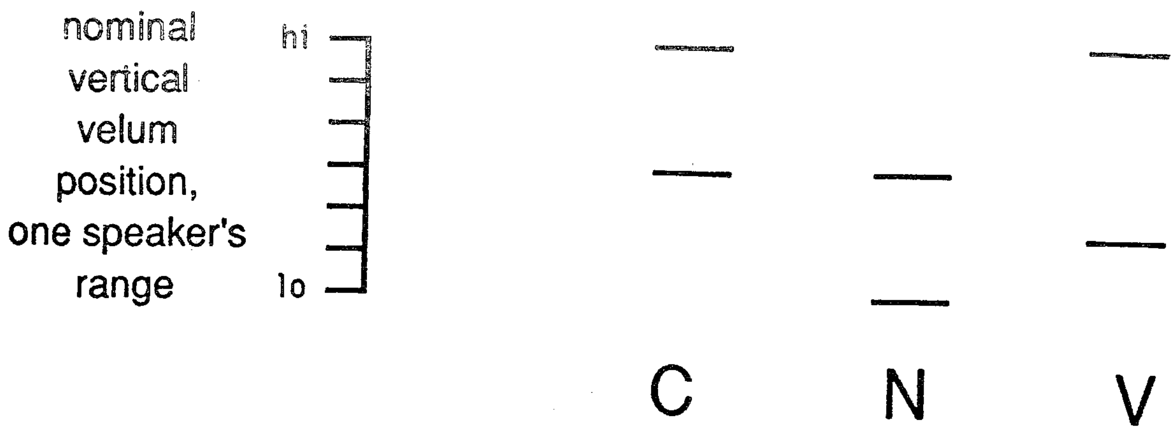
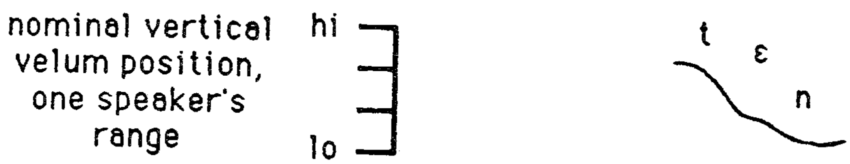


Figure 2. Ranges of values for vertical position of a point on the velum in sentences, based on figures in Vaissiere (1983).



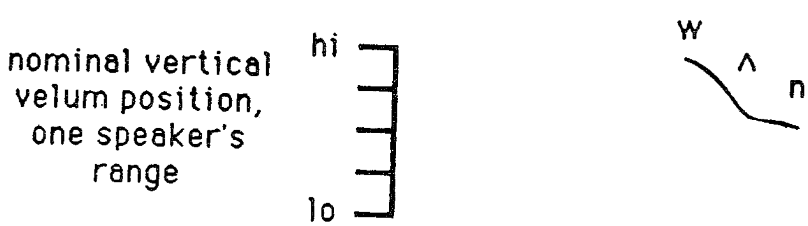


Figure 3. Vertical position of a point on the velum in CVN sequences from sentences, after Kent et al. (1974).

paths through these sequences. Vowels, with their wide windows, easily accomodate 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.

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 5. It is also apparent in Vaissiere (1983)'s 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. Vaissiere 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.

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 almost any A and C will satisfy B's window. Thus B's values, and in fact its entire trajectory, will usually depend

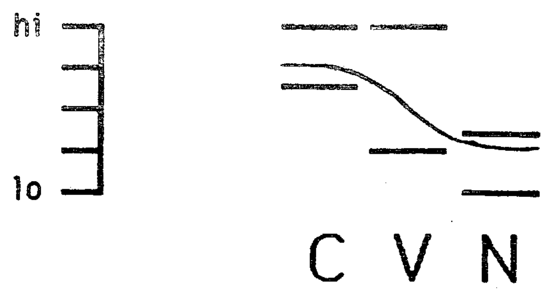


Figure 4. Schematic velum height windows, based on data in Figure 2 and other data from Vaissiere (1983), with contour.

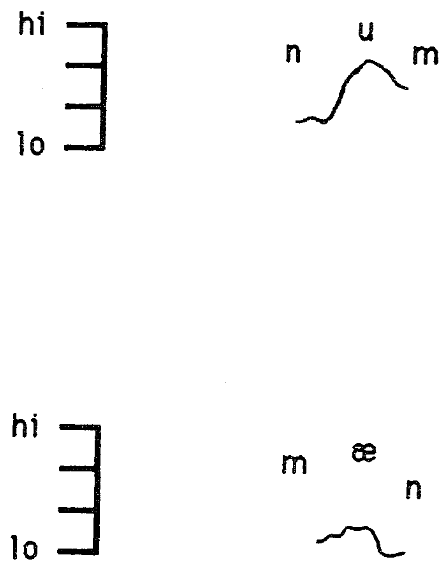


Figure 5. Observed raising of velum for vowel between two nasal consonants, after Kent et al. (1974).

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 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

---

[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 1700 to 2000 Hz, depending on the speaker.

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 <u>reduces, not introduces</u>, 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 indeterminism 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.

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 data of Amerman, Daniloff & 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 has 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 & Ohde (1981), in their review of coarticulation, do: they cite this study, plus Gay (1977), and Sussman, MacNeilage & Hanson (1973), as showing "feature spreading and shifts in target position" for the jaw.

Is jaw position during consonants, then, an anticipation 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 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

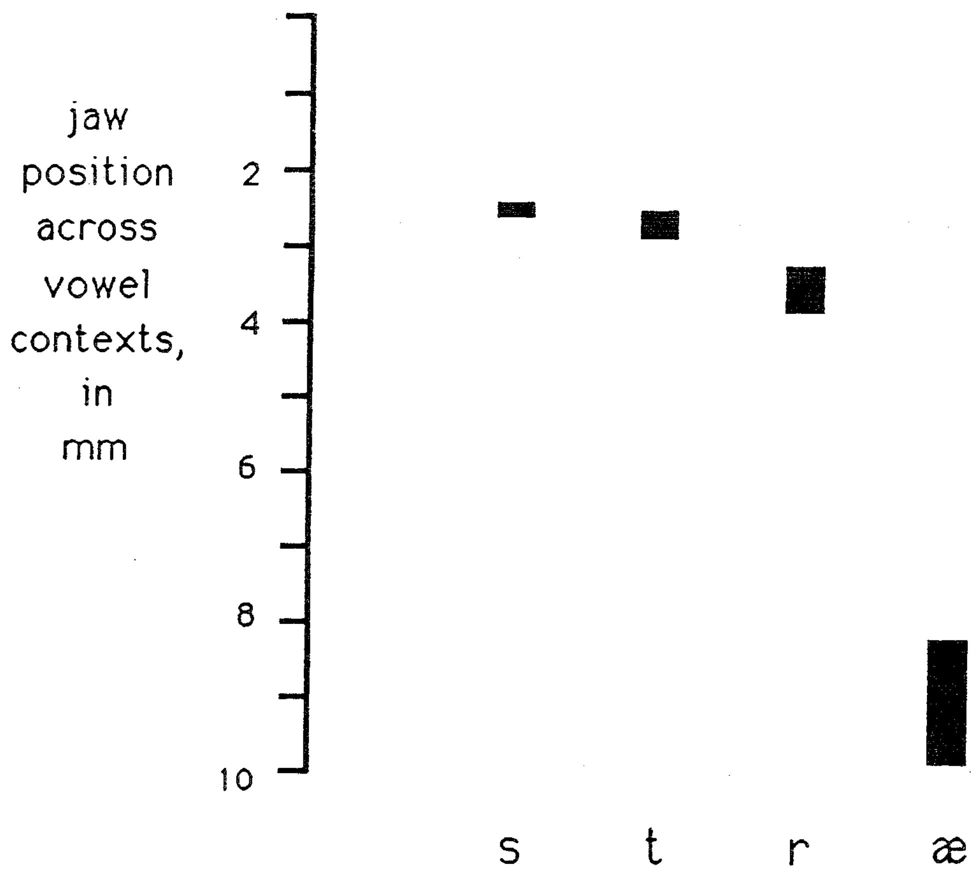


Figure 6. Range of mean extreme values for jaw position for four segments, from data of Keating et al. (1987).

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 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 8: high and narrow for /s/, similar but slightly wider for /t/, middle and medium for /r/, and low and wide for /æ/. In the figure, the 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.

3. General Discussion and Conclusion

The window model is a proposal about how successive segments are accomodated 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.

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

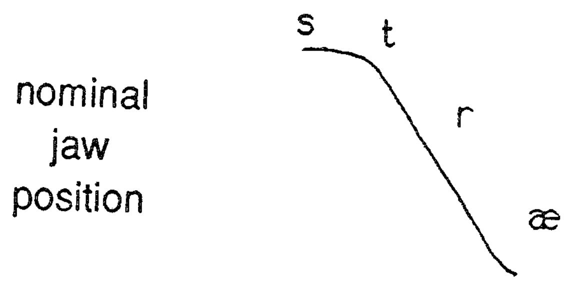


Figure 7. Vertical position of a point on the jaw in /stræ/, after Amerman et al. (1970).

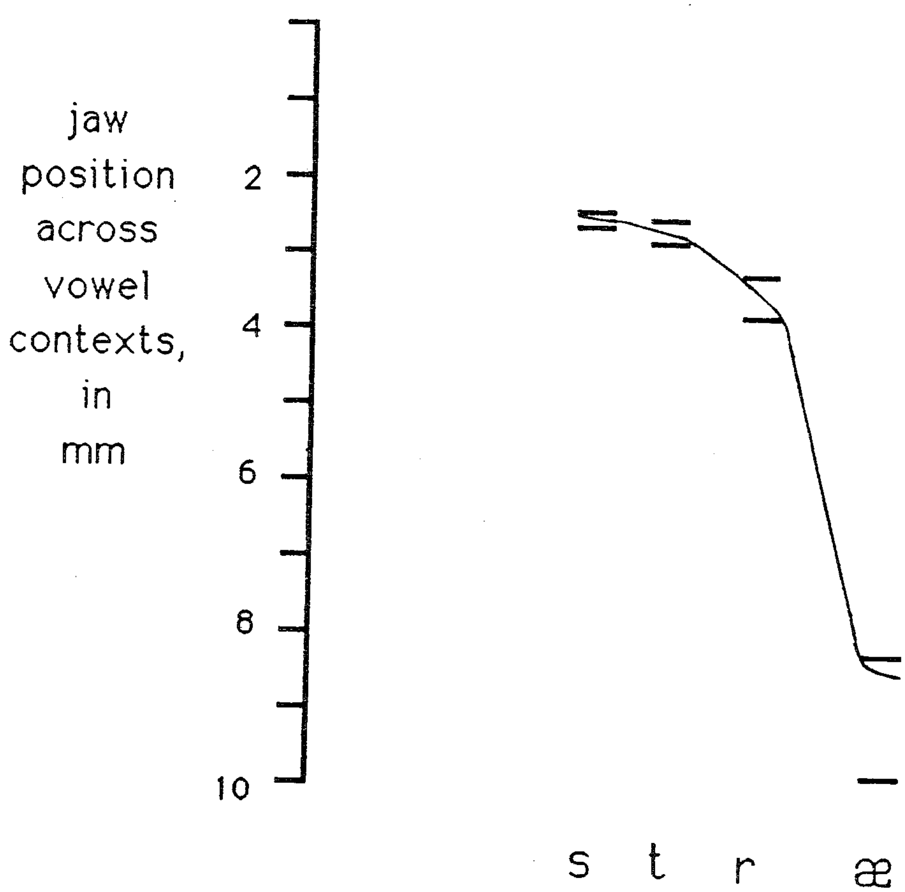


Figure 8. Sequence of schematic jaw windows for /stræ/, with contour.

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 either of 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 features a segment is specified for, the more narrow windows it is likely to have 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 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, non-linguistic constraints on variation may be found.

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 & Al-Bamerni 1976, Bladon & Nolan 1977) proposed

---

[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 throughout this paper.

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 incompatibly specified segments. The difference is that Lindblom's blocking was gradient in nature; segments would <u>more or less</u> block coarticulation. However, Lindblom did not quantify his dimension; indeed, it is not obvious how this could be done.

More recently, Manuel & Krakow (1984) and Manuel (1987) have discussed variability in terms of what they call *output constraints*. The idea here is that the size and distribution of a phonemic inventory determines the limits on each 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 & 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 & 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

Acknowledgements

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.

References

Allen, Jonathan, Sheri Hunnicutt, & Dennis Klatt (1987). *From text to speech.* Cambridge, UK: Cambridge University Press.

Amerman, James, Raymond Daniloff, & Kenneth Moll (1970). Lip and jaw coarticulation for the phoneme /æ/. *J. Sp. Hear.* Res. 13:147-161.

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

Bladon, R. A. W. & A. Al-Bamerni (1976). Coarticulation resistance of English /l/. *J. Phon.* 4:135-150.

Bladon, R. A. W. & Francis Nolan (1977). A videofluorographic investigation of tip and blade alveolars in English. *J. Phon.* 5:185-193.

Card, Elizabeth (1979). *A phonetic and phonological study of Arabic emphasis.* Ph.D. dissertation, Cornell U.

Fowler, Carol (1980). Coarticulation and theories of extrinsic timing. *J. Phon.* 8:113-133.

Gay, Thomas (1977). Articulatory movements in VCV sequences. *J. Acoust. Soc. Am.* 62:183-193.

Ghazeli, Salem (1977). *Back consonants and backing coarticulation in Arabic.* Ph.D. dissertation, U. Texas.

Goldsmith, John (1976). *Autosegmental Phonology.* Ph.D. dissertation, MIT, distributed by Indiana University Linguistics Club.

Keating, Patricial (983). Comments on the jaw and syllable structure. *J. Phon.* 11:401-406.

Keating, Patricia (1985). Phonological patterns in coarticulation. Paper presented at LSA Annual Meeting, Seattle. Ms in preparation under new title.

Keating, Patricia, & Marie Huffman (1984). Vowel variation in Japanese. *Phonetica* 41:191-207.

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

Kent, Raymond, Patrick Carney & Larry Severeid (1974). Velar movement and timing: Evaluation of a model for binary control. *J. Sp. Hear. Res.* 17:470-488.

Kent, Raymond & Frederick Minifie (1977). Coarticulation in recent speech production models. *J. Phon.* 5:115-133.

Koopmans-van Beinum (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. *The Production of Speech*, ed. by Peter MacNeilage. New York: Springer-Verlag.

MacNeilage, Peter (1970). Motor control of serial ordering of speech. *Psych. Rev.* 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 & 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, & R. Daniloff (1971). Investigation of the timing of velar movements during speech. *J. Acoust. Soc. Am.* 50:678-694.

Ohman, Sven (1966). Coarticulation in VCV utterances: spectrographic measurements. *J. Acoust. Soc. Am.* 39:151-168.

Ohman, Sven (1967). Numerical model of coarticulation. *J. Acoust. Soc. Am.* 41:310-320.

Pierrehumbert, Janet (1980). *The phonology and phonetics of English intonation.* Ph.D. dissertation, MIT.

Schouten, M. & L. Pols (1979). Vowel segments in consonantal contexts: a spectral study of coarticulation - Part I. *J. Phon.* 7:1-23.

Sharf, Donald, & Ralph Ohde (1981). Physiologic, acoustic, and perceptual aspects of coarticulation: Implications for the remediation of articulatory disorders. *Speech and Language: Advances in Basic Research and Practice*, Vol. 5, ed. by Norman Lass. New York: Academic Press.

Steriade, Donca (1982). *Greek prosodies and the nature of syllabification.* Ph.D. dissertation, MIT.

Sussman, Harvey, Peter MacNeilage & R. Hanson (1973). Labial and mandibular movement dynamics during the production of bilabial stop consonants: Preliminary observations. *J. Sp. Hear. Res.* 16:397-420.

Tatham, Marcel (1984). Towards a cognitive phonetics. *J. Phon.* 12:37-47.

Vaissiere, Jacqueline (1983). Prediction of articulatory movement of the velum from phonetic input. Bell Labs. ms.

Underspecification in Phonetics

Patricia A. Keating

0. Introduction

It is often assumed in work on phonological underspecification of segments that while representations may at first be underspecified, they end up fully specified. Various kinds of rules are posited to ensure that in output forms all features have values for all segments. In this paper I will consider an alternative view: that underspecification may persist into phonetic representations. I begin by reviewing some of the relevant phonological phenomena and mechanisms, and the history of underspecification in phonetic studies. I then show how phonetic data may reflect the presence or absence of feature values in surface forms. Finally, the extent to which surface specification depends on segmental contrasts is examined.

1. Underspecification in phonological representations

1.1. Phonological behavior of underspecified segments. Two types of arguments are often offered in support of phonological underspecification: what I will call arguments from variability and arguments from transparency. An example of an argument from variability is one based on vowel harmony, where a vowel's value for a certain feature varies, depending on values of other vowels in the word. Because the vowel that varies seems to have no particular original value of its own, it may be analyzed as being unspecified for that feature, and subject to a rule of harmony. That is, the segment is analyzed as being underspecified because it takes on more than one value for a feature; a non-contrasting segment can undergo a feature-assigning rule. Steriade (1987) dismisses this kind of argument because a segment could be originally specified, but later undergo a delinking rule. However, Clements (1987) and others are more willing to argue from such cases. By contrast, Steriade stresses the value of arguments from transparency. In these cases, a segment is analyzed as being unspecified for a feature because it seems to have no value at all. An example of an argument from transparency would be that a consonant does not trigger, undergo, or block harmony -- that it is completely invisible to the harmony process. (It should be noted that these two situations are not incompatible: a given segment could behave transparently with respect to one rule, then acquire a feature value by a later harmony rule.)

1.2. Phonological rules to supply unspecified feature values. What happens to underspecified segments? At least three broad types of rules are able to provide their missing feature values. One kind applies context-freely; two others apply context-sensitively.

Fill-in rules. This category comprises rules that fill in feature values without reference to context beyond the segment in question, and includes rules that either refer to other feature values of the segment, or refer to no other information. For example, a fill-in rule may introduce the feature value [+voice] for sonorants, or the feature value [-low] for [+high] segments; or if only [+voice] is underlying, a fill-in rule may introduce [-voice] for any segment lacking a value for this feature. Thus this broad category includes Archangeli & Pulleyblank (1986)'s default and complement rules, and Steriade (1987)'s redundancy rules.

Sometimes, despite lack of contrast, a segment consistently has a particular phonetic quality, so that it behaves as if it had a value for the particular feature. An example would be [s], which always appears to be phonetically [+high], though this feature value is not contrastive. The tongue body simply needs to be high to position the blade appropriately. (At the same time, [s] lacks the extreme fronting or backing that would make it palatalized or velarized.) Another example would be the relation between backness and redundant rounding for vowels in many languages: back vowels tend to be round, and front vowels unrounded. This kind of relation is called "enhancement" by Stevens, Keyser & Kawasaki (1986), meaning that the redundant value of one feature is chosen to enhance the distinction conveyed by the other feature. Thus the value [+high] enhances the value [+strident] for the alveolar [s], while [+round] enhances [+back] for vowels. The crucial point is that enhancement rules, like fill-in rules generally, supply feature values by table look-up, on the basis of the individual segment or segment classes.

Position rules. At least two kinds of context-sensitive rules are possible.

In both cases a segment clearly seems to have a non-contrastive feature value, but that value depends on the context. I call "position rules" those rules that fill in feature values on the basis of a segment's position in the string of segments, in syllable structure, etc., but without reference to melodic feature values of neighboring segments. Examples of such rules would include those that govern velarization of /l/ (i.e. assign [+back]) solely on the basis of syllable position, and those that assign the many allophones of English voiced and voiceless stops (whatever features may be involved). The feature values assigned by position rules would then be available for later rules in just the way that values given by fill-in rules would be.

Context rules. I call "context rules" those rules that fill in feature values on the basis of neighboring segments' feature values. These are rules of assimilation, including harmony, and dissimilation. Examples of an allophonic context rule would be velarization of /l/ dependent on the following vowel's backness, or fronting of velars before front vowels. A nice case of velar fronting is seen with the Russian fricative /x/, which assumes a quite steady fronted quality when followed by /i/; this case will be discussed below. Furthermore, a few of Stevens et al. (1986)'s enhancement rules are context rules, even if not obviously assimilatory or dissimilatory; a redundant value for one segment enhances a contrastive value in an adjacent segment. For example, they suggest that vowel lengthening enhances [+voiced] in consonants.

The difference between position and context rules is in the nature of the output representations. A context rule typically supplies feature values by spreading values already linked to other segments; a position rule does not.

2. Underspecification in phonetic representations

Consider now another possibility for underspecified segments: that they remain underspecified even in surface representations. In such cases, no segmental feature rules apply; the only rules that apply are quantitative phonetic rules, and these rules respect the lack of a feature value for an underspecified segment. Clements (1987) gives this as a possible kind of argument for underspecification.

The idea that even phonetic representations may contain underspecified segments is not new. Within speech science, during many years of work on coarticulation, some kind of phonological underspecification has been assumed, and in some proposals the underspecification persists into phonetic representations. More recently, the approach to intonation of Pierrehumbert (1980) and later work posits underspecification of syllables for tones. I begin discussion of surface underspecification by reviewing these proposals.

2.1. Coarticulation. Phoneticians working on coarticulation, or the articulatory overlap of segments, have assumed since at least the 1930's that coarticulation occurs in part because segments may lack inherent specification for particular articulations. That is, a given segment is characterized by some articulations, but is neutral with respect to other articulations. For example, an alveolar consonant is characterized by a tongue blade articulation, but is neutral regarding lip rounding. Three influential proposals about coarticulation from the 1960's relied on such a distinction between specified and unspecified articulations. One proposal was from Kozhevnikov & Chistovich (1965): they proposed that the articulation of lip rounding in Russian was begun at the onset of the syllable, where "syllable" simply meant a vowel plus any number of preceding consonants. Presumably such coarticulation over a whole syllable would not be possible if segments had various contradictory articulatory specifications, and so this proposal depended on underspecification of consonants for lip rounding.

A second influential contribution was Henke's (1966) computer model of the articulation of English stop + vowel sequences. The model took as input a sequence of segmental phonemes, associated each phoneme with an articulatory goal, and gave as output a sequence of vocal tract configurations (like those shown in X-ray tracings) at one-msec intervals. Henke's main concern appears to have been the design and implementation of a computer program that could access and integrate various kinds of information needed to move articulators from one goal to the next. However, the work is best known for its "look-ahead" mechanism for anticipatory coarticulation. Henke rejected the view that segments always have complete targets; a stop consonant will always have a goal for the stop occlusion, but not necessarily other articulatory goals. With "look-ahead", Henke proposed that as soon as the stop contact is made (but no earlier), the stop looks ahead to the vowel's targets for other articulators. Henke found that anticipation of this very limited kind sufficed for the X-ray data on CV syllables he was modeling; Gay (1977)'s X-ray data on tongue movements are similar to Henke's in this respect. Henke's proposal is thus similar to Kozhevnikov & Chistovich's; differences arise because they examined different segment sequences.

The third proposal, by Öhman (1966, 1967), was addressed to coarticulation between vowels. Öhman's model, like the others, assumed that segments are not specified for all articulators. However, Öhman's model incorporated a very explicit and restrictive version of this idea, namely that vowels and consonants are articulated by independent articulators, called "channels of articulation". Though Öhman's model was in terms of articulators, not phonological features, this is much like saying that vowels and consonants are specified for different sets of features. There is a set of features for which vowels are specified but consonants unspecified, and another set of features for which the reverse holds. I call this proposal complementary underspecification, since the sets of features are complementary across segments. Given complementary underspecification, a given feature usually does not have values on adjacent segments, and this, in Öhman's view, is the source of vowel interactions: when articulatory values are connected up, segments without values will be invisible. With respect to vowel articulations, the influence of one vowel will extend through an adjacent consonant and right into the edges of the next vowel.

In all of these formulations, the basic descriptive dimension is the articulator and its gestures or goals, rather than features per se. Thus all three incorporate underspecification relative to fairly shallow representations. Interestingly, this is not true of subsequent, more overtly phonological, work on coarticulation. In later work (e.g. Moll & Daniloff 1971, Benguerel & Cowan 1974, Kent et al. 1974), other researchers directly equated articulations with binary features[1]. In particular, Henke's look-ahead mechanism for anticipatory coarticulation was extended to provide a "feature-spreading" mechanism[2]. Under feature-spreading, binary feature values were copied from upcoming specified segments to unspecified segments, for an unbounded span of segments. The result was full specification of all segments for all features (except presumably word-final segments). To see how the feature copying worked, consider three segments S1, S2, S3, and three features F, G, H, as in (1).

(1)

	S1	S2	S3		S1	S2	S3
F	0	0	+		+	+	+
G	+	0	–	--->	+	–	–
H	0	–	+		–	–	+

Thus segments are underspecified in their input forms, but not in their output forms[3].

In sum, all of these accounts of coarticulation relied on underspecification at the level of representing segment contrasts; the underspeci-

---

[1] These binary features were not necessarily the same as those posited by linguists on the basis of phonological behavior; instead they were based directly on vocal tract dimensions. Thus, for example, Sharf & Ohde (1981) seem to assume a "feature" for jaw position.

[2] Like others, I have in the past mistakenly ascribed this view to Henke.

[3] The model and its look-ahead mechanism say nothing about what happens after that, for example, exactly when an anticipated property (or any property) would be seen in the physical signal. That would depend on additional assumptions or proposals about the phonetic component that would interpret feature values as temporal events (for example, Henke's own model). Little attention was given to such a component, and as a result the binary feature spreading model was found wanting in its ability to appropriately predict details of timing of coarticulation (Kent et al. 1974; Kent & Minifie 1977). Other difficulties are discussed by Gay (1978), Sharf & Ohde (1981), and Kent (1983).

fication assumed in this work was across the board absence of a feature value for a class of segments. For example, in English the feature [nasal] would have a plus value for nasal stops, a minus value for oral stops, and potentially no value for any other segment. However, such representations were often conceived in terms of articulations rather than phonological features. To the extent that articulatory plans are phonetic representations, then some of these accounts also incorporated phonetic underspecfication.

