The American English Flapping Rule and the Effect of Stress on Stop Consonant Durations

Citation for published version:
Turk, A 1992 'The American English Flapping Rule and the Effect of Stress on Stop Consonant Durations' Cornell University Press, Ithaca, N.Y, pp. 103-134.

Working Papers of the Cornell Phonetics Laboratory

No. 7 • March 1992

Department of Modern Languages and Linguistics
Morrill Hall
Cornell University
Ithaca, N.Y. 14853-4701
telephone (607) 255-7394
fax (607)-255-7491

The Cornell Phonetics and Phonology Group (listed alphabetically):

Laura Bennua, Graduate Student in Linguistics
Ann Bradlow, Graduate Student in Linguistics
Eugene Buckley, Visiting Assistant Professor of Linguistics
Abigail Cohn, Assistant Professor of Linguistics
Kevin Connely, Graduate Student in Linguistics
Nick Clements, Professor of Linguistics and Cognitive Studies
Beverley Goodman, Graduate Student in Linguistics
Jeong-Im Han, Graduate Student in Linguistics
Susan R. Hertz, Senior Research Associate
Marie Huffman, Visiting Research Associate
Paul Iverson, Graduate Student in Psychology
Michael Jessen, Graduate Student in Linguistics
Allard Jongman, Assistant Professor of Linguistics
Hyunsoon Kim, Graduate Student in Linguistics
Becky Letterman, Graduate Student in Linguistics
Corinne Moore, Graduate Student in Linguistics
Annie Rialland, Adjunct Associate Professor (also C.N.R.S., Paris)
Susana Sainz, Graduate Student in Linguistics
Joan Sereno, Visiting Fellow in Linguistics and Psychology
David Silva, Graduate Student in Linguistics
Alice Turk, Graduate Student in Linguistics
Munther Younes, Senior Lecturer in Near Eastern Studies

Technical and Administrative Staff:

Mark Barsamian, Audio-Video Technician
Joseph DiGoia, Student Audio-Video Technician
Matt Kangas, Student Technician
Harsha Khattri, Student Technician
Luke Kaven, Systems Programmer/Analyst
Mark Kotik, Student Audio-Video Technician
Jason Laflamme, Audio Technician
Marge Leonard, Administrative Aide
Bruce Wootton, Computer Hardware Technician

Previous Issues

WPCPL No. 1 (December 1983, 90 pp.)..$5.00
papers by Mary Beckman, Susan R. Hertz, Osamu Fujimura, Atsuko Shoji, Stuart Millikin, and Katharine Davis

WPCPL No. 2 (April 1988, 176 pp.): Research in Laboratory Phonology..$8.00
G.N. Clements: The Role of the Sonority Cycle in Core Syllabification
Susan R. Hertz: The Delta Programming Language: an Integrated Approach to Nonlinear Phonology, Phonetics, and Speech Synthesis
John Kingston: Articulatory Binding

WPCPL No. 3 (June 1988, 109 pp.): Stress, Tone and Intonation.......$7.00
Susana Sainz: A Noncyclic Analysis of English Word Stress
Chi-lin Shih: Tone and Intonation in Mandarin

WPCPL No. 4 (December 1990, 262 pp.)..$9.00
Rukayyah S. Herzallah: Aspects of Palestinian Arabic Phonology: a Nonlinear Approach (1990 Ph.D. dissertation)

WPCPL No. 5 (September 1991, 123 pp.): Phonetic and Phonological Studies on Vowel Features... $9.00
Khathatso Evelyn Khabanyane: The Five Phonemic Vowel Heights of Southern Sotho: an Acoustic and Phonological Analysis
G.N. Clements: Vowel Height Assimilation in Bantu Languages
G.N. Clements: Place of Articulation in Consonants and Vowels: a Unified Theory

WPCPL No. 6 (October 1991, 356 pp. + xi)...................... ....$9.00
Elizabeth W.-Y. Leung: The Tonal Phonology of Llogoori: A Study of Llogoori Verbs (1986 M.A. dissertation)

This issue of the *Working Papers* was edited by A. Bradlow. Copies of this report are distributed free of charge to phonetics, phonology and speech laboratories interested in exchanging their own reports with us, and are otherwise available from us for the at-cost price of $9.00 per issue, surface postage included (inquire for air mail rates to addresses outside the U.S. or Canada).

Phonological Primes: Gestures or Features?*

George N. Clements

1. Introduction

In their studies of the articulatory patterns that underlie speech, Browman and Goldstein (see e.g. Browman and Goldstein 1989, 1990, this issue) have brought to light some of the important respects in which the physical activity of speaking contributes to the way phonological systems are structured. From this point of view, their work has several interesting implications. One is that much of what has usually been considered as lying in the domain of discrete phonological rules may be better understood from the point of view of nondiscrete or gradient properties of articulatory organization. Thus, in various papers Browman and Goldstein have produced evidence suggesting that many types of prosodically-conditioned reduction, contextual allophony and casual speech variation reflect dynamic properties of the activity of speaking and are better modeled at the level of physical speech production than at that of more abstract categorical representation. Somewhat more ambitiously (and controversially), they propose that the dynamic properties of speech production, often viewed by phonologists as having little interest for the study of grammatical organization, play a large or even predominant role in shaping the structure of what is interpreted: the phonological system of rules and representations itself.

In this issue of *Phonetica*, Browman and Goldstein propose that a gesture-based model of phonology and phonetics can provide an alternative to models taking segments and features as their basic units. Their explicit intention is to show that "gestures are basic units of contrast among lexical items as well as units of articulatory action" and to "help clarify the differences among gestures, features, and segments." This commentary will address both of these goals. It will examine several areas in which articulatory phonology as presently conceived by Browman and Goldstein appears insufficient to account for some of the generalizations that are usually thought to lie in the domain of phonological (and phonetic) theory, and will suggest ways in which a theory of this sort might be extended in order to accommodate these generalizations. It will finally consider the status of articulation-based models of phonetic interpretation in phonological theory as a whole.

---

* Editors note: This paper will appear in a forthcoming theme issue of *Phonetica* . All instances of "this issue" in the text and references refer to this issue of *Phonetica.*

2. On the notion "gesture"

As Browman and Goldstein describe them, gestures are abstract characterizations of the formation and release of local constrictions in the vocal tract. The constrictions themselves are defined in two ways. First, gestural scores display the temporal duration of individual gestures as well as the extent of their mutual overlap. Second, tract variables specify the location and degree of the constriction formed by the coordinated set of articulators that create it. In this conception, gestures are defined not in terms of articulator movements as such, but in terms of temporary, local constrictions of the vocal tract.

There is an important division of labor between the linguistic gestural model, whose function is to construct the score, and the task dynamic model, whose function is to execute it. However, Browman and Goldstein offer only brief and informal discussion of how these two models relate to each other. Their account of phonological rules and casual speech processes implies both that gestures can be rephased by the linguistic gestural model, affecting the structure of the score, and that parameter values can be assigned to tract variables by the task dynamic model, affecting the dynamic properties of gestures. Since adjustments at both of these levels can affect the nature of the output, one would like to know more about how they are allowed to interact, and how their interaction is constrained. In particular, assignment of parameter values (which eventually include stiffness and damping ratios, see Browman and Goldstein 1990) has the potential of bring two or more related gestures "out of synch" and thus of producing output inconsistent with the score. I discuss related problems concerning gestural coordination in section 4.

Since gestures not only drive the task dynamic model but are also primitives of the representational system, they play a role similar to that of features in more familiar phonological models. Specifically, Browman and Goldstein make use of five gestures, which are defined with respect to the lips, the tongue tip, the tongue body, the velum, and the glottis. Each of these is assigned to a separate tier, and may be present or absent in a given gestural score. Thus their function is entirely analogous to that of the articulator nodes labial, coronal, dorsal, nasal, and laryngeal in models such as that of Sagey (1986). In addition, the task-dynamic model supplies the dynamical parameter values which determine the degree, shape and location of the constriction formed by each set of articulators associated with a gesture. Although these values are not inherently categorical, they are said to function as such since their ranges are subject to the constraints provided by quantal articulatory-acoustic relations and/or adaptive dispersion principles. Elsewhere (see Browman and Goldstein 1989), they have been allowed to take on a small set of discrete values, forming a subset of the places and manners of articulation recognized by traditional phonetic description. Thus, the dynamical parameters provide information analogous to that provided by binary-valued features such as [continuant], [anterior], [distributed] and so forth. Finally, constriction degree values - corresponding roughly to features such as [continuant], [sonorant], and [consonantal] - are computed from local and global configurations of the vocal tract at any given point in time by the subsystem of tube geometry (Browman and Goldstein 1989).

This brief review should be sufficient to show that the gestures and parameter values proposed in Browman and Goldstein's articulatory phonology capture roughly the same set of contrasts defined by phonological feature theory. The next two sections will compare gestures with phonological features more closely. Section 3 will argue that as far as their intrinsic properties are concerned, gestures differ from features primarily, if not uniquely, in that they are allowed to take on gradient values, a fact which renders them less appropriate than features for the representation of lexical contrasts. Section 4 will offer reason to believe that gesture-based representations, or scores, must introduce a richer notion of hierarchical structure if they are to capture a full range of phonological regularities.

3. Intrinsic Properties of Gestures and Features

In most current versions of feature theory, including the one assumed here, features are defined in terms of acoustic and aerodynamic as well as articulatory properties. Thus, for example, [spread glottis] is defined both in terms of an acoustic effect (aspiration) and the articulatory means used to produce that effect (glottal opening). Of course, feature theories differ among themselves in terms of the set of features proposed, and the properties which are taken as definitional of any given feature.

It might appear that gestures differ from features in being defined in exclusively articulatory terms. However, this is not quite accurate. The term "gesture," in Browman and Goldstein's system, is not used in its ordinary-language sense of "movement of a body part to convey emphasis or meaning," but is used in a technical sense to refer to local constrictions in the vocal tract, as we have seen. In addition, by allowing gestural parameter values to be constrained by acoustic models such as quantal theory and dispersion theory (as well as by aerodynamic models such as tube geometry), Browman and Goldstein assign acoustic and aerodynamic considerations a potentially significant role in their model. This requires a reevaluation of their claim that gestures "do not correspond to features:" to the extent that gestures are defined or constrained in part by acoustic and aerodynamic considerations, the intrinsic difference between gestures and features, as general descriptive categories, is reduced. The major differences will lie primarily in the range of values that may be assigned to them at any given level of representation, and in the way they are organized in representations, or scores.

One fundamental respect in which gestures differ from features lies in the "quantitative variation in a gesture's dynamic parameters" that we observe under prosodic or other contextual conditions, discussed by Browman and Goldstein in some detail. Dynamic parameters, including duration,[1] may be assigned any value within their (nonbinary) range. Thus, for example, gestural shrinkage has the effect of scaling down the metrical properties of a gestural event, as in the gradient reduction of the glottal gesture that reduces or eliminates aspiration in speech output. From their discussion, it is clear that Browman and Goldstein do not regard gradient operations of this sort as constrained by quantal considerations, which presumably operate at the more abstract levels of system organization and lexical contrast. In contrast, features are usually defined as binary, or in some cases one-valued, and thus contrast categorically rather than quantitatively.

In this respect, features seem to provide a more adequate unit than gestures for expressing regularities at the abstract level of lexical contrast. Underlying lexical contrasts do not require the full range of gradient parameter values needed for the description of output regularities. Instead, gestural parameters of constriction degree and location, etc., regularly behave in a categorical fashion in lexical representations and early phonological rules, and durational information is usually reduced to a simple long vs. short contrast. To explain this observation, gesture-based models must restrict the parameters defining gestural events to effectively binary values; but this is exactly what is claimed by feature theory, in which only categorial distinctions (plus vs. minus specfications of binary features, presence vs. absence of privative features) are available.

Just the same considerations hold with respect to gestural phasing relations and overlap. To account for the diversity of ways in which gestures can be timed with respect to each other, Browman and Goldstein allow several types of phasing relations, varying according to where separate gestures are aligned with each other, as well as three degrees of overlap (minimal, partial, complete). But again, this range of choices projects to an excessively large number of theoretical lexical contrasts. As far as overlap is concerned, only two types appear to be required: full vs. partial. No further distinction, such as one between partial and minimal overlap, seems required for the expression of lexical contrasts. Thus, for example, tone languages do not distinguish between two types of contour tones differing only in the relative duration assigned to their first and second components, nor are any languages known to have phonemic contrasts between "partially" and "minimally" prenasalized stops. This restriction of contrasts in overlap to two is exactly what is predicted by current feature-based phonological theories, which allow features to be related to each other in either a one-to-one or a many-to-one relation (see e.g. Clements 1985, in press, Sagey 1986, McCarthy 1988). Thus the difference between plain nasal stops and prenasalized stops can be represented as follows (simplifying irrelevant detail):

[n]:	[^{n}d]:
root	root
|	/ \
[+nasal]	[+nasal] [-nasal]

In the first of these representations, the single feature [+nasal] is aligned, or associated, with the other features of the segment, represented by the root node. In the second, two sequenced features [+nasal], [-nasal] are so aligned. Since association is a discrete relation, this mode of representation allows no way of distinguishing two pairs of sequenced features in terms of differences in their mutual temporal overlap with other features, and thus predicts (correctly) that such distinctions are irrelevant at the level of lexical representation.[2]

A stronger case might be made for lexical contrasts involving phase relations, i.e. the relative *timing* of overlapping gestures. For example, a lexical distinction between pre- and post-aspirated stops could be expressed quite naturally in terms of the relative phasing of the glottal opening gesture with stop arrest or closure in the first case, stop release in the second. Another candidate discussed by Browman and Goldstein is the distinction between voiceless and voiced aspirated stops in languages like Hindi. They suggest that the difference between these two categories can be treated as a built-in difference in phasing relations, with the glottal gesture timed later in voiced aspirated stops than in voiceless ones.

However, the theory still predicts too large a range of possible lexical contrasts in phasing relations. Browman and Goldstein currently recognize three points at which consonantal gestures can be coordinated with other gestures: onset of movement toward the target, achievement of the target, onset of movement away from the target. Allowing that two gestures can be aligned with each other at any pair of these points, this allows up to nine ways in which two consonantal gestures can be coordinated in lexical representations; and as still further gestures overlap, the number of possible phasing relations among them increases exponentially. In contrast, current feature-based theories of phonology place strong inherent constraints on the ways in which any pair of features or nodes can be coordinated in lexical representation, as we have seen, and provide no way of representing differences in phasing relations in lexical representations.

If gestures are to be used to express lexical contrasts, then, strong constraints must be placed on the way gestures can be coordinated with each other. One direction currently being explored by some phoneticians working in feature-based frameworks involves introducing a level of discrete structural positions representing "articulatory landmarks" such as stop closure and release into the representational system, to which glottal features (and perhaps others) can associate (Huffman 1989, Keating 1990). The introduction of such landmarks into gestural phonology would result in a more constrained theory of gestural coordination, though it is still unclear whether they are required for the expression of lexical contrasts. A somewhat different, though not incompatible approach draws on the view that features are defined in part in terms of acoustic and aerodynamic goals, and that these goals must be effectively realized in the output, at least in the absence of further cues to their presence. Thus, glottal features would be defined in part in terms of goals such as vocal fold vibration, aspiration, etc., rather than uniquely in terms of the glottal constriction. In this view, the fact that glottal opening in aspirated stops is not completely overlapped by the stop closure would follow automatically from the requirement that the functional goals associated with features must be effectively realized, in this case forcing the alignment of glottal opening with the arrest or release phase of a stop consonant to produce audible aspiration (i.e. with an "acoustic landmark" in the sense of Stevens 1991, or a "transition" in the sense of Hertz 1991). Similarly, the lower amplitude and extra phasing lag of the glottal opening gesture in voiced aspirated stops as compared to voiceless ones can be explained by the fact that these properties are conducive to achieving the goal of glottal vibration during the period of stop closure (Davis 1991). These and other approaches are likely to receive continuing attention in feature theory, and might have useful implications for the development of gesture-based models.

4. The hierarchical organization of phonological units

By taking gestures as primitives, Browman and Goldstein are able to describe a large range of phenomena in terms of relations among successive and overlapping gestures (constrictions). However, since the location and degree of the constrictions do not occupy separate tiers in gestural scores, Browman and Goldstein's system strongly predicts that they will always be synchronized with each other and with the activity of the articulator whose movement they constrain. Here again the gesture-based model appears to make incorrect predictions.

In many languages, we find that certain segments assimilate place of articulation - a term which I use to designate the active articulator and its constriction location - from a neighboring segment, without assimilating its constriction degree. A common process of this type consists of the assimilation of nasals to following consonants. Thus in Yoruba, the syllabic nasal which forms the progressive aspect prefix assimilates in place to a following stop, fricative, or nasal, and optionally (or speaker-variably) to a liquid. Before nonassimilated liquids and the glides /w, j, h/, on the other hand, it is realized as syllabic [ŋ] (Ward 1952; R. Ṣonaiya, personal communication).

a. obstruents:			nasal place of articulation:
N + b:	m-bɛ	'be well'	bilabial
N + f:	ɱ-fɔ	'be washing'	labiodental
N + t:	n̪-tɛ	'be spreading'	dental-alveolar
N + d:	n-dɛ	'be setting a trap for'	alveolar
N + s:	n-se	'be cooking'	alveolar
N + ʃ:	ń-ʃe	'be doing'	alveo-palatal
N + ɟ:	ń-ɟa	'be fighting'	alveo-palatal
N + k:	ŋ-ka	'be reading'	velar
N + g:	ŋ-ge	'be cutting'	velar
N + kp:	ŋm-kpa	'be killing'	labio-velar
N + gb:	ŋm-gbona	'be getting hot'	labio-velar
b. sonorants:			
N + m:	mmu	'be drinking'	bilabial
N + n:	nna	'be beating'	alveolar
N + l:	nlɔ ~ ŋlɔ	'be going'	alveolar or velar
N + r:	nra ~ ŋra	'be buying'	alveolar or velar
N + j:	ŋjɔ	'be coming out'	velar
N + w:	ŋwa	'be coming'	velar
N + h:	ŋhɔ	'be scratching'	velar

([ŋm, kp, gb] are doubly articulated stops). Since the prefix's place of articulation is always predictable from the following consonant (the prefix does not occur before vowels), it should not be specified in its lexical representation. Its place of articulation is either assigned from the following consonant, or realized with a default velar constriction. In all cases, the nasal is realized with full closure.