2.2. Intonation. Pierrehumbert (1980 and later work) proposed explicit deep and surface representations for intonation patterns using tonal elements. A major result of this work was that the number of tones needed at either level to specify the intonation or fundamental frequency pattern was usually far less than the number of syllables, or potential tone-bearing units, in the string. The same result holds in Pierrehumbert & Beckman (1987) for Japanese, where surface representations have traditionally been characterized by as many tonal elements as there are units to bear them, provided by fill-in rules and/or spreading rules. Pierrehumbert & Beckman instead show that quantitative phonetic rules are responsible for the surface patterns that the earlier feature-manipulating rules had been proposed to generate.

Pierrehumbert & Beckman note that surface tonal underspecification is a deep similarity between Japanese and English. In both languages, as a result of surface underspecification, rules that interpolate fundamental frequency contours between target fundamental frequency values play a major role in determining contour shapes. There is an inverse relation between degree of specification and importance of quantitative rules: the fewer the tones, the more of the contour is provided by the phonetic rules; thus the quantitative rules for fundamental frequency are much more powerful than was envisioned in prior generative work.

In sum, the idea of surface underspecification is not new in phonetics, and indeed is central to current work in the area of intonation. For segmental features, however, it has generally been simply assumed, and has never been discussed in relation to proposals about phonological representations.

3. Phonetic effects of phonetic underspecification: phonetic transparency

Suppose a segment exits from the phonology with no value for a feature, having failed to receive one either lexically or by a later rule. That segment will then be transparent in the phonetics to any rule sensitive to that feature. In particular, when phonetic rules build trajectories between segments, an unspecified segment will contribute nothing of its own to the trajectory. In this section I describe cases that illustrate this point. There are two ways in which the phenomenon can be examined. First, the underspecified segment itself can be studied: is it, on the relevant dimension(s), entirely dynamic or transitional in nature, as would be expected if it has no feature value of its own to contribute? Second, the effects of the underspecified segment on its neighbors can be studied: it should be effectively transparent and should permit its neighbors to interact with each other.

3.1. Segment-internal effects. The transparency of an underspecified segment is perhaps clearest in the case of /h/. As a glottal approximant, /h/ has no intrinsic oral feature values, and is usually described by phonologists as a voiceless version of the following vowel, the result of feature-spreading assimilation. But in connected speech, /h/ is not assimilated to a following vowel. Rather, in most instances /h/ is transparent; the acoustics of intervocalic /h/ show that rather than acquiring feature values, /h/ is simply interpolated right through. Figure 2 shows wideband spectrograms in which, for example, the second formant of the /h/s in three languages clearly interpolates between the surrounding vowels. The apparent assimilatory effect on the /h/ is a dynamic, transitional one, not a static one. /h/ starts out unspecified for oral features, and it remains so. That is, sometimes "assimilation" is a lower-level phonetic effect, rather than spreading of phonological feature values. With /h/, all of the oral features are involved, so the effect is quite striking.

Though the transitional effects are less striking with other segments, where fewer features may be unspecified, they can still sometimes be clearly seen in acoustic displays. Thus Figure 3 shows movement during the second formant of [s], reflecting changing backness and/or rounding, which is a

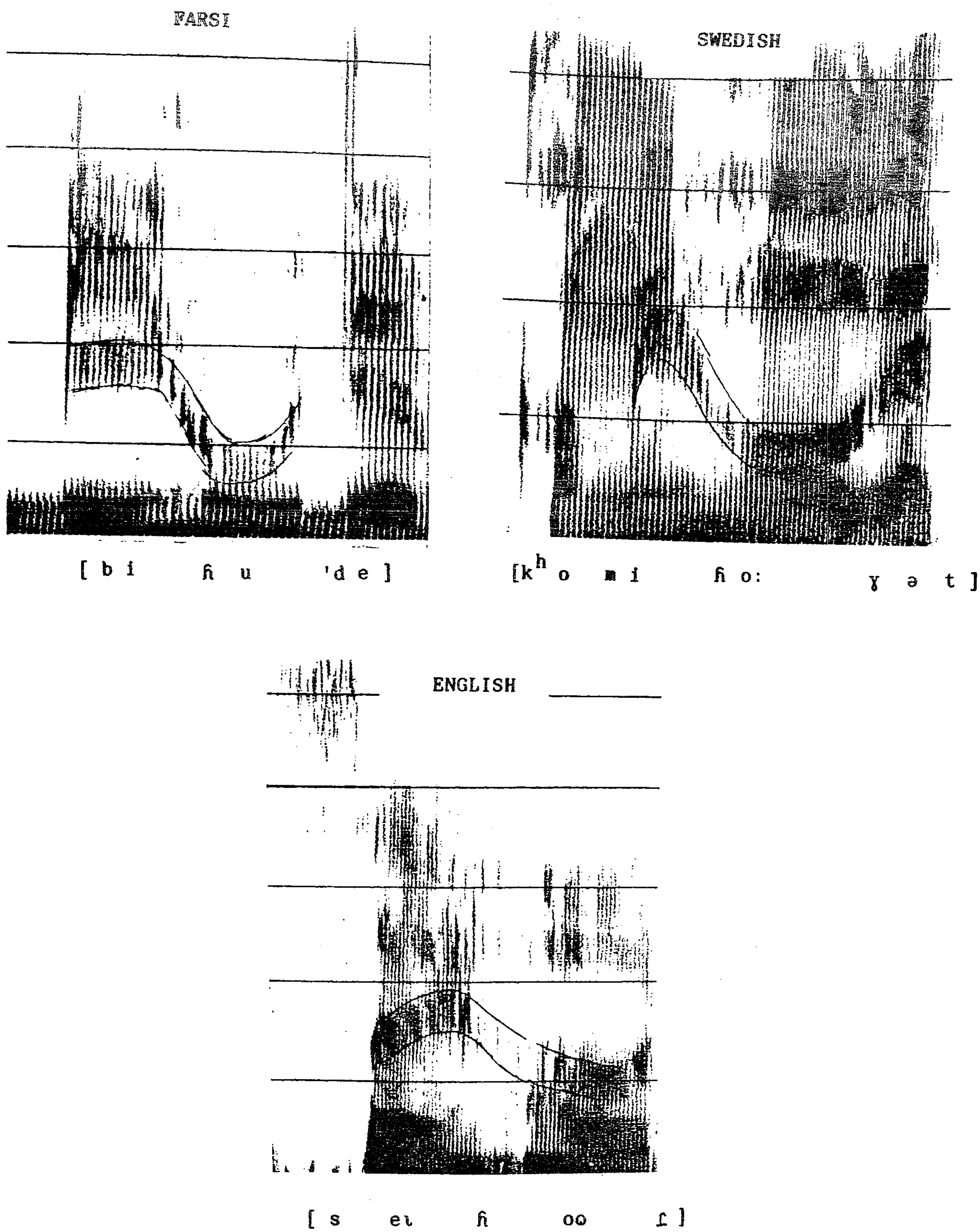


Figure 2. V+/h/+V sequences in words of three languages -- Farsi, Swedish, English -- with the second formant (F2) outlined. /h/ is voiced [ɦ] in each. Calibration lines are at 1000 Hz intervals.

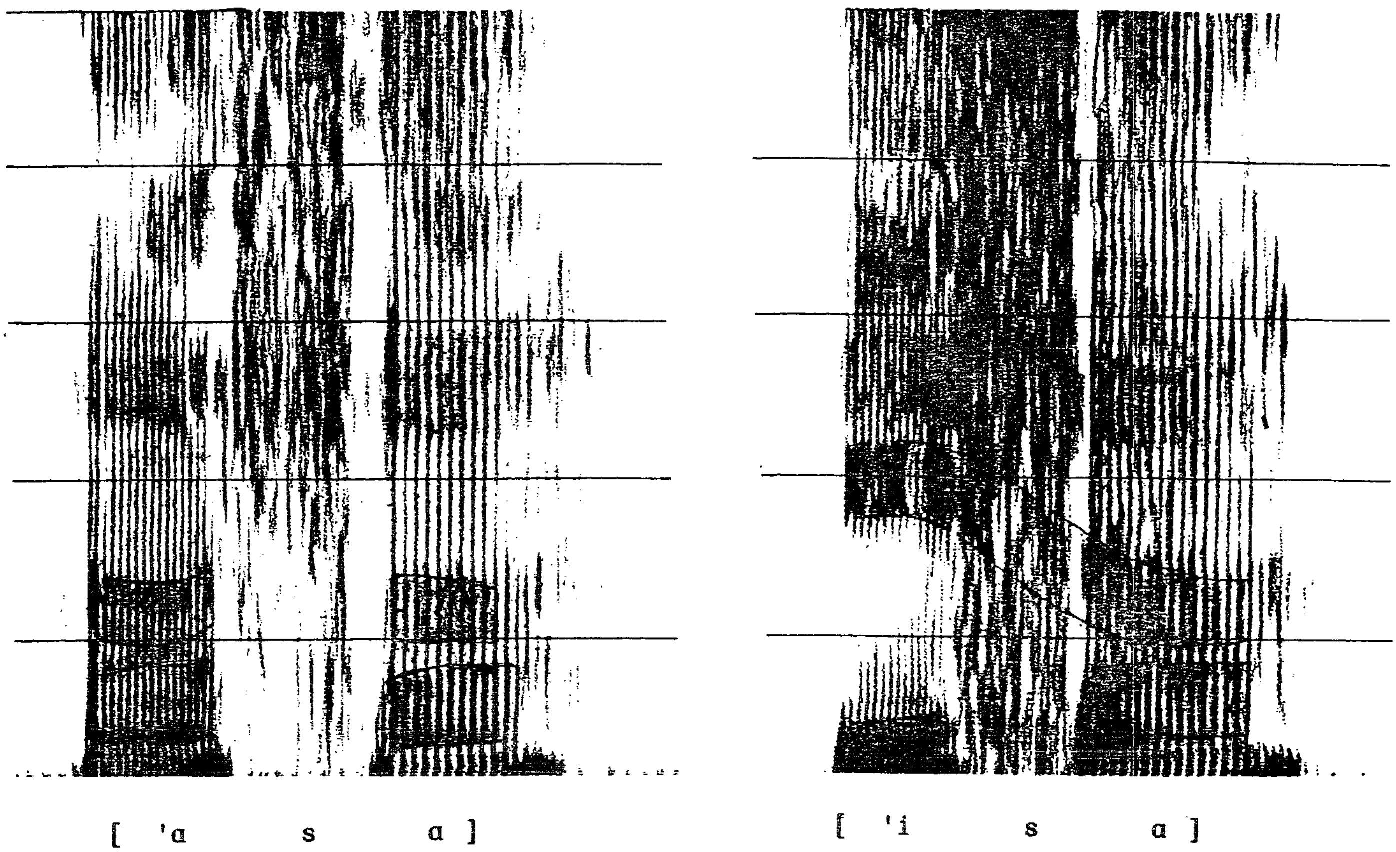


Figure 3. English nonsense items /ɑsɑ/ and /isɑ/, with the first and second formants (F1 and F2) outlined. F1 is the same for V2 in both; F2 is higher after /i/. Calibration lines are at 1000 Hz intervals.

transitional effect of the adjacent vowels.

3.2. Phonetic effects on context. Just as in the phonology, a missing feature value should be apparent from a segment's failure to block a rule from applying across it. A useful test for phonetic underspecification of consonants (though one that sometimes requires judicious interpretation) can be found in the presence or absence of vowel interactions across consonants. Vowel-to-vowel coarticulation, or the effect of one vowel on another even across an intervening consonant, was described by Ohman (1966). He found that in English and Swedish there was a small but consistent effect of vowels on the transitions of other, stressed vowels; others have since found stronger effects on the centers of unstressed vowels in English (Fowler 1981, Bell-Berti & Harris 1976, Magen 1984, Huffman 1986), and on vowels in other languages (Magen 1984, Manuel & Krakow 1984, McAdams 1987). Ohman explained vowel-to-vowel coarticulation through consonants as a consequence of the underspecification of consonants for vowel articulations. In his data (contrary to Henke's) the tongue body made a single smooth gesture from vowel to vowel, as if for a diphthong, overlapping with any consonant gestures. On his account, the influence of the first vowel extends right through the consonant (which the vowel gesture does not see) and into the next vowel, and vice versa in the other direction. Though sometimes subtle, vowel-to-vowel effects can be quantified by measuring variation in vowel formant frequencies across contexts. Many examples of these effects are available in the references cited; in addition, they can be seen in Figure 3 above.

4. Comparison of output types

Particular acoustic patterns can be observed during segments that plausibly lack certain feature values. Thus these acoustic patterns suggest tests that can be applied to determine whether a segment has a particular feature value. However, it is important to compare the results expected as outputs of the various rule types, so as to make clear how they are to be distinguished.

4.1. Context rules vs. phonetic rules. It is not trivial to distinguish the output of context rules, which provide feature values, from the output of phonetic rules, which build continuous contours through transparent segments.

The difference seems clear enough in principle: if a segment acquires a feature value from an adjacent segment, it will share a phonetic property with that segment across most or all of its duration; if a contour is built through a segment it will have a more or less continuously changing, transitional, quality from beginning to end that will depend on context to either side. The examples given above of phonetic rules display this quality. In practice, the intended distinction may not always be obvious; however, an example of a context rule whose output is clearly different from the output of phonetic rules is provided by fronting of the velar fricative /x/ in Russian. Although most Russian consonants contrast as either palatalized or not, /x/ does not, but instead depends on context for its value for [back].

If a segment acquires a feature value from a neighbor, then it should display the relevant property much as the neighbor does. Fricatives allow us to examine the time course of any assimilatory effect. To this end, Figure 4 compares Russian /x/ in mirror-image contexts before and after /i/. The second formant, which reflects backness, is indicated on the spectrograms in this figure. The two tokens of /x/ are very different. In /ixa/, the fricative shows a continuous change from more front to more back, like other examples of phonetic rules considered above. By contrast, in /axi/ the fricative is extremely fronted throughout its duration. This steady quality of the /x/ very different from the preceding /a/, and necessitates an extreme transition, especially of the second formant, between the two segments, and supports the idea that the fricative is [-back] over its entire duration, rather than having only a transitional front quality.

4.2. Fill-in rules. Fill-in values, once assigned, should presumably have the same status as underlying contrastive feature values. One test of this would be whether a fill-in value on a consonant blocks vowel-to-vowel coarticulation across that consonant. As discussed earlier, the consonant [s] in English is usually made with a very high tongue body position. Tongue height is reflected acoustically mainly in the frequency of the first formant, but since during voiceless fricatives this formant is not visible, only effects on neighboring segments can be examined. The lowest two formant frequencies were examined for effects of vowel-to-vowel coarticulation across

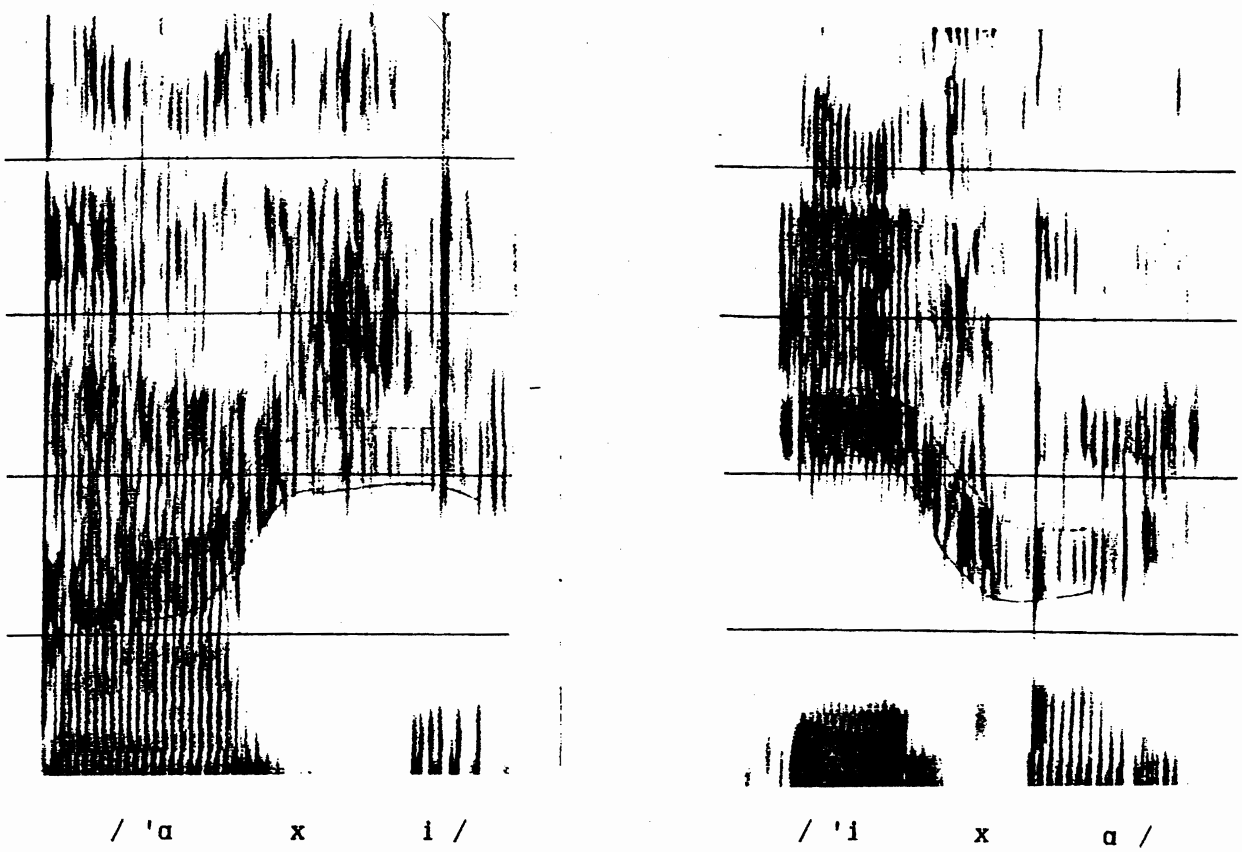


Figure 4. Russian nonsense items /ɑxi/ and /ixɑ/, with the second formant (F2) outlined. The F2 trajectories are not mirror images. Calibration lines are at 1000 Hz intervals.

[s] in VCV nonsense items. The vowels were /i/, /a/, and /u/, with the first vowel receiving main stress. Two repetitions of each were measured for two speakers of American English, using wideband spectrograms to measure the formant frequencies at the end of V1 and the beginning of V2. If vowel-to-vowel coarticulation occurs, then the formant frequency for one vowel will depend in part on the identity of the other vowel. For F2, the amount of coarticulation found was similar to what is reported in the literature across other consonants, with carryover effects stronger than anticipatory. In tokens where F2 can be seen faintly during the [s], it is clearly moving from the first to the second vowel's value. On the other hand, F1 is consistently low in the vowels around the [s] and shows little if any coarticulation. An example can be seen in Figure 3 above. That is, F1, unlike F2, seems to have a fairly fixed value for [s]. (It must be noted that stops consonants have been found to allow vowel-to-vowel coarticulation in F1, so the effect with [s] is not a general property of F1.) Interpreting these data in terms of feature values, we can say that [s] behaves as if it has a value for [high], namely [+high], this value presumably the result of a fill-in rule; and that [s] behaves as if it had no value for [back].

4.3. Three-way comparison. For a given segment, three different output representations may emerge from the phonology, with different acoustic reflexes. These possibilities are illustrated by the schematic representations shown in (5), using VCV sequences.

(5)	(a)	V	C	V	(b)	V	C	V	(c)	V	C	V
		|	|	|		|	\	|		|		|
		F	F	F		F		F		F		F

In (a), each segment has its own value for the feature F, and so each will have its own phonetic quality. There should be no vowel-to-vowel effects in either direction, because the feature value on the consonant will block them; there may be extreme transitions between segments that differ in values for F. In (b), the consonant has acquired a value for F from V2, by a spreading rule. There will be no effect of V1 on V2, because the consonant's value for F again serves to block interactions. However, there will appear to be an effect of V2 on V1, because the consonant's feature value will affect V1.

The transitions between V1 and C will look just like those of case (a), but since in (b) C's value depends on V2, V1's variation will depend on V2 as well as C. Thus there will be apparent asymmetrical vowel-to-vowel effects, but these are instead purely local. Finally, in (c), there will be vowel-to-vowel effects in both directions, and the consonant will lack its own phonetic quality, showing instead only a gradual transitional quality.

Blocking vowel-to-vowel coarticulation, as in (a), simply shows that a segment has a value for a feature in the output from the phonology. It does not indicate whether that value was present underlyingly, or whether it is a later fill-in value. The phonology of the language (the system of contrasts, behavior of the segment in other rules) still has to be consulted to discover the source of the feature value.

5. Role of contrast in phonetic behavior

In the several versions of underspecification theory elaborated recently, systems of contrast do not by themselves determine what features will be underspecified for what values, but they enter in as one factor. For segment classes in which a given feature is distinctive, either one or both values may be given in underlying representations. The kinds of phonetic data described above can bear on the nature of underspecification, providing evidence about which segments are underspecified for which features. Evidence from vowel-to-vowel interactions shows that some, but not all, segments behave as if they have feature values on the surface. In this section I present an example where segment contrasts affect coarticulation, and then discuss the possibility that even certain contrasting segments may lack feature values.

5.1. Underspecification depending on contrast within a segment class. In general, when a segment bears a feature value by virtue of participating in a contrast, we expect that feature value to block phonetic vowel-to-vowel interactions; and when a segment does not contrast and has no feature value, we expect it to allow such interactions. This should be so even within a segment class and within a language. More attention to contrast seems to resolve an apparent contradiction in the literature about vowel-to-vowel coarticulation in Russian. Öhman's model with channels of articulation predicts no coarticulation across Russian stops that contrast in palatalization, and both Öhman and Purcell (1979) found none. However, the same prediction is made for contrasting fricatives, yet Derkach, Fant, & Serpa-Leitão (1970) report anticipatory vowel-to-vowel coarticulation across Russian voiceless fricatives. This result appears to be a counterexample to Öhman's model (though Derkach et al. did not make this interpretation), and to the claim that phonetic data can bear on questions of feature specification. However, Derkach et al. do not report their results in great detail, and they appear to have treated all the fricatives as a single group, not differentiating the various places of articulation, even though some places have pairs contrasting in palatalization while others do not. The surface contrasts in voiceless frictives are shown in (6). The palato-alveolar /ʃ/ is usually described as redundantly [+back] (it takes back vowel allophones), while the velar /x/ is described as allophonically variable. Even speakers who allow a contrast in the velar stops have none for the fricative.

(6)	f	s	ʃ	x
	f'	s'		

Examination of my own data on individual Russian fricatives reveals that the various fricatives do not behave identically. In particular, vowels before the velar fricative show strong effects of following vowels. There may also be lesser effects across the palato-alveolar, but my data are unclear on this point. The clear effects observed with /x/ are due to the assimilation of the fricative to a following /i/, as discussed above. Because the /x/ is affected by the /i/, and because the effect is so strong and prolonged, the /x/ itself influences the preceding vowel, as seen in the previous figure. An extreme transition is required from the vowel to the /x/, and this gives rise to strong measured vowel interactions. That is, what looks like a vowel-to-vowel (through the consonant) effect in this case is actually a vowel to consonant to vowel effect. It is not really an appropriate situation to examine for vowel-to-vowel effects, since phonetically the consonants are not the same at all. Conceivably this effect alone accounts for Derkach et al.'s finding; grouping all fricatives together obscures this principled difference. The status of the palato-alveolar with respect to these phonetic effects remains to be determined.

5.2. Underspecified contrasts. When two segments contrast within a class, must both the plus and minus values of the relevant feature be used? Phonologists distinguish cases in which both feature values characterize a contrast, from those in which only one feature value is present. Thus Steriade (1987) includes [round] as necessarily single-valued, and [voice] and [lateral] as underlyingly single-valued in some languages. According to Steriade, it is rare for a feature to be both distinctive and clearly single-valued within a language, but a few fairly clear cases nonetheless are known.

The same hypothesis can be entertained for phonetic representations, and in fact has been the subject of discussion in the traditional coarticulation literature. In some cases, both members of a contrast behave as if they possess a feature value. In articulatory terms, this is equivalent to saying that one member has one articulation, while the other member has a contrary articulation, such as lip rounding vs. lip spreading. But in other cases, only one member of the contrast behaves as if it possesses a feature value. This is equivalent to saying that only one member has an articulation, such as lip rounding vs. no rounding.

Again, the occurrence of vowel-to-vowel coarticulatory effects can be taken as an indication that a segment lacks its own feature value. For example, consonants with contrastive secondary articulations, differentiated by their values for the feature [back], can be examined to see whether each consonant blocks vowel interactions, as would be expected if each has its own feature value. If only one member of the contrast is specified for [back], then that segment should block vowel interactions, while the other member should be transparent to them. Published data on Russian stop consonants, and other data on fricatives, were discussed above. These data suggest that in Russian both members of the contrasting pairs of consonants are specified for [back], since all contrasting consonants seem to block vowel-to-vowel coarticulation.

However, other data on secondary articulations also suggest the opposite: that in some languages, only half of the contrasting consonants are specified for [back]. One case is that of pharyngealization in Arabic. Dialects differ, but the basic idea is that certain coronal obstruents (plus sometimes other consonants) are paired as having or not having the tongue body in a low back position. The question here is whether the consonants that are not pharyngealized have some other property of their own by way of contrast. My own preliminary data suggest that they do not, that is, that they are simply non-pharyngealized and otherwise variable, and therefore allow vowel interactions. The same finding is presented, with clearer data, by Hussein (1987). That is, a feature such as [back] may in some cases (such as Russian) occur in a given segment class with both of its values, and in other cases (such as Arabic) occur with only one value. This in effect restates traditional descriptions of the phonetic nature of these contrasts.

In general, then, the occurrence or lack of vowel interactions across a consonant is informative and suggestive. Nonetheless, it must be noted that a single instance of such effects cannot be definitive. There are many factors that may affect the occurrence or strength of vowel interactions, including the possibility of language-specific prosodic constraints and rate of speech effects. The intrinsic properties of various consonants will affect the degree of vowel interactions (Huffman 1987). Furthermore, languages may differ in how coarticulatory effects are distributed within a sequence of segments. Therefore it is overall patterns of coarticulatory effects that provide the best evidence.

6. Conclusions

Phonetic data provides detailed information about sound structure that allows us to detect and distinguish different types of assimilatory effects, or the lack of them, in segments and in their contexts. Underspecification theory provides a mechanism for describing where such effects should occur: different acoustic patterns correspond to differences in surface feature specification. Acoustic data reveal that a non-contrasting segment may behave as if it has acquired a feature value through assimilatory rules of feature-spreading; or as if it has acquired one without reference to context; or as if it lacks any such value on the surface.

The claim that phonetic evidence can be used to help diagnose feature specifications has an important implication for the learnability of highly underspecified representations. The occurrence of coarticulatory patterns in speech, to the extent that they are audible, could be useful evidence to a language-learner as to what features are specified in the phonetic form for particular segments, and how. With such phonetic evidence, the learner would not have to accomplish a complete phonological analysis of the language, so as to replace fully specified entries with underspecified ones. Instead, immediately accessible surface information could form the basis for an initial distinction between the three representations shown in (5) above. Of course, this information only reflects surface specification vs. underspecification, and therefore provides evidence only about what could have been originally unspecified. Additional analysis by the learner would be required to arrive at appropriate underlying representations. The inventory of underlying segment contrasts provides limits on what must minimally be specified; phonetic evidence about surface representations provides substantially more information to the learner.

Acknowledgements

This work was supported by the NSF under grant BNS 8418580. Some of this material was presented at various conferences: in Lyons, France at the June 1985 colloquium on multi-linear phonology, in Seattle at the Dec. 1985 LSA meeting; at the UCLA conference on segment structure in Oct. 1987; and in talks elsewhere. I thank those who offered suggestions or comments on those occasions, as well as my colleagues in the UCLA Phonetics Lab.

References

Archangeli, D. & D. Pulleyblank (1986). *The content and structure of phonological representations.* Ms., forthcoming from MIT Press.

Bell-Berti, F. & K. Harris (1976). Some aspects of coarticulation. *Haskins Laboratories SR* 45/46.

Benguerel, A.-P., & H. Cowan (1974). Coarticulation of upper lip protrusion in French. *Phonetica* 30. 41-55.

Clements, G. N. (1985). The geometry of phonological features. *Ph. Yrb.* 2: 225-252.

Clements, G. N. (1987). Toward a substantive theory of feature specification. Presented at UCLA Conference on Segment Structure.

Derkach, M., G. Fant, & A. de Serpa-Leitão (1970). Phoneme coarticulation in Russian hard and soft VCV-utterances with voiceless fricatives. *STL-QPSR* 2-3: 1-7.

Fowler, C. (1981). Production and perception of coarticulation among stressed and unstressed vowels. *J. Sp. Hear. Res.* 46: 127-139.

Gay, T. (1977). Articulatory movements in VCV sequences. *J. Acoust. Soc. Am.* 62: 183-193.

Gay, T. (1978). Articulatory units: segments or syllables? *Syllables and segments.* A. Bell & J. B. Hooper, eds. Amsterdam: North-Holland.

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

Huffman, M. (1986). Patterns of coarticulation in English. *UCLA Working Papers in Phonetics* 63: 26-47.

Hussein, L. (1987). Coarticulation in VCV sequences in Arabic. *J. Acoust. Soc. Am.* 81, Suppl. 1: S68.

Kent, R. (1983). The segmental organization of speech. *The production of speech.* P. MacNeilage, ed. New York: Springer.

Kent, R., P. Carney, & L. Severeid (1974). Velar movement and timing: Evaluation of a model for binary control. *J. Sp. Hear. Res.* 17: 470-488.

Kent, R. & F. Minifie (1977). Coarticulation in recent speech production models. *J. Phon.* 5: 115-133.

Kozhevnikov, V. & L. Chistovich (1965). *Speech: Articulation and perception.* Translated and distributed by Joint Publications Research Service.

Magen, H. (1984). Vowel-to-vowel coarticulation in English and Japanese. *J. Acoust. Soc. Am.* 75, Suppl. 1: S11.

Manuel, S. & R. Krakow (1984). Universal and language particular apsects of vowel-to-vowel coarticulation. *Haskins Laboratories SR* 77/78: 69-78.

McAdams, B. (1987). The relation between vowel variation and vowel-to-vowel coarticulation across languages. *J. Acoust. Soc. Am.* 81, Suppl. 1: S66.

Moll, K. & R. Daniloff (1971). Investigation of the timing of velar movements during speech. *J. Acoust. Soc. Am.* 50: 678-694.

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

Öhman, S. (1967). Numerical models of co-articulation. *J. Acoust. Soc. Am.* 41: 310-320.

Pierrehumbert, J. B. (1980). *The phonology and phonetics of English intonation.* MIT dissertation.

Pierrehumbert, J. B. & M. Beckman (1987). *Japanese Tone Structure.* Forthcoming in LI Monograph Series. MIT Press.

Purcell, E. (1979). Formant frequency patterns in Russian VCV utterances. *J. Acoust. Soc. Am.* 66: 1691-1702.

Sharf, D. & R. Ohde (1981). Physiological, acoustic, and perceptual aspects of coarticulation: Implications for the remediation of articulatory disorders. *Speech and language: Advances in basic research and practice* 5. N. Lass, ed. New York: Academic.

Steriade, D. (1987). Redundant values. To appear in *CLS* 23, Parasession on Metrical and Autosegmental Phonology.

Stevens, K., S. J. Keyser, & H. Kawasaki (1986). Toward a phonetic and phonological theory of redundant features. *Symposium on invariance and variability of speech processes.* J. Perkell & D. Klatt, eds. Hillsdale, NJ: Lawrence Erlbaum.

Quantitative Characterization of Degree of Coarticulation in CV Tokens*

Abigail C. Cohn

0. Introduction

Central to our understanding of phonetics is a characterization of the role of coarticulation. There is evidence that certain aspects of coarticulation must be language specific. It has been argued by researchers such as Ladefoged (1972) that languages may differ systematically in terms of their coarticulation. In spite of this, many researchers find it convenient to ignore this possibility and assume that there are not significant cross-linguistic differences with respect to degree or type of coarticulation. To my knowledge no work has been done to systematically quantify coarticulatory differences between languages. In this study, the development of a metric for degree of influence of a vowel on a preceding stop consonant has been undertaken in order to do just that.

The metric proposed in this study is a template fitting procedure using short-term spectra following the basic methodology of Blumstein and Stevens (1979). In this talk, results will be reported for Kana, an Ogoni language spoken in Nigeria, and Russian. Kana gives the auditory impression of having a high degree of coarticulation. These two languages have the same three basic places of articulation in stops: labial, dental, and velar; but differ in that Russian also has a phonemic secondary articulation distinction, palatalization. Palatalization involves the superposition of an /i/-like articulation onto a consonant. Thus in the three places of articulation, an additional contrast is made between palatalized and non-palatalized sounds in Russian. It is hypothesized that because of this phonological difference in which the tongue body is assigned a role in consonant production, Russian will demonstrate significantly less effect of a vowel on a preceding consonant than Kana.

---

*Paper presented at the Acoustical Society Meeting, November 1987, Miami Florida. This work was supported by NSF Grant No. BNS-8418580 to Patricia Keating.

1. Methodology

In this study, coarticulation is defined by how alike the onset spectra of a given consonant are in different vowel contexts. Our general procedure is to characterize a consonant in one vowel context, and then see how well that generalizes to the other vowel contexts. The methodology is schematized in Figure 1.

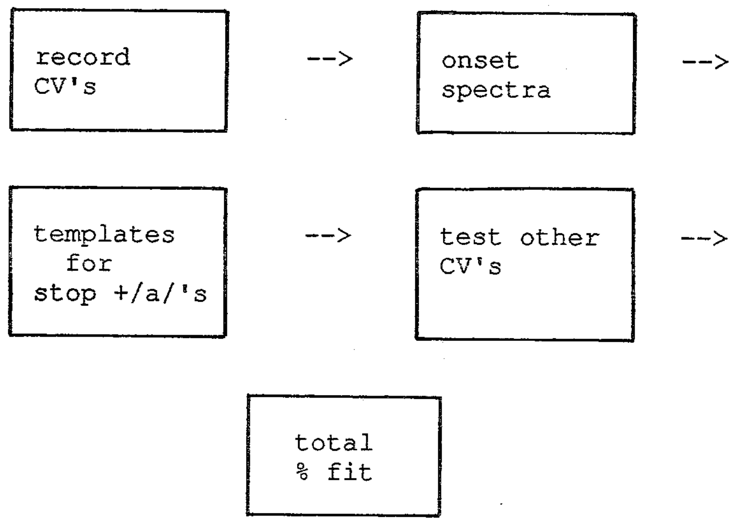


Figure 1. Schematization of Methodology

First, consonant vowel tokens were recorded; onset spectra were made; based on these, templates for stop + /a/ tokens were constructed; the templates were then tested against the other stop vowel tokens, in order to determine a total percent fit.

Several speakers of each language were recorded saying CV sequences of the shape stop-vowel for stops at each place of articulation combined with all vowels. This included labial, dental, and velar stops for both languages. The vowels and relevant stops for each language are listed in Figure 2. Not all such stop-vowel utterances were words, but in all cases they were phonotactically possible.

There were three females and four male speakers of Kana and two female and two male speakers of Russian. Two tokens were analyzed for each CV type except in the case of Russian stop + /a/ tokens, where four were analyzed in order to increase the number of tokens for templating. The total number of tokens analyzed in Kana was 294 and in Russian was 136. Short term LPC spectra of the burst were made, with a 26 ms window for all tokens, using the CSpeech speech analysis program on an IBM AT.

	KANA	RUSSIAN
Stops	b	p
	t d	t
	k g	k
Vowels	i u	i u
	e/ɛ o/ɔ	e o
	a	a

Figure 2. Vowels and Relevant Stops in Kana and Russian

For each language, templates were made of the spectra when the vowel was /a/ for each place of articulation, thus for both Kana and Russian labial, dental, and velar templates were constructed. The templates constructed by Blumstein and Stevens in their work on invariance for English were used as a starting point, but in all cases significant modifications were needed. Both inclusion and exclusion of tokens were used as criteria, with exclusion being the most important. This procedure was continued until adequate templates were constructed at each place of articulation for each language. Bar graphs with correct inclusion, marked with diagonal lines, and exclusion figures for each template in Kana and Russian are presented in Figure 3.

To take an example, for the Kana labial template (Figure 3, top left), the % correct inclusion for labial stop + /a/ tokens was 100%. The percent correct exclusion of dental stop-a tokens from the labial template was 90% and correct exclusion of velar stop + /a/ tokens was 100%. One outcome of this approach is that templates for the same place of articulation in different languages may differ substantially.

For Russian, only the non-palatalized stop-vowel sequences were templated to make the comparison with Kana as constrained as possible. Were the palatalized stops also included, the templates would have been much more general and one would have therefore expected greater percent fits for other vowels. To ensure that any percentage difference observed between Russian and Kana is indeed due to degree of coarticulation, only the non-palatalized consonants were used.

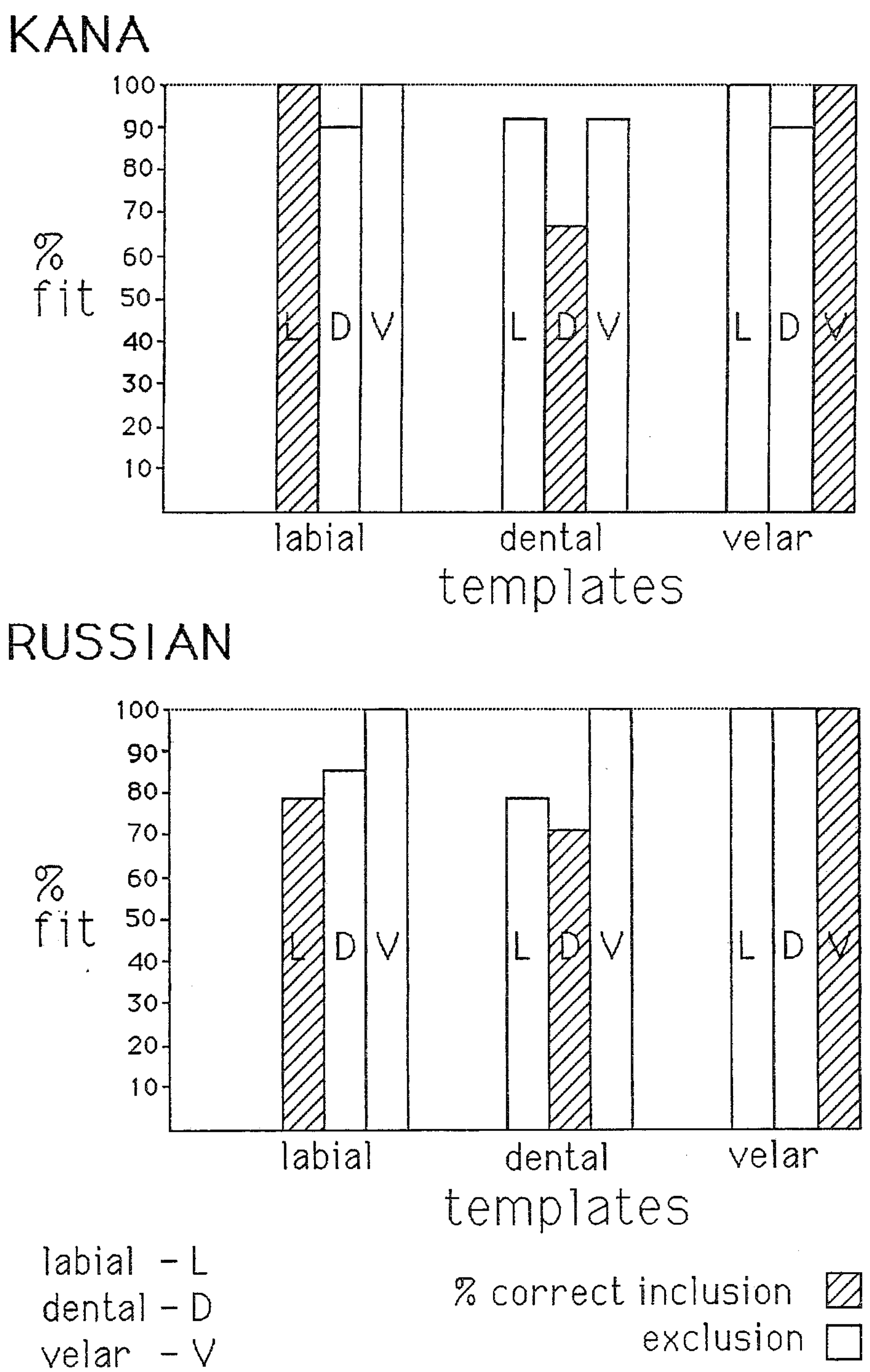


Figure 3. % Correct Inclusion and Exclusion for Templates

Once acceptable templates were constructed, the burst spectra of the other stop vowel sequences were tested against these templates to determine a percent fit. Each token was categorized as either fitting or not fitting. For a fit, all requirements of the template had to be met. In many cases, a token almost fit; these cases were counted as not fitting. Let us consider an example of template fitting, shown in Figure 4.

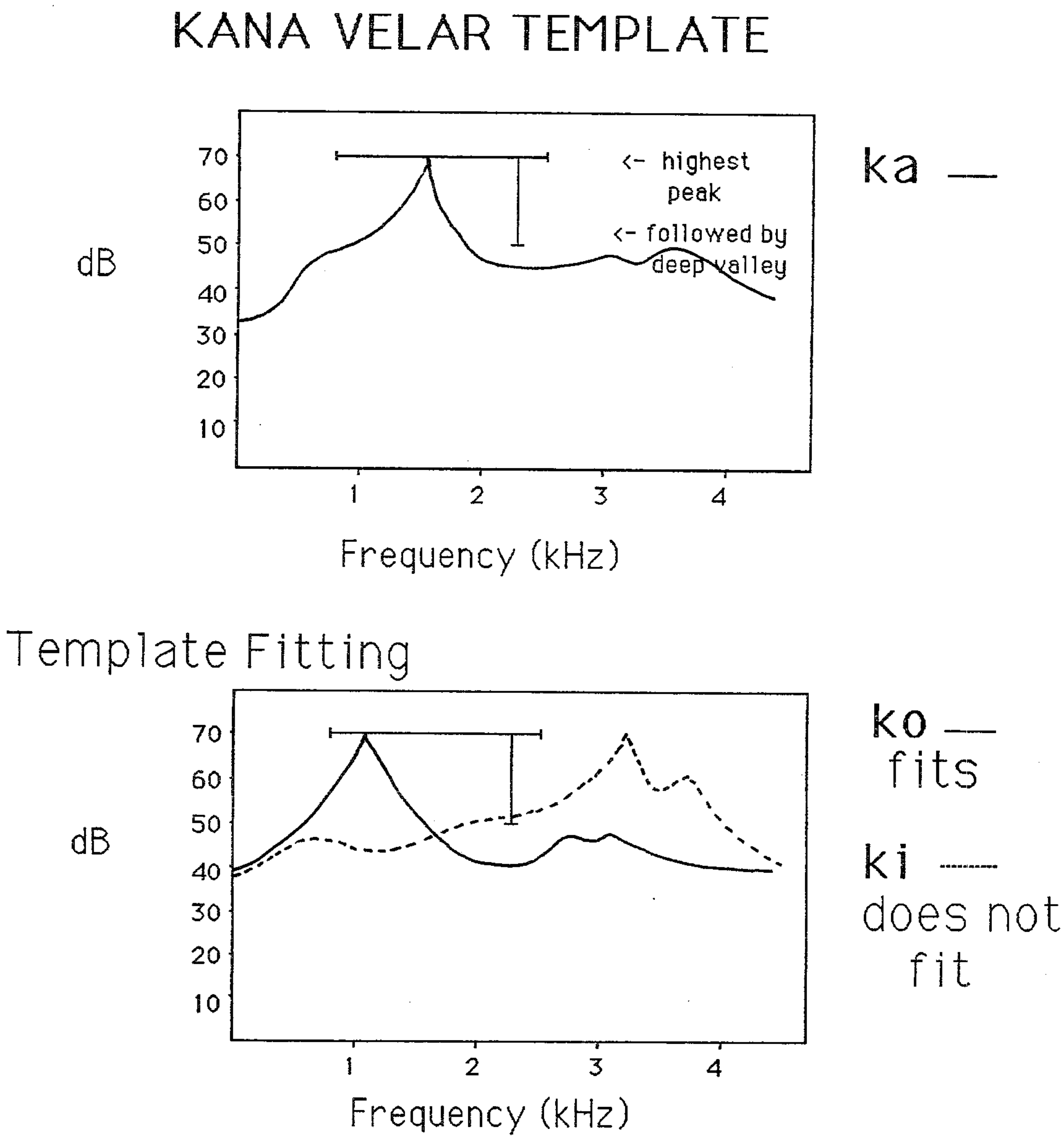


Figure 4. An example of template fitting

In the top half of Figure 4, the velar template for Kana is shown. The basic requirements are as follows: The highest peak must be in the range between 800-2500 Hz. A deep valley of at least 20 dB must follow the peak. An example *ka* token, used for developing the template, is shown. In the lower half of the figure, two tokens are fitted against the template: *ko*, the solid line, fits; *ki*, the dashed line, is rejected.

A total percent fit was calculated for each language; also percent fits were calculated for each stop-vowel type, for each place of articulation, and for each vowel. The total percent fit allowed a gross comparison of effect of a vowel on a preceding stop between languages, whereas the other figures allowed consideration of the interaction of place of articulation and vowel quality.

It should be noted explicitly that although the methodology in this study is similar to that used by Blumstein and Stevens in their work on invariance, the goals of this research are orthogonal to questions of invariance in language. Blumstein and Stevens noted vowel effects in their work, although their goal was to look beyond these effects. In contrast, our goal is to focus on precisely such effects. Additionally, no theoretical claims are intended by the shape of our templates. Also, before turning to specific results, it must be noted that these results are preliminary because only a limited number of tokens were templated and checked and the templates were constructed in an ad hoc fashion. Other acceptable templates might have led to slightly different results.

2. Results

First, we turn to the results for Kana. Kana gives the auditory impression of having a high degree of coarticulation, thus we would expect to see a low percent fit. This is indeed the case. Total percent fit is 43%, that is 102 out of 236 non-a tokens fit the respective templates. In the graph in Figure 5, results are presented by type, with % fit on the vertical axis and place of articulation along the horizontal axis.

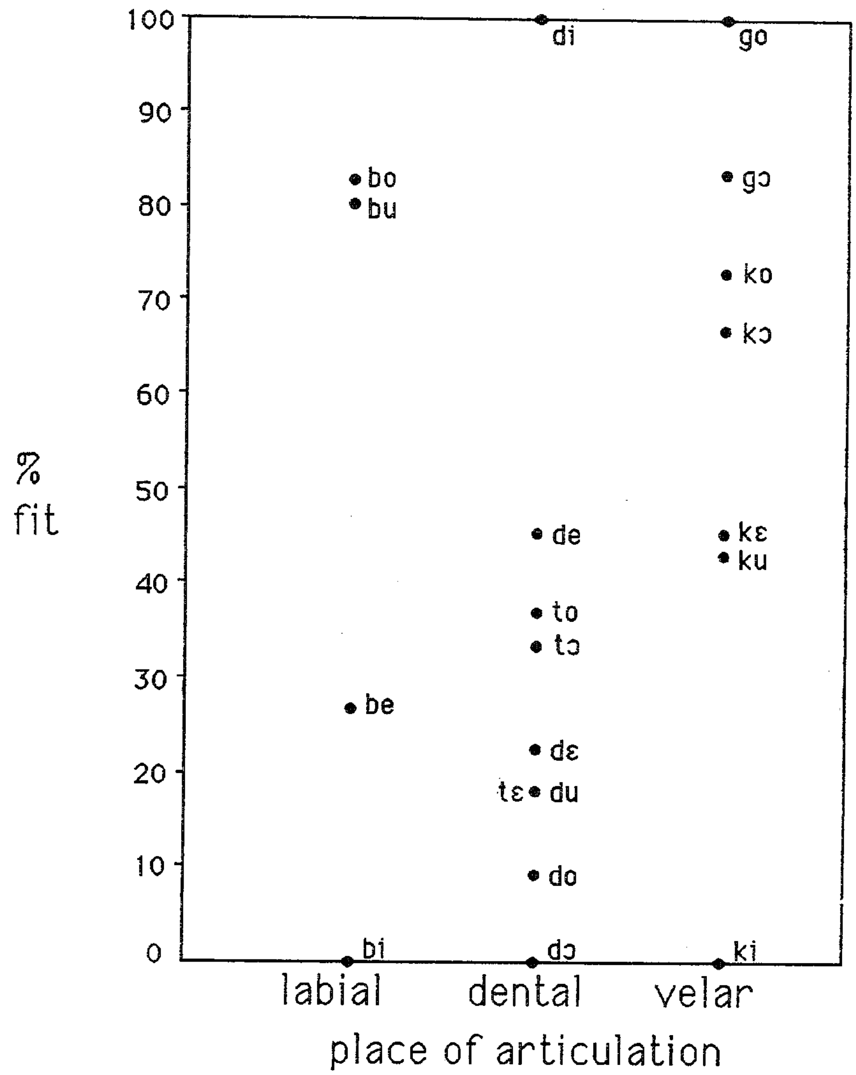


Figure 5. Kana % fit by type

The range of percent fits by type is extreme: 0 to 100%. The relationship of fit with respect to place of articulation and vowel quality is a systematic one. Labials show a good fit for back vowels but poor for front vowels. Dentals show higher percent fits for high front vowels, and worse fits for back as well as non high vowels. Velars show better fits for non-high back vowels and poorer fits for high vowels. At least for Kana, the fit appears to be related to the accentuation of the burst by characteristic formant values.

Turning now to Russian, the total percent fit for Russian is 73%. As hypothesized, percentage fit is much greater for Russian, demonstrating significantly less effect of a vowel on a preceding consonant than in Kana. The results are also presented by type, in the graph in Figure 6.