A gestural interpretation of facts such as these must allow the constriction location of the following consonant to overlap the nasal gesture, but in the current model this cannot be done without simultaneously overlapping its constriction degree. In previous discussion of a similar case, Browman and Goldstein suggest that if the assimilated nasal is indeed realized with complete oral closure, the following consonant's oral gesture must change its constriction degree when it overlaps the velum lowering gesture (Browman and Goldstein 1989, 242). However, this proposal only displaces the problem, since if the overlapping gesture is produced with complete closure, the following consonant (which shares the gesture) should also be realized with complete closure. Similar cases of assimilated stop + continuant sequences can be cited from other languages, posing a genuine problem for Browman and Goldstein's characterization of gestures.

One may always question whether such descriptions are based on accurate observations. I know of no quantitative or instrumental study of nasal closure in nasal + consonant sequences in Yoruba. However, at least one such description of nasal clusters has been supported by instrumental analysis. Shona, as described by Doke (1931), has the homorganic nasal sequences [mb mv nd nz ɳʐ ɲdʒ ŋg], where the symbol [ɳ] designates a heavily rounded alveolar nasal appearing only before the voiced alveolar-labialized fricative [ʐ]. As far as the labial sequences are concerned, Doke states (p. 54):

> The bilabial nasal in Shona is formed just as in English, by complete contact of the lips, the air passing through the nose . . . Apart from the use of **m** immediately before vowels, it appears in the compounds **mb** (homorganic), **mv** (semi-homorganic), . . . Shona does not employ the denti-labial nasal **ɱ**, as do Zulu, Lamba, Bemba, etc., homorganically before **f** or **v**, but the full bilabial nasal in the combination **mv**.

Although the sequence [mv] is only "semi-homorganic" in Shona, we probably want to describe it as fully homorganic at the phonological level just as we do the other nasal clusters, and account for the fact that labial stops are predictably bilabial by adjusting the constriction location in the articulatory realization (without more information, of course, we cannot propose a definitive analysis). As far as the alveolar sequences [nz ɳʐ] are concerned, Doke does not explicitly state whether the nasals are produced with complete closure. However, he reproduces palatograms for these sequences from several Shona dialects, all of which show complete contact between the tip or blade of the tongue and the alveolar ridge. This closure must be attributed to the nasal, since (as other palatograms show) the fricative alone leaves an unobstructed passage through the center of the vocal tract. In Shona, then, nasals unambiguously assimilate constriction location but not constriction degree from a following continuant.[3]

There is also evidence that segments can assimilate constriction degree, but not constriction location. In Browman and Goldstein's model, vowel height is modelled in terms of constriction degree, which has the values [narrow], [mid], and [wide] (Browman and Goldstein 1989, 225-6), while place of articulation depends on the constriction location (palatal, velar, uvular, etc.) of the dorsal articulator. We would thus expect vowel height and place of articulation to assimilate as a single unit, never separately. In a number of languages, however, vowel height assimilates separately from place of articulation, and vice-versa (see Odden 1989 for a review of cases). The following examples are from Kimatuumbi, in which noninitial vowels in the stem assimilate to the height, but not the place of articulation of the initial vowel, provided both are nonlow. (Underlying noninitial vowels are represented below with upper-case letters, indicating their archisegmental status.)

underlying:	surface:	example (stem):	
i + I	i + i	-yipilya	'thatch with for'
i + U	i + u	-libulwa	'be ground'
u + I	u + i	-utika	'be pullable'
ɪ + U	ɪ + ʊ	-tɪkʊlya	'break with'
ʊ + I	ʊ + ɪ	-ʊʊgɪlwa	'be bathed'
ʊ + U	ʊ + ʊ	-kʊmbʊlya	'beat with'
ɛ + I	ɛ + ɛ	-chɛɛngɛya	'make build'
ɔ + I	ɔ + ɛ	-bɔɔlɛlwa	'be de-barked'
ɔ + U	ɔ + ɔ	-bɔmɔlwa	'be destroyed'

Once again, these facts are problematical for Browman and Goldstein's account of the gesture.

A solution to these problems can be found if place of articulation and constriction degree are allowed to occupy separate tiers of their own, where they can spread to other points in the representation independently of each other. This conception "unpacks" the notion of gesture without undermining it, since gestures can still be defined in terms of the lowest node superordinate to place of articulation and constriction degree. Suppose, to be specific, that we define consonantal and vocalic gestures in terms of the following configurations:

```
          cons                      voc
          /  \                      / \
    stricture \           stricture    \
             place                     place
```

The "cons(onantal)" and "voc(oidal)" nodes in these figures designate gestures, and features are assigned to the stricture and place nodes to characterize them in terms of constriction degree and place of articulation. "Cons" and "voc" are not abstract labels, but define the range of stricture values appropriate to consonants and vocoids, respectively, and thus function similarly to the feature [±consonantal]. They link to the higher-level root node, to which their values may percolate. It is assumed that all nodes in this figure are assigned to different tiers in a representation (or score), effectively segregating consonants and vowels.

The "place" node dominates the set of oral tract articulator features labial, coronal, dorsal, and perhaps radical, or their gestural counterparts (L, TT, TB, etc.). These features are further specified for constriction location and shape by features such as [anterior] and [distributed]. The "stricture" node dominates stricture features, or their gestural counterparts: [±continuant] in the case of consonants, and vowel height features in the case of vocoids. Vocoids might also be redundantly assigned [+continuant], if the evidence warrants it. Clements (1990) presents evidence that vowel height features may spread independently of each other, and must thus be arrayed on further independent tiers.

These representations have the properties we need to characterize gestures in a way compatible with Browman and Goldstein's general conception, while allowing stricture and place of articulation to spread or overlap independently of each other. However, they appear to require modification of Browman and Goldstein's framework in two respects. First, in their current presentation Browman and Goldstein do not recognize a separate tier for "place of articulation," as is required by the representations above. However, there is much evidence that phonological rules may target the full set of oral place (articulator) features as a whole, rather than targeting only individual features. Thus in the Yoruba examples above, the rule assigning place of articulation to the nasal prefix applies to all oral places of articulation without exception, and spreads both components of the doubly articulated stops [kp, gb]. This is expressed in feature frameworks by spreading the place node leftward, but cannot be directly expressed in a gesture-based framework which does not provide a place node.

Second, in their current presentation Browman and Goldstein do not explicitly allow units on different tiers to be linked by association lines or other devices for indicating their membership in higher-level units.[4] Indeed, they appear to take a skeptical view toward the recognition of higher-level groupings of gestures, or coordinative structures, stating that "the only hierarchical unit for which we currently have evidence is that of the oral gestures in a (syllable-initial) consonant cluster." But as we have just seen, there is reason to group the components of the doubly-articulated stops [kp, gb] of Yoruba into a hierarchical structure (the place node), since both components spread or overlap as a unit. And once we analyze gestures into separate place and constriction degree components, it is necessary to indicate the connection between these components in some fashion. In the feature-based representations proposed above, this is done by connecting them with association lines, although other devices can be imagined.

The use of such associations or connections generalizes to much of the other data discussed by Browman and Goldstein. For example, they analyze the casual speech deletion of schwa in words like beret in terms of the overlapping of the two gestural components of [r] (tongue tip constriction and tongue root constriction) with the labial gesture. But the reason the two gestures of [r] overlap or spread to the labial as a single unit is most likely that they characterize a single phonological segment, represented in current phonological models by association to the root node. If the two gestures are regarded as merely accidental constellations with no intrinsic connection between them, the fact that they behave as a single temporal unit goes unexplained. Browman and Goldstein have presented many examples of sets of gestures that function as temporal units, but in none of these cases do the coordinated gestures fail to constitute a segment under most types of phonological analysis. Thus it would appear that there is motivation for grouping gestures into higher-level hierarchical units such as the segment, even for the treatment of casual speech phenomena.

Similar evidence can be given for recognizing further hierarchical units, such as the mora, the syllable, and and the phonological word. In short, if articulatory phonology is to be a viable candidate for a general theory of phonology and phonetics, it must provide an explicit way of expressing hierarchical relations among phonological units.[5]

5. Discussion

We have made two general points. First, to the extent that gestures and features differ in their empirical properties, features seem more appropriate to express the discrete nature of lexical contrast: at this level, what we require is not a photograph of the vocal tract, but a rough map. Second, if gesture-based models are to be adequate to the expression of many phonological (and perhaps phonetic) generalizations, they require additional hierarchical organization of the sort postulated in current feature-based models of representation.

What, then, is the relation between gestural phonology and feature-based phonology? Are the two completely incompatible? I have suggested above that if the notion "gesture" is suitably revised along the lines suggested above, gesture-based phonologies will differ little if at all from feature-based phonologies at the more abstract and systematic levels of phonological representation at which units behave in a discrete fashion. The apparent incompatibility between the two modes of representation can perhaps be resolved in terms of Browman and Goldstein's further observation that "increase in overlap among gestures in fluent speech is a general gradient process that can produce *apparent (perceived) discrete alternations*" (my italics). What this remark suggests is that speech is produced in a gradient fashion, but perceived (and thus represented) categorically. If this view is correct (and there is much evidence that it is), then the status of gesture-based models in a global theory of phonology and phonetics may be clarified. Gesture-based models (and others with similar goals) are models of speech production, which address the complex and important problem of "how to bridge the gap between the discrete segments of the phonological system and the fluid change in time and space that is the final result of phonetic processes," to quote Huffman's apt characterization (1989, 139).

In sum, Browman and Goldstein have developed an elegant and comprehensive theory of articulatory structure and its relation to phonological structure which can serve as a basis for proposing and testing specific hypotheses about how abstract phonological representations are related to physical phonetic output. Indeed, it has already proven its value in this respect. On the other hand, it has not successfully handled certain aspects of phonological patterning in which current feature theories seem to offer a closer fit to linguistic reality. To the extent that the future development of their approach is able to address problems of this sort successfully, it will take a step toward relieving phonological feature theory of the task of accounting for phenomena that lie outside its proper domain, and help lay the basis for a better definition of the way in which abstract phonological structure is transmitted through the medium of speech.

Notes

[1] The status of duration in Browman and Goldstein's current model is not altogether clear. On the one hand, they state that "quantitative temporal information is provided not by specifying time directly but by specifying the parameters of the gestural regimes and their phasing" (Browman and Goldstein 1990, 310). On the other, the present paper states that a gestural score "displys the duration of the individual gestures as well as the overlap among the gestures." If the gestural score constitutes the input to the task dynamic model which contributes to determining temporal information, as Browman and Goldstein propose, temporal information should not be available to the score. For purposes of the present discussion I take their more recent statement as the definitive one.

[2] See Steriade (1990) for further discussion.

[3] An exceptional case was noted for a speaker of the Ndau dialect, for whom "when **nz** was initial, the **n** did not effect complete contact, there being a space showing that in reality it was nasalized **z** (z̃) which was produced. Complete contact was effected when not initial, as in **hanzu**" (Doke 1931, 265). However, the same speaker realized /tʃ/ as [ʃ] word-initially, and incomplete contact was found before the (similarly fricativized) realization of word-initial /ɲdʒ/, suggesting that this speaker used incomplete closure quite generally in initial position.

[4] They have elsewhere suggested, however, that association lines might be used to express phasing relations between overlapping <u>successive</u> segments (Browman and Goldstein 1989).

[5] Browman and Goldstein have elsewhere (1989) tentatively proposed that autosegmental and prosodic structure can be interfaced with gestural scores on distinct planes, but this suggestion is not taken up in the present paper.

References

Browman, C.P. and L. Goldstein (1989) "Articulatory Gestures as Phonological Units," Phonology 6.2, 201-51.

Browman, C.P. and L. Goldstein (1990) "Gestural Specification using Dynamically-defined Articulatory Structures, Journal of Phonetics 18.3, 299-320.

Browman, C.P. and L. Goldstein (this issue) "Articulatory Phonology: an Overview," Phonetica.

Clements, G.N. (1985) "The Geometry of Phonological Features," Phonology Yearbook 2, 225-252.

Clements, G.N. (1990) "The Hierarchical Representation of Vowel Height," unpublished ms. Ithaca, N.Y.: Cornell University.

Clements, G.N. (in press) "Place of Articulation in Consonants and Vowels: a Unified Theory," to appear in B. Laks and A. Rialland, eds., L'Architecture des Représentations Phonologiques, Paris: Editions du C.N.R.S.

Davis, K. (1991) The Acquisition of the Voicing Contrast by Speakers of English and Hindi, unpublished Ph.D. dissertation. Ithaca, N.Y.: Cornell University.

Doke, C.M. (1931) A Comparative Study in Shona Phonetics. Johannesburg: University of the Witwatersrand Press.

Hertz, S.R. (1991) "Streams, Phones, and Transitions: Toward a New Phonological and Phonetic Model of Formant Timing," Journal of Phonetics 19.

Huffman, M. (1989) Implementation of Nasal: Timing and Articulatory Landmarks, Ph.D. dissertation (also UCLA Working Papers in Phonetics 75), Los Angeles: Dept. of Linguistics, UCLA.

Keating, P.A. (1990) "Phonetic Representations in a Generative Grammar," Journal of Phonetics 18.3, 321-34.

McCarthy, J. (1988) "Feature Geometry and Dependency: a Review," Phonetica 45,84-108.

Odden, D. (1989) "Vowel Geometry," unpublished ms. Columbus, Ohio: the Ohio State University.

Sagey, E. (1986) The Representation of Features and Relations in Nonlinear Phonology, Ph.D. dissertation. Cambridge, Ma.: MIT.

Steriade, D. (1990) "Gestures and Autosegments: Comments on Browman and Goldstein's 'Gestures in Articulatory Phonology'," in J. Kingston and M.E. Beckman, eds., Papers in Laboratory Phonology I: Between the Grammar and the Physics of Speech, pp. 382-97. Cambridge: Cambridge University Press.

Stevens, K.N. (1991) "Phonetic Evidence for Hierarchies of Features," paper presented at the Third Conference on Laboratory Phonology, Los Angeles: U.C.L.A.

Ward, Ida C. (1952) An Introduction to the Yoruba Language. Cambridge: Cambridge University Press

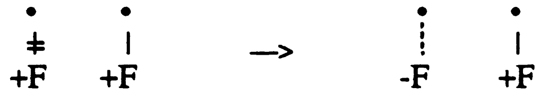
The Consequences of Dissimilation in Sundanese*

Abigail C. Cohn

1. Introduction

Studies of phonological assimilation have played a central role in the development of current phonological theory. As widely discussed in the literature, assimilation is an extremely common phonological process cross-linguistically and therefore an adequate phonological theory should represent it simply and naturally. This has led to the current view of assimilation as spreading (Goldsmith 1976, Clements 1976, Hayes 1986, among others). Much less work has addressed itself to the issue of dissimilation, but recently it has been suggested that dissimilation should be analyzed as delinking followed by default fill-in (Odden 1987, Poser 1987, McCarthy 1988, Yip 1988). This approach is schematized in (1).

(1) Dissimilation as delinking



Missing value filled in by default: [ØF] —> [-F]

In the case of two tier adjacent identical feature specifications, one of the specifications is delinked. The missing value is then filled in by a default rule. Dissimilation, thus consists of two independent processes, delinking and default fill-in.

It has been argued that dissimilation is motivated by the Obligatory Contour Principle (OCP), the principle that adjacent identical elements are prohibited. (McCarthy 1985, 1988, Kaisse 1988, Yip 1988). Yip (1988, p. 73), following earlier work by McCarthy, takes the position that "all rules involving identity of target and trigger with an output in which they are no longer identical and adjacent are OCP-triggered rules." It is thus an OCP violation which requires delinking of one of the offending specifications in a sequence of identical feature specifications. When this is followed by default fill-in, this yields dissimilation. "Another possible result of deleting a feature matrix occurs if deletion is

---

*Parts of this work were presented at the Cornell University DMLL Colloquium, December 1990 and the 65th meeting of the Linguistic Society of America, Chicago Illinois, January 1991. Thanks to both audiences for stimulating discussion. Thanks also to Nick Clements, Beverley Goodman, and Bruce Hayes for helpful comments.

followed by application of redundancy rules that insert the opposite value of the deleted feature(s). . . . This of course is dissimilation, and such rules are widely found in natural languages." (Yip 1988, p. 80).

Yip discusses what kind of language data would falsify the claim that such delinking is indeed motivated by the OCP:

> The kind of case that would, I think, require a weakening of this claim would be a language with the following properties:
>
> (i) Dissimilation of F: αF —> -αF / __ αF
>
> (ii) Demonstrable morpheme-internal αFαF sequences, as opposed to doubly linked αF
>
> (ii) would show that the OCP did not operate on αFαF sequences. It thus could not act to trigger a rule like (i). (Yip 1988, p. 73).

Turning Yip's prediction around, if dissimilation is motivated by the OCP, then other aspects of the phonology, as well as underlying phonotactic patterns, should display the same restrictions. Thus we need to examine rules of dissimilation in the broader context of the lexical structure and phonology of the languages in which such rules obtain. (See Goodman this volume for a similar argument.)

Sundanese (an Austronesian language, spoken in West Java, Indonesia) displays a case of dissimilation in the form of the plural marker. At first blush, the factors conditioning the shape of this formative appear to violate the OCP; but, upon closer inspection, we will see that the constraints motivating the dissimilation hold more generally in the lexical and phonological representations of Sundanese and thus Yip's prediction is borne out.