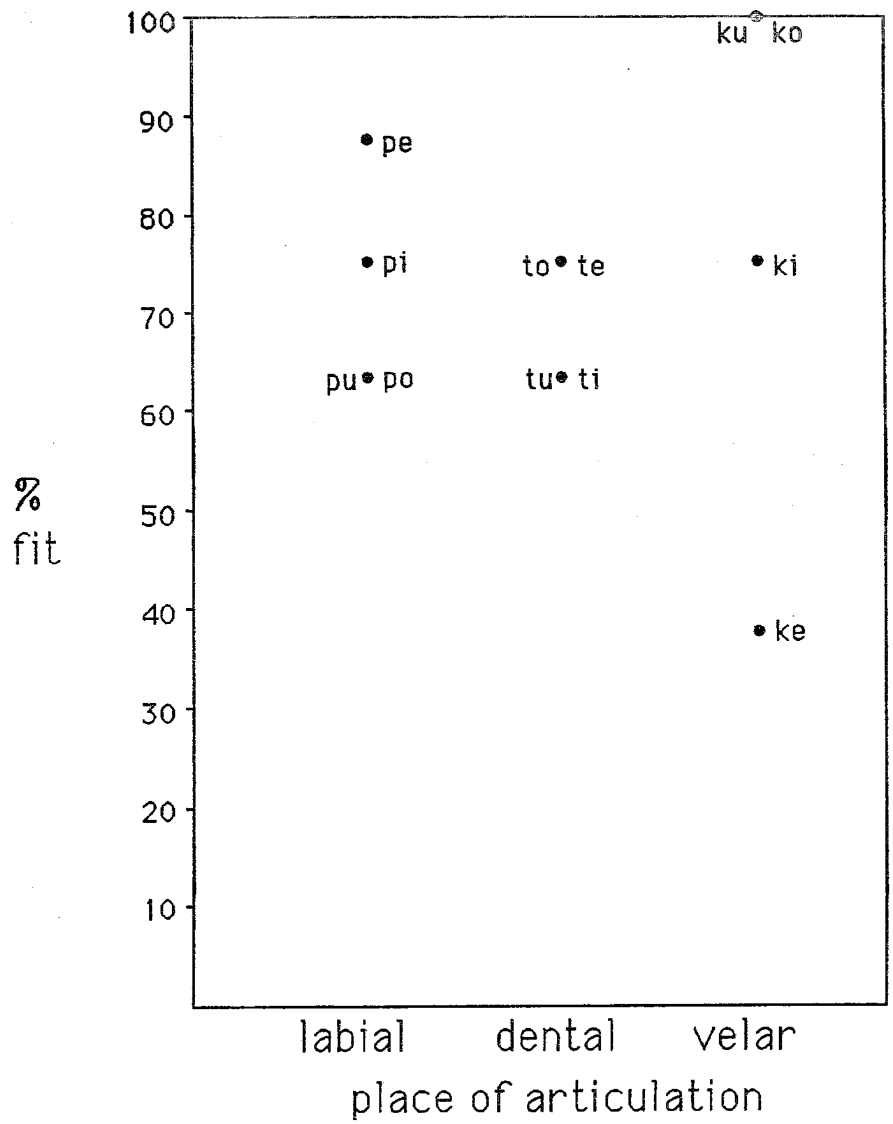


Figure 6. Russian % fit by type

The range of percent fits for CV types is much smaller; with the exception of *ke* at 38%, all types fall between 63 - 100%. Unlike in Kana, the relative differences of percent fit for particular types are not systematic.

3. Discussion

Based on two languages, Kana and Russian, we observe that languages do differ in the degree of influence of a vowel on a preceding stop. This difference has been quantified using a metric of template fitting based on onset spectra, with a low percent fit indicating that vowels have a great effect on preceding consonants and a high percent fit indicating that they do not. The hypothesis that significant and systematic differences exist between languages with respect to degree of coarticulation is supported. The relevance of such a metric for quantifying degree of coarticulation is clear. Once aspects of coarticulation are quantifiable, this information will be easily incorporated into research in speech synthesis as well as general phonetic research.

No claim is made that this is the only way to construct such a measure; rather, this is an accessible and systematic way to do so. Analyzing sprectra with respect to change over time could be considered as well, as proposed in recent work by Lahiri et al.(1984). At this time, additional languages are being considered to see how they come out with respect to the metric. It is argued that Kana and Russian differ as greatly as they do, due to the articulatory constraints imposed on Russian because of palatalization. An interesting question for further research is what ways the phonology of a language directly affects the degree and type of coarticulation observed.

References

Blumstein, Sheila E. and Kenneth N. Stevens (1979). "Acoustical invariance in speech production: evidence from measurements of the spectral characteristics of stop consonants". *JASA* 66(4):1001-17.

Ladefoged, Peter (1972). "Phonetic prerequisites for a distinctive feature theory". in *Papers in Linguistics and Phonetics in Memory of Pierre Delattre*, A. Waldman ed. pp. 273-286 The Hague: Mouton.

Lahiri, Aditi, Letitia Gewirth, and Sheila E. Blumstein (1984). "A reconsideration of acoustic invariance for place of articulation in diffuse stop consonants: evidence from a cross-language study". JASA 76(2):391-414.

Phonetic Rules of Nasalization in French*

Abigail C. Cohn

0. Introduction

Most languages use the feature [nasal] contrastively in consonants, that is they have phonemic contrasts between sounds such as /d/ and /n/ (e.g. English). Some languages use the feature [nasal] contrastively in vowels as well. Such is the case in French, which has a surface contrast between oral and nasal vowels. In English, it has often been observed that there is extensive nasalization of vowels adjacent to a nasal consonant. Such contextual effects are assumed to be tolerated precisely because [nasal] is not contrastive for vowels. In other words, extensive coarticulation is allowed. The converse is usually assumed as well, that is, that in a language where [nasal] is contrastive in both consonants and vowels, contextual effects should be very restricted, limited only to transitions at the edge of segments.

Based on these assumptions, we might expect the phonetic output, outlined in Figure 1, for a vowel-nasal-vowel sequence in two languages such as English and French. (V = oral vowel, $\tilde{V}$ = (contrastively) nasal vowel, C = oral consonant, N = nasal consonant.)

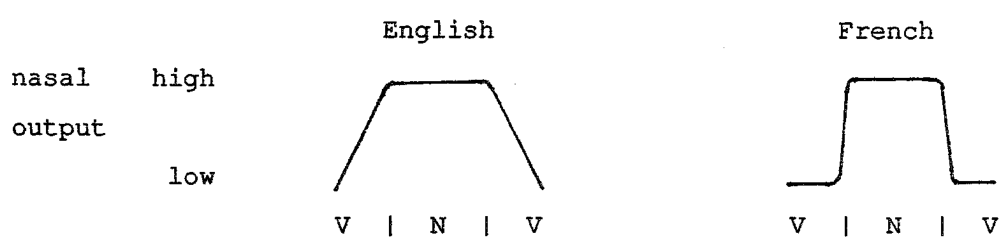


Figure 1. Hypothetical phonetic output

*Paper presented at the Linguistic Society of America Meeting, December 1987, San Francisco

In the first case, where there is no possible contrast, there is no limit on coarticulation and some level of nasalization might occur throughout the vowel, whereas, in the second case, where there is a possible contrast, only limited effects of hooking up the segment would be expected. Rather surprisingly, the latter is not at all the observed pattern for French. I will show that nasalization (indicated in this study by amount of air flowing through the nose, that is, nasal airflow) occurs to varying degrees on phonologically oral vowels adjacent to nasal consonants. Such effects in French have been observed in previous studies, (e.g. Zerling 1984, and Benguerel et al. 1977); much of the previous work is nicely summarized by Van Reenan (1982). But the results in previous studies were limited to a few tokens without adequate consideration of the effect of change over time.

This study undertakes a systematic characterization of contextual and contrastive nasalization of vowels in French. These two sources of nasalization will be characterized in terms of quantity (amplitude of flow) and timing (of onsets and offsets). The results have important implications for the types of rules that must exist at the phonetic level of representation. In the discussion that follows, I assume that the output of the phonology for the feature [nasal] is in terms of +'s, -'s, and 0's. We need to consider how to get from this discrete segment by segment representation to the continuous phonetic output.

1. Methodology

In order to investigate the relationship between the phonological and phonetic output, the phonetic output for the feature [nasal], that is amount of nasalization, was measured using nasal airflow. Two male speakers of French (both from Paris) were recorded saying six repetitions of each utterance. Filtered nasal flow and the audio signal were printed out on paper. The data presented here are filtered nasal flow traces.

All test items were carefully chosen to create minimal sets and to restrict as much as possible effects of such factors as vowel height, place and manner of articulation in consonants, stress, and so forth. With one exception (described below), only monosyllabic forms were used. All items are existing words in French. All tokens were said in the frame *dite* _____ *deux fois* 'say ____ two times.' /dit ___dø fwa/.

Nasal vowels were compared first with nasal consonants and then with oral vowels both preceding and following a nasal consonant. Five comparisons were made. These are schematized in Figure 2 and will be detailed in the discussion of the results.

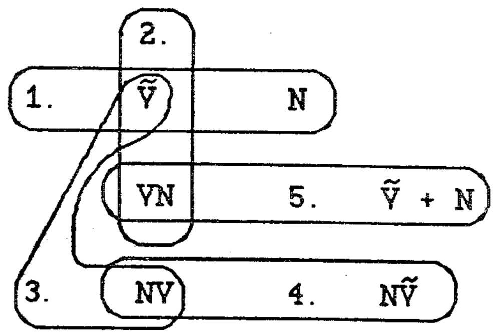


Figure 2. Schematization of comparison

2. Results

The results presented here are for the first speaker; preliminary analysis of the second speaker supports these findings. The results were very consistent across tokens. In discussing the results, I will present individual tokens, deemed representative. To show how consistent the results are, I first present five tokens of *bon* 'good' /bɔ̃/, in Figure 3.

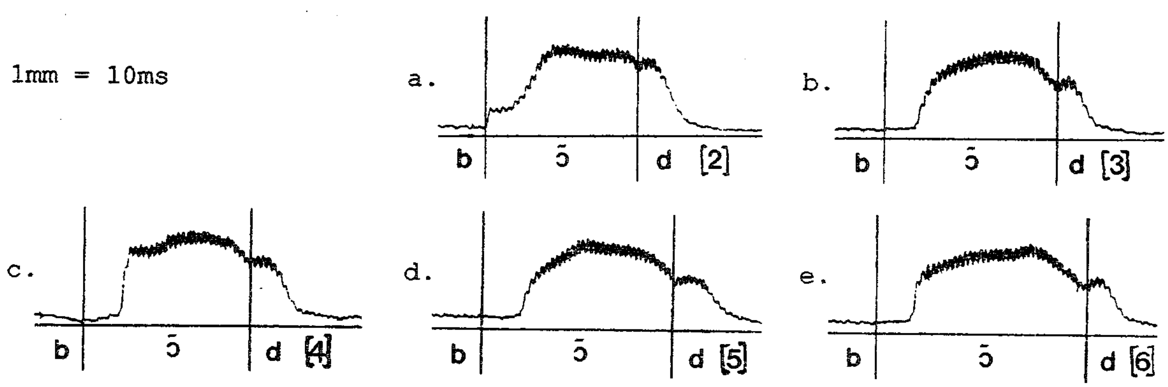


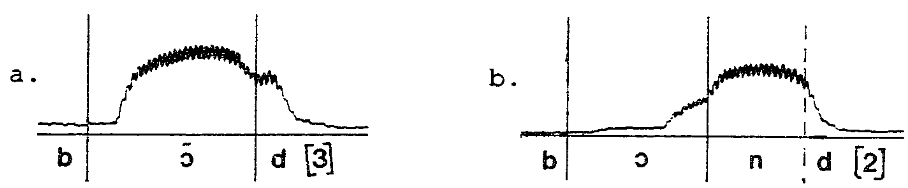
Figure 3. Five repetitions of *bon* 'good, m.' /bɔ̃/

In the tracings, the vertical lines indicate segment boundaries as interpreted from the acoustic wave signal, measured from the paper printout and spectrograms made of the audio signal. Vowel onset was measured from the stop release. For the most part, segmentation was straightforward; ambiguous segmentation is marked with a dashed line. In the traces in Figure 3, we see that the timing of the nasal airflow, starting almost at the baseline during the consonant /b/, increases rapidly after the onset of the vowel. The flow quickly reaches a maximum flow which is maintained till near the end of the vowel. Tokens b., d., and e. are extremely similar, with a. and c. differing somewhat, but these differences are negligible in terms of the kinds of differences we will be observing.

Turning now to the comparisons: First, as shown in Figure 4, how does a nasal vowel compare with a nasal consonant, exemplified by the nasal vowel /ɔ̃/ in *bon* 'good, m.' in 4.a. and the nasal consonant, /n/ in *bonne* 'good, f.' in 4.b.

bon 'good'
/bɔ̃/

bonne 'good'
/bɔn/



/bɔ̃/

/bɔn/

Figure 4. Ṽ vs. N

In terms of amplitude over time, that is the shape of the waveform, nasal vowels are similar to nasal consonants. Both cases reach a high level of flow early in the segment and maintain it throughout. In these cases, the decrease in flow occurs during the phonemically oral stop /d/ of the frame sentence. One difference is the timing of the onset of flow. The onset for the nasal consonant is during the last part (approximately one third) of the preceding vowel. For the nasal vowel, it is after the onset of the vowel and about twice as rapid.

In the next two cases, we compare a nasal vowel to contextual nasalization. First, we look at the case of a nasal vowel and anticipatory nasalization, that is nasalization of a phonemically oral vowel preceding a nasal consonant: $\tilde{V}$ vs. VN. Here we return to the forms presented in Figure 4, looking again at the nasal vowel in 4.a. and now the phonemically oral vowel in 4.b.

We see nasalization on the vowel preceding a nasal consonant, which differs in certain ways from the nasalization of the nasal vowel. There is no significant flow till about two thirds of the way through the vowel. At which time there is a turning point and a distinct increase in nasal flow. By the onset of the nasal, the flow is about half of what it is during the center of the nasal consonant--less than half of that observed for the nasal vowel. This looks transitional in nature and might be viewed as a local transition between the oral vowel and the nasal consonant, although it should be noted that the onset is earlier and slower (60ms) than that observed for the onset of the nasal vowel (about 30ms).

In Figure 5, we contrast a nasal vowel with carryover nasalization, that is, with nasalization of a phonemically oral vowel following a nasal consonant, exemplified by the nasal vowel /ɛ̃/ in *daim* 'deer' in 5.a, and the oral vowel /e/ in *nez* 'nose' in 5.b.

daim 'deer' /dɛ̃/ *nez* 'nose' /ne/

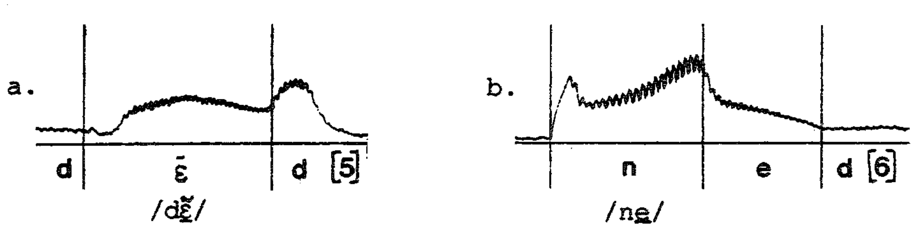


Figure 5. $\tilde{V}$ vs. NV

Here we see a different vowel. The amplitude of flow is lower for /ɛ̃/ than for /ɔ̃/, but the overall shape and timing is comparable. In 5.b., the carryover case shows quite marked nasal flow throughout the duration of the vowel. It has not reached the baseline by the end of the segment. The most obvious differences between this and the nasal vowel are the marked decline at the beginning of the vowel (in contrast to a rapid increase for the nasal vowel) and its general sloping nature. The overall similarity between contrastive and contextual nasalization is indeed surprising, but will be seen to be even more so when compared with a nasal vowel preceded by a nasal consonant, the next case.

In French, a contrast between a nasal vowel and an oral vowel is maintained after a nasal consonant. This contrast is exemplified in Figure 6, with the nasal vowel /ɛ̃/ in *nain* 'dwarf' in 6.a. and the oral vowel /e/ in *nez* 'nose' in 6.b. (the same as 5.b.).

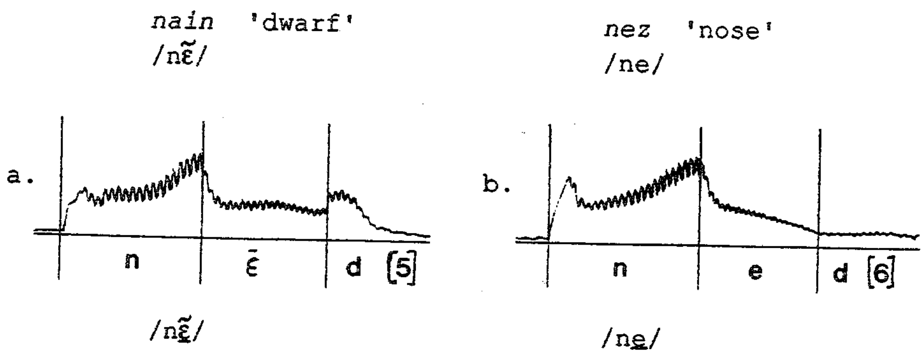


Figure 6. NṼ vs. NV

In this case, the forms are very similar. Both show a comparable drop in flow after the nasal. The nasal vowel and oral vowel are differentiated only by the flat character of flow on the nasal vowel and the slightly sloping character of that on the oral vowel. They also differ as far as the degree of flow on the following stop, the /d/ of the frame sentence. Perceptually this may be as salient (or more salient) a cue to the distinction. This case, in particular, points to the inadequacy of looking at either timing or amplitude alone.

Turning now to the final case: Within words in French, there is no contrast between a nasal vowel and an oral one preceding a nasal consonant (ṼN sequences do not occur word internally), but there is a contrast across word boundaries. In this one case, we consider two syllable utterances. This is exemplified by the nasal vowel /ɔ̃/ in *bon nez* 'good nose' in 7.a. and the oral vowel /ɔ/ in *bonnet* 'cap' in 7.b.

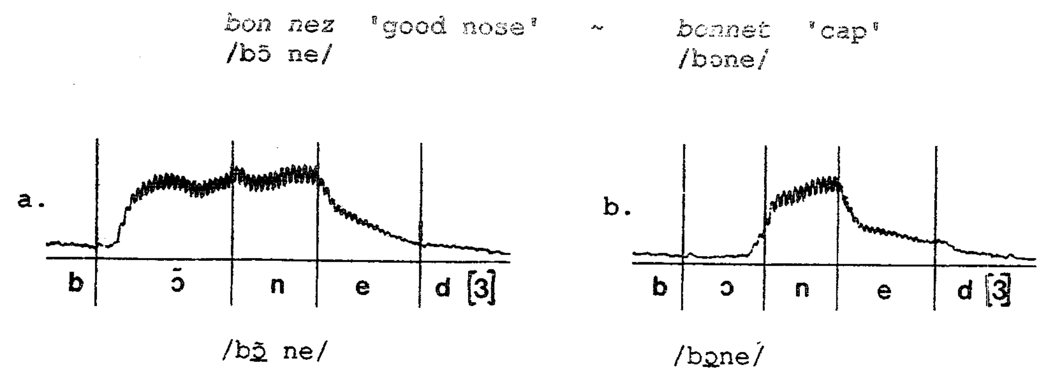


Figure 7. $\tilde{V}$ + N vs. VN

Here we see that the vowels look very much as they did in Figure 4, that is the character of nasalization of a nasal vowel is not affected by the following nasal, except toward the end of the vowel (where a higher level of flow is maintained). The oral vowel is quite similar to the oral vowel in 4.b., although the onset of flow is somewhat later (more transitional). So, in this case, where there is no contrast within the word, the phonetic output remains very distinct.

3. Discussion

The most striking of the above observations is that these contextual effects are not just transitional ones. The anticipatory effects could be characterized in this way, although the onset is not as fast as it could be (mechanically); but the carryover effects certainly cannot be characterized as merely transitional. The offset is sloping, but there are no sharp turning points and there is not much difference in amplitude from the case of a nasal vowel following a nasal consonant. Thus the contrastive use of the feature [nasal] on vowels does not in and of itself restrict the range of coarticulation. These low level effects are distinct from the feature specifications in the phonology and they have the characteristic phonetic quality of being gradient (as described by Keating to appear).

Since the phonological representation alone is not sufficient to determine the phonetic output, it must be concluded that such effects are accounted for in the phonetics. The nature of these effects suggests that they are due to language specific phonetic rules. The next step is to pursue the formal properties of the phonetic rules to best account for such data. I believe that such effects can be effectively characterized as a series of targets hooked up by interpolations rules, following the model developed for intonation by Pierrehumbert (1980). A study of these rules as well as an analysis of more long distance effects of nasalization in French remain as topics for further research.

References

Benguerel, A.-P., Hirose, H., Sawashima, M., and T. Ushijima. (1977) "Velar coarticulation in French: a fiberscopic study," *Journal of Phonetics* 5:149-158.

Keating, Patricia. (to appear) "The window model of coarticulation: articulatory evidence," in *Papers in Laboratory Phonology 1*. J. Kingston and M. Beckman, eds., Cambridge: Cambridge University Press.

Pierrehumbert, Janet. (1980) *The Phonology and Phonetics of English Intonation*. MIT Ph.D. dissertation.

Van Reenan, Pieter. (1982) *Phonetic Feature Definitions, their integration into phonology and their relation to speech, a case study of the feature nasal*. Dordrecht, Holland: Foris Publications.

Zerling, J.-P. (1984) "Phénomènes de nasalité et de nasalisation vocaliques: étude cinéradiographique pour deux locuteurs," *Travaux de l'Institut de Phonétique de Strasbourg* No. 16:241-266.

Timing of contextual nasalization in two languages*

Marie K. Huffman

1. Introduction

The increasing evidence of language-specific coarticulatory effects (Ohman (1966), Lubker and Gay (1982), Magen (1984), Manuel and Krakow (1984)) has highlighted our need to identify the range of coarticulatory phenomena, and explore possible constraints on coarticulation. My work is concerned with the way in which the phonological status of the feature [nasal] in various languages affects timing patterns in contextual nasalization. It has often been assumed that if a language has a phonemic contrast for a feature, coarticulatory effects for that feature will be minimal. Conversely, it is assumed that lack of contrast will allow extensive coarticulation. For ease of reference, I will call this the contrast hypothesis. This paper investigates the extent to which the contrast hypothesis holds for nasalization in two languages, one which has an oral/nasal contrast on vowels, and one which does not. I will show that the role of contrast in timing of coarticulatory nasalization is somewhat less direct than is usually assumed, and can only be understood by looking at anticipatory and carryover effects together; that, by looking at the distribution of nasalization in a broader temporal context.

2. Language Background

The languages we are considering are both tone languages of West Africa. The first, Akan, is spoken in Ghana. Akan has an oral/nasal contrast in vowels. Interestingly, unlike in many languages, in Akan it is possible to get this contrast even in the context of a nasal consonant. So for example, as seen in Figure 1, the forms for "he doesn't give" and "he doesn't come" are a surface minimal pair for nasalization on V_2. Thus Akan offers us an opportunity to observe contextual nasalization in exactly the kind of context where the contrast hypothesis predicts that it will be constrained by the presence of a contrast. These forms are the result of consonantal assimilation in nasalization when a negative prefix, which is a homorganic nasal, is added

* This paper is a slightly modified version of a paper presented at the 114th meeting of the Acoustical Society of America, Miami, Florida.

Akan

òmmá "he doesn't come"   òmmã́ "he doesn't give"

òbá "he comes"   òmã́ "he gives"

Wote _____ daa

Agwagwune

ámá "mouth"

Dor _____ girige

Figure 1.

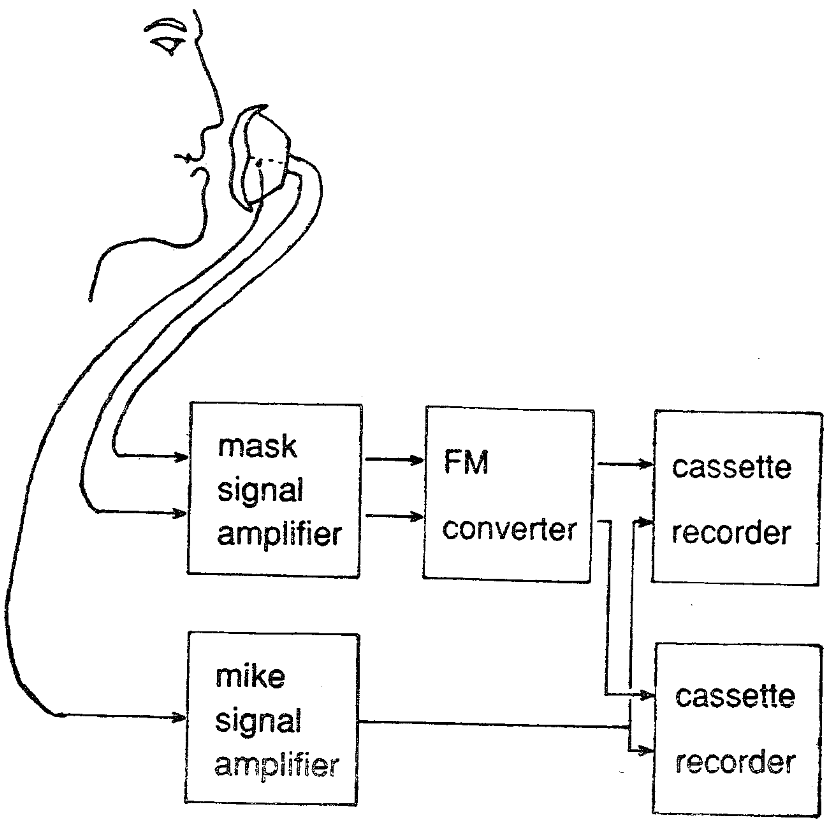


Figure 2. Experimental setup.

to the verb forms shown further down in the figure. I recorded speakers saying these verbs in the frame sentence shown, where five repetitions each appeared, pseudorandomized, in a list along with comparable, completely oral items.

For comparison with the Akan case, I will discuss data from one speaker of Agwagwune, a little-studied language of Nigeria which has no oral/nasal contrast in vowels. In general terms of oral/nasal contrasts, Agwagwune is like English and might be expected to show extensive contextual nasalization, as is reported for English. The Agwagwune word and frame sentence are shown at the bottom of Figure 1. Again, five repetitions of this item were recorded, along with completely oral control words.

3. Recording

Because no completely satisfactory metric has been developed for measuring nasalization, presently methodology is determined in large part by one's goals. Although the velotrace has been demonstrated to be a very useful tool in tracking velum movement, as illustrated on Tuesday by Rena Krakow, my interest in gathering data on a number of different languages has led me to choose a less invasive technique, which subjects would be more comfortable with; namely, airflow measurements. I am using a ratio of nasal to oral flow as an indication of nasalization. The experimental setup is illustrated in Figure 2. The subject speaks into a Rothenburg split mask. A simultaneous audio recording, essential for segmentation, is made with a small microphone attached on the outside of the mask. The oral and nasal flow signals are rectified and smoothed by taking an RMS envelope; then they are frequency modulated. The flow signals and the audio signal are then recorded on cassette tapes.

4. Analysis

The data analysis was done on IBM pcs using the CSPEECH system and supplemental programs written by Lloyd Rice. The flow recordings were demodulated and then digitized, with the audio signal on a second channel. The audio signal was used to segment and line up the nasal and oral flow signals. Since we are taking nasal flow as indicative of nasalization, we must control for spurious changes in nasal flow that may be due to variations in overall flow. This was done by computing the ratio of nasal to oral flow. When the

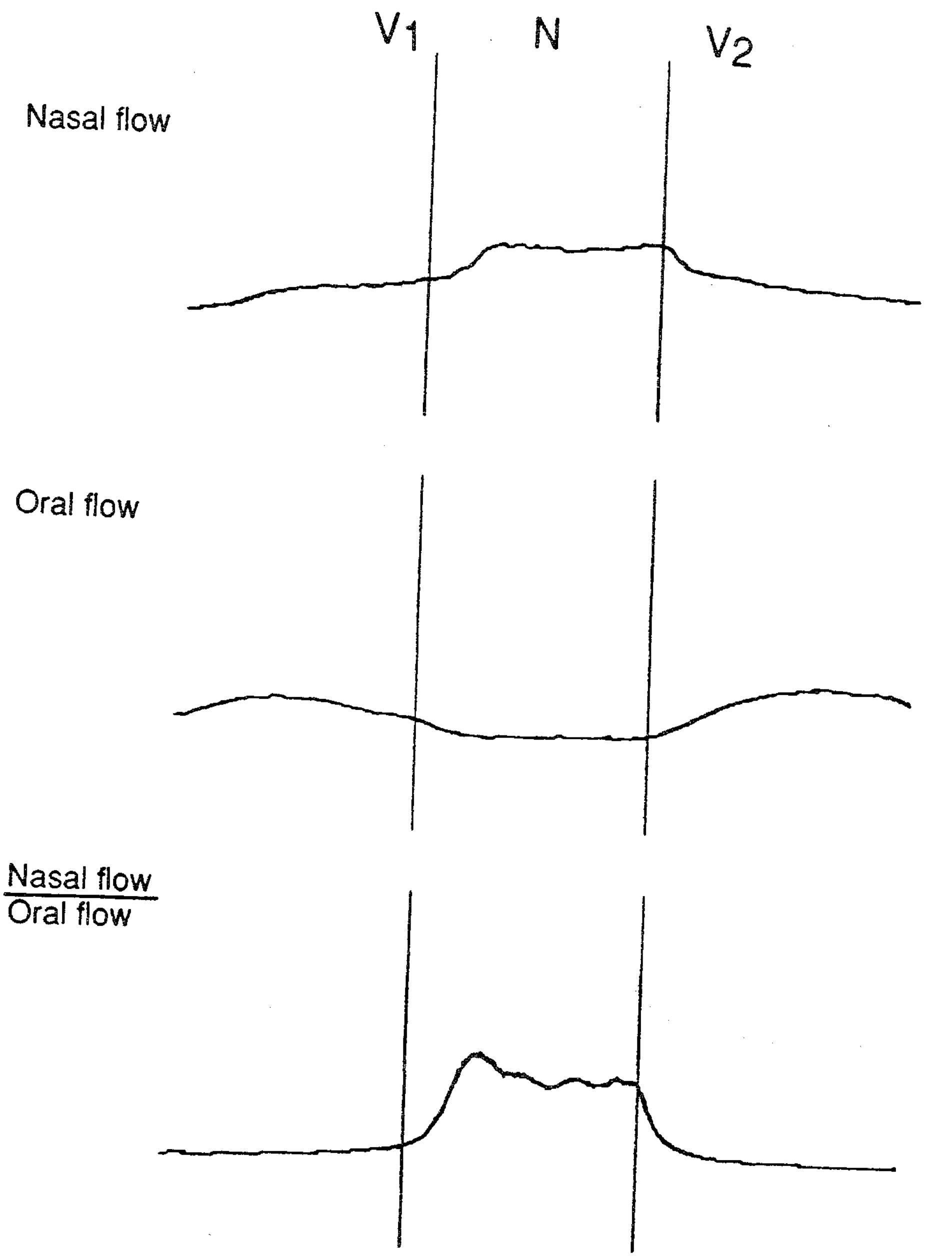


Figure 3. Example flow wave-forms. Vertical lines indicate segment boundaries.

oral flow was zero, this division was not possible without a mathematically indeterminate result, so nasal flow was divided by 1; that is, it was left unchanged. Finally, the output is multiplied by a gain factor, for visual convenience. This procedure is illustrated in Figure 3. From the resulting ratioed flow waveforms, a time-averaged waveform was computed for the five repetitions of each experimental item.

5. Results

Turning to the results for Akan (Figure 4), let us look at contextual nasalization where it contrasts with phonemic nasalization. In these and subsequent figures, the dashed line indicates a baseline determined from system levels in the absence of speech or breath. Arrows show where flow reaches this zeroline. The vertical lines mark the average durations of the vocalic and consonantal portions. Looking first at the final vowels, let's see whether contextual nasalization, case 1, is really minimized in a context where a contrast does occur, as with the nasal vowel of case 2. In case 1, at the onset of V_2 flow drops off fairly sharply, reaching zero after 58 ms, or at about 60% of the vowel duration. For case 2, at the onset of V_2 flow decreases less sharply, reaching zero later in the vowel, at 78 ms, or about 85% of the vowel duration. So while there is certainly less extensive nasalization on V_2 in case 1, the contextual nasalization case, the nasalization could not be called minimal. Of additional interest is the fact that nasalization on the phonemically nasal vowel, case 2, does not extend throughout the duration of the vowel, despite the fact that this might help maintain the oral/nasal contrast in this context.

To fill in the rest of the picture, let's consider contextual nasalization on V_1. Here we have two instances of contextual nasalization on an oral vowel. An oral/nasal contrast is not possible in this position in Akan, so contextual nasalization need not be limited, under the contrast hypothesis. In fact, what we find is moderate amounts of nasalization, of an order comparable to the carryover cases. However, since these vowels are shorter than the final vowels, a similar duration of nasalization occupies a larger proportion of their total durations, and so in this sense nasalization is more extensive.

Turning briefly to Agwagwune, Figure 5 shows the time-averaged waveform for five repetitions of /ama/. Looking first at nasalization on V_2, notice

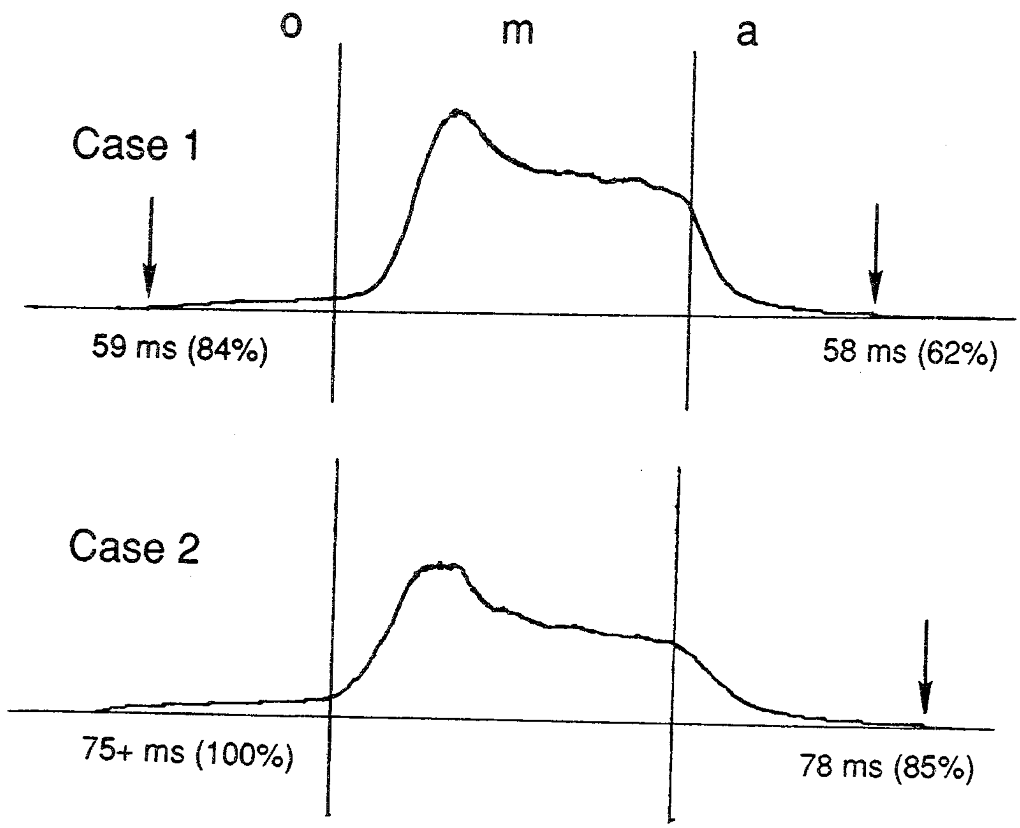


Figure 4. Time-averaged flow waveforms of five repetitions each for the Akan words "omma" (case 1) and "ommã" (case 2).

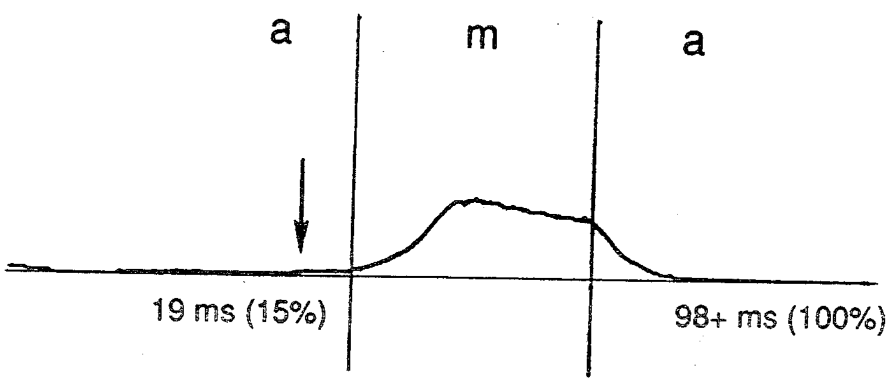


Figure 5. Time-averaged flow waveform for five tokens of Agwagwune "ama".

that flow lasts throughout the the vowels's duration. Note also that the absolute and relative amounts of nasalization on V_2 are greater than what was seen in Akan, even on contrastively nasal vowels. So if we look just at V_2, then Agwagwune behaves as expected according to the contrast hypothesis, in being a language with no contrast showing extensive nasal coarticulation. But now look at nasalization of V_1. This anticipatory nasalization is minimal. In fact, flow begins only just before closure for the nasal consonant. While this is data for only one speaker, it illustrates that in a language without a contrast, not all contexts will necessarily show extensive coarticulatory effects.

Consider what we have seen so far. One language, Akan, has moderate carryover and anticipatory contextual nasalization. The other language, Agwagwune, has extensive carryover coarticulation and minimal anticipatory coarticulation. These differences do not fall out from the differences in contrastive use of the feature [nasal] in the two languages. That is, the contrast hypothesis does not predict these patterns. However, if we look at anticipatory and carryover effects together, we can see that there is a pattern to the timing behavior described earlier. In a gross sense, there is a difference between the languages in how nasalization is distributed in time over similar kinds of segments. As sketched in Figure 6, nasalization is centered earlier in Akan than in Agwagwune. This results in more anticipatory nasalization and slightly less carryover nasalization in Akan than in Agwagwune. This difference in distribution of nasalization correlates with a difference between the languages in the location of peak nasal flow (indicated by X's). Looking again at Akan (Figure 4), peak nasal flow occurs about 1/3 of the way into the consonantal closure, while in Agwagwune (Figure 5), we see that peak nasal flow occurs almost directly in the middle of the consonantal closure. Thus location of peak nasal flow affects the gross patterns of contextual nasalization in the two languages.

However, choice of location of peak nasalization is not an arbitrary difference between the two langugages. I suggest that nasalization is centered earlier in Akan precisely because there can be an oral/nasal contrast on the vowel following the consonantal closure. This makes it possible to reduce carryover contextual nasalization (compared to, say, the Agwagwune case) and get a less nasal vowel when this is needed, as in case 1 for Akan. Of course if a nasal vowel is intended in this position, then the duration of nasalization can be extended, as for case 2 in Akan.

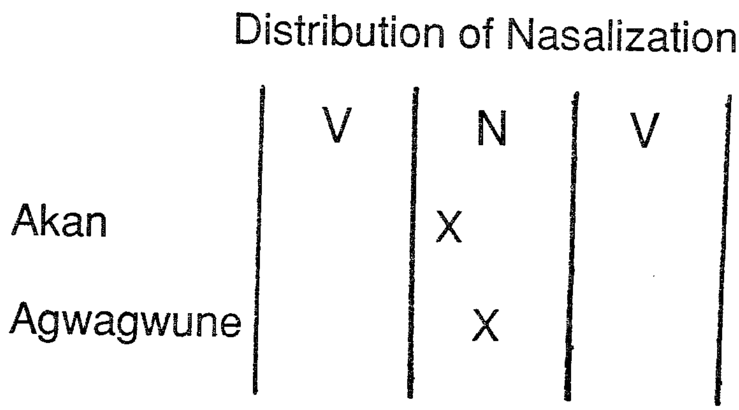


Figure 6. Comparison of how nasalization is distributed over time in Akan and Agwagwune. X indicates peak nasal flow.

6. Conclusions

In summary, I am claiming that the role of contrast in constraining contextual nasalization is less global and somewhat less direct than is often assumed. Languages will differ in where peak nasalization occurs on a nasal consonant. This choice is determined largely by whether a contrast in vowel nasalization occurs in the language, and if so, where it occurs in relation to the nasal consonant. This choice will affect gross patterns of contextual nasalization. It should be emphasized that additional factors may come into play in determining where peak nasalization will occur. For example, principles of articulatory coordination might limit the set of possible locations of peak nasalization. Furthermore, to interpret the aerodynamic data with confidence, we need to better understand the relation of velum movement to peak nasal flow. However, assuming that this relation is comparable for similar segments across languages, additional data like that presented today will increase our understanding of the principles governing timing of nasalization.

Acknowledgements

This work was supported by NSF grant BNS-8418580 to Patricia Keating, and by a fellowship from the UCLA Department of Linguistics.

References

Lubker, James F., and Thomas Gay. (1982). "Anticipatory labial coarticulation: Experimental, biological and linguistic variables". *J. Acoust. Soc. Am.* 71: 437-448.

Magen, Harriet (1984). "Vowel-to-vowel coarticulation in English and Japanese". *J. Acoust. Soc. Am.* 75: S41.

Manual, Sharon Y., and Rena A. Krakow (1984). "Universal and language particular aspects of vowel-to-vowel coarticulation". *Haskins Labs. Stat. Rep. Speech Res.*, SR77/78: 69-78.

Ohman, Sven E.G. (1966). "Coarticulation in VCV utterances: Spectrographic measurements". *J. Acoust. Soc. Am.* 39: 151-168.

Palatals as complex segments: X-ray evidence*

Patricia A. Keating

What is the palatal place of articulation? Figure 1a shows a sample of the kind of phonetic evidence available to help answer such a question: for a palatal nasal stop [ɲ] in Czech, we have a tracing from a mid-sagittal X-ray, a tracing from a photo of the lips, and palatography. What phonetic feature values best describe this articulation? Several contradictory and sometimes confusing feature descriptions have been offered in the literature, though without much discussion. In broad terms, there have been four kinds of descriptions of palatals. First is the traditional description of places of articulation (e.g. Ladefoged 1971), in which palatal is one of many places of articulation located along the palate, between post-alveolar and velar points of articulation. Second is the Jakobson, Fant, & Halle (1963) description, championed more recently by Lahiri & Blumstein (1984), which is based primarily on acoustics but also has articulatory correlates. Here palatals, together with palatoalveolars like English [ʃ], comprise one of four major areas of articulation, defined as acoustically acute and compact. Third is the description of palatals offered by Chomsky & Halle (1968), echoed more recently by Ladefoged & Maddieson (1986): that palatals are related to velars in being made with the tongue body or dorsum, with palatals more front than velars. Last is the description accepted by Halle & Stevens (1979), and adopted in much recent work on feature geometry (e.g. Sagey 1986): that palatals are primarily made with the front part of the tongue, just behind the blade, and thus can be described as coronals, in particular non-anterior, distributed, coronals.

To my knowledge phonetic evidence has not informed these diverse proposals about palatals. Instead the phonetic description of palatals has been more or less shoe-horned into various theories. My purpose in this talk

---

*Paper presented at the December 1987 Annual Meeting of the Linguistic Society of America in San Francisco. This work is supported by the NSF under grant BNS 8418580.

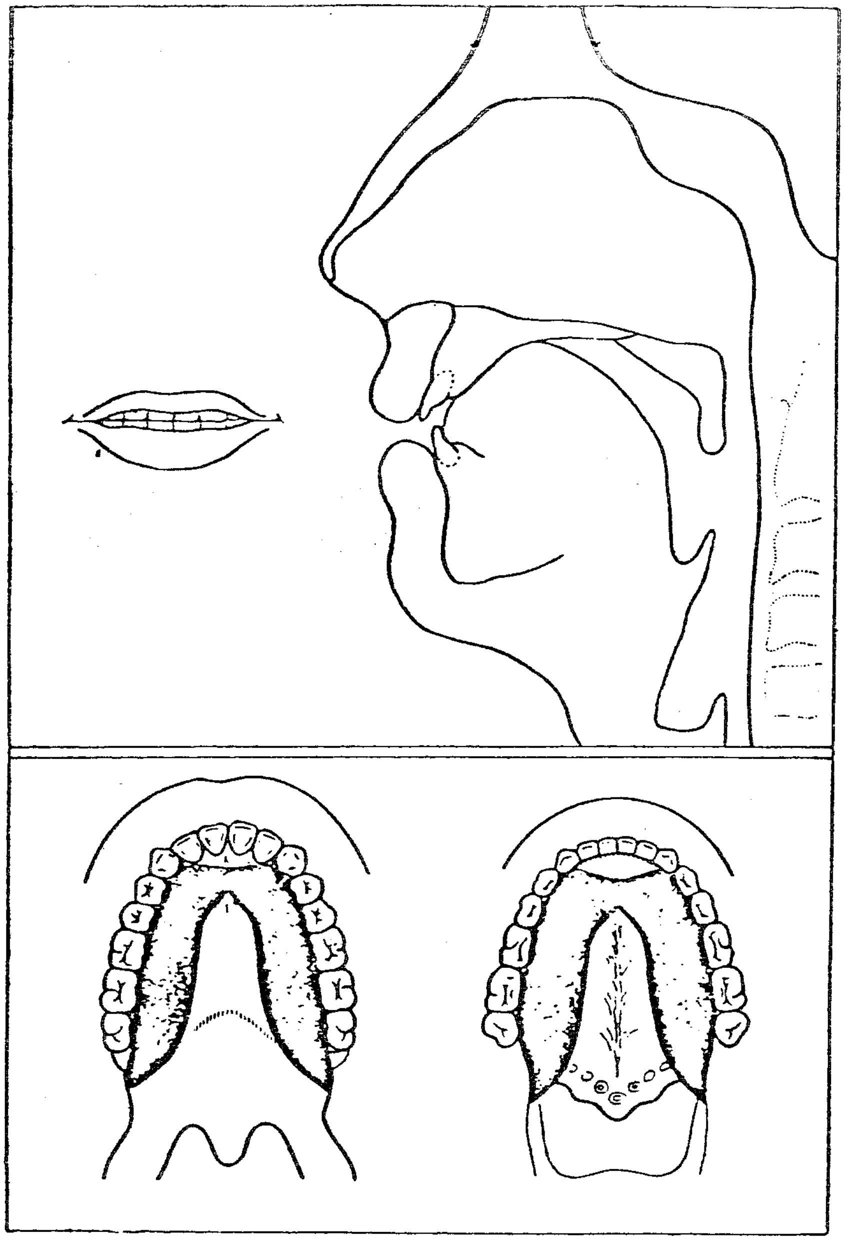


Figure 1a. Physiological data on a Czech palatal nasal [ɲ]. Clockwise from upper left: front view of lips, side-view X-ray, linguogram, palatogram.

is to examine the articulation of palatal consonants so as to arrive at an accurate phonetic description of the palatal place of articulation. My method will be to compare palatals with other places of articulation. In this way we will end up with a description not only of palatals, but also of the relation of palatals to coronals and to velars.

The data discussed here consist of tracings of midsagittal X-rays of various consonants from various languages. These tracings are taken from the literature, and are available to me largely as xeroxes in the UCLA X-Ray Database, described in Dart (1987). To begin, let me introduce you to the relevant aspects of an X-ray tracing by examining our first figure, the Czech palatal nasal, taken from Hála (1962), repeated as Figure 1b. First, where is the constriction or contact, the narrowest passageway? The contact is marked by hatches, and will be marked like this in all the other figures. What part of the tongue and what part of the palate are involved? How long (in the sagittal plane) is the constriction? It is a very long contact between the entire front of the tongue and much of the hard palate. The corner or protuberance at the back of the alveolar ridge (marked by an arrow) is contacted, but the part of the ridge just behind the upper teeth is not. Second, where is the tip and front of the blade, marked by stipling? This sound is clearly "tongue tip down"; in fact, in this instance the tip is securely behind the lower teeth. Note that because the tip is so low there is no empty space under the blade of the tongue or behind the lower teeth. Ken Stevens has discussed the importance of a cavity under the tongue for the acoustics of non-anterior coronal articulations, but palatals are different in this respect. Finally, what is the overall shape and orientation of the tongue? In this case, the tongue body, especially at about the level of the uvula, is very much fronted. It is pitched forward to make possible the combination of tip down and wide contact that is characteristic of palatals.

At this point, let me note that it is important to be able to see all of these aspects of articulation, but that this requirement makes some of the available X-rays less than helpful. It happens that, for palatals, the best documented language is Czech, for which there are data from six different speakers.

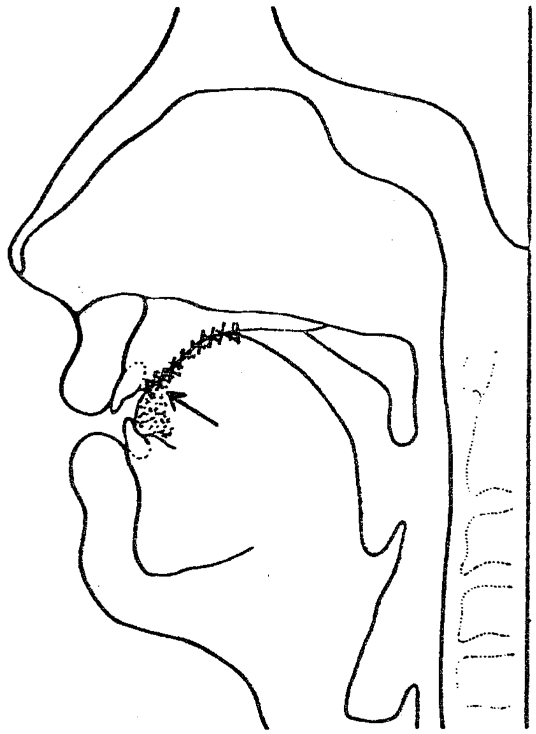


Figure 1b. X-ray from Figure 1a. Hatches show area of contact; arrow points to protuberance at back of alveolar ridge; stipling shows tip and blade front.

We began by looking at a palatal stop, since contact is the easiest articulation to see. However, for most people the prototypical palatal consonant is the glide [j], related to a high front vowel articulation. Figure 2, taken from Wierzchowska (1980), compares an example of [i] with an example of [j]. The glide is like the stop we just saw in having a very long narrowing between the tongue and the palate, with no tip involvement and no space under the tongue. However, the tongue is necessarily less forward for the glide than it was for the stop. The comparison with the vowel shows that the vowel's narrowest section is in the same broad area as that of the glide. Further narrowing of the constriction will result in a fricative, which will again have a very long constriction. Figure 3, taken from Martens (1970), makes this point by comparing the vowel [i] with a palatal fricative in German, in this case from a sequence within a single word: the fricative constriction is more front along the palate than the vowel narrowing, and the tongue is higher as well, to make the closer fricative articulation. Note that the tongue does not just move up, as textbooks often have it; that would result in a much less even constriction. If you continue this process all the way to contact, you get the palatal stop. The length of the constriction is so great because the tongue keeps its vowel-like shape. To get a shorter constriction, one more localized along the palate, the overall tongue shape would have to change more.

The length of the constriction seen for palatals means that the palatal articulation is not really a point of articulation in the way that other consonants can be located along the palate. Palatals also do not involve a single part of the tongue. They use the very back of the blade, and the large front part of the dorsum.