In Sundanese, a formative *=ar=* or *=al=* marks the plural, as exemplified in (2), where = = indicates infixation.

(2)	a.	kusut	ar	—>	k=ar=usut
		'messy'	'pl.'		'messy, pl.'
	b.	dahar	al	—>	d=al=ahar
		'eat'	'pl.'		'eat, pl.'
	c.	visualisasi			
		'visualize'			
		di-visualisasi-kɨn	ar	—>	di-v=ar=isualisasi-kɨn
		pass-visualize-v.s.	'pl.'		'visualized, pl.'

As exemplified in (2a & b), the plural marker, either *=ar=* or *=al=*, is usually infixed after the first consonant of the root. The process is highly productive, as exemplified in (1c) by the pluralization of a recent borrowing such as *visualisasi*. Verbs, adjectives, and a few nouns exhibit such infixed forms.[1] The observed pattern of allomorphic alternation is triggered by both assimilation and dissimilation. As this morphophonemic process is productive and regular, it should be accounted for by rule. The pattern of infixation also merits attention, as it is directly related to an adequate account of the allomorphic alternation.

The structure of the paper will be as follows. In §2, we consider the process of infixation. In §3, the pattern of allomorphy is presented and it is shown that the allomorphy results from both assimilation and dissimilation. An analysis is proposed. Yet, on the face of it, the analysis of dissimilation is not motivated by the OCP as predicted and it poses questions regarding the issues of markedness and underspecification. We address these issues in §4 and conclude in §5. The data discussed here are from work with two native Sundanese speakers from Bandung, the capital of Sunda (West Java), where what is considered to be the standard dialect is spoken. This process has also been briefly described in the literature on Sundanese (see Eringa 1949 and Robins 1959).

2. Infixation

Sundanese has a rich system of affixation, including prefixes, suffixes and infixes. Both prefixes and suffixes are of course extremely common cross-linguistically and are fairly straightforward to account for formally. True infixes, on the other hand, are much rarer and pose problems in terms of their structural representation. Sundanese has three infixes, all of the shape VC (V = vowel, C = consonant): *=ar=/=al=*, *=in=*, *=um=*. Of these, only the plural marker is productive and we therefore focus our attention on it. Noteworthy is the difference between CV and VC affixes at the beginning of a root in Sundanese, as exemplified in (3):

---

[1] Although the infix is typically referred to as a plural marker, Ewing (1991) has shown that the formative actually creates distributive forms. I will nevertheless continue to refer to it as a plural marker. The reader is referred there for discussion of both the semantics and pragmatics of such forms.

(3)a.	CV affixes at the beginning of a word				
i.	CV + V initial root = prefix	di 'passive'	atur 'arrange'	—>	di-atur 'arranged'
ii.	CV + C initial root = prefix	di 'passive'	dahar 'eat'	—>	di-dahar 'eaten'
b.	VC affixes at the beginning of a word				
i.	VC + V initial root = prefix	ar 'pl.'	anjɨn 'you'	—>	ar-anjɨn 'you, pl.'
		ar 'pl.'	ayɨm 'patient'	—>	ar-ayɨm 'patient, pl.'
ii.	VC + C initial root = infix	ar 'pl.'	damaŋ 'well (adj.)'	—>	d=ar=amaŋ 'well, pl.'
		ar 'pl.'	poho 'forget'	—>	p=ar=oho 'forget, pl.'

In (3) we observe that a CV affix at the beginning of a word is always prefixed, whether the root starts with a vowel (3ai) or a consonant (3aii). In contrast, a VC affix at the beginning of a word is prefixed with a vowel initial root (3bi), but infixed after the root initial consonant with a consonant initial root.[2] A generalization emerges: The placement of infixes in Sundanese is prosodically conditioned and has the net effect of maximizing open (CV) syllables and avoiding unallowable CC sequences. (Anderson 1972 makes a similar observation.) An adequate formal account of infixation in Sundanese needs to incorporate this generalization.

McCarthy and Prince (1986) argue that infixation should be analyzed as melodic extraprosodicity. Following this view, they sketch out an analysis of infixation in Sundanese, whereby the initial C is extraprosodic; the infix is really a prefix; and the "Onset Rule", which resyllabifies a consonant to become a syllable onset, applies twice. Their analysis is exemplified in (4) (following McCarthy and Prince 1986, p. 48).

```
(4) a.  sg.      pl.
        [niʔis]  [nariʔis]
        'to cool oneself'

    b.  σ                  c.  a r
        |\                     |  \
        ar                     σ  σ    σ
                               \  |   /|\
                               (n) i  ʔ i s      [nariʔis]
```

---

[2] Root initial consonant clusters are very rare and I have yet to find such a root from which a plural form can be constructed.

The infix is proposed to have the shape in (4b). (Note that the proposed representation seems to assume underlying syllabification.) The root initial /n/ is extraprosodic, and the Onset Rule applies twice, yielding the derived representation in (4c). Basic to this analysis is Planar Segregation (see McCarthy 1989) which assumes that separate morphemes constitute independent phonological planes. In support of their analysis, McCarthy and Prince (1986) claim that "The representation of the affixal melody on a separate tier is independently required by Sundanese nasal harmony. . . " (p. 48).[3]

There are a number of problems with this view. (1) The basic insight that the infix is located prosodically, not melodically is missed. The pattern of infixation has to do with syllable structure, not melodic structure as implied by McCarthy and Prince's view of melodic extraprosodicity. All consonant initial verbal and adjectival roots in the language would need to have an extraprosodic initial consonant. Clearly a generalization is being missed.[4] (2) A widely held view is that there can be no phonological interaction between separate planes, before Plane Conflation (see McCarthy 1989 and Yip 1988 for discussion). Yet several phonological rules of Sundanese refer to both the infix and the rest of the form, including nasal harmony. These processes include both lexical and post-lexical ones. This would require Plane Conflation before the relevant phonology, but there is no evidence that the intervening unconflated stage exists in the derivation. (3) The facts of nasal harmony do not constitute an argument in favor of the plural infix being represented on a separate tier (plane). As argued by Cohn (1990), the facts of Sundanese nasal harmony can be accounted for in a straightforward manner by a cyclic analysis within the framework of Lexical Phonology. Nasal harmony is shown to be a lexical rule, applying both before and after infixation. Such an analysis accounts directly for the apparent overapplication of the rule.

I propose, instead, a single plane analysis of Sundanese infixation. The infixes in Sundanese are basically like prefixes, except that they are located before the first mora of the root. Infixation results in reapplication of the rules of syllabification. In order to see how the analysis works, we need to establish explicit principles of syllabification. Following Hayes (1989), I will assume the following syllabification algorithm:

---

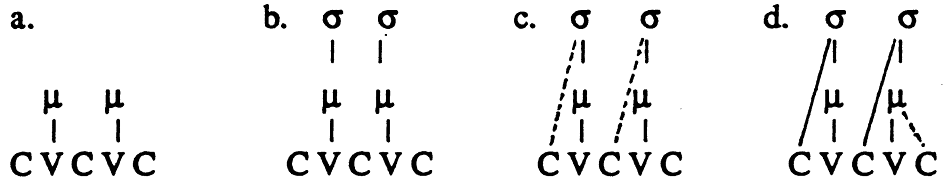
[3] I understand the use of *tier* here to the same as *plane* as discussed by McCarthy (1989).

[4] See also Anderson (1991) for a similar criticism of McCarthy and Prince's view of infixation.

(5) Syllabification (following Hayes 1989)

Basic syllable structure in Sundanese is (C) V (C)

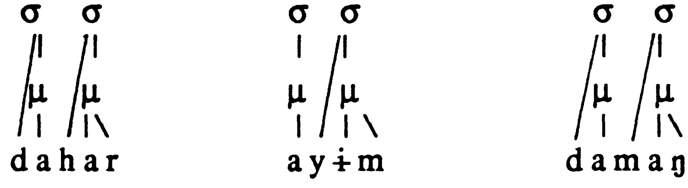
a. Assign a mora to each vocalic element.
b. Each mora is dominated by a syllable.
c. Associate a preceding consonant directly to the syllable node (as "onset").
d. Associate a following consonant to the preceding mora (into the "rime").



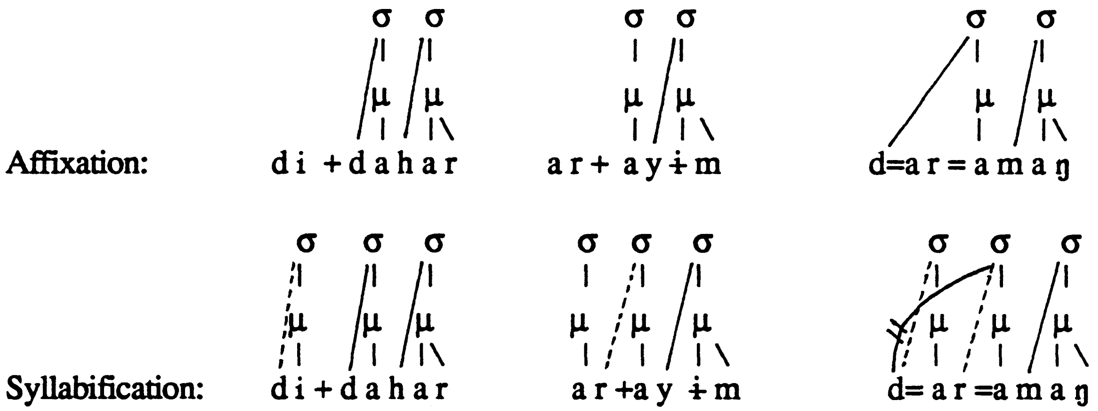
Following Hyman (1985) and McCarthy and Prince (1986), the mora is argued to be the basic unit of phonological weight, basic to the representation of stress, tone, and the syllable. As widely noted, certain aspects of syllabification are universal and others language specific. In a language with no underlying moras, a mora is assigned to each vocalic element by rule (5a). Sundanese allows only one mora per syllable, as there are no long vowels and coda consonants do not affect the weight count, thus each mora is dominated by a syllable (5b). Hayes (1989) captures the fact that onsets (typically) do not contribute to phonological weight, by associating onset consonants directly to the syllable node (5c). Finally in a language where coda consonants do not contribute weight, an unsyllabified consonant following a vocalic element is associated to the preceding mora (5d).

Following this view of syllabification, the plural marker of Sundanese can be correctly positioned, occurring initially with a vowel initial root or following the first consonant of a consonant initial root, as shown in the derivations in (6).

(6) Sample derivations	a.	b.	c.
UR:	dahar	ayɨm	damaŋ



Syllabification:



Output: [didahar] [arayɨm] [daramaŋ]
$ = syllable division

Assuming that syllabification is either an "everywhere process" or applies cyclically, the root is syllabified at the beginning of the derivation, as shown in each of the three examples. Affixation then occurs, triggering a reapplication of the syllabification algorithm. In (6a), *di-* is prefixed and syllabified as an additional syllable. In (6b) *=ar=* is positioned before the first mora of the root. The vowel is assigned a mora and a syllable, while the consonant is syllabified as the onset of the first syllable of the root. In (6c) *=ar=* is again located before the first mora of the root, necessarily detaching /d/ from the first root syllable. The vowel is again assigned both a mora and a syllable and then Rule 5c is applied twice, making /d/ the onset of the first syllable of the word and /r/ the onset of the first syllable of the root. This is parallel to McCarthy and Prince's application of the Onset Rule, but this approach differs in the fact that no underlying syllabic structure is assumed for the infix and all aspects of its syllabification follow from the general rules of syllabification in the language. The fact that *=ar=/=al=* is located before the first mora of the root might be specified as part of the lexical entry of the affix, or it might be specified as part of the more general prosodic and phonotactic patterns for VC root initial affixes in Sundanese.

We see then that the simple assumption that infixes are located before the first mora of the root in Sundanese, together with independently motivated principles of syllabification, account directly for the observed location of the plural marker in Sundanese. If we refer directly to the prosodic structure, there is no motivation for representing the infix in Sundanese on a separate plane. This accounts directly for the morphological domains in which phonological rules are observed to apply in Sundanese and avoids an unmotivated stage of the derivation before Plane Conflation. The placement of the infix in Sundanese lends support to Hayes' (1989) claim that onset consonants link directly to the syllable node and not to the initial mora.

3. Allomorphy of the plural marker

As exemplified above in (2a & b), the plural marker may take the shape of *=ar=* or *=al=*. This allomorphic alternation is a systematic one and can be characterized in terms of the canonical root pattern of Sundanese. The consonant and vowel inventories of Sundanese are presented in (7a & b) respectively and the canonical root patterns are schematized in (8).

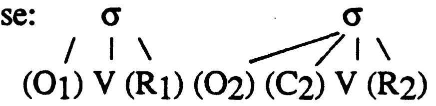
(7) a. Sundanese consonant inventory

p	t	c	k		
b	d	j	g		
m	n	ɲ	ŋ		
		s[5]			
	l/r				
w		y		h	(ʔ)

b. Sundanese vowel inventory

	front		back
high	i	ɨ	u
mid	e	ə	o
low		a	

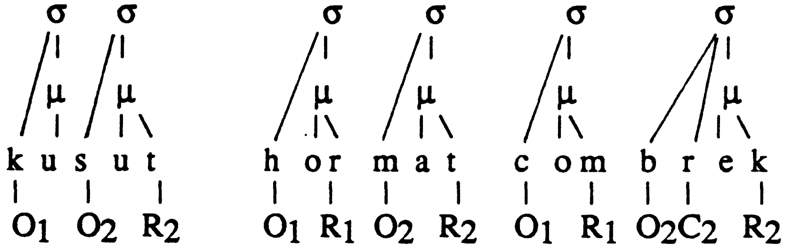
(8) a. Canonical root pattern of Sundanese:



O = onset
R = rime

O_1, O_2 = any consonant
R_1 = nasal homorganic to the following stop, /r/ and a few others marginally
R_2 = most consonants, except palatal [-continuant] consonants
C_2 = /r/, /l/ after a stop (quite rare)

b.



As schematized in (8a), most roots are disyllabic, with each syllable consisting minimally of a vowel. Both syllables have an optional onset (O_1, O_2), typically consisting

---

[5] /s/ patterns as a palatal in Sundanese.

of a single consonant; although in the second syllable, a stop may be followed by a liquid (C_2). Both syllables may have an optional consonant in the rime (R_1, R_2). These root patterns are exemplified in (8b).

As shown in the chart in (9), the allomorphic alternation between *=ar=* and *=al=* is conditioned by the presence of either /r/ or /l/ in the root. In all other cases, the *=ar=* variant is used.

(9)	/r/		/l/	
				a.
O_1	rVCVC	r=ar=VCVC	lVCVC	l=al=VCVC
	rahɨt 'wounded'	r=ar=ahɨt	lɨtik 'little'	l=al=ɨtik
	rɨwat 'startled'	r=ar=ɨwat	ləga 'wide	l=al=əga
		b.		
R_1	CVrCVC	C=al=VrCVC	CVlCVC	
	hormat 'respect'	ŋa-h=al=ormat	not attested (so far)	predicted:
	pərceka 'handsome'	p=al=ərceka		C=ar=VlCVC
O_2	CV(C)rVC	C=ar=V(C)rVC	CV(C)lVC	C=ar=V(C)lVC
	di-kirim 'sent (pass.)'	di-k=ar=irim	gɨlis 'beautiful'	g=ar=ɨlis
	curiga 'suspicious'	c=ar=uriga	ŋuliat 'stretch'	ŋ=ar=uliat
		c.		
C_2	CV(C)CrVC	C=al=V(C)CrVC	CV(C)ClVC	C=ar=V(C)ClVC
	combrek 'cold'	c=al=ombrek	ŋajləŋ 'jump'	ŋ=ar=ajləŋ
	motret 'take a picture'	m=al=otret	ŋoplok 'flop down'	ŋ=ar=oplok
		d.		
R_2	CVCVr	C=al=VCVr	CVCVl	C=ar=VCVl
	bocor 'leaking'	b=al=ocor	mahal 'expensive'	m=ar=ahal
	bɨŋhar 'rich'	b=al=ɨŋhar	gətol 'diligent'	g=ar=ətol
		e.		
O_3	CVCVrV	C=al=VCVrV		
	ŋumbara 'go abroad'	ŋ=al=umbara		
	siduru 'sit by a fire'	s=al=iduru		

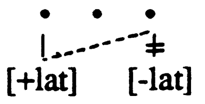
Consider first roots with /l/'s in them. If an /l/ occurs anywhere in the root, except root initially, the infixed form appears with the *=ar=* variant. But in the case of a root initial /l/, as seen in box a, the infixed form appears with the *=al=* variant. In the case of roots with /r/, if either of the first two syllables starts with an /r/ (O_1 or O_2), the infixed variant appears with the *=ar=* variant. But if /r/ occurs in the rime of the first syllable or second syllable (boxes b & d respectively), or as the second member of a complex onset to a second syllable (box c), or as the onset of a third syllable (box e); then the infixed form occurs with the *=al=* variant. The patterns observed for the two speakers in the present study concur with the observations made by Robins (1959, p. 343): "The variants *-ar-*/*-al-* are contextually determined; *-al-* is used with forms whose initial consonant is *l*, and with those containing a following *r*, except as initial consonant of the second syllable. Words of any other structure regularly infix *-ar-*. . . ."

With /l/ initial roots, we have a case of assimilation, that is, the infix assimilates to the initial /l/ (box a). Whereas, when an /r/ occurs in the root, except as the beginning of the first or second syllable, there is dissimilation between the two /r/'s (boxes b-e). I will first propose an analysis to account for these facts, then in the next section (§4), I will turn to the question of why these rules should obtain.