Now let us compare palatals to velars. The main argument of Chomsky & Halle for relating palatals to velars was the claim that when velars are fronted, palatals result. Some textbooks echo this claim in saying that the way to make a palatal is to say English words like "key". However, this claim is clearly incorrect. Figure 4, taken from Skalozub (1966) and Pacesová (1969), contrasts a palatalized velar from Russian with a palatal from Czech; they are very different indeed. The velar contact is much shorter, partly overlapping with where the back-most contact for the palatals

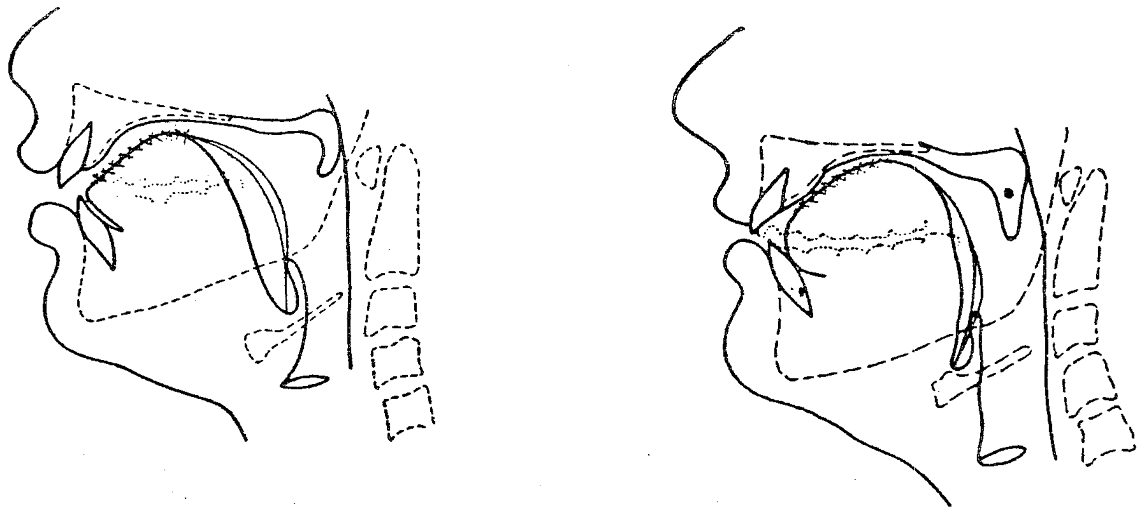


Figure 2. Polish vowel [i] and glide [j].

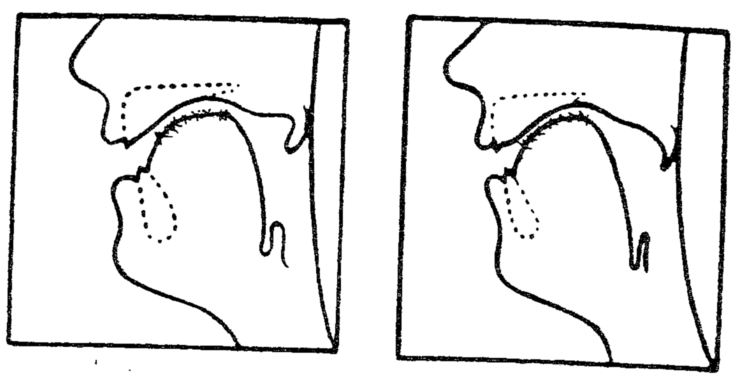


Figure 3. German [i] and palatal fricative [ç].

is seen. Though the tip is down for the velar, it is much further back, and there is a very large front cavity. If a palatalized velar is not a palatal, what is it? Figure 5, from Oliverius (1974), compares a palatalized and a plain velar in Russian. The palatalized velar is simply a fronted velar, with the same length of contact as the back velar, but moved forward on the palate. Figure 6, taken from Koneczna & Zawadowski (1956) and Brunot & Bruneau (1933), makes this clear by comparing a palatalized velar from Russian with two velars from French, before vowels [i] and [u]. Compare the location and length of contact, and the overall shape of the tongue. The Russian palatalized velar looks like the more fronted French velars (though the overall tongue shapes are not identical). Thus velars may be front or back, but fronting a velar does not turn it into a palatal. Figure 7 gives a schematic summary of this point. In this figure we are looking only at the roof of the mouth, so please ignore the tongue. The hatch marks along the front show the part of the palate used for palatals. The hatch marks along the back show the parts of the hard and soft palates used for velars. The two areas overlap somewhat on the hard palate. A palatal consonant covers the whole palatal region at once with a very long contact. A velar consonant has a much shorter contact, somewhere within the velar region.

I said earlier that palatals are not the result of just raising the tongue from the position for [i]. Instead, this is more like how fronted velars are formed, with one addition: fronted velars, like other velars, generally pull the blade back, so as to form a rounded tongue hump. This suggests a subtle but interesting point: the articulatory relation of palatals to high front vowels is not parallel to the relation between velars and high back vowels. We have already seen that if you take a high front vowel and drive it to contact, you get a palatal. The same may not be true of high back vowels; if you drive one to contact, you should see a much longer contact than in fact you see with velars. Velars may have a somewhat more humped up tongue shape than do back vowels, to allow the shorter constriction.

Given our description of palatals, what feature values should represent them and distinguish them from other places of articulation? First, they are coronal, if that node is taken to include the part of the tongue beyond

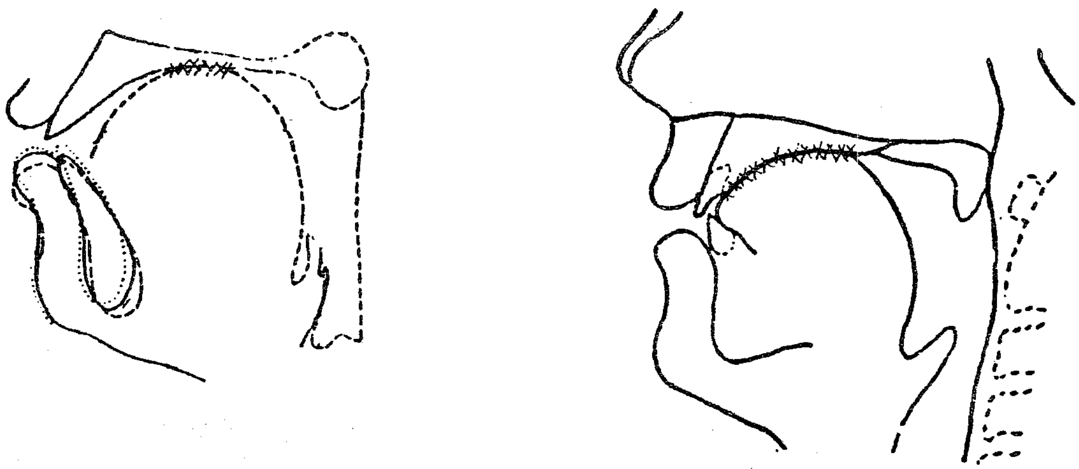


Figure 4. Russian palatalized velar $[k^j]$ and Czech palatal [c].

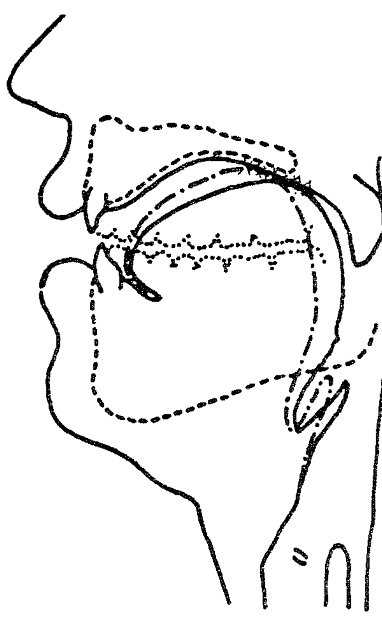


Figure 5. Russian velars: plain (solid) and palatalized (dot-dash).

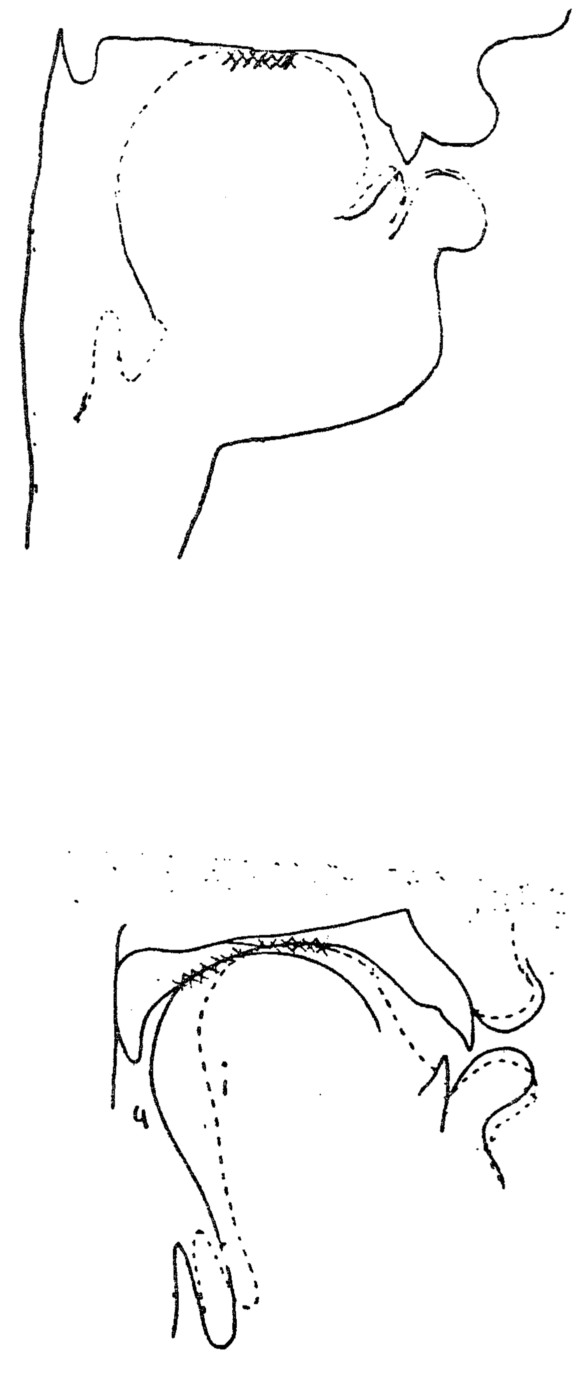


Figure 6. Russian palatalized $[k^j]$, and French [k] before $\left\{ \begin{array}{l} \text{i (dashed)} \\ \text{u (solid)} \end{array} \right\}$

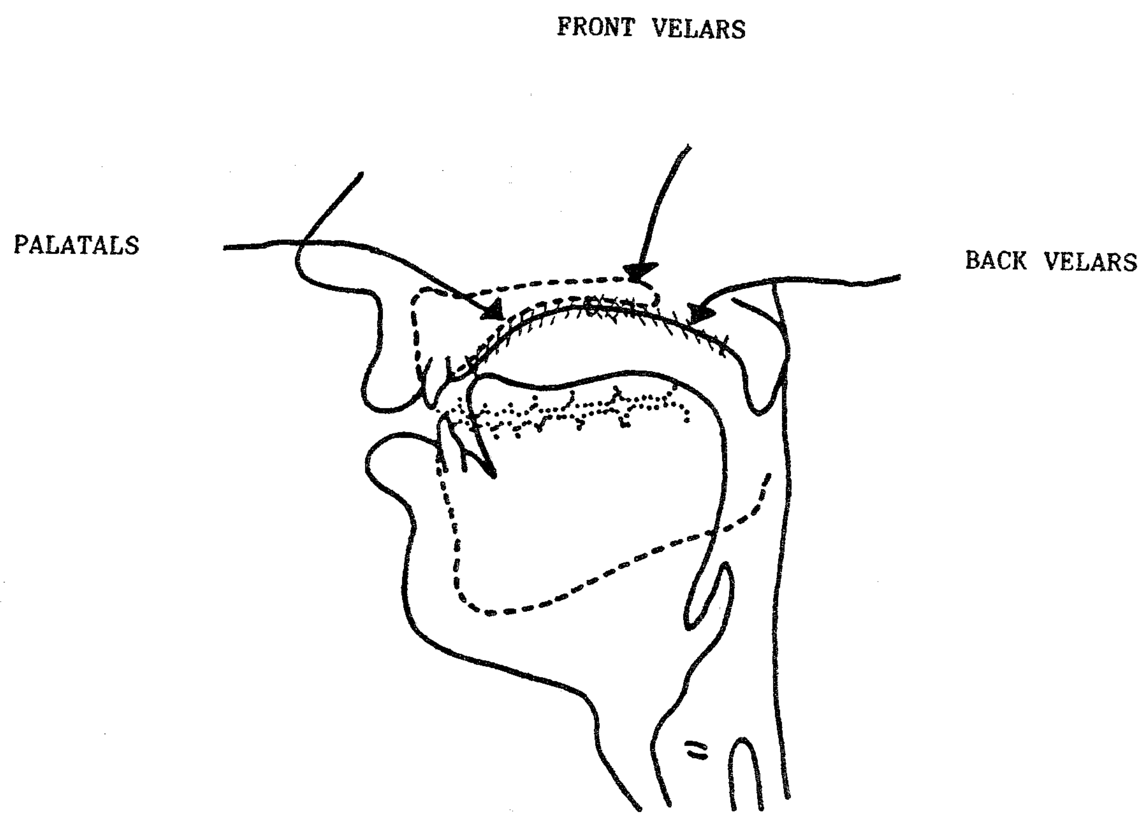


Figure 7. Areas of articulation along palate.

the very front, as discussed by Halle & Stevens (1979): that is, from the tip up to the dorsum. Although this part of the tongue is not "raised", as is sometimes taken to be the criterion for coronals, it is involved in forming the constriction. As coronals, palatals must be non-anterior, since the upper teeth and the front of the alveolar ridge are never constricted. They also must be distributed, meaning that they have a long constriction; in fact, they are the most distributed of any place of articulation. Palatal constrictions are typically on the order of twice as long as those of velars, and three times as long as those of anterior coronals.

Let me briefly mention how palatals are distinct from the other nonanterior, distributed, coronal places of articulation, shown in the next two figures. Figure 8, taken from Hála (1962), shows Czech palato-alveolars, a fricative and a stop from an affricate. Unlike palatals, these have a cavity under the tongue blade, and a relatively short constriction. Figure 9, taken from Ohnesorg & Švarný (1955), shows Mandarin alveolo-palatals, again a fricative and a stop from an affricate. (These are very much like the alveolo-palatals of Polish, except that the Polish ones vary across speakers as phonetically anterior or nonanterior). Like the palato-alveolars, these have a cavity under the tongue blade (though it's small for the affricate), and a constriction which, though rather long, is nonetheless shorter than a palatal constriction, as it does not extend as far back; and in fact it generally varies across tokens. Furthermore, here at least the overall orientation of the tongue is not like that of front vowels or palatals. That is, these are consonants in which the blade is raised up to form a simple constriction.

Palatals are more than this, being articulated with tongue surface further behind the blade area, forming a constriction extending further back on the palate. This is because palatals are articulated like "consonantal" front vowels. Chomsky & Halle were wrong in saying that palatals are the same thing as fronted velars, but they were right in saying that the palatal articulation involves a high front tongue body. Under current geometric feature theory, there is no difficulty in representing palatals as coronal and at the same time related to high front vowels. A possible (partial) representation is shown in (10) below. Note that, for the palatal obstruents,

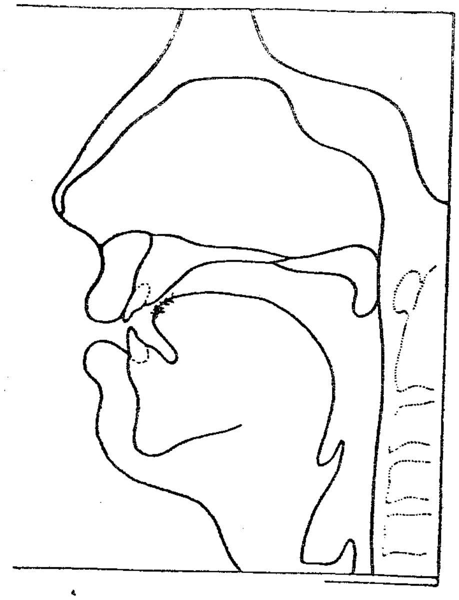


Figure 8. Czech palato-alveolars: fricative [ʃ] and stop from affricate [tʃ].

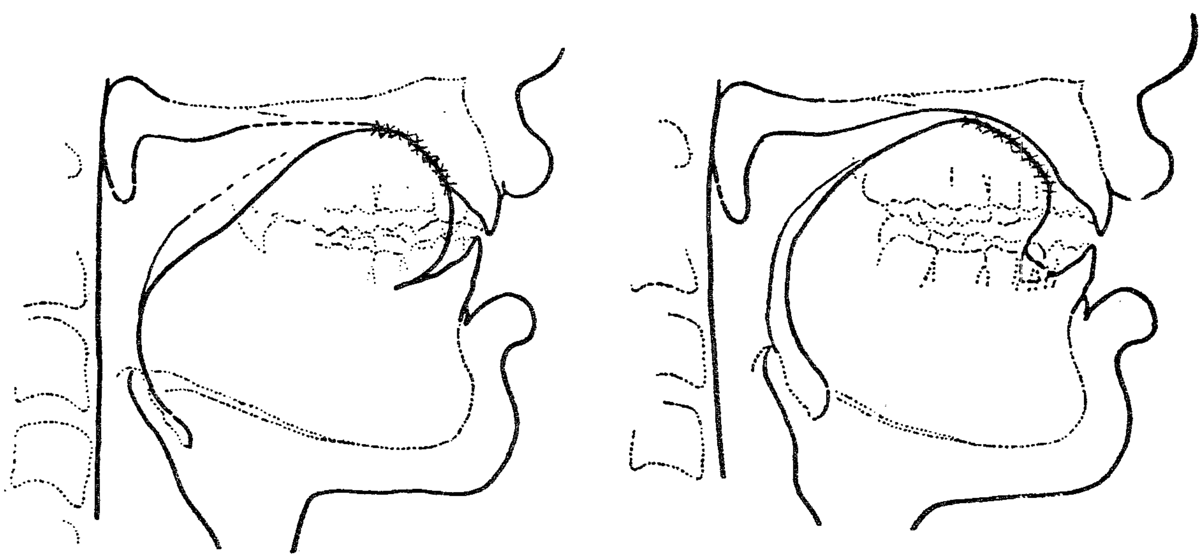
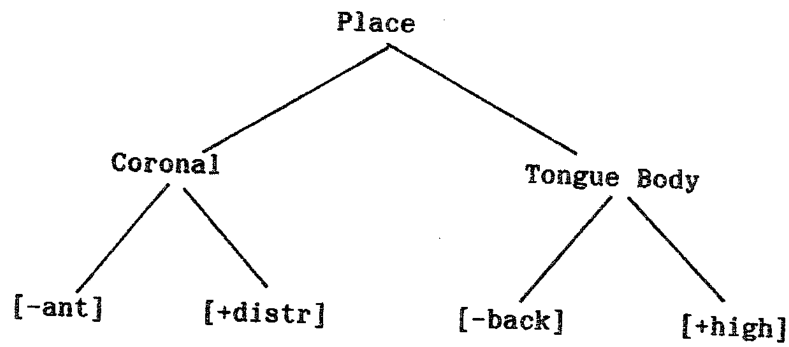


Figure 9. Mandarin Alveo-palatals: stop from affricate [tɕ] and fricative [ɕ].

neither part of the palatal articulation is sonorant-like. This feature, among other characteristics, might be exploited to distinguish palatals from secondarily palatalized consonant types.

(10)



So palatals are articulated by more than one class of articulators. They are therefore complex segments, equivalent to a double articulation, as found, for example, in labial-velars. This move has several welcome consequences. First, it distinguishes palatals from front velars, as is clearly necessary. Front velars have no coronal component. Second, it distinguishes palatals from other non-anterior coronals, which are less similar to high front vowels, and are represented only as coronals. To date there has been no way to distinguish these coronals. Finally, the representation of palatals as complex segments predicts their rarity relative to other places of articulation, at least for obstruents. An obstruent constriction involving two different articulators, the blade and the body, is complex in articulation and should be equally complex in its representation.

References

Brunot, F. & C. Bruneau (1933). *Précis de grammaire historique de la langue française.* Paris: Masson.

Chomsky, N. & M. Halle (1968). *The sound pattern of English.* New York: Harper and Row.

Daneš, F., B. Hála, A. Jedlička, and M. Romportl (1954). *O mluvenem slove.* Prague: Statní Pedagogické Nakladatelství.

Dart, S. (1987). A bibliography of x-ray studies of speech. *UCLA WPP 66*: 1-97.

Hála, B. (1923). *K popisu pražské výslovnosti (Studie z experimentální fonetiky).* Prague: Rozpravy České Akademie Věd a Umění, Vol. 3, No. 56.

Hála, B. (1962). *Uvedení do fonetiky češtiny: Na obecně fonetickém základě.* Prague: Nakladatelství Československé Akademie Věd.

Hála, B.& M. Sovak (1955). *Hlas-rec-sluch (Základni věci z fonetiky a logopedie).* Prague: Státní Pedagogické Nakladatelství.

Halle, M. & K. N. Stevens (1979). Some reflections on the theoretical bases of phonology. In *Frontiers of speech communication research*, ed. B. Lindblom & S. Ohman. London: Academic Press.

Jakobson, R., G. Fant & M. Halle (1963). *Preliminaries to speech analysis.* Cambridge, MA: MIT Press. Fourth printing. First published 1951.

Koneczna, H. & W. Zawadowski (1956). *Obrazy rentgenograficzne glosek rosyjskich.* Prace językoznawcze, 9. Warsaw: Panstowowe Wydawnictwo Naukowe.

Ladefoged, P. (1971). *Preliminaries to linguistic phonetics.* Chicago: University of Chicago Press.

Ladefoged, P. & I. Maddieson (1986). *Some of the sounds of the world's languages* (preliminary version). *UCLA WPP* 64.

Lahiri, A. & S. Blumstein (1984). A re-evaluation of the feature coronal. *J. Phonetics* 12.2: 133-146.

Martens, P. Optische Unterstutzung akustischer Sprachsignale. In B. Hála, M. Romportl, & P. Janota (eds.), *Proceedings of the Sixth International Congress of Phonetic Sciences* (Prague 1967). Prague: Československá Akademie Věd.

Ohnesorg, K., & O. Švarný (1955). *Études experimentales des articulations chinoises.* Prague: Rozpravy Ceskosloveské Akademie Věd, Vol. 65, No. 5.

Oliverius, Z.F. (1974). *Fonetika russkogo jazyka.* Prague: Státní Pedagogické Nakladatelství.

Pacesová, J. (1969). *Fonetika a ortoepie češtiny.* Brno: Nákladem JAMU

Sagey, E. (1986). *The representation of features and relations in non-linear phonology.* MIT dissertation.

Skaličková, A. 1974. *Srovnávací fonetika angličtiny a češtiny.* Prague: Československá Akademie Věd.

Skalozub, L. G. (1966). *Uprazhenija po fonetike russkogo jazyka.* Kiev University, Laboratory of Experimental Phonetics.

Styczek, I. (1973). *Badania eksperymentalne spirantow Polskich s,š,ś ze* *stanowiska fizjologii i patologii mowy.* Warsaw: Polska Akademia Nauk.

Wierzchowska, B. (1980). *Fonetyka i fonologia języka polskiego.* Warsaw: Ossolineum.

Vowel variation by consonantal context in Kana*

Keiko Nagano with Kelly Burch

Introduction

To date there has been no acoustic phonetic analysis of Kana, an Ogoni language of Nigeria. This study presents acoustic data on Kana vowels in various consonant contexts. Kana has a seven oral vowel system with /i/, /e/, /ɛ/, /a/, /ɔ/, /o/, and /u/, both long and short, which is common in related West African languages. Compared to English vowels, the vowels of Kana are fairly evenly spaced on a vowel-chart, but the mid tense vowels /e/ and /o/ are close to the high vowels.

In this study the vowels of Kana are represented by their first two formant frequencies (F1 and F2), in a vowel space defined by F1 vs. F2-F1. Basic formant frequency data for the Kana vowels are provided. In addition, the vowels are analyzed in four consonantal contexts: bilabial, alveolar, velar and labio-velar, with consideration of the acoustical and articulatory characteristics of these contexts. The possibility of systematic changes across vowels depending on their contexts is examined. The following hypotheses are tested:

First, because the production of bilabial consonants involves closure of the lips, and therefore a lowering of all formants, it is hypothesized that the formants of a vowel following a bilabial will be lowered. Second, because alveolar consonants have as an acoustic correlate a locus of F2 at about 1700 to 1800 Hz, it is hypothesized that the second formant of a vowel following an alveolar will be influenced by this locus value. This means that vowels whose F2 is otherwise lower than this, will have a raised F2, and vice versa. It is known for English vowels that coronal consonant contexts tend to result in centralizing of vowels in the F2 dimension (Stevens & House 1963; but Schouten & Pols 1979 found no clear effect in Dutch). Third, since

---

*This article is based on Keiko Nagano's 1987 senior essay in Linguistics, with editing by Kelly Burch and Patricia Keating.

velar consonants are produced further back in the mouth than the other consonants, it is hypothesized that a vowel following a velar will be backed, and thus a lowered value of F2 (or F2-F1). Finally, since a labio-velar is an articulatory combination of a labial and a velar, it is hypothesized that a vowel following a labio-velar will be affected by both of these articulatory components of the consonant.

Procedure

Four male speakers of Kana were recorded by Dr. Patricia A. Keating at the University of Port Hancourt, Nigeria, in 1984. Each speaker produced a set of real Kana words. Each word consists either of one of the ten consonants (/b/, /d/, /t/, /z/, /s/, /l/, /g/, /k/, /gb/, kp/) followed by one of the seven short oral vowels (/a/, /ɛ/, /e/, /i/, /o/, /ɔ/, /u/), or of one of the alveolars or the glide /j/ followed by a long vowel. Most words have a mid tone, but low and high tones also occur. It must be noted that since only real words were used, not all logically possible CV combinations occur. For each token the English gloss was read by the experimenter, followed by the Kana word spoken twice.

At UCLA, the formant frequencies of the steady states for all of the vowel tokens were measured using CSpeech, a speech analysis program for the IBM PC by Dr. Paul Milenkovic of the University of Wisconsin. The steady state of a vowel was judged from waveform and running LPC spectra displays, and was always towards the end of the vowel in these CV utterances. Formant frequencies were measured by manual peak-picking from LPC spectra and/or FFT spectra where necessary. F2 and F3 of some of the back vowels were difficult to measure; for these cases, just the FFT analysis and first differenced LPC spectra were used. (First-differencing boosts energy in the higher frequencies.) Measurements for repetitions of each word were averaged, and various graphs were plotted. For each speaker five averages of F1 and F2 for each vowel were calculated: following bilabials, alveolars, velars, labio-velars and all consonants.

Results

The graphs indicate how a preceding consonant affects the F1 and the F2-F1 of a vowel. Each dot represents an averaged vowel in a consonant context, with the dots for a given vowel connected on the graph and identified by a nearby vowel symbol. The place of articulation of the preceding consonant is indicated by the symbols p, t, k, kp next to each dot. For the sake of clarity of the graphs, and because of different patterns found for different vowels, subsets of the data are presented on each graph.

The graph in Figure 1 shows the front vowels, across consonant contexts, averaged for all 4 speakers, while Figures 2 through 5 show the data for the individual speakers. For each of the four speakers, context affects front vowels' F1 more than F2-F1. For example, the graphs of speakers 1 and 2 show that for /i/ only F1 changes with consonantal change. There is a variation in F1 of about 60 Hz across the four contexts. For example, the F1 of speaker 1 is at 230 Hz for bilabials, 310 Hz for alveolars and 337 Hz for velars. Similar variation is found for the other three speakers. Lower vowels vary more than higher vowels. There is no overlap among the vowels; for example, F1 for /i/ varies in F1 from 258 Hz to 301 Hz, while F1 for /e/ varies from 320 Hz to 381 Hz.

As one example of effects of consonant context on front vowels, /i/ has its lowest F1 (an average of 258 Hz) when following a bilabial, a higher F1 (an average of 273 Hz) when following an alveolar and a still higher F1 (an average 301 Hz) when following a velar. The corpus does not happen to include /i/ in the context of a labio-velar, but the F1's of most other vowels following a labio-velar are higher than the F1's of vowels following an alveolar (with the exception of speaker 2's /ɛ/, which has an average F1 of 411 Hz after labio-velars and an average F1 of 469 Hz following alveolars). The lowering of F1 after bilabials is in accord with the hypothesis. The high F1 after velars may be due to a possible low jaw position for velars relative to alveolars, as has been observed for other languages.

The back vowels (Figures 6-10) vary more in the F2-F1 than the F1 dimension. The vowels after alveolar consonants tend to have the highest F2-F1 (/u/ 1017 Hz, /o/ 1107 Hz, /ɔ/ 1094 Hz and /a/ 1425 Hz). Vowels after labio-velar consonants tend to have the lowest F2-F1 (/u/ 718 Hz, /o/ 902 Hz, /ɔ/ 934 Hz, and /a/ 1272 Hz), though labials and velars also have low F2-F1

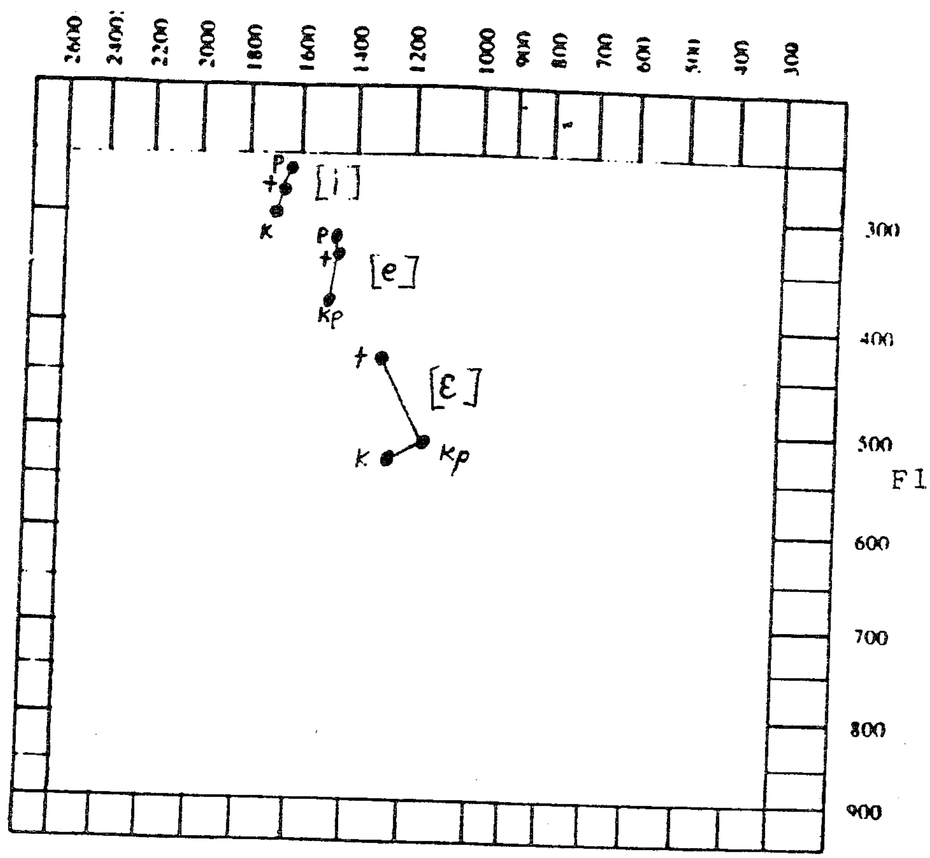


FIG. 1

ALL SPEAKERS

FIGS. 2-5 SHOW THE FOUR SPEAKERS INDIVIDUALLY

F2-F1

FIG. 2

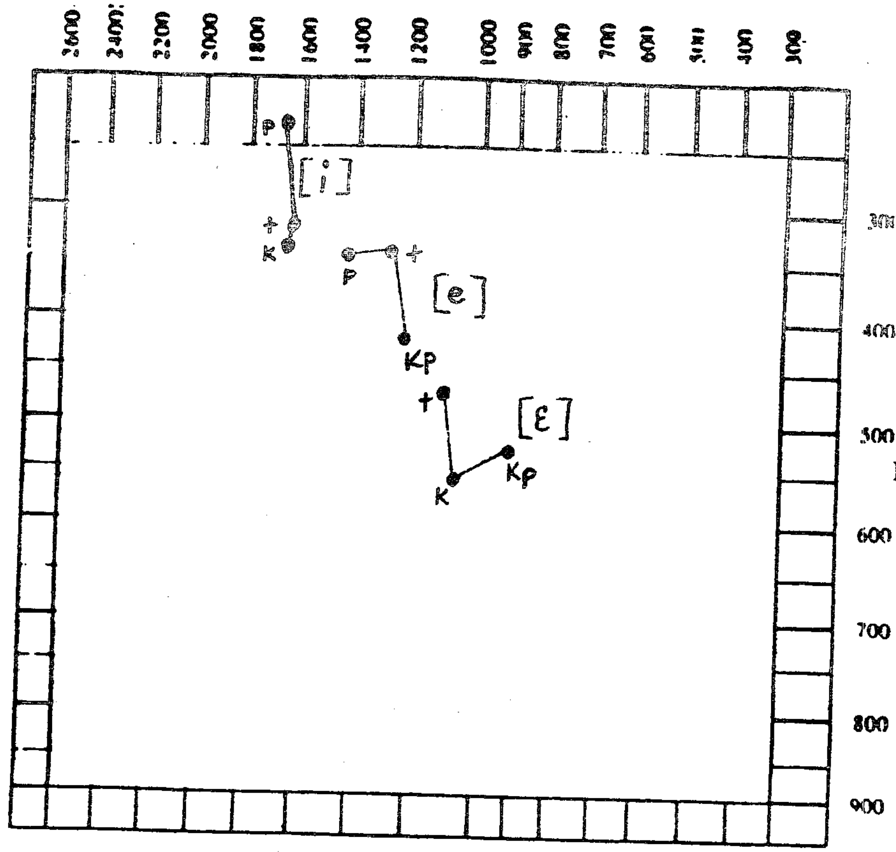


FIG. 2

F2-F1

FIG. 3

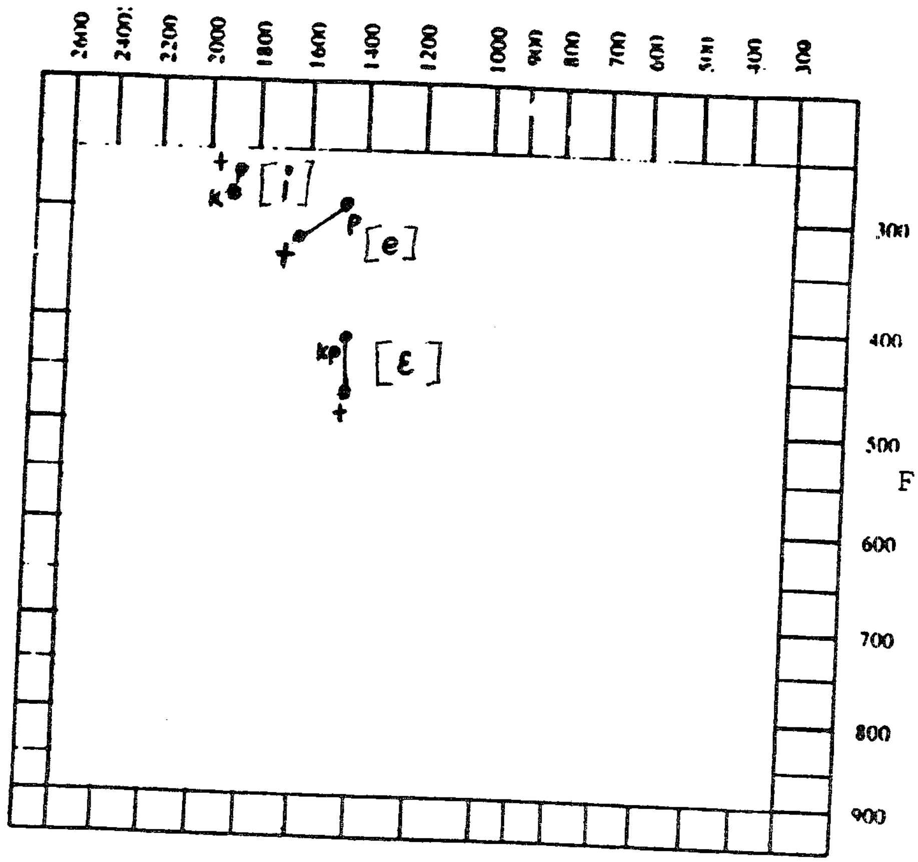


FIG. 3

FIG. 4

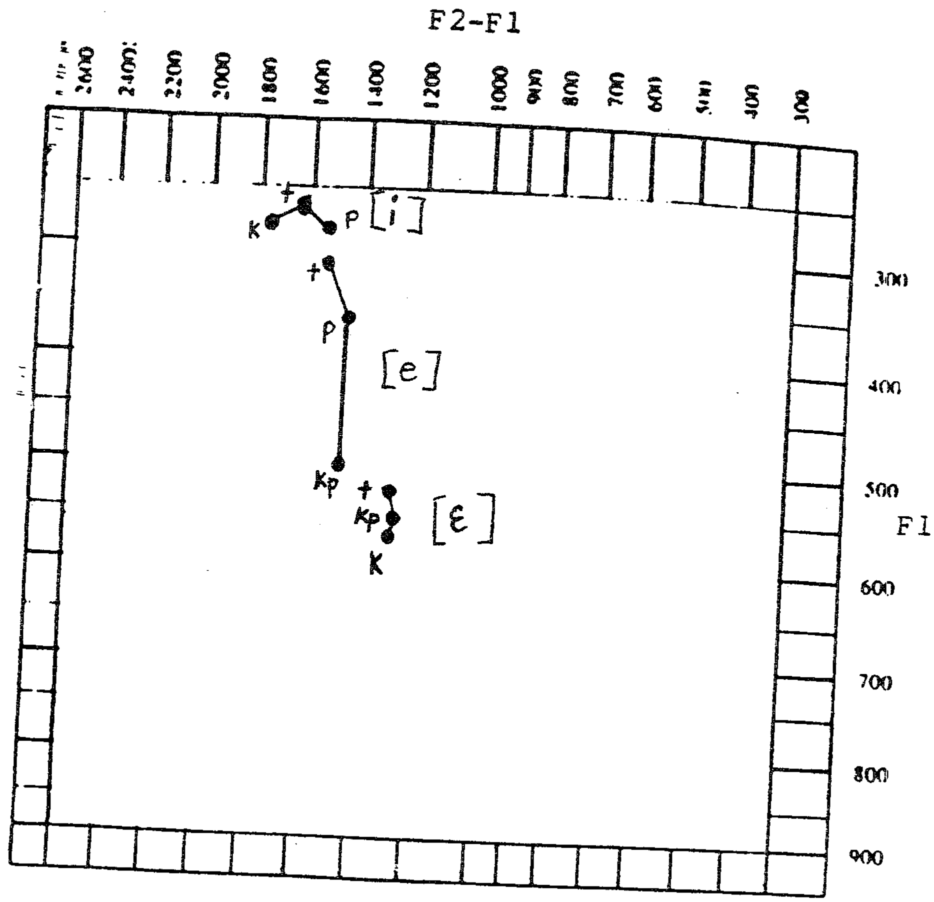


FIG. 4

FIG. 5

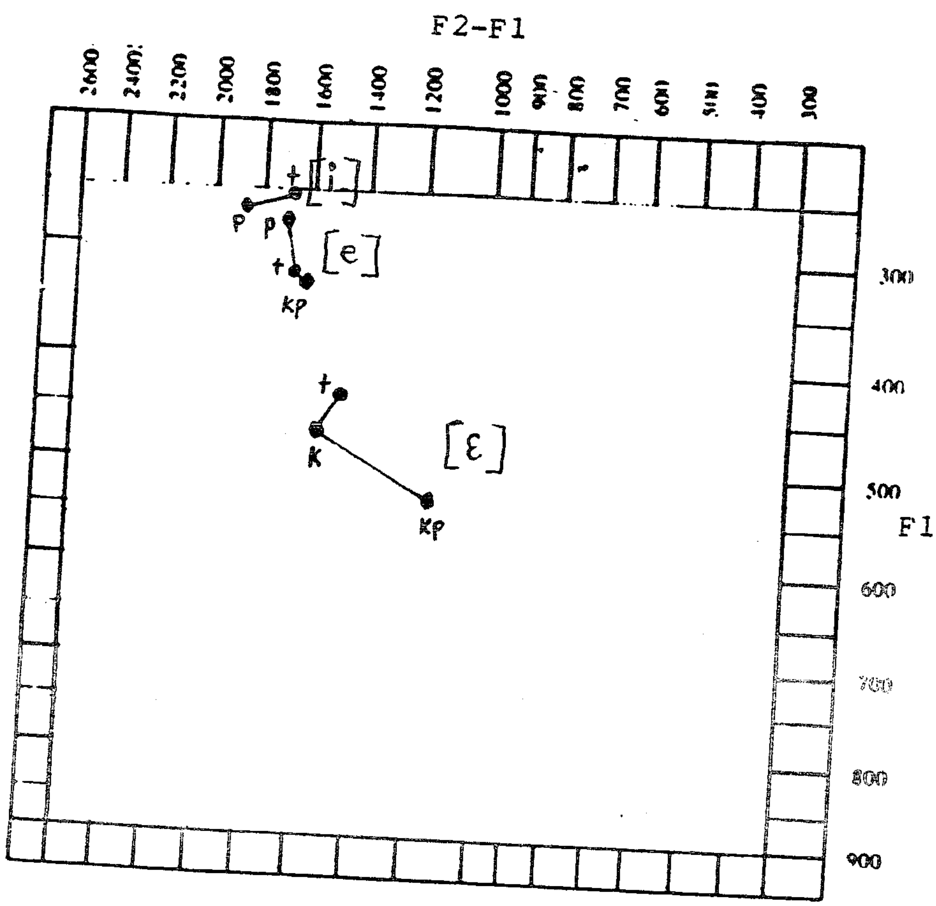


FIG. 5

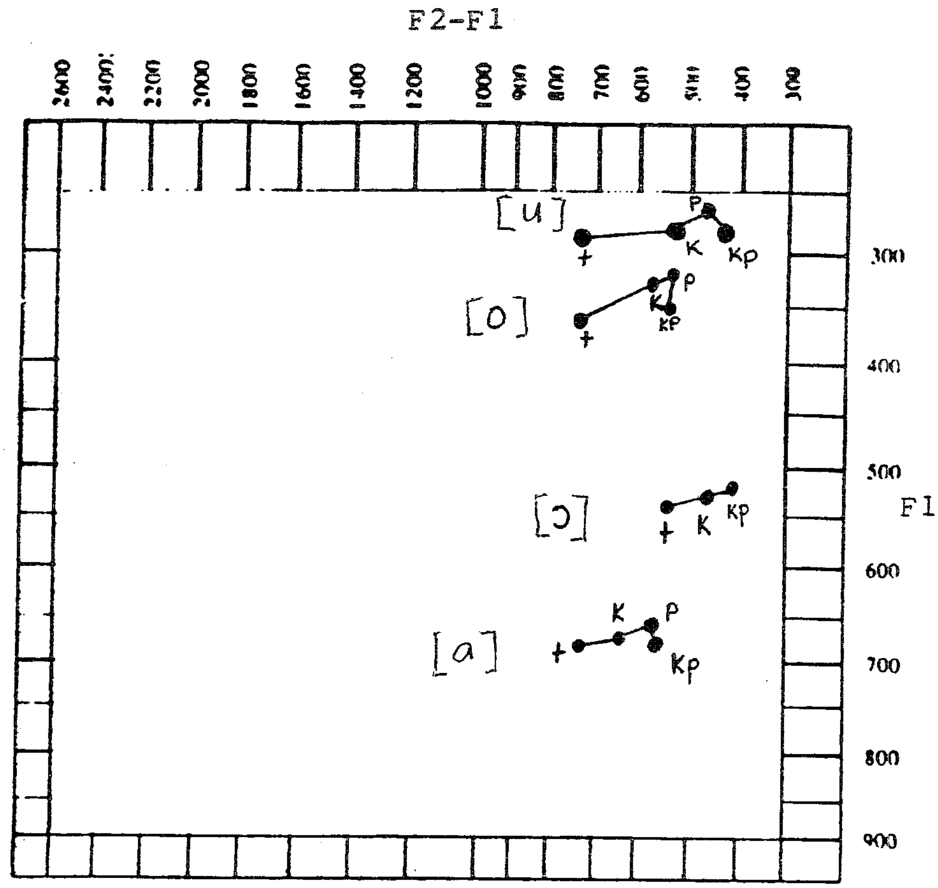


FIG. 6

ALL SPEAKERS

FIGS. 7-10 SHOW THE FOUR SPEAKERS INDIVIDUALLY

FIG. 7

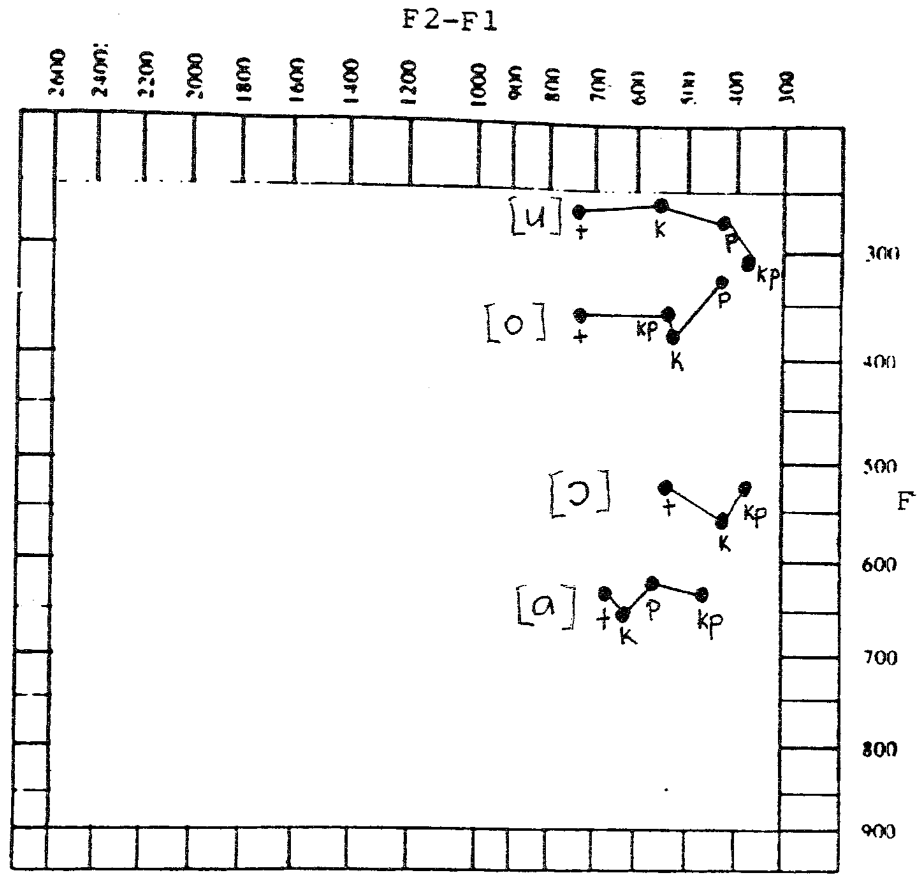


FIG.7

FIG. 8

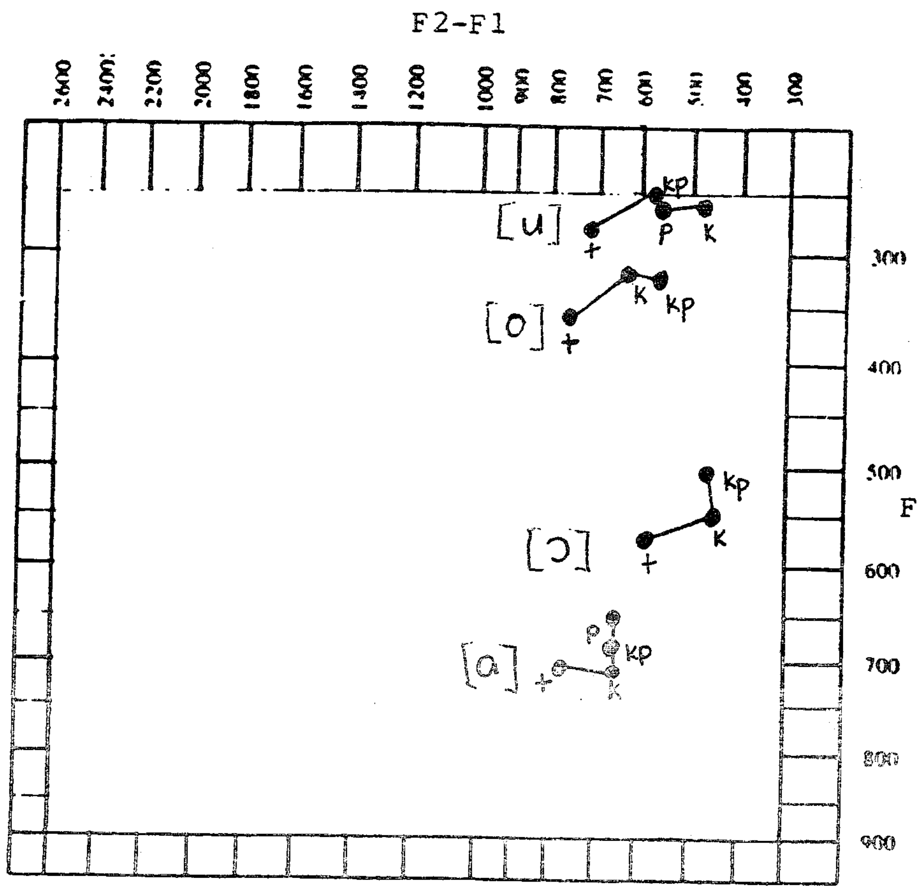


FIG. 8

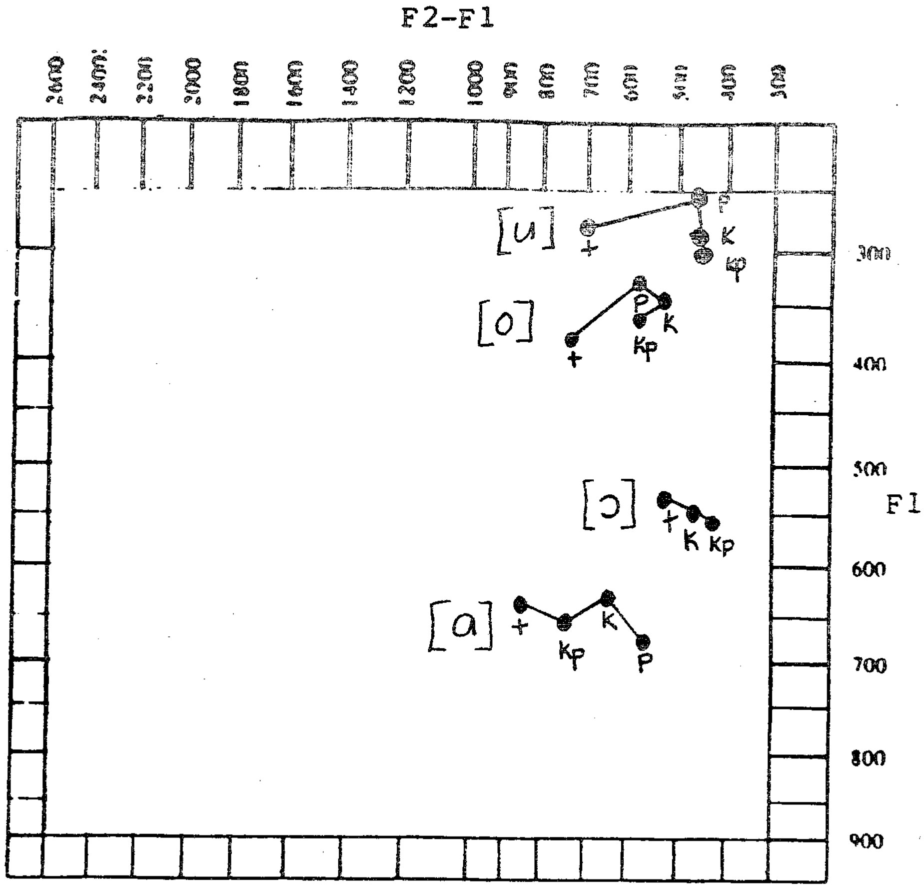


FIG. 9

FIG. 9

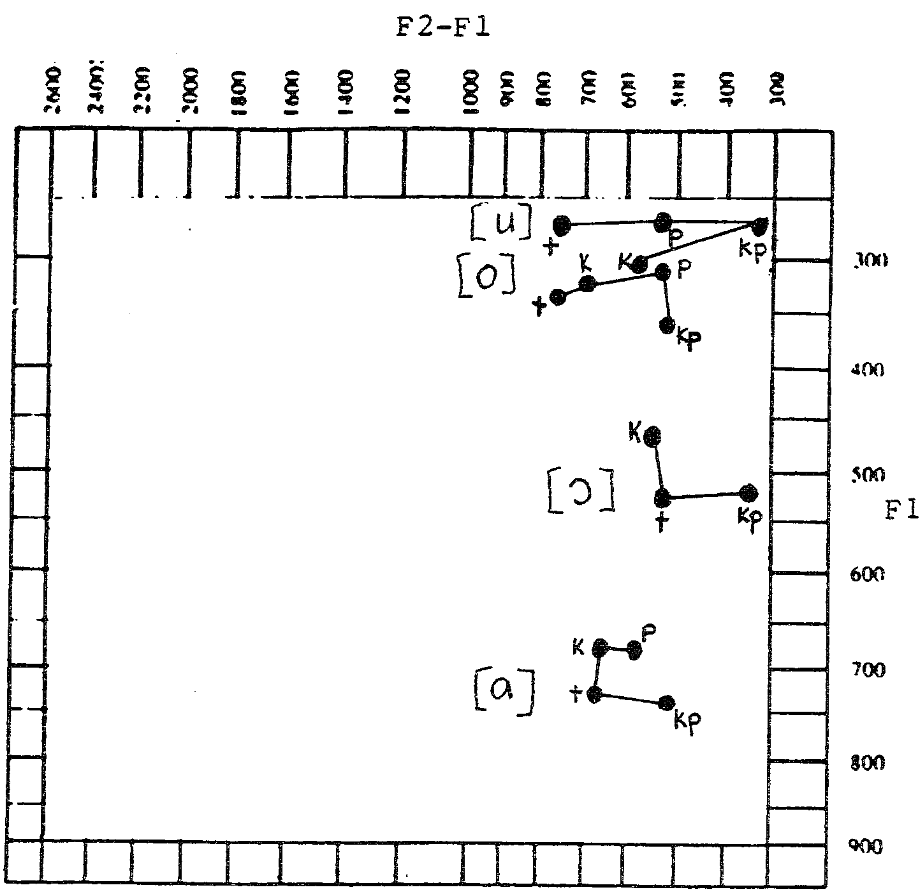


FIG. 10

FIG. 10

values. That is, back vowels are fronted after alveolars and slightly backed after labio-velars compared to the other consonant contexts. These effects, which are in accord with the hypotheses, are larger for the high vowels. Furthermore, despite this variation, there is again no overlap between the different back vowels. For example, /u/ shows an F1 variation of 268 Hz to 290 Hz while /o/ shows an F1 variation of 322 Hz to 363 Hz.