In order to account for the observed pattern of allomorphy, I assume that the underlying form of the infix is *=ar=* and propose two rules to account for the assimilation and dissimilation: Lateral Assimilation and /r/ Dissimilation respectively. I assume that only the liquids /l, r/ are specified for the feature Lateral in Sundanese.[6]

Following the view of assimilation as spreading, I propose a rule of Lateral Assimilation to account for the assimilation of the /r/ of the infix to a root initial /l/:

(10) Lateral Assimilation: r —> l / l = V __ ≐



applies to /r/ of the plural marker

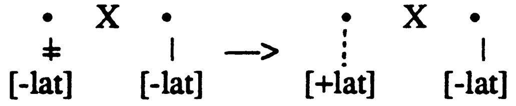
When the /r/ of the infix is preceded by an /l/, the [+lateral] specification of the /l/ spreads to the right, with concomitant delinking of [-lateral].

As discussed in §1, dissimilation has been argued to consist of delinking of two tier adjacent identical feature specifications, followed by default fill-in of the deleted feature specification. Following this approach, /r/ Dissimilation can be represented as follows.

---

[6] /l, r/ also differ in their specifications for Continuant (see Cohn 1990). I assume that appropriate Continuant specifications result automatically from a change in specification for the feature Lateral.

(11) /r/ Dissimilation: r —> l / =a __ = X r where X = intervening segmental material



applies to /r/ of the plural marker

Condition: except if the trigger is the onset of the second syllable

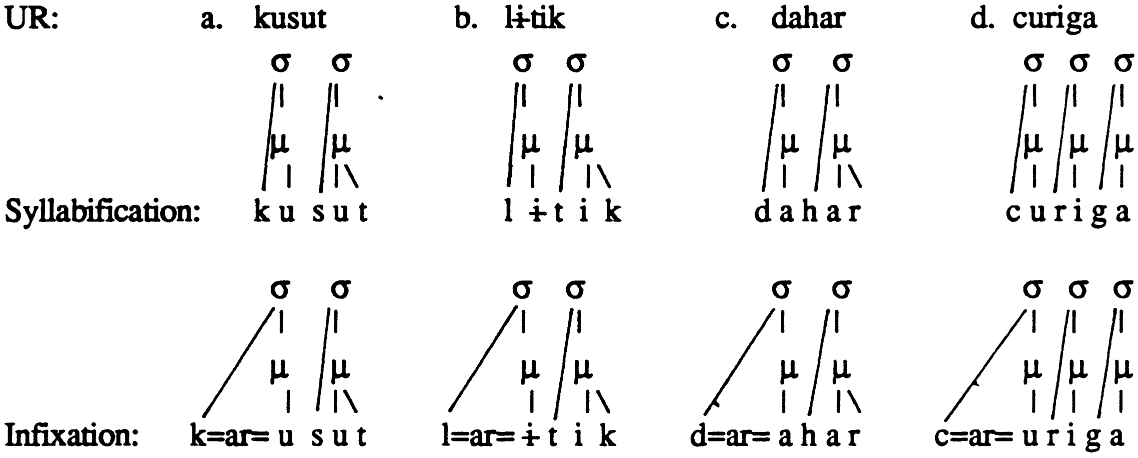
a. The first [-lateral] is delinked (due to tier adjacent [-lateral] specifications)

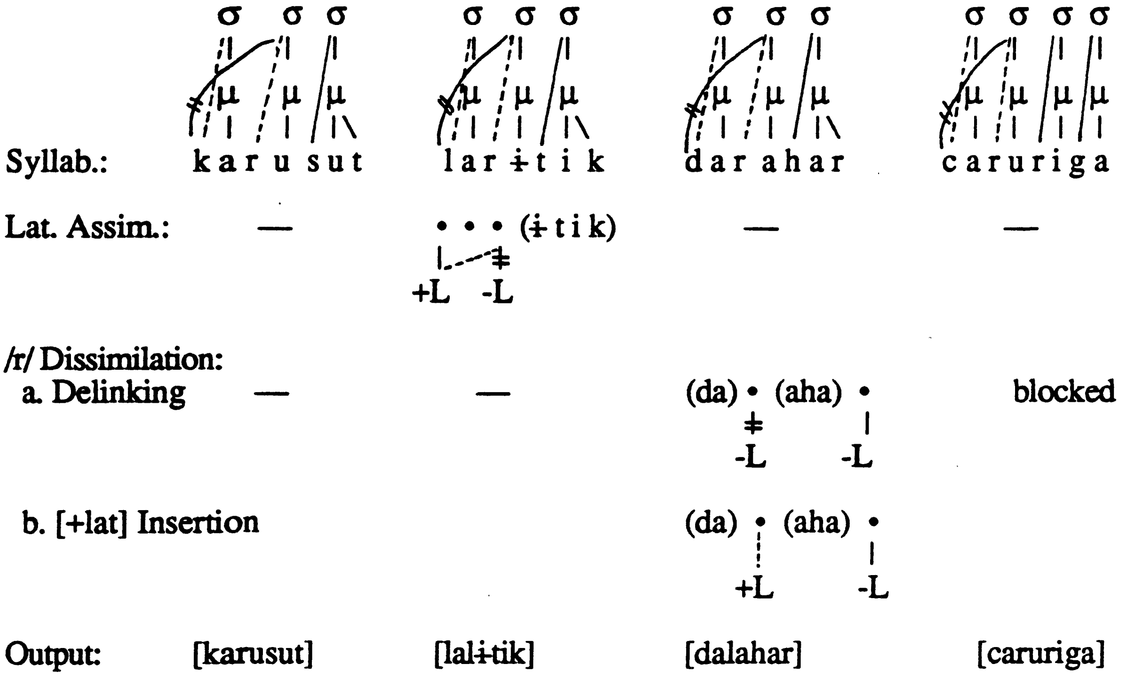
b. $\begin{bmatrix}+\text{cons}\\ +\text{son}\\ -\text{nasal}\\ \varnothing\text{lateral}\end{bmatrix} \varnothing \text{—>} [+\text{lateral}]$

When the /r/ of the plural infix is followed by a tier adjacent [-lateral] specification, the first lateral specification is delinked. However Delinking is blocked if the second [-lateral] specification occurs as the onset of the second syllable. Stated this way, the rule does not apply if the /r/ of the infix is preceded by an /r/, an issue that we return to below. Delinking then is followed by a rule filling in a [+lateral] specification for liquids unspecified for Lateral. Note, however, that the filling-in of [+lateral] is not a general default rule in Sundanese, as /r/ appears to be less marked than /l/, another issue we return to below.

Sample derivations of both Lateral Assimilation and /r/ Dissimilation are presented in (12).

(12) Sample derivations





In each case, the root is first syllabified, following the principles of syllabification discussion above in §2. The plural marker *=ar=* is then inserted before the first mora of the root. This results in resyllabification as shown. In (12a), as there are no /r/'s or /l/'s in the root, neither Lateral Assimilation or /r/ Dissimilation is applicable, resulting in the form [karusut]. In (12b), the root initial consonant is an /l/, which triggers the application of Lateral Assimilation. The [+lateral] specification of the root initial /l/ spreads to the /r/ of the infix, triggering delinking of the [-lateral] specification, resulting in the form [lalɨtik]. In (12c), Lateral Assimilation is not applicable; but /r/ Dissimilation is, due to the root final /r/. The [-lateral] specification of the /r/ of the infix is delinked and a [+lateral] specification is inserted, giving the form [dalahar]. In (12d), an /r/ occurs in the root, but the /r/ is the onset of the second syllable, so /r/ Dissimilation is blocked, yielding the form [caruriga]. We see, then, that the two proposed rules, Lateral Assimilation and /r/ Dissimilation, account for the observed allomorphy of the plural marker in Sundanese. But some questions remain; we address these in the next section.

4. Problems and issues

Although the allomorphy of *=ar=/=al=* is accounted for straightforwardly under the proposed analysis, some basic issues remain. (1) How does this view of dissimilation relate to questions of markedness and underspecification? (2) How general are the rules of /r/ Dissimilation and Lateral Assimilation? Do they apply only to the /r/ of the plural marker? (3) Why tolerate closer /r/'s, but not ones farther away? We address the first question in §4.1 and the remaining questions in §4.2.

4.1. Dissimilation: markedness and underspecification

Adequate formal accounts of dissimilation need to address issues of markedness and underspecification. In order to consider the nature of these issues, we turn to liquid dissimilation in Latin. As discussed by Steriade (1987),

> . . . the adjectival suffix /-alis/ (as in /nav-alis/ 'naval') becomes /-aris/ when preceded by a stem containing /l/: /sol-aris/ 'solar', /milit-aris/ 'military', /Lati-aris/ 'of Latium'. Dissimilation fails only when the stem /l/ is separated by the suffix by an intervening /r/: /flor-alis/ 'floral', /sepulchr-alis/ 'funereal', /litor-alis/ 'of the shore'. (p. 351)

These facts can only be accounted for if both + and - specifications are present within the class of liquids at the time that the rule applies, since an intervening /r/ blocks dissimilation between two /l/'s. As argued by Steriade, these facts can be accounted for directly under the view of Contrastive Underspecification, whereby both values of a feature are specified within a contrasting class (in this case the class of liquids) and no value is specified otherwise, as exemplified in the following derivation:[7]

```
(13) Phono Input:            m i l i t - a l i s          l i t o r - a l i s
                                 |       |                |     |   |
                                +L      +L               +L    -L  +L

     Liquid Dissimilation:   m i l i t - a l i s                 —
     a. Delinking                |       ‡
                                +L      +L

     b. Default fill-in      m i l i t - a r i s          l i t o r - a l i s
        [   ] —> [-L]        |   | |     |  |             | |  |     |  |
                            -L +L-L    -L  -L            +L-L -L    +L -L

     Output:                 [militaris]                  [litoralis]
```

Following a view of Radical Underspecification (Kiparsky 1982, Archangeli and Pulleyblank 1986, among others), whereby only one value is specified underlyingly, a default fill-in rule must apply before dissimilation to block the application of dissimilation in cases such as [litoralis]. But the rule must only provide default values to /r/, not to any other consonants, since only an intervening /r/, not any other intervening consonant, blocks

---

[7] Steriade does not provide explicit derivations of dissimilation in Latin, but her discussion is consistent with the view of dissimilation presented here.

the application of the rule. Since delinking is followed by default fill-in, then separate default fill-in rules must apply before and after delinking as exemplified in (14).

(14) Phono Input:	m i l i t - a l i s | | +L +L	l i t o r - a l i s | | +L +L
Default fill-in: $\begin{bmatrix}+\text{cons}\\ +\text{son}\\ -\text{nasal}\end{bmatrix} \longrightarrow [-L]$	—	l i t o r - a l i s | | | +L -L +L
Liquid Dissimilation: a. Delinking	m i l i t - a l i s | ǂ +L +L	—
b. Default fill-in [ ] —> [-L]	m i l i t - a r i s | | | | | -L +L-L -L -L	l i t o r - a l i s | | | | | +L-L -L +L -L
Output:	[militaris]	[litoralis]

As argued by Steriade (1987), these facts constitute a strong argument in favor of Contrastive Underspecification; since there is no independent evidence of the stage before the first default rule applies.[8]

As observed by Odden (1987, p. 237), following the assumption that "default rules introduce unmarked feature values. . . . This entails that the result of dissimilation should be relatively less marked." For the feature Lateral, [+lateral] is usually assumed to be the marked value (as are + values in general) and thus liquid dissimilation in Latin is consistent with this prediction.

The facts of dissimilation in Sundanese are quite similar to those of liquid dissimilation in Latin, except that in Sundanese, the dissimilation is triggered by /r/'s, not /l/'s and it is thus /l/, not /r/ that gets filled in by default. This difference could be accounted for, if we assume that markedness relations are language specific (though contra the basic notion of markedness) and that Sundanese differs from Latin in that it is /r/, and not /l/, that is marked in the former. But this view is not tenable, as within the phonology of Sundanese, it is /r/ that appears to be less marked, e.g. it has a broader phonotactic distribution. It is not then a difference in markedness per se that accounts for the difference

---

[8] Archangeli (1988) argues that, although the above facts constitute a convincing argument, other arguments suggest the opposite conclusion. The reader is referred there for consideration of these other arguments.

between dissimilation in Sundanese and Latin. In the case of Latin, dissimilation results in unmarked segments, as predicted by Odden, while in Sundanese, dissimilation applies between unmarked segments resulting in marked ones. Kaisse (1988) makes a similar observation based on the facts of Continuant Dissimilation in Modern Greek. Dissimilation occurs between two adjacent obstruents that agree in continuancy, resulting in the delinking and feature fill-in of both values of the feature Continuant. Thus, although an attractive suggestion, the view that feature specifications lost due to the delinking of tier adjacent identical specifications are filled in by general default rules cannot be maintained.

The fact that dissimilation may occur between unmarked values poses additional problems for a Radical Underspecification account of dissimilation. Following Radical Underspecification, it is generally assumed that it is marked values that are specified underlyingly. Thus in Sundanese, only [+lateral] would be specified underlyingly. But before dissimilation applies, default fill-in must provide [-lateral] specifications to /r/'s, but not to other consonants, since it is between /r/'s that dissimilation occurs. Yet following delinking, it is [+lateral] specifications that must be provided to account for the observed outputs. Thus again two fill-in rules are required, the first to assign [-lateral] specifications to /r/, the second to reassign [+lateral] specifications to those /r/'s which lose their lateral specifications.

We conclude then that the facts of dissimilation in Sundanese lend additional support to the conclusion that both values are specified within a contrasting class, as dissimilation itself requires the presence of [-lateral] specifications. The facts of Sundanese, however, argue against the view that dissimilation necessarily occurs between more marked values, resulting in less marked ones. But following Contrastive Underspecification, this poses no formal problem, since both values are present underlyingly and plausibly either could be filled-in by rule after delinking applies. In the case of Latin, it is the unmarked [-lateral] specification that is inserted after delinking, while in Sundanese is the [+lateral] specification which is filled in.

4.2. Dissimilation and the OCP

Dissimilation in Latin and Sundanese differ, at least superficially, in another respect. In Latin the process of dissimilation is seen to be strictly local (on the lateral tier). While in Sundanese dissimilation appears to apply between /r/'s which are farther apart and not ones that are closer together, e.g. /s=a**r**=idu**r**u/ —> [saliduru], but /**r**=a**r**=hɨt/ —> [rarahɨt]. This brings us back to the two remaining questions posed at the beginning of this section, repeated from above: (2) How general are the rules of /r/ Dissimilation and Lateral

Assimilation? Do they apply only to the /r/ of the plural marker? (3) Why tolerate /r/'s which are "close" to each other? These cases appear to be violations of the OCP, unless they constitute doubly linked structure. We turn now to these questions.

Other cases of dissimilation of /r/'s occur in Sundanese as exemplified in (15):

(15) Other cases of dissimilation of /r/'s

a. in borrowed forms: *rapor*, also *lapor* or *rapot*
'report' (from Dutch)

direktur also *dalektur*
'director'

b. other morphologically complex forms:

i. *para putra* —> *pala putra*
'pl' 'young male'

ii. *baraŋ + siar* —> *balaŋsiar*
'thing' 'seek' 'seek a livelihood' (Eringa 1949, p. 95)

As exemplified in (15a), dissimilation may occur in borrowed forms which contain two /r/'s. Dissimilation also occurs in other morphologically complex forms. As noted by Eringa (1949), in certain cases it may occur across word boundaries, as exemplified in (15bi) or between a prefix and a root, as exemplified in (15bii). *baraŋ-* is the only other indigenous affix in Sundanese which has an /r/ in it. In these cases, dissimilation is optional, but this is not surprising as the rule may be less apt to apply across weaker phonological boundaries, such a word boundary. Additionally, the prefix *baraŋ-* may function more like a compound than a prefix. The occurrence of dissimilation in such cases, although optional, suggests that dissimilation in Sundanese is a more systematic process in the language, not limited only to the plural marker.

As discussed in §1, if Dissimilation is indeed motivated by the OCP, then morpheme structure constraints would be expected to parallel the observed rules and any tier adjacent identical feature specifications which occur morpheme internally must be doubly linked structures, rather than a sequence of identical values. In order to test this claim, let's examine the distribution of morpheme internal liquids in Sundanese.

In considering constraints first on /r/'s, then on /r/'s and /l/'s, quite striking phonotactic constraints emerge. The following generalizations are based on rough calculations from the *Kamus Umum Basa Sunda* (Lembaga Basa & Sastra Sunda 1985), the most complete dictionary of Sundanese, with an estimated 17,000 entries. These observations are based on a consideration of /r/ and /l/ initial forms.

Of approximately 960 /r/ initial entries, 105 have more than one /r/. These 105 cases fall into the following three patterns:

(16) Phonotactic constraints on /r/'s (. . . = additional segments)

a.	$rV_1rV_2\ldots$	67	57 $V_1 = V_2$ = copying of 1st syllable	e.g.	*rara* 'braid'
b.	rVCrVC	20	19 $V_1C_1 = V_2C_2$ = copying of monosyll. form	e.g.	*ragrag* 'fall'
c.	r . . . r	18	17 are recent borrowings	e.g.	*radar*

As summarized in (16a), in the majority of /r/ initial cases with two /r/'s (67 forms), the form consists of a /rVrV . . . / sequence, where the second /r/ is the onset of the second syllable. In 57 of these 67 cases, the first and second vowels are the same, thus constitute a "phonological copy" of the first syllable.[9] As shown in (16b), in 20 cases the /r/'s are onsets to the first and second syllable, both of which are heavy. In 19 of these 20 cases, the vowel and consonant of the rime in the first and second syllable are the same, thus the forms consist phonologically of a copy of a monosyllabic form. Finally, as summarized in (16c), in 18 cases, there is a second /r/ in the form, not occurring as an onset to the second syllable. Of these 18 cases, 17 are clearly recent borrowings. Based on /r/ initial forms, we conclude then that roots with two /r/'s in the indigenous vocabulary are almost non-existent, unless the form involves a phonological copy of the first syllable. In addition there are no cases with /r/'s in both the onset and rime of the same syllable. Thus

```
*   σ
  / | \     appears to be disallowed in any case.
  r V r
```

---

[9] I use the term "phonological copy" to refer to a sequence of phonological material that is the same as another sequence. I do not mean to imply that these sequences are phonologically derived through a process of copying. I use the term to distinguish from the morphological process of reduplication. The former may or may not be due to the latter. Morphologically, partial reduplication of an initial syllable occurs in Sundanese, serving a number of functions. Although some of the cases of phonological copy may result from such partial reduplication, it is unlikely that all of them do. Since a monolingual dictionary was used for this investigation, further study would be needed to determine the morphological structure of these forms. But of the cases where I know the gloss, these do not appear to be morphologically derived forms in any obvious sense.

Constraints also obtain on the distribution of /r/'s and /l/'s. Although /r/'s and /l/'s may cooccur in some positions within the word, clear restrictions apply in cases with /r/ and /l/ in the onset of the first two syllables.

(17) Phonotactic constraints on /r/'s and /l/'s

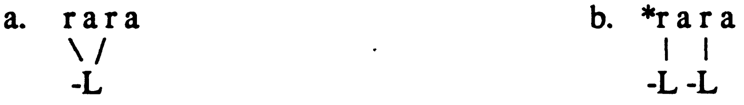
a.	rVlV . . .	4	4 are recent borrowings	e.g. *rəlatip*
b.	lVrV . . .	25	4 are recent borrowings	e.g. *lori*
			12 have an alternate form rVrV	e.g. *loris* ~ *roris* 'check'

As summarized in (17a), of the approximately 960 /r/ initial roots, the pattern rVlV . . ., with /r/ as the onset of the first and /l/ as the onset of the second syllable, only occurs in four forms and all four of these forms are recent borrowings. As shown in (17b), of the approximately 990 /l/ initial entries, 25 are of the shape lVrV . . . ., where /l/ is the onset of the first and /r/ the onset of the second syllable. Of these 25 cases, four are recent borrowings and of the remaining cases, twelve have alternate forms of the shape rVrV . . . . Thus we see that forms of the shape #[αlat] V [-αlat] V . . . are rare and when they occur they often have an alternate form of the shape rVrV . . . .

The following generalizations emerge. (1) In both monomorphemic forms and morphologically complex ones, two /r/'s are avoided unless they are in the onset of adjacent syllables, usually a phonological copy of an initial syllable. (2) Additionally, there is a tendency to avoid two unlike liquids in the onsets of the initial two syllables of a form. In answer to our second question above, we see that constraints on the distribution of two /r/'s and /r/'s and /l/'s hold much more generally than just in the context of the plural marker. /r/ Dissimilation applies (optionally) in other morphologically complex forms and across word boundaries. Basic phonotactic patterns also conform to such constraints. The behavior of Lateral Assimilation is less systematic, but there is a strong tendency against dissimilar liquids occurring as onsets of the initial two syllables of the form. Thus other aspects of the phonology of Sundanese display the same restrictions exhibited by the allomorphic alternation of the plural marker.

This leads us to our third question. Yip argues that if dissimilation is indeed motivated by the OCP, then other aspects of a phonological system should display similar constraints. Thus the fact that sequences of /r/'s are constrained more generally in Sundanese is as predicted by Yip. Yet if dissimilation is motivated by the OCP, it should always apply between tier adjacent like specifications (as seen to be the case above in Latin); while in Sundanese, as observed above, dissimilation is blocked between closer /r/'s, but applies between more distant ones. How can we account for this surprising situation? First we note that in the more distant cases of dissimilation in Sundanese, even if the /r/'s are separated by as much as two syllables, they are still tier adjacent. But why are /r/'s in the onset of two adjacent syllables tolerated? If we assume that precisely in these cases the /r/'s are doubly linked instead of constituting two discrete [-lateral] specifications, rather than being counterexamples to Yip's predictions, these cases would be exactly as predicted. This suggests that underlyingly, a sequence of identical liquids in two adjacent syllable onsets must be doubly linked, as illustrated in (18a):

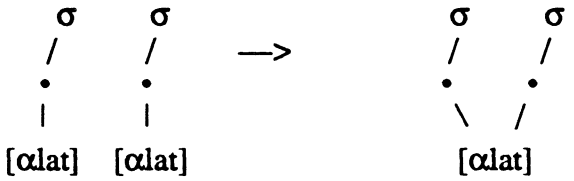
(18)



Such a representation is plausible, since there is no reason to assume that vowels are specified for the feature Lateral (either underlyingly or derivationally) and thus the structure in (18b) would be an OCP violation.

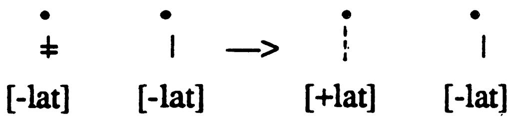
If in the case where two /r/'s occur as the onset of adjacent syllables in morphologically complex forms, these also constitute linked structures, we have an explanation of why just in these cases /r/ Dissimilation is blocked. If the two adjacent /r/'s constitute a linked structure, no OCP violation would occur. But if these cases involve linked structures, these must be created by rule. To account for this, I propose a rule of Lateral Node Merger:

(19) Lateral Node Merger



Identical lateral specifications in adjacent syllable onsets are merged into a single lateral feature specification. Lateral Node Merger applies before /r/ Dissimilation, thereby bleeding it. The existence of Lateral Node Merger simplifies our rule of /r/ Dissimilation as it eliminates the need for the condition specifying that the rule does not apply if the trigger is the onset of the second syllable. In addition, as shown above, /r/ Dissimilation is not restricted to the /r/ of the plural marker. /r/ Dissimilation is restated in (20):

(20) /r/ Dissimilation:

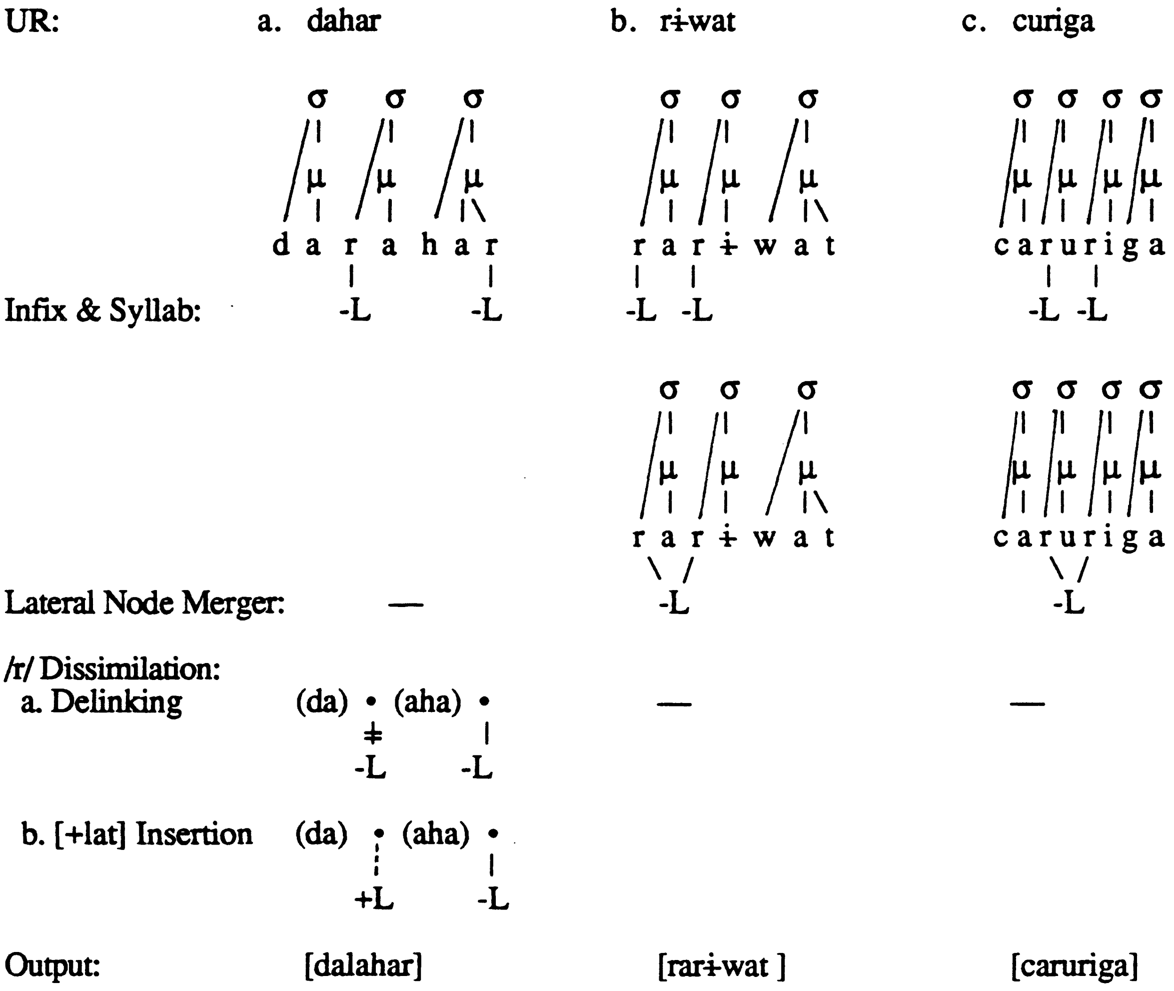


a. The first [-lateral] is delinked (due to tier adjacent [-lateral] specifications)

b. $\begin{bmatrix} \text{+cons} \\ \text{+son} \\ \text{-nasal} \\ \text{Ølateral} \end{bmatrix} \longrightarrow \text{[+lateral]}$

Sample derivations exemplifying both Lateral Node Merger and /r/ Dissimilation are presented in (21):

(21) Derivations:

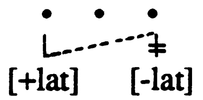


In (21a) the environment for Lateral Node Merger is not met; the tier adjacent [-lateral] specifications create an OCP violation, triggering the application of /r/ Dissimilation. But, as exemplified in (21b), with an /r/ initial root, the outcome of infixation is a sequence of two /r/'s in the onset of adjacent syllables, which meets the structural description of Lateral Node Merger, thereby bleeding the application of /r/ Dissimilation. Similarly if the onset of the second syllable is an /r/, as shown in (21c), this also results in a sequence of two /r/'s in adjacent onsets, again triggering Lateral Node Merger.

Lateral Node Merger does not apply when the /r/ of the root is in the rime of the first syllable, e.g. /pərceka/, /p=al=ərceka/, since the structural description is not met. I believe that this relates to a more general constraint whereby sequences of two /r/'s within a syllable are disallowed in Sundanese: $\begin{array}{c} {*\sigma} \\ /\ |\ \backslash \\ \mathrm{r\ V\ r} \end{array}$. Finally, Lateral Node Merger does not apply if the /r/ of the root appears in the second syllable as part of a complex onset, e.g. /motret/, /m=al=otret/. This suggests that there is something special about the structure of these complex onsets, though I do not have a formal mechanism to propose to block the application of Lateral Node Merger in these cases. We conclude that the interaction of Lateral Node Merger and /r/ Dissimilation accounts for the observed pattern of allomorphy of the plural marker when an /r/ is present in the root (as summarized above in (9)).

Finally consider the case where the *=al=* allomorph appears in /l/ initial roots, e.g. /lɨtik/, /l=al=ɨtik/. These forms were accounted for with the proposed rule of Lateral Assimilation, repeated in (22).

(22) Lateral Assimilation: r —> l / l = V __ =



applies to /r/ of the plural marker

The derivational outcome of the rule of Lateral Assimilation is parallel to the tendency to avoid sequences of two unlike liquids in the onsets of the two initial syllables of a word (*#[αlat] V [-αlat] V . . .) discussed above for both /r/ and /l/ initial roots. The structure and output of Lateral Assimilation look suspiciously similar to the rule of Lateral Node Merger, except that the lateral specifications are opposite. This raises the possibility that Lateral Assimilation and the parallel constraint on underlying forms could be accounted for with a more general statement of Lateral Node Merger, involving any two lateral specifications, whether the same or the opposite. Two observations argue against such a generalized rule of Lateral Node Merger. First, the underlying constraint on *#[αlat] V [-αlat] V . . . sequences is less systematic than that observed for forms with two /r/'s. Second, the *=al=* variant of the plural marker appears with /l/ initial roots, but not for roots where /l/ appears as the onset of the second syllable; the environment of Lateral Assimilation is more limited than the environment of Lateral Node Merger. Thus, although the parallel is interesting, these differences argue that the two rules cannot be collapsed and that we need to maintain both Lateral Assimilation and Lateral Node Merger as distinct rules.[10]

In summary, we see, assuming a rule of Lateral Node Merger, that the patterns of allomorphy observed in the plural marker are as expected, if the OCP holds both underlyingly and derivationally in Sundanese. These results lend strong support to the view that rules of dissimilation are motivated by the OCP.

5. Conclusions

We have examined in some detail the formal properties of the plural marker *=ar=* of Sundanese. We have considered both its behavior as an infix and the allomorphic alternation between the *=ar=* and *=al=* variants. The unusual property of the plural marker, whereby it usually appears as an infix, but can also occur as a prefix, results in maximizing preferred syllable types. Its placement is accounted for straightforwardly within a moraic framework, in which the affix is located before the first mora of the root. The allomorphy of the plural marker involves both assimilation and dissimilation, accounted for in the proposed analysis by three rules: /r/ Dissimilation, Lateral Node Merger and Lateral Assimilation. The facts of Sundanese are compatible with the current view of dissimilation as delinking and subsequent feature fill in. Yet Sundanese provides strong evidence that dissimilation does not always result in less marked segments as had been previously suggested. Furthermore the case of Sundanese lends further support in favor of Contrastive Underspecification, rather than Radical Underspecification. Finally, examination of the general phonological patterns of Sundanese reveals that Sundanese offers clear support for the view that dissimilation is an OCP driven process, in that both the rule of /r/ Dissimilation and more general phonotactic patterns behave alike.

---

[10] Anderson (1991) suggests that the *=al=* allomorph in /l/ initial roots is not a general phonological rule of Sundanese, but specific to the plural marker. This may be the case and could possibly be determined by closer inspection of sequences of two /l/'s in Sundanese.

References

Anderson, S. R. (1972) "On Nasalization in Sundanese," *Linguistic Inquiry* 3: 253-268.

Anderson, S. R. (1991) "Wackernagel's Revenge: Clitics, Morphology and the Syntax of Second Position," ms., The Johns Hopkins University.

Archangeli, Diana (1988) "Aspects of Underspecification Theory," *Phonology* 5: 183-207.

Archangeli, Diana and Douglas Pulleyblank (1986) *The Content and Structure of Phonological Representations.* Ms., forthcoming from MIT Press.

Clements, G. N. (1976) *Vowel Harmony in Nonlinear Generative Phonology.* Distributed by the IULC (1980).

Cohn, Abigail C. (1990) *Phonetic and Phonological Rules of Nasalization.* UCLA Ph.D. dissertation, also appeared as *UCLA Working Papers in Phonetics* 76.

Eringa, F. S. (1949) *Loetoeng Kasaroeng Verhandelingen van het Koninkliijk Instituut voor Taal-, Land- en Volkenkunde,* Deel 8. (pp. 94-95).

Ewing, Michael (1991) "Plural Concord in Sundanese," talk presented at the 6th International Conference on Austronesian Linguistics, Honolulu, HI.

Goldsmith, John (1976) *Autosegmental Phonology.* MIT Ph.D. dissertation. Published by Garland Press, New York, 1979.

Goodman, Beverley (this volume) "Takelma Dissimilation and the Form of the OCP."

Hayes, Bruce (1986) "Assimilation as Spreading in Toba Batak," *Linguistic Inquiry* 17: 467-500.

Hayes, Bruce (1989) "Compensatory Lengthening in Moraic Phonology," *Linguistic Inquiry* 20: 253-306.

Hyman, Larry M. (1985) *A Theory of Phonological Weight.* Dordrecht: Foris Publications.

Kaisse, Ellen (1988) "Modern Greek Continuant Dissimilation and the OCP," ms., University of Washington.

Kiparsky, P. (1982) "Lexical Morphology and Phonology," In The Linguistics Society of Korea (eds.) *Linguistics in the Morning Calm.* Seoul: Hanshin Publishing Co. pp. 3-91.

Lembaga Basa & Sastra Sunda (1985) *Kamus Umum Basa Sunda,* cetatan kelima. Bandung, Indonesia: Penerbit Tarate Bandung.

McCarthy, John (1985) "Features and Tiers: the Structure of Semitic Roots," talk presented at MIT.

McCarthy, John (1988) "Feature Geometry and Dependency: A Review," *Phonetica* 45: 84-108.

McCarthy, John (1989) "Linear Order in Phonological Representation," *Linguistic Inquiry* 20: 71-99.

McCarthy, John and Alan Prince (1986) "Prosodic Morphology," ms., University of Massachusetts, Amherst and Brandeis University.

Odden, David (1987) "Dissimilation as Deletion in Chuckchi," *ESCOL '87*: 235-246.

Poser, William J. (1987) "An Autosegmental Theory of Dissimilation," talk presented at WCCFL V1, Tucson Arizona.

Robins, R. H. (1959) "Nominal and Verbal Derivation in Sundanese," *Lingua* 8: 337-369.

Steriade, D. (1987) "Redundant Values," *CLS* 23: 339-362.

Yip, Moira (1988) "The OCP and Phonological Rules: a Loss of Identity," *Linguistic Inquiry* 19: 65-100.

Takelma Dissimilation and the Form of the OCP*

Beverley Goodman

0. Introduction

In *The Sound Pattern of English* (SPE), Chomsky and Halle (1968) propose that the form of dissimilation rules is, in general, as in (1).

(1) $[\alpha F] \rightarrow [-\alpha F] / \_\_ [\alpha F]$ (p. 178)

Since dissimilation rules change the feature value $[\alpha F]$ to $[-\alpha F]$ in the context of $[\alpha F]$, it has been proposed that they are a result of the Obligatory Contour Principle (McCarthy 1985, Odden 1987, Yip 1988). McCarthy (1986) proposes that the OCP, in addition to applying to tonal sequences, also applies to melodic sequences as stated in (2).

(2) Obligatory Contour Principle: At the melodic level, adjacent identical elements are prohibited (p. 208)

In this paper I examine two rules of dissimilation in Takelma: Coronal Dissimilation and Nasal Dissimilation. These rules, I claim, are the result of the Obligatory Contour Principle (henceforth OCP). With respect to these two dissimilation rules the question arises as to whether the OCP functions only in cases of dissimilation or whether it plays a more general role in the grammar of Takelma. In addressing the question of the role of the OCP in Takelma, I focus on the non-symmetrical behavior of coronal sonorants and coronal obstruents. The coronal obstruents neither trigger nor undergo the dissimilation rule. One might suggest that this asymmetrical behavior can be characterized by the underspecification of the coronal feature in the case of obstruents. However, I will show that underspecification cannot account for the fact that Coronal Dissimilation targets only coronal sonorants. Rather, I argue that the OCP in Takelma is conditioned by a segment's value for sonorancy. I show that the dissimilation of coronal sonorants is part of a more general prohibition on adjacent coronal specifications with like values for the feature [sonorant] at both the underlying level and during phonological derivations.

Dissimilation in Takelma results in alternations in the final consonant of noun characteristic suffixes as shown in (3).

---

* This paper was presented at the 65th annual meeting of the Linguistic Society of America. I would like to thank the members of that audience for useful comments. I would also like to thank Abby Cohn and John McCarthy for beneficial discussion.

(3)	a. /pep+Vn/	->	[pepen]	'rushes'
	b. /hel+Vn/	->	[helam]	'board'
	c. /ʃim+Vn/	->	[ʃimil]	'dew'
	d. /k^{w}an+Vn/	->	[k^{w}alam]	'road'

As shown in (3a), the suffix final nasal consonant is [n] following stems which contain non-nasal obstruents. However, if the suffix follows a stem which contains a [+sonorant] consonant, various alternations are evidenced. As shown in (3b) if the stem contains an [l] the suffix final nasal surfaces as [m]. In (3c), the suffix final nasal surfaces as [l] following a stem ending in an [m]. In (3d) the stem contains an [n] underlyingly and the suffix-final nasal surfaces as [m] while the stem-final nasal surfaces as [l]. In this paper I will present an analysis of this set of alternations within a non-linear theory of phonological structure which incorporates the OCP as a constraint on hierarchical phonological representations.

The structure of the paper is the following. In section 1, I discuss the role of the OCP in dissimilation rules. In section 2, I present the facts of Takelma dissimilation. I propose an analysis of both Coronal Dissimilation and Nasal Dissimilation. I claim that the OCP disallows adjacent [+sonorant] coronal and nasal specifications. I also discuss the status of default values and the role of structure preservation. In section 3, I examine the underlying distribution of coronal sequences. I show that both the dissimilation rules and the underlying distribution of coronal sequences provide evidence that the OCP in Takelma is conditioned by a segment's value for [sonorant]. In section 4, I discuss how the facts of dissimilation might be analyzed within an approach which claims that at least some coronal consonants are unspecified for place features and will argue that such an approach is not tenable. Section 5 concludes the paper.

1. The Form of Dissimilation Rules

Following McCarthy's (1985) early proposals, Yip (1988) proposes that all rules involving identity of target and trigger--of which dissimilation is a clear example--are the result of the OCP functioning in the particular language under examination. The OCP, as a universal trigger, renders such rules less marked universally or less costly to the grammar. Alternatively, one might view the OCP as a universal constraint which prohibits the surfacing of disallowed sequences and therefore the language implements a phonological process to alleviate such violations.

1.1. OCP Triggered Rules and Feature Structure

The relationship of dissimilation and the OCP becomes especially interesting in the context of hierarchical segmental representations such as those proposed in Clements (1985) and Sagey (1986). Models of feature geometry provide a new perspective on dissimilation rules. The relevant aspects of hierarchical representation, for the purposes of this paper, are the characterization of the root node as [sonorant], the representation of place features and the feature [nasal] as illustrated in (4).

```
(4)            X                     Skeleton
               |
             [son]                   Root Node
              / |
          Nas   •                    Place Node
              / | \
          Lab Cor Dor                Articulator Nodes
```

I assume that segments are represented by a timing or skeletal slot, depicted here as X. All features characterizing segment quality are dominated by a root node. The root node itself is inherently characterized by the features [sonorant] and [consonantal] following McCarthy (1988). All relevant segments in the case of Takelma dissimilation are [+consonantal] so this is omitted from all following diagrams. The feature [nasal] occurs higher in the representation than the Place node which dominates the terminal articulator features [labial, coronal, dorsal] which are assumed to be single valued (or privative) specifications (following Sagey 1986).

The second important component of the theoretical framework assumed for the treatment of dissimilation in this paper is the OCP. The OCP, as it affects representations such as those in (4), is schematized in (5) where the linear sequence [α F] [α F] is disallowed.

```
(5)           *X      X              Skeleton
               |      |
               .      .              Root
               |      |
             [αF]   [αF]             Tier-adjacent feature specifications
```

The OCP may affect any aspect of hierarchical feature organization. Thus in (5), [α F] may be any two identical feature specifications. Content nodes, such as [labial], [voice], etc. are the relevant part of the structure for determing whether or not the OCP applies. However, the elimination of a disallowed sequence may affect abstract nodes (for example, place node merger). Such a representation not only characterizes existing dissimilation rules, but predicts that only tier-adjacent identical specifications will result in dissimilation. Steriade (1987) illustrates, in the context of translaryngeal spreading rules, that hierarchical feature representations themselves impose certain locality conditions on phonological spreading rules. Dissimilation rules are expected to show parallel properties. A dissimilation rule which affects adjacent identical coronal specifications scans the coronal tier. Other place specifications, such as labial or dorsal, will be transparent to the dissimilation process since these features occupy distinct tiers. The Takelma dissimilation facts confirm these predictions.

1.2. Underspecification and Default Rules

Hierarchical segment structure and a universal constraint such as the OCP result in two distinct types of rules. First, the OCP may result in a rule which delinks an existing node. An adjacent node may then spread to this node resulting in assimilation (see Yip's (1988) discussion of Berber). The second type of rule is one which delinks a feature or node as the the first step in cases of dissimilation (Odden 1987). Once delinking applies, the segment acquires a feature specification through the insertion of a default value.

Within most versions of autosegmental phonology two fundamental rules types are recognized: spreading and delinking. In (6), I schematize the two cases.

```
(6) a.    *X X X --delinking-->X X X --spreading-->X X X
           | | |               | | |               | | |
           • • •               • • •               • • •
           | | |               |   |               |/  |
           α β β               α   β               α   β

    b.    * X X--delinking-->X X--default insertion-->X X
            | |              | |                      | |
            • •              • •                      • •
            | |              |                        | |
            α α              α                        α β
```

The result of delinking is a segment uncharacterized for any feature on some tier. One way a segment which has no specification can obtain feature content is for an adjacent feature to spread to the unspecified segment as schematized in (6a). Note in this context that after the deletion of the first feature [β], the second feature [β] would be free to spread leftward analogous to the rightward spreading of [α]. Thus, it is theoretically possible that the delinking of [β], will be obscured by the subsequent spreading of the segment which participated in the OCP violation in the first place.

In (6b), I illustrate delinking followed by feature fill-in. However, delinking and subsequent feature insertion depend upon a theory of underspecification and default values. An issue raised by dissimilation rules is, therefore, how nodes are specified once delinking has applied. It has been suggested by Odden (1987) that one would not expect OCP-triggered rules of feature delinking to be followed by rules that insert arbitrary feature values. Rather, the expected case would be that the delinking of a feature results in a general default specification which can be motivated on independent grounds on a language particular basis.

Alternatively, the fill-in specifications might be expected to be constrained by Structure Preservation (Kiparsky 1982). That is, the choice of appropriate fill-in values is limited by the segmental inventory. As will be shown, the Takelma dissimilation facts present evidence that the value which is filled in following delinking is structure preserving. There is no evidence that the value filled in after dissimilation is the general default value for Takelma. However, the phonemic inventory of Takelma limits the possibility of what values may be filled in for partially specified segments.

Thus, if the result of dissimilation as a series of elementary operations is the change of a sequence of identical melodic elements into a non-identical sequence, then one can view such rules as relating in a direct and obvious way to dynamic OCP effects.

2. Takelma Dissimilation

The focus of this section is on two rules of dissimilation in Takelma, an extinct Penutian language described by E. Sapir in his 1922 dissertation.[1] A lexicon of Takelma appears as an appendix to a collection of texts also authored by Sapir (1909).

In (7) I give the consonant and vowel inventories of Takelma.[2]

(7) Takelma Consonant and Vowel Inventories:

(a) Vowels: i, ii u, uu
e, ee o,oo
a
Diphthongs: V+i,u

---

[1] Sapir (1922) describes the process of dissimilation of Takelma noun characteristics. His generalizations are incorporated into the analysis proposed in this paper.

[2] Note that in the consonant inventory, both alveolar and palatal obstruents are given. It is not clear, however, that this is a phonemic contrast. Sapir claims that the alveolar [s] and the palatal [ʃ] may be non-distinct, surface realizations of one phoneme. He does not explicitly claim that this is the case for [ts] and [tʃ]. However affricates may parallel the fricatives. There is one minimal pair in the lexicon which contrasts [ts] and [tʃ]. Within the context of a limited lexicon, the exact status of the alveolar and palatal affricates is, therefore, unclear. I return to this point in section 4. For the time being, I parenthesize the palatal ([-anterior, coronal]) fricative and affricate.

(b) Consonants:

	Labial	Coronal +Ant	Coronal -Ant	Velar
Stops: Plain	p	t		k
Asp.	p^h	t^h		k^h
Glott.	p'	t'		k'
				ʔ,h
Affricates		ts'	(tʃ')	
Fricatives		s	(ʃ)	x
Liquid		l		
Nasals	m	n		
Glides	w	y		

2.1. Dissimilation of Noun Characteristics

As examplified above, Takelma dissimilation occurs when a noun is suffixed with what Sapir terms a noun characteristic. Takelma noun characteristics are suffixes that occur on the noun before all nominal increments, i.e., pronominal suffixes and locatives. According to Sapir, the noun characteristic has no clear grammatical function. Takelma noun stems occur in their underlying, or non-suffixed, form only in their absolutive function or when incorporated into the verb. In their non-suffixed form, most Takelma noun stems are monosyllabic of the shape CV(C).

In the examples in (8), the noun characteristic surfaces as [-Vn]. In examples (8a-c), we see that the stem final consonant may be a labial stop or glide. In (8d-g), the underlying form of the noun ends in a coronal stop or glide. (I assume that the glides are vocalic articulations mapped to non-peak slots in syllable construction. As such they play no role in consonantal dissimilation.) In (8h-k), the underlying form of the noun ends in a velar stop or fricative. Finally, the examples in (8 l, m) illustrate nouns which end in either a consonant cluster or a labialized velar consonant.

(8) Case 1: Noun Characteristic Surfaces as [-Vn]:

Stem final labial consonant, glide:

a. /wuup'+Vn/ -> [wuup'un] 'eyebrows'

b. /pep+Vn/ -> [pepen] 'rushes'

c. /yiw+Vn/ -> [yiwin] 'speech'

Stem final coronal consonant, glide:

d. /yut'+Vn/	->	[yut'un]	'white duck'
e. /xt+Vn/	->	[xtan]	'eel'
f. /k^wit'+Vn/	->	[k^wit'in]	'wrist'
g. /p'iy+Vn/	->	[p'iyin]	'deer'

Stem final velar consonant:

h. /tak+Vn/	->	[takan]	'turtle'
i. /wik+Vn/	->	[wikin]	'red lizard'
j. /kak'+Vn/	->	[kak'an]	'house ladder'
k. /tʃ'ax+Vn/	->	[tʃ'axan]	'blue striped lizard'

Stem final consonant cluster / labialized velar:

l. /yuxk+Vn/	->	[yuxkan]	'trout'
m. /ʃuk^w+Vn/	->	[ʃuk^wan]	'root basket'

In all the examples above, the surface form of the suffix has a final [-n]. Furthermore, in almost all these examples, the vowel of the suffix is a copy of the vowel of the underlying stem. There are two cases where this generalization does not hold. In example (8e), the stem has no underlying vowel and the suffix vowel, in this case, surfaces as an [a], the epenthetic vowel in Takelma (Goodman 1987). In examples (81) and (8m) the suffix vowel is not a copy of the stem vowel but is, again, [a], the default vowel. These examples contain a consonant cluster or a consonant with a secondary articulation and again surface with the default vowel.[3]

Based on these patterns, I assume that the underlying form of the noun characteristic is /-Vn/, but surface alternations occur. (The reasons why the underlying representation of the suffix-final nasal is proposed to be fully specified for the feature coronal will be discussed in Section 4.)

The examples in (9) illustrate that when the stem contains an [l], the suffix consonant surfaces as [m].

(9) Case 2: Noun characteristic surfaces as [-Vm] after a stem containing an [l]:

(a) Adjacent /..l+Vn../ sequences:

i. /hel+Vn/	->	[helam]	'board'
ii. /kel+Vn/	->	[kelam]	'river'
iii. /tʃ'el+Vn/	->	[tʃ'elem]	'hail'

---

[3] Although an interesting issue, the distribution of copy vowels versus the default vowel [a] is beyond the scope of this paper. I will assume that, in general, the duplication of the stem vowel takes place through spreading. In cases where spreading is blocked, the default [a] surfaces.

iv. /tʃ'ul+Vn/	->	[tʃ'ulum]	'wart'
v. /hapil +Vn/	->	[hapilim]	'empty'
vi. /yul+Vn/	->	[yulum ~ yulam]	'eagle'
vii./kul+Vn/	->	[kulum]	'oak'

(b) Non-adjacent /..l(V)C+Vn../ sequences

i. /lap^{h}+Vn/	->	[lap^{h}am]	'frog'
ii. /lek+Vn/	->	[lekem]	'kidney'
iii. /lox+Vn/	->	[loxom]	'manzanita'
iv. /tolk^{h}+Vn/	->	[tolk^{h}am]	'anus'

The examples in (9a) illustrate that the noun characteristic surfaces as a vowel plus labial nasal after a stem-final [l]--a coronal sonorant. The underlying coronal nasal dissimilates to the non-coronal [m]. The examples in (9b) show that any occurrence of [l] in the stem triggers the dissimilation of the coronal nasal. The sequential cooccurence of any two coronal sonorants is prohibited in suffixation contexts. In such cases dissimilation occurs. Thus, the examples in (9) illustrate a case of place dissimilation where the second of two coronal specifications on sonorant consonants is changed to a labial specification.

The third case, where the stem contains an [m] and the suffix surfaces as [-Vl], is exemplified in (10a-c).

(10) Case 3: Noun Characteristic Surfaces as [-Vl] after a stem containing [m]:

a. /tʃ'am+Vn/	->	[tʃ'amal]	'mouse'
b. /ʃim+Vn/	->	[ʃimil]	'dew'
c. /meh+Vn/	->	[mehel]	'basket for cooking'

In the examples in (10), the [m] and the [n] dissimilate with respect to the feature [nasal]. If the stem contains a nasal, the suffix consonant surfaces as the non-nasal coronal sonorant [l]. After dissimilation, these cases contain neither a sequence of coronal sonorants nor a sequence of nasal consonants on the surface.

In the final set of examples, the suffix surfaces as [-Vm] when the stem ends in a coronal nasal [n] as illustrated in (11). Additionally, the underlying stem-final [n] surfaces as [l].

(11) Case 4: Noun characteristic surfaces as [-Vm], stem /n/ dissimilates to [l]:

a. /k^{w}an+Vn/	->	[k^{w}alam]	'road' (cf. [k^{w}an], absolutive)
b. /xan+Vn/	->	[xalam]	'urine' (cf. [xan], absolutive)

These cases illustrate two cases of dissimilation: the suffix nasal dissimilates in place of articulation and the stem final consonant dissimilates with respect to nasality.

2.2. The OCP and Takelma Dissimilation

In this section I present an analysis of the Takelma facts assuming that the OCP, as defined on sequences of coronal and nasal sequences, is the driving force behind the rules of dissimilation in Takelma. First, the underlying /n/ of the suffix dissimilates to an [m] when preceded by a [coronal, +sonorant] consonant, either oral or nasal. Second, the result of suffixation is never a sequence of two nasal consonants. If the stem contains a nasal, the suffix final [n] dissimilates to an [l] or if the suffix [n] has dissimilated to [m], the [n] of the stem dissimilates to [l]. In this analysis, I propose that [l] is the non-nasal counterpart of the nasal coronal sonorant. That is, [l] need not be characterized for the feature [lateral] at all.

The tiers upon which the OCP is defined and where dissimilation applies are the coronal tier and the nasal tier. The OCP as it applies to segmental sequences is stated in (12).

(12) Statement of the OCP for Takelma

a. *[coronal] [coronal]; condition: [αsonorant]

b. *[nasal][nasal]

The statement in (12a) says that two tier-adjacent [coronal] consonants with identical values for [sonorant] are disallowed in Takelma.[4] The second part of the OCP statement in (12b) disallows a sequence of two tier adjacent [nasal] specifications. The unique aspect of the analysis presented in this paper is the proposal that [α sonorant] conditions the application of the OCP on the coronal tier. The [α sonorant] condition holds vacuously on the nasal tier.

The claim inherent in the proposal that the OCP is conditioned by [α sonorant] is that, on a language paraticular basis, some feature may override the absolute prohibition expressed by the OCP. The feature [sonorant], I claim, is such a feature. A natural question in this context is: what does it mean to claim that the OCP is conditioned? Yip (1989) in a survey of morpheme structure constraints shows that the OCP, in the general case, disallows sequences of consonants from the same place of articulation. One might claim that this is the default setting for the OCP. However, it appears that some languages allow various feature specifications to override the general statement of the OCP. Yip presents cases where [±distributed] or [±anterior] as dependent features complicate the

---

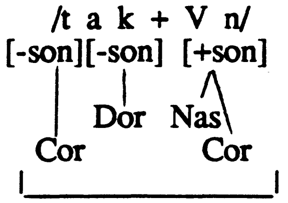
[4] Dresher (1989) suggests that the OCP in Arabic should be conditioned by the feature [+sonorant]. He proposes that the domain of the OCP is [+sonorant] for coronal consonants. McCarthy (1988) also discusses the fact that the Arabic Morpheme Structure Constraint affecting place of articulation is sensitive to the [sonorant] property of the segments it affects.

statement of the OCP. If the OCP is viewed as a constraint on underlying representations which disallows sequences of segments which are not articulatory and perceptually distinct, then languages may vary in what counts as making two otherwise identical segments distinct enough to satisfy the OCP. From this point of view, it is not suprising that a segment's value for sonorancy would be one way to do this. Under most views of segment specification, all consonants are classified as to whether they are [+sonorant] or [-sonorant]. This distinction characterizes obstruents versus sonorants. There are many cases of phonological rules which refer to one of these groups to the exclusion of the other. Typically only the [+sonorant] consonants serve as syllable nuclei. It is extremely marked for a language to allow syllable peaks to be occupied by obstruents. On the other hand, in the general case of consonant voicing assimilation, the consonants which most typically participate in these rules are obstruents since this is the class of segments which generally involve contrastive voicing. Because a segment's value for [sonorant] is fundamental and phonologically important, it is not unexpected that the distinction in this value may render consonantal sequences distinct enough to override OCP violations. The facts of Takelma, both the dissimlation rule and the underlying distribution of coronal sequences, provide strong evidence for such a view.

2.2.1. The Representation of Non-dissimilating Cases

First, consider the representation of noun stems and the characteristic suffix in cases where dissimilation does not apply. As shown in (8) above, dissimilation does not apply when the coronal consonants in the stem and suffix differ with respect to their values for sonorant. I illustrate this case in (13).

(13) Case 1. No Dissimilation. Underlying /-Vn/ surfaces with an [n].



-> [takan] 'turtle' (= 8h)

No OCP violations; coronal specifications are not [α sonorant]

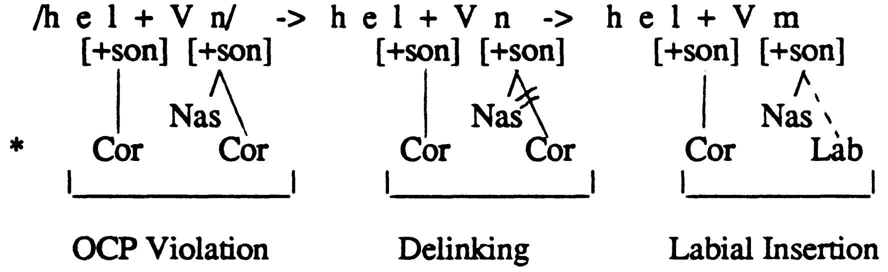
No rules apply to forms of this type. Although there are two coronal specifications, they differ in their value for sonorant and the OCP fails to apply. The [n] of the suffix is represented as specified for the feature coronal. One of the central claims made in this analysis is that full specification provides the better and more predictive anlaysis of the

Takelma facts. The question of coronal specification at the point dissimilation applies is examined in Section 4.

2.2.2. Coronal Dissimilation

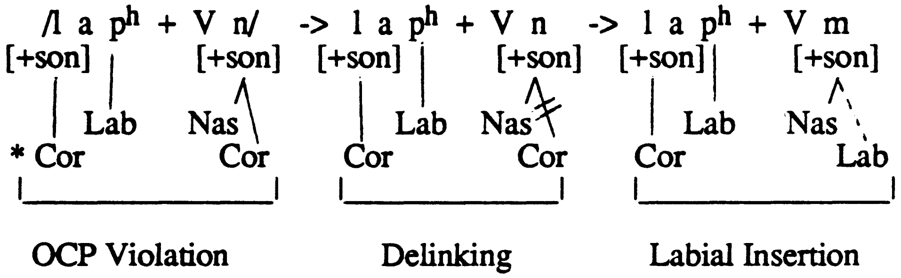
In (14) and (15) I give derivations which illustrate the dissimilation of coronal consonants that agree in their value for [sonorant]. In this instance, there are two sub-cases; one where the dissimilating segments are separated by only a vowel illustrated by the data in (9a) above. In the second case, (9b) above, the dissimilating consonants are separated by vowels and dorsal or labial consonants. I refer to the second case as "long-distance" Coronal Dissimilation.

(14) Case 2 a. Coronal Dissimilation. Underlying /-Vn/ surfaces with an [m].



-> [helam] 'board' (= 9a.i)

(15) Case 2 b. Long distance coronal dissimilation.



-> [lap^{h}am] 'frog' (= 9b.i)

(16) Coronal Dissimilation:

Operation: Delink second

Dissimilation is proposed to be the result of the OCP and the rule specifies simply that the second coronal specification delinks. Since the rule applies on the coronal tier, labial (and dorsal) consonants are transparent. Thus, two instances of the [coronal] specification are adjacent under the assumption that articulator nodes are one valued as proposed, for example, in Sagey (1986) and the two subcases--local and long-distance--receive the same analysis. The combination of privative articulator specifications and the notion of tier-adjacency results in intervening non-coronal place specifications having no effect on the application of the dissimilation rule. Strikingly, there are no examples in Sapir's Grammar or the Lexicon which involve cases of the rule of long-distance Coronal Dissimilation applying across an intervening coronal consonant, i.e., there are no examples of the form /lVt+Vn/. Since an example of this type is relevant to the proposals in this paper, one might wonder why they are conspicuously absent. Interestingly, as I will illustrate in section 3, there are simply no [l] initial forms in the Takelma lexicon or grammar.

The final step in the examples illustrated above is to fill in the correct value for the place of articulation of the suffix-final nasal. The following rule is proposed.

(17) [Ø Place] -> [Labial]

Aside from the rule of coronal dissimilation in noun suffixation contexts, there is no independent evidence that [labial] is the default place specification in Takelma. However, there is no dorsal nasal. Therefore, if the rule is structure preserving (Kiparsky 1982), then once the [coronal] specification of the nasal is delinked the only alternative place specification available for a nasal segment is [labial].

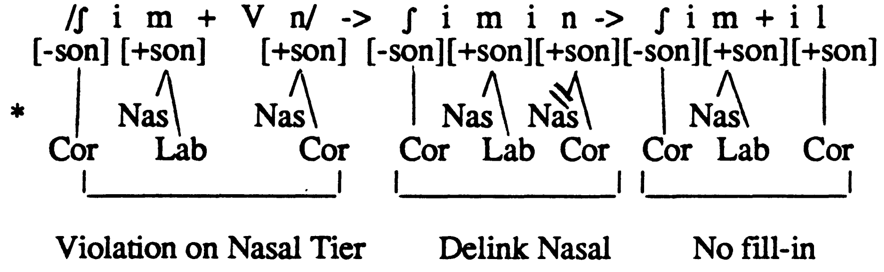
In sum, I claim that the articulator feature [coronal] is present in both the stem and the suffix nasal underlyingly. Due to the OCP prohibition, the second of two [+sonorant] coronal specifications is delinked and the feature labial is filled in by rule.

2.2.3. Nasal Dissimilation

Cases which involve only Nasal Dissimilation, illustrated above in (10) receive a straightforward analysis under the proposals developed so far.

Nasal Dissimilation delinks the nasal specification in cases which constitute a violation of of the OCP on the nasal tier.

(18) Case 3. Nasal dissimilation. Underlying /-Vn/ surfaces as [-Vl].

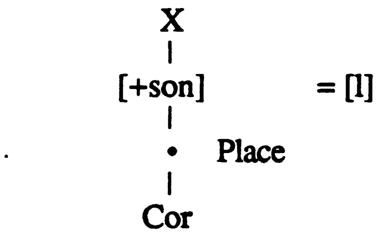


= [ʃimil] 'dew' (10b)

(19) Preliminary: Nasal Dissimilation: Delink [nasal]

In these cases, the second of two nasal specifications delink. Once the rule of Nasal Dissimilation has applied to forms such as those illustrated in (18), the derived structure of the non-nasal coronal sonorant is as illustrated in (20).

(20)

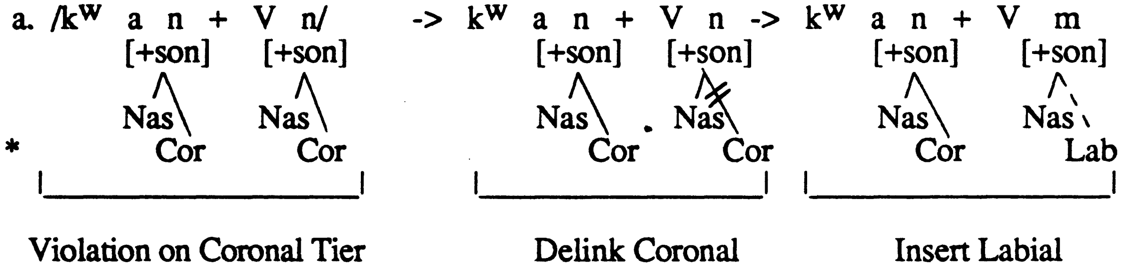


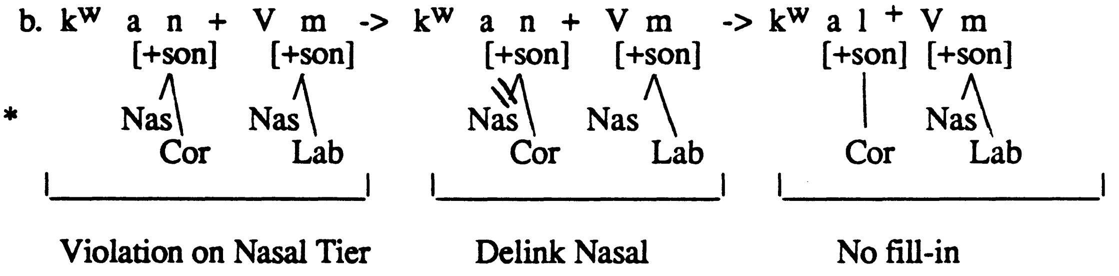
I assume that this is the representation of [l] in Takelma.

2.2.4. Cases which Involve both Coronal and Nasal Dissimilation

The data in (11) above illustrate the application of both Coronal Dissimilation and Nasal Dissimilation within the same form. Such cases are represented as shown in (21). Examples such as these require the reformulation of the Nasal Dissimilation rule.

(21) Case 4. Both Coronal Dissimilation and Nasal Dissimilation apply.





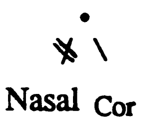
-> [kʷalam] 'road' (= 11 a)

In (21a) the application of Coronal Dissimilation is illustrated and in (21b) the application of Nasal Dissimilation is illustrated. In (21a) the OCP violation on the coronal tier is alleviated by deleting the second occurrence of the [coronal] articulator node under the [α sonorant] condition. The feature [labial] is filled in for the suffix-final nasal consonant, this being the only other place of articulation at which nasals occur.

Following the analysis developed to this point, we would expect the second nasal specification to delink in (21) while, in fact, the OCP violation on the nasal tier is alleviated by delinking the first occurrence of [nasal] (21b). Briefly consider the possibilities involved in this derivation. Delinking the second [nasal] specification results in intermediate /kwan+al/. If Coronal Dissimilation applies and delinks the place specification of [l] the segment is uninterpretable since there is no non-nasal labial sonorant in Takelma. Therefore, this derivation is ruled out. Structure preservation then plays an important role in both Coronal Dissimilation and Nasal Dissimilation. To capture this fact, I propose to reformulate Nasal Dissimilation as in (22) encoding the structure preserving nature of the rule directly. The statement of Nasal Dissimilation says that [nasal] can only be delinked from coronal consonants.

(22) Nasal Dissimilation

Operation: Delink



3. The Morpheme Structure Constraint

In addition to claiming that dissimilation rules are triggered by the OCP, Yip (1988) also makes the following claim:

> (23) ...All rules involving identity of target and trigger with an output in which they are no longer identical and adjacent are OCP-triggered rules. The kind of case that would...require a weakening of this claim would be a language with the following properties:
>
> (i) Dissimilation of F: [αF] ->[-αF] / __[αF]
>
> (ii) Demonstrable morpheme-internal [αF] [αF] sequences, as opposed to doubly linked [αF]. (p. 73)

While Yip herself does not pursue this prediction, the expectation is that if a language evidences a rule of dissimilation which is arguably due to the OCP, then it follows that such a language should also have a morpheme structure constraint which disallows morpheme-internal sequences of the dissimilating feature(s) (see also Cohn, this volume).

The analysis developed above depends on the claim that the OCP operates in Takelma on tier-adjacent coronal specifications and tier-adjacent nasal specifications. According to Yip's proposals, these are just the sequences which should either be multiply-linked or absent underlyingly. Given the proposal that the OCP in Takelma is conditioned by a segment's value for the feature [sonorant], the predictions are even more specific.

First, we predict that sequences which do not agree in sonorancy are freely permitted. The second prediction is that all underlying monomorphemic sequences of [coronal] which agree in sonorancy are either subject to dissimilation or are multiply linked, since sequences of the feature coronal agreeing in sonorancy are disallowed on the surface in Takelma. Therefore, if such sequences are present underlyingly they must consist of only one occurrence of the feature linked to two segments. The third prediction is that there will be no sequences of coronal specifications which agree in sonorancy but which differ in [±anterior] since these sequences should share a single coronal specification but cannot if they differ in anteriority. This set of predictions is borne out in an interesting way by the distributional facts of Takelma.

In a corpus of 575 nouns and adjectives in the lexicon, there are 61 cases which contain an underlying coronal-vowel-coronal sequence. This is a large number compared to other places of articulation.[5] However, in 47 of these cases (77%) the sequence consists of a non-sonorant coronal and a sonorant coronal consonant with the order [obstruent] - [sonorant] in most cases. I summarize the distributional facts in (24).

(24)

Coronal	Vowel	Coronal	Total
[α son]		[-α son]	47
[+ son]		[+ son]	2
[- son, α ant]		[- son, α ant]	8
[- son, α ant]		[-son, -α ant]	4
			61

Examples illustrating the allowed sequences of coronal-vowel-coronal differing in sonorancy are given in (25).

---

[5] There are 12 labial-vowel-labial sequences. Five of these are obstruent-vowel-obstruent, five are obstruent-vowel-sonorant, there is one sonorant-vowel-sonorant and one sonorant-vowel-obstruent sequence within the labial group. There are 10 dorsal-vowel-dorsal sequences; all are obstruent-vowel-obstruent since there are no dorsal sonorants in Takelma.

(25) Underlying sequences of coronal differing in sonorancy (partial list):

a.	/tan/	'rock'	j.	/ʃin/	'wood coals'
b.	/tel/	'yellow-jacket'	k.	/tola/	'hollow tree'
c.	/nos/	'next door'	l.	/tolax/	'things,utensils'
d.	/ʃal/	'foot'	m.	/ʃilek'ʷ/	'acorn-pestle of stone'
e.	/sel/	'black paint, writing'	n.	/tʰelma/	'acorn pestle'
f.	/sens/	'bug'	o.	/t'ela/	'louse'
g.	/tʰan/	'squirrel'	p.	/tʃ'ana/	'about to die'
h.	/t'ela/	'shinny stick'	q.	/tʃ'ulm/	'wart'
i.	/tʃ'il/	'red'	r.	/ts'an/	'porcupine'

Given the proposal that the OCP in Takelma disallows all coronal sequences except those which disagree in sonorancy, this is the expected underlying distribution of such sequences. Thus the first prediction, that coronals differing in sonorancy should occur freely, is borne out.

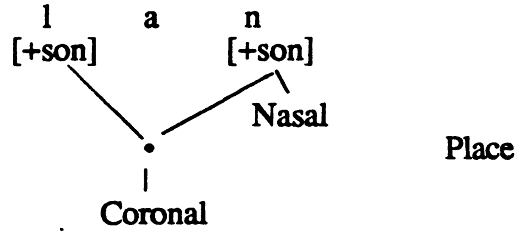
Turning to the second prediction, we see that, while the overall preference for coronal-vowel-coronal sequences is for only one of the coronal consonants to be [+sonorant], cases where both coronal consonants are [+sonorant] are the most infrequent. Since such sequences must be linked to only one coronal specification, their infrequency may be attributed to their complex underlying representation. With respect to the third prediction, we note that coronal sequences which disagree with respect to the feature [anterior] are rare in Takelma nouns and adjectives. In general, coronal obstruents agree in their value for [anterior] systematically in the small number of lexical items available. The four cases in (24) above which constitute violations to this generalization come from the semantic field centering around 'smallness,' i.e., [t'oʃo] 'small, a little' and, by semantic criteria, constitute only one exceptional case.

The examples below in (26) illustrate the apparent underlying violations of the OCP. In all cases, it is possible to represent the sequences in question as involving multiply linked segments.

The sequences illustrated below are infrequent in the corpus of nouns and adjectives examined. However, these cases illustrate how coronal and nasal sequences are allowed if the [coronal] and [nasal] nodes are multiply linked in underlying representations. The cases that cannot be multiply linked are the cases where coronal obstruents differ in their value for [±anterior] and these are true exceptions.

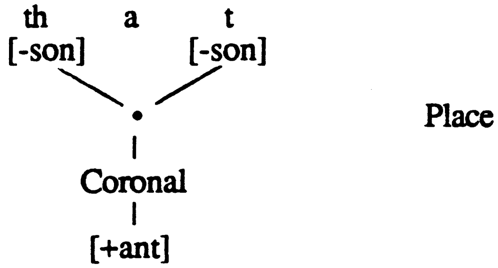
(26) a. Coronal-Vowel-Coronal; [+sonorant]

/lan/ 'fishing net'



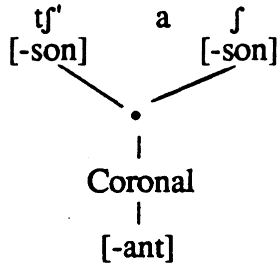
b. Coronal-Vowel-Coronal;
[-sonorant]

/t^{h}at/ 'father's sister'



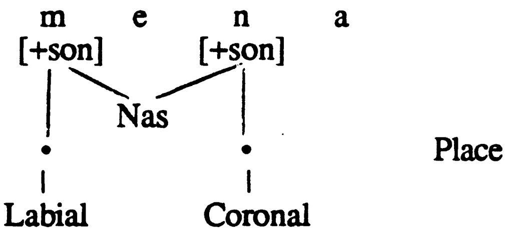
c. Coronal-Vowel-Coronal;
[-sonorant]

/tʃ'aʃ/ 'blue jay'



d. Nasal-Vowel-Nasal

/mena/ 'brown bear'



In this context, one might wonder why there is a difference between the result of OCP violations underlyingly and derivationally. Underlyingly, sequences of features which violate the OCP are multiply linked and in this way removed as violations. Derivationally, however, such multiple linkings are not an option; hence disallowed sequences result in the application of the dissimilation rules. We can assume that the rule of dissimilation in Takelma is subject to the Strict Cycle Condition (Kiparsky 1982). Therefore, it is expected that it will apply only in derived environments and not root-internally.

3.1. Summary

The analysis of Takelma Dissimilation proposed above claims that the OCP operates on adjacent coronal and nasal specifications under the [α sonorant] condition. The analysis also proposes that structure preservation is a guiding principle for the specification of fill-in values. Further, under the assumption that place of articulation features are privative, we explain the failure of labial and dorsal consonants to block dissimilation. We have seen that the proposals developed in the context of the dissimilation rules extend in an interesting way to the underlying distribution of coronal sonorants and obstruents. However, the fact that only the coronal sonorants participate in the rule of Coronal Dissimilation raises, quite naturally, the question of whether all coronal consonants are fully specified for the coronal feature, an issue we turn to in the next section.

4. Underspecification and Takelma Dissimilation

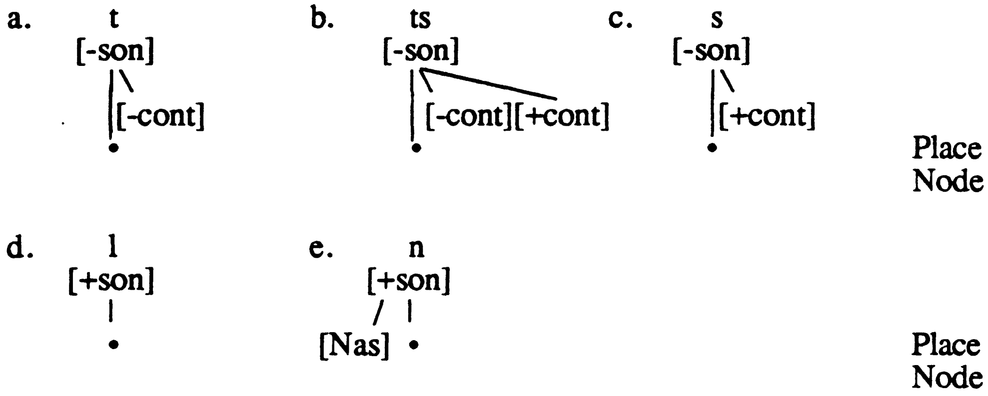
As Mester and Ito (1989) point out in a review of the issues, underspecification plays a central role in the theory of Autosegmental Phonology. In determining whether or not a segment is specified for some feature the criteria of transparency or opacity and a segment's failure to trigger or to undergo some process are often invoked. We have seen that coronal obstruents neither trigger nor undergo dissimilation on the coronal tier. We have proposed to account for this fact by claiming that the OCP is sensitive to a segment's value for sonorant. However, it would be equally plausible to assume that the assymetry is due to the absence of a coronal specification in the case of the coronal obstruents. The facts of Takelma are especially interesting in the context of underspecification because it is precisely the coronal place of articulation which has been claimed to be the unmarked or default case (Avery and Rice 1988). We now turn to an examination of the underspecification of coronal in Takelma.

There are three possibilities for coronal underspecification in the case of Takelma dissimilation. First, all coronal consonants may be unspecified for place of articulation. Second, one might claim that only coronal obstruents are unspecified for coronal thus accounting for their failure to trigger the dissimilation of the noun characteristic. Third, perhaps only the suffix-final nasal is unspecified for the coronal feature. I discuss each of these possibilities in the following sections.

4.1. Total Coronal Underspecification

Consider an analysis where all coronal consonants are unspecified for place of articulation. Avery and Rice (1988) develop such an approach to underspecification. They claim that the underspecification of coronal is inventory-driven. This means that if all coronal consonants can be distinguished without the use of the coronal articulator feature then that feature must be suppressed in underlying representations. Such a proposal results in representations of the Takelma coronal consonants as illustrated below with the place node unspecified.

(27)



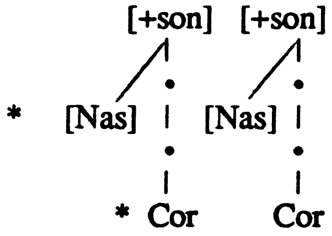
At this level of representation the OCP will have no effect on any of these segments on the coronal tier since none of them have a coronal feature. Therefore, at least some of these segments must be specified for [coronal] before the rule of dissimilation can apply.

Assume that the coronal sonorants acquire place features via the following rule: $\emptyset \rightarrow [\text{coronal}] / \underset{[+\text{son}]}{\underline{\qquad}}$. A later context-free default rule will fill in the value for [coronal] on non-sonorants. If the rule that fills in [coronal] applies to all consonants unspecified for place of articulation--non-sonorants as well as sonorants--the underspecification analysis becomes identical with the one proposed here, including the conditioning of the OCP by the feature [sonorant].

However, if we maintain the idea that there are two rules involved in specifying the articulator feature coronal, then the underspecification approach can distinguish between the sonorant and non-sonorant coronal consonants. The OCP, as stated, will correctly apply to all sequences of coronal sonorants. If these are the only coronal consonants specified for [coronal], the [α sonorant] condition on the OCP is unnecessary; it is replaced by the context-sensitive rule which inserts [coronal] for consonants unspecified for place but specified for [+sonorant].

The representation of a case which involves both Coronal Dissimilation and Nasal Dissimilation is illustrated in (28) after the rule which inserts [coronal] in [+sonorant] contexts applies.

(28) a. k^w a N + V N -> [k^walam] 'road'



At this point, there are two options available. If Nasal Dissimilation applies first, the second occurrence of [nasal] will be delinked. Coronal Dissimilation then applies and the second occurrence of [coronal] is delinked and replaced by [labial]. This results in the incorrect form: *[k^wamal]. Alternatively, if [coronal] delinking and [labial] insertion apply first and the statement of the rule of Nasal Dissimilation in (22) is adopted, then the first [nasal] specification is delinked because it is the only occurrence of [nasal] which is also characterized by [coronal] as required by the rule. If the rules apply in this order, the correct [k^walam] results. Note, however, that the formulation of Nasal Dissimilation itself requires the coronal sonorants to be specified for place of articulation.

To sum up, an approach which underspecifies [coronal] must have two separate rules which insert the value for place of articulation, one context-sensitive insertion rule and one context-free rule. Second, the rules of Coronal Dissimilation and Nasal Dissimilation must apply in that order. In contrast, if underspecification of [coronal] is not adopted but rather the OCP in Takelma is conditioned by [α sonorant], there are no default rules, context-sensitive or otherwise. Further, in an analysis which specifies coronal underlyingly, only one ordering, Coronal Dissimilation followed by Nasal Dissimilation, is possible.

4.2. Partial Coronal Underspecification

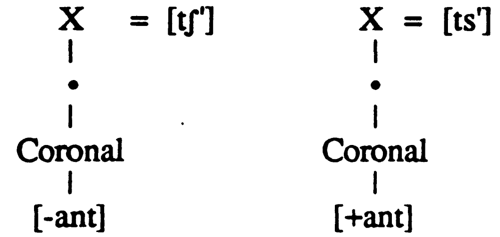
Consider then an analysis where only the [+sonorant] coronals are specified. This would account for the failure of coronal obstruents to trigger the dissimilation of the suffix final coronal nasal. An approach which accounts for the behavior of coronal obstruents by claiming that they are unspecified for the dissimilating feature predicts that all underlying constrasts within the coronal obstruent series can be captured without specifying the coronal place of articulation.

This prediction is, as mentioned earlier, difficult to test. Takelma may or may not contrast affricates in terms of [± anterior]. There is one minimal pair in the lexicon which I give below.

(29) /tʃ'aya/ 'hide' vs. /ts'aya/ 'wash'

Adopting the claim that [anterior] is a dependent of the coronal articulator node (Sagey 1986 and others) leads to the representations for the coronal affricates given in (30).

(30)



Within any theory of underspecification, at least one of these two segment types must be specified for the feature coronal underlyingly. We would then expect at least one of these segments to pattern with the coronal sonorants with respect to the rule of dissimilation. However, this sort of argument against the underspecification approach is considerably weakened by the scarcity of relevant minimal pairs.

If the underspecification of coronal obstruents is, at least in part, motivated by their high-frequency across languages and is intended to reflect their unmarked status then we would also expect that the unmarked coronal obstruents will be frequent in lexical items which contain a coronal-vowel-coronal sequence. The distributional facts of Takelma, however, show that the coronal obstruents and the coronal sonorants are equally frequent in lexical items. The underspecification approach offers no explanation for the fact that the most highly preferred sequence of coronal consonants in Takelma are just those which disagree with respect to their value for [sonorant]. This is especially striking when one takes into account the fact that five out of the seven coronal consonants in the Takelma inventory (suppressing the questionable palatals) are coronal obstruents.

Thus, under an analysis that underspecifies coronal, the distributional facts of Takelma are, while accomodated, unexplained.

4.3. Suffix Underspecification

Finally, consider an analysis where only the suffix nasal is unspecified for place of articulation and the feature [coronal] is inserted by default. In cases where the suffix nasal surfaces as [m], the rule of [coronal] insertion must be blocked. The claim would be, then, that the insertion of default values does not take place if the result creates an OCP violation. However, such an analysis hinges on the default rules of coronal insertion applying only to the suffix nasal. The specification for [coronal] in the case of sonorants must be present in the stem in order to block the insertion of [coronal] in the suffix.

There are, however, problems with an analysis which blocks the insertion of [coronal] in the suffix if the stem contains a coronal consonant. One of these is that the rule of [coronal] insertion must "look ahead" to see that the result of rule application is not well-formed. Alternatively, suppose that the rule which inserts [coronal] applies freely and creates an OCP violation. The [coronal] specification is therefore delinked. This allows the [labial] specification to be inserted. Both these alternatives are somewhat unsatisfactory. The first because rules must be given the power to look ahead; the second because it is more complicated than assuming that the [coronal] specification of both the stem consonant and the suffix-final consonant is simply present underlyingly.

5. Conclusions

In this paper we have proposed that the alternations of the Takelma noun characteristic suffix involve two cases of dissimilation: nasal dissimilation and coronal dissimilation. We adopt an approach where both nasal and the articulator features coronal, labial, dorsal are specified in underlying representation. We propose that the OCP in Takelma operates on the nasal and coronal tier and is conditioned by a segment's value for sonorant. In cases where coronal consonants agree in their value for sonorant one of the offending coronal specifications delinks. The labial specification is filled in for nasal consonants. Because nasals in Takelma can be only coronal or labial, the rule that fills in labial is structure preserving.

Analyzing Takelma dissimilation as the result of the OCP applying to sequences of [nasal] and to sequences of [coronal] place specifications governed by like values for sonorancy simplifies the rule statements and allows dissimilation to be related to more general distributional properties of underlying representations. Based on the facts of Takelma dissimilation and the underlying distribution of coronal obstruents and sonorants, I conclude that underspecification provides a less explanatory and more complicated analysis of the selective behavior of coronal sonorants.

References

Avery, Peter and Keren Rice (1989) "Segment Structure and Coronal Underspecification," *Phonology* 6: 179-200.

Chomsky, Noam and Morris Halle (1968) *The Sound Pattern of English.* New York: Harper & Row.

Clements, George N. (1985) "The Geometry of Phonological Features," *Phonology* 2: 225-252.

Cohn, Abigail (this volume) "The Consequences of Dissimilation in Sundanese."

Dresher, B. Elan (1989) "Comments on McCarthy, 'Gutteral Phonology,'" talk given at the MIT Conference on Feature and Underspecification Theories.

Goodman, Beverley (1987) "Takelma Verbal Morphology," *NELS 18*, pp. 156-174.

Kiparsky, Paul. (1982). "Lexical Phonology and Morphology," *Linguistics in the Morning Calm*, Hanshin Publishing Company, Seoul, Korea.

McCarthy, John J. (1985) "Features and Tiers: The Structure of Semitic Roots," talk presented at MIT.

McCarthy, John J. (1986) "OCP Effects: Gemination and Antigemination," *Linguistic Inquiry* 17: 207-264.

McCarthy, John J. (1988) "Feature Geometry and Dependency: A Review," *Phonetica* 45: 84-108.

Mester, A. and Junko Ito^ (1989) "Feature Predictability and Underspecification: Palatal Prosody in Japanese Mimetics," *Language* 65: 258-293.

Odden, David (1987) "Dissimilation as Deletion in Chukchi," *ESCOL '87* : 235-246.

Sagey, Elizabeth (1986) The Representation of Features and Relations in Non-Linear Phonology, Ph.D. dissertation, MIT, Cambridge, Mass.

Sapir, Edward (1909) *Takelma Texts.* Anthropological Publications of the University Museum, Vol. II, No. 1, Univ. of Pennsylvania, Philadelphia, Pa.

Sapir, Edward (1922) The Takelma Language of Southwestern Oregon, in F. Boas, ed., *Handbook of American Indian Languages.*

Steriade, Donca (1987) "Locality Conditions and Feature Geometry," *NELS 17,* pp. 595-618.

Yip, Moira (1988) "The Obligatory Contour Principle and Phonological Rules: A Loss of Identity," *Linguistic Inquiry* 19: 65-100.

Yip, Moira (1989) "Feature Geometry and Cooccurrence Restrictions," *Phonology* 6: 349-374.

The Bilabial Fricatives in Ewe: Innovation or Retention? *

Hounkpati B. C. Capo

0. Abstract

Eʋe is usually offered as a classical example of a language with a phonemic contrast between bilabial and labio-dental fricatives. As more and more data have become available on neighboring languages, a Gbe unit has been set up as one of the intermediary nodes between Kwa (ultimately dominated by Niger-Congo via Atlantic-Congo via Volta-Congo) and Eʋe, such that it also dominates Fɔn, Gen-Mina, Gun, etc. The obvious question is whether the famous contrast was attested in Proto-Gbe. While exploring alternative views, the present paper argues that there was a contrast between the antecedents of these sounds, but that the bilabial fricatives, as presently attested in Eʋe, have been innovated. It also touches upon issues related to the hierarchical representation of features and their contents.

1. Introduction

From a typological point of view, the bilabial fricatives [*ƒ*, ʋ] are less widely attested than, let us say the bilabial stops; in addition, in most languages that exhibit them, they are usually not distinctive, as they can be derived from the bilabial stops (as in Spanish, see Alarcos 1961) or from the labio-dental fricatives. Yet, right from the earliest description of Eʋe, bilabial fricatives have been recognized as phonemic since they contrast with other labials in the language. This peculiar characteristic of Eʋe makes one wonder whether it is an areal feature or a genetic one; whatever the answer is, some of our

---

* A first draft of this paper was read at a meeting of the Tuesday Colloquium at the Department of Modern Languages and Linguistics, Cornell University. I am very grateful for comments received at that time, especially from Carol Rosen, Abby Cohn and Nick Clements. Thanks are also due to the Fulbright Foundation for enabling me to spend a year at Cornell as a Scholar-in-Residence (1990-91).

received ideas in typological linguistics may be reconsidered as well as the theoretical apparatus usually proposed to account for them.

Geographically, Eʋe is mainly spoken in the southern part of Eastern Ghana (Volta Region) and Togo, yet no other Ghanaian and Togolese language described so far has been reported as having phonemically contrastive bilabial fricatives. Genetically, it has been traditionally classified as a Kwa language: here again , it seems that only some Edoid languages (now classified as Benue-Congo, whereas Eʋe remains Kwa: see Elugbe 1989, and Williamson 1989a) have the bilabial fricatives and apparently not from the same sources. In recent times, however, it has been forcefully argued that Eʋe has very closely related neighbors such as Fɔn, Gen, Gun, together with which it constitutes a Gbe node, an idea which was taken for granted in earlier accounts where "Ewe" was actually used as synonymous of Gbe even though only Eʋe was described. Even here, the bilabial fricatives do not characterise Gbe as a whole (contrary to the impression given in earlier descriptions of "Ewe"), but only the Vhe section, made up of standard Eʋe, Awlan, Pecí, Wací, Kpándo, etc. In fact Capo (1981 and 1988) crucially uses