Thus front vowels vary more in the F1 dimension, while back vowels vary more in the F2-F1 dimension. In neither case does the variation result in overlap in the vowel space. Consonants affect vowels in a systematic way, but the effects on front vowels differ from those on back vowels.

Now consider the effects on vowels by consonant type.

1. Bilabials. Figure 11 shows the vowels after bilabial consonants, for all speakers. It was hypothesized that the closure of lips for bilabials would affect the lowering of all formants of the following vowel. Values for both F1 and F2 are often lower than average after bilabials. The graphs for the group and for the individual speakers show that the whole vowel set shifts up and back to lower F1 and F2-F1 values relative to the other consonant contexts. High vowels have an especially low F1 for all four speakers.

In speaker 1's case, front vowels are pronounced with very low F1, for example /i/ (230 Hz). The F1 of /e/ is also low (348 Hz). The F2-F1 of /e/ is lower than that of the velar's. In the back vowels, /a/ has the lowest F1 (658 Hz) of the four contexts. In addition to the lowering of F1 in /o/ and /u/, these vowels have very low F2.

In speaker 2's case, there are not enough vowels occurring in the bilabial context to compare with vowels in other contexts. However, all three vowels are higher and toward the right on the graph indicating lower F1 and F2-F1.

Speaker 3 shows little difference in the F1 of front vowels, but some difference in the F2 (/i/ 116 Hz and /e/ 33 Hz below their average values). However, there are differences in both the F1 and the F2 of back vowels. For

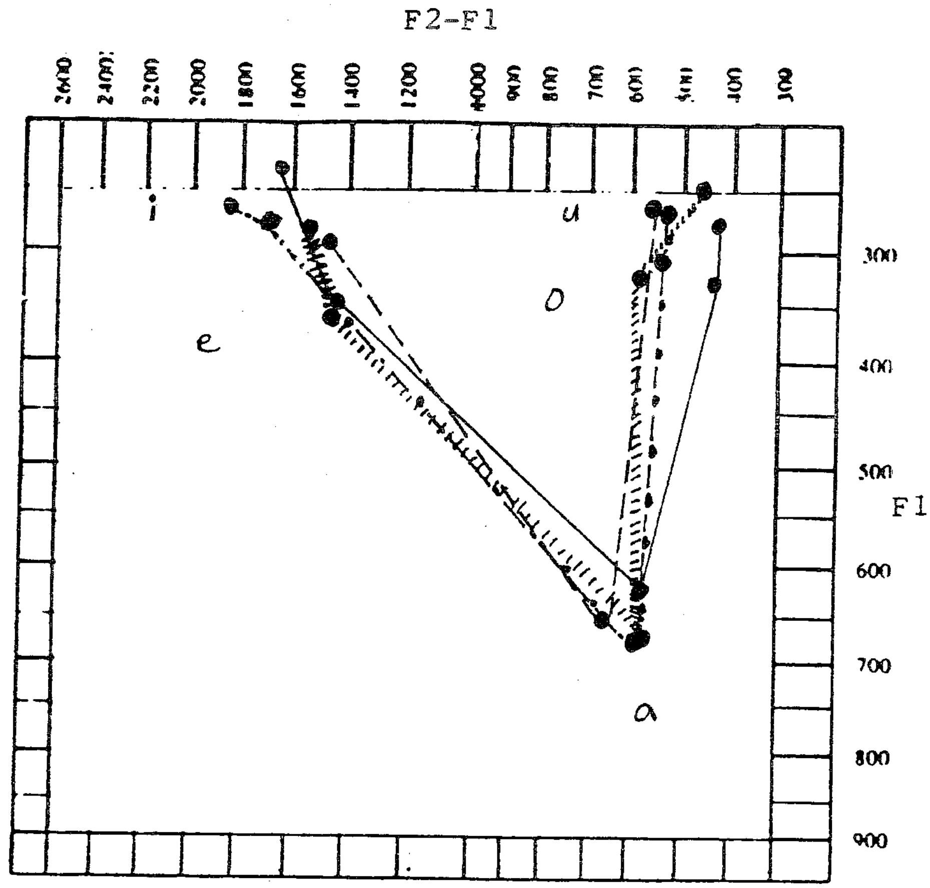


FIG. 11

————	SPEAKER 1
— — — —	2
,,,,,,,,,,,,	3
—·—·—	4

/u/ and /o/, F1 and F2 are lower than average (for /u/ F1 is -25 Hz, F2 is -86 Hz and for /o/ F1 is -28 Hz, F2 is -47 Hz), and in the graph they are close to velars and labio-velars but far from alveolars. Speaker 3's /a/ has its lowest F2 (1255 Hz) and (surprisingly) its highest F1 (683 Hz) in the bilabial context resulting in it lowest F2-F1 (572 Hz). Therefore /a/ in this context is produced lower and farther back than /a/ in the other contexts.

For speaker 4, the difference in F1 between high tense and mid tense vowels is small in the bilabial context compared to other contexts. With the exception of /e/, whose F2 is slightly lower than average, there are no differences in the F2-F1 of front vowels among the consonants. Nonetheless, the F1 and F2 of this subject's back vowels are lower, with the F2 lowering greater than the F1 lowering.

2. Alveolars. It was hypothesized that vowels in the alveolar consonant context would be centralized in the F2 dimension. Figure 12 shows that it is in the alveolar context that there is the smallest difference in F2-F1 between the front vowels and the back vowels. The vowels in the alveolar context have the narrowest 'U' shape. In the graph the F2-F1 values of front vowels move backward (due to lowering) and those of back vowels move forward (due to raising). Therefore, all the vowels are centralized. For the back vowels, for which F2 is generally below 1000 Hz, the raising of F2 is marked, about 100 Hz compared to other consonant contexts.

There is also some centralization of F1, especially for the back vowels. Most vowels' F1 is close to 500 Hz. This indicates that the high vowels are lowered (raising F1) and that the low vowels are raised (lowering F1).

3. Velars. It was hypothesized that vowels in the velar consonant context would be backed, resulting in a lowered F2. Figure 13 shows that the hypothesis is supported in the case of the back vowels but not that of the front vowels. The back vowels in the velar contexts have low F2-F1 values, second only to the labio-velar contexts (see below). The front vowels, on the other hand, generally have their highest F2-F1 values in the velar context.

FIG. 12

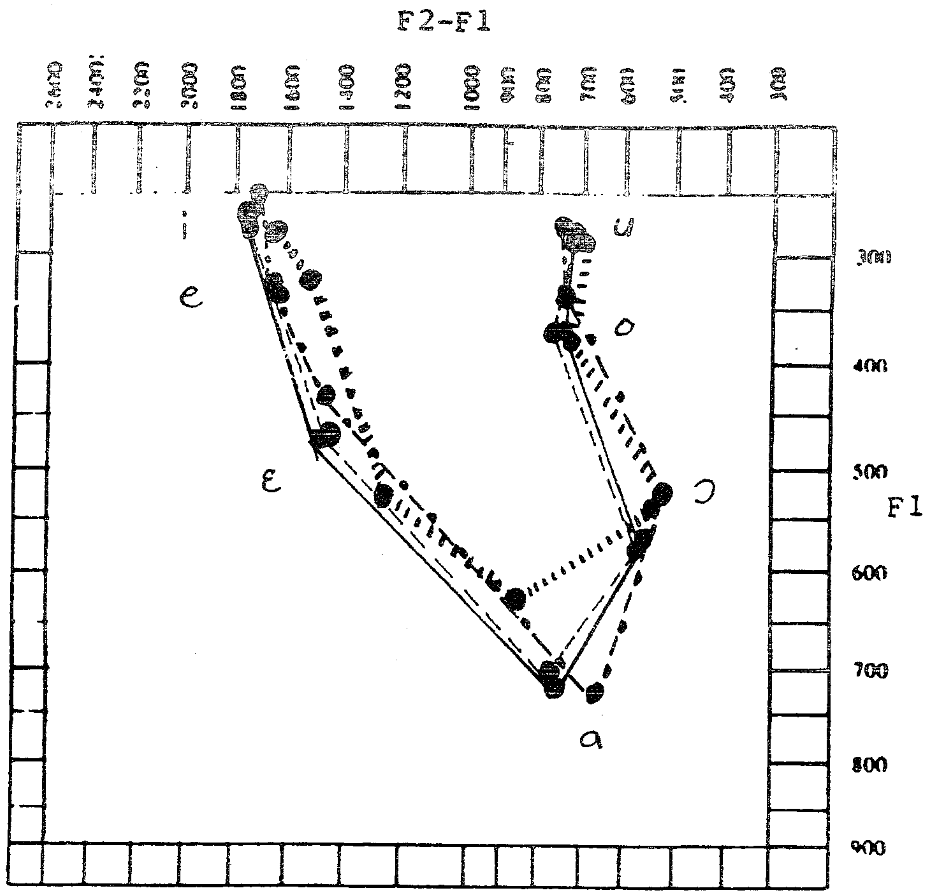


FIG. 12

FIG. 13

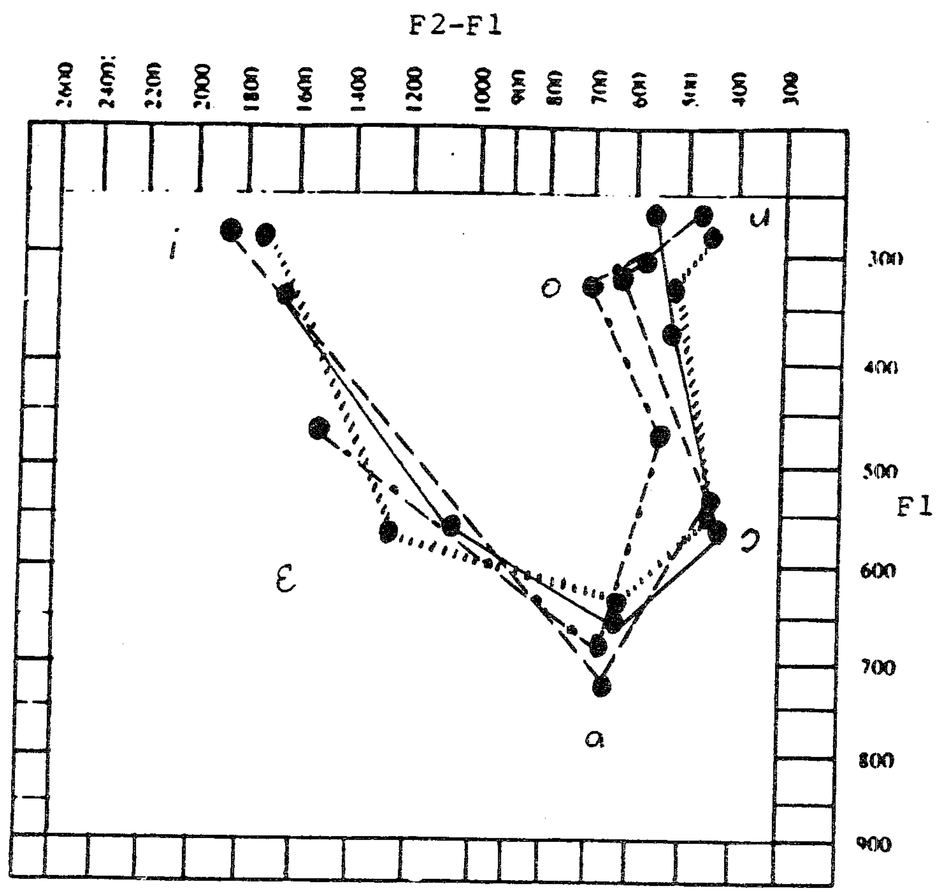


FIG. 13

It appears that back vowels in the velar context are either pronounced slightly farther back in the mouth (or, less plausibly, that the lips are more rounded). The front vowels are not similarly influenced by preceding velar consonants. This may be because the velar consonants are fronted after front vowels, as A. Cohn's work on the Kana consonants (in this issue of WPP) indicates.

4. Labio-velars. Because the articulations of two consonants are involved in the production of labio-velars it was hypothesized that vowels in the labio-velar context would either be influenced by the velar, by the labial or by both. The results are shown in Figure 14.

There is no clear lowering of vowel F1 in labio-velar contexts, as there was after bilabials. For example, the differences between labio-velars and average values show a higher F1 for speaker 1 and speaker 3 but a lower F1 for speaker 2 and speaker 4. However, the F2 in most of the cases is lower than the average, though more variably so than after bilabials. Lowering of F2-F1, especially for back vowels, would suggest influence of the velar component of labio-velars.

There is a substantial lowering of speaker 1's F2-F1. The F2-F1 of his /o/ does not change as much as the other vowels', making a large F2-F1 contrast between /u/ and /o/ in the graph. His /o/ and /7/ have formant frequencies that are very close to his /o/ and /7/ in the velar context. Other vowels in the labio-velar context are about 100-200 Hz lower than in the velar context. So, the formant frequencies of speaker 1's vowels are similar in the labio-velar context to those in the velar context, with the vowels in labio-velar context having slightly lower F2-F1 than those in the velar context.

The data for speaker 2 was insufficient to allow comparison of labial and labio-velar contexts. However, for F1 and F2-F1 of the back vowels, there is a difference between the velar and labio-velar contexts of less than 50 Hz.

In speaker 3's case the difference between F1 in the velar and in the

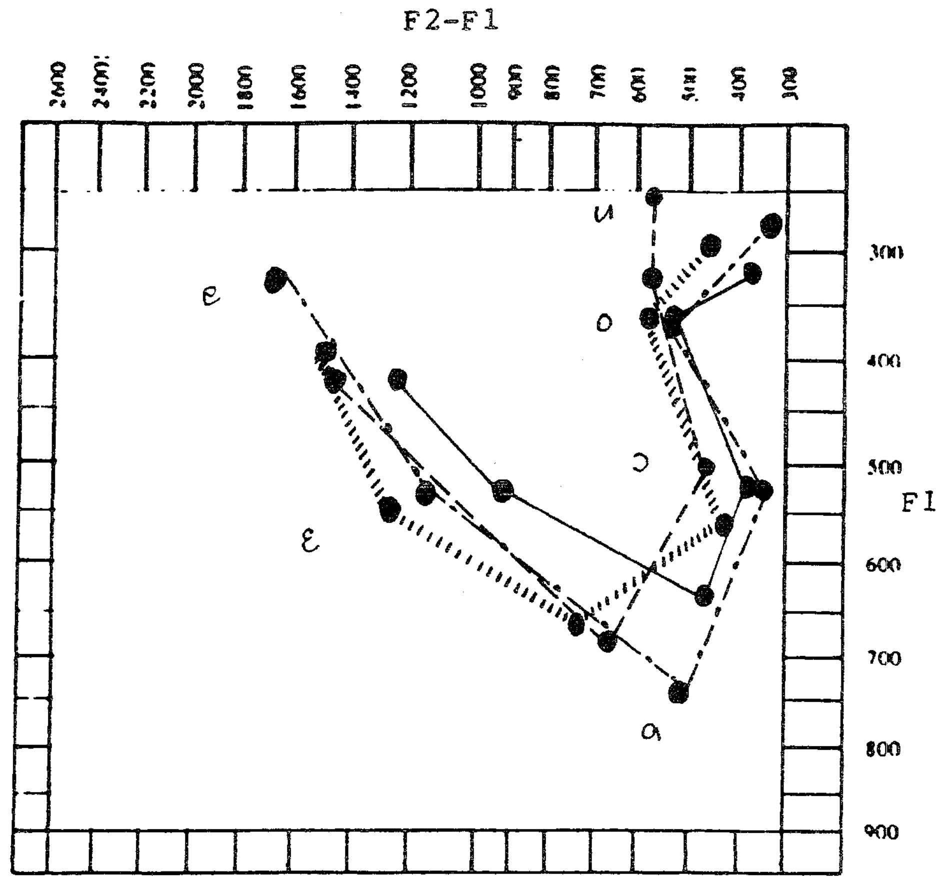

F2-F1
2600
2400
2200
2000
1800
1600
1400
1200
1000
900
800
700
600
500
400
300
F1
300
400
500
600
700
800
900
u
e
o
ɔ
ɛ
a

FIG. 14

labio-velar contexts is less than 20 Hz. And the difference of the F2-F1 is less than 50 Hz in /ɛ/, /o/, /ɔ/ and /u/. Speaker 3's /u/ and /o/ in the labio-velar context also have a big difference in F2-F1. /u/ and /o/ appear to move in opposite directions (F2-F1 of /u/ 457 Hz, and /o/ 578 Hz).

In speaker 4's case, although all the vowels in the labio-velar context have lower F2-F1 than the vowels in the velar context, the labio-velar graph is shaped like the velar graph.

Finally, Figure 15 shows the averaged vowels in the different contexts. Both the labio-velar and the velar contexts have a tendency to backness (F2-F1 lowering) and a big gap between /u/ and /o/ vowels. There is a clear differnce between the F1 of /i/ and /e/ and F2-F1 of /u/ and /o/.

Discussion

Goldstein (1983) provides a useful comparison for the present study of vowel variation. Goldstein used an articulatory synthesizer to study vowel variation as a function of random variations in tongue positioning, whereas in this study similarities and differences in the variation due to the particular consonant context are considered in real vowels produced by people. To compare Golstein's data and the data in this study, graphs were made by plotting F2 on the x-axis and F1 on the y-axis, though these graphs are not presented here.

Goldstein reported, "For front vowels, height is the only direction of movement; back vowels can also be fronted." (However, lip rounding effects can not be isolated in such graphs). Goldstein further stated that a vowel like [i] has an acoustically stable palatal constriction, but there is a change in the size (i.e. narrowness) of the constriction. For Kana, this change in size may be influenced by the preceding consonant.

In the Kana data, the front vowels vary in F1 more than in F2 for all four speakers, in accord with Goldstein's basic result. However, compared to Goldstein's data, the F2 of front vowels in Kana varies more. This is presumably because there are other things (such as lip position) affecting the formants besides the size of the primary oral constriction. Goldstein

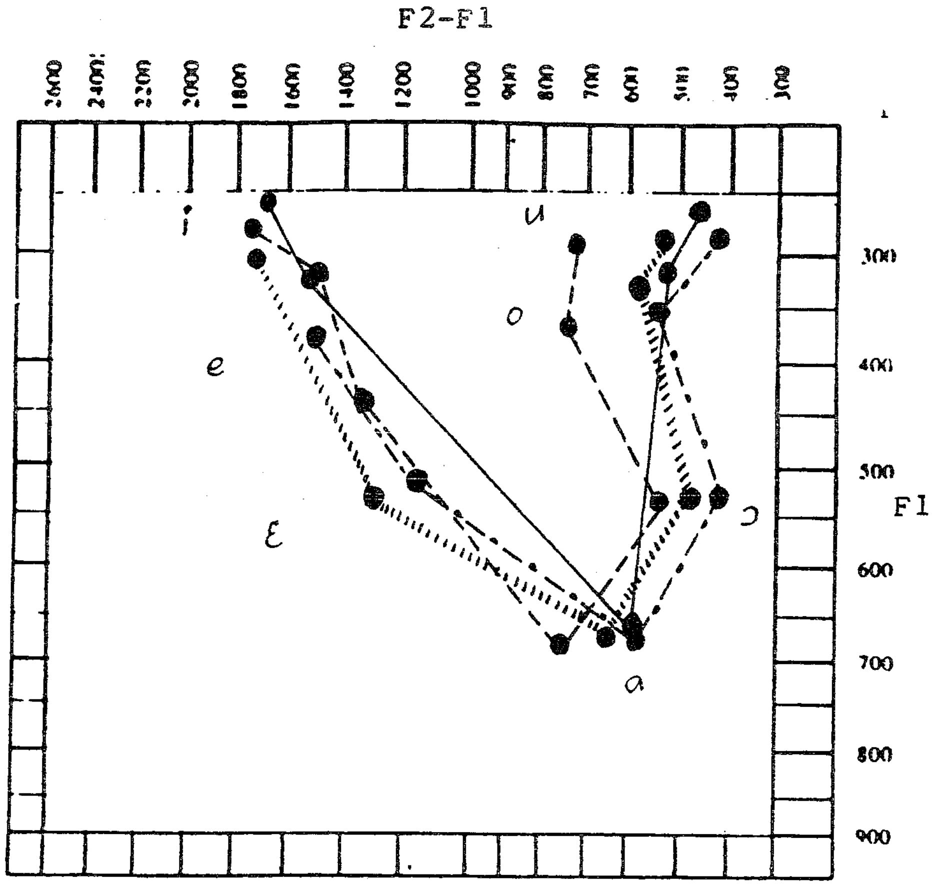


FIG. 15

BILABIALS

ALVEOLARS

VELARS

LABIO-VELARS

found that the F2 of front vowels does not vary. The Kana data support this result.

Goldstein also discussed the back vowels saying, "In addition to F1 variance the back vowels vary in F2." The differences in F2 across consonant contexts are large for the back vowels in Kana. The high back tense vowel /u/ has the largest variation in F2. As the vowels get lower, the variation in F1 becomes more pronounced. The low back vowel /a/ has equal changes in F1 and F2. Goldstein said that /u/ and /o/ differ in constriction size. So, the variability of back vowels is not exclusively along the vowel height dimension; and the back vowels show some fronting as well as height changes in vowel shifts.

Conclusion

For Kana, vowel variation in the four consonant contexts is noticeable and systematic when examined in the F1 vs. F2-F1 acoustic space. Each consonant affects a following vowel's formant frequencies by its acoustic and articulatory characteristics. In some cases, the influence is not the same for front vowels as for back vowels or for high vowels as for low vowels. In particular, front vowels vary more in F1 while back vowels also vary in F2.

References

Goldstein, L. (1983). Vowel shifts and articulatory-acoustic relations. In A. Cohen & M.P.R. v.d. Broecke (eds.), *Abstracts of the Tenth International Congress of Phonetic Sciences*:267-273.

Schouten, M.E.H. & L.C. Pols (1979). Vowel segments in consonantal contexts: a spectral study of coarticulation -- part I. *J. Phon.* 7:1-23.

Stevens, K.N. & A.S. House (1963). Perturbation of vowel articulations by consonantal context: an acoustical study. *J. Speech Hear. Res.* 6:111-128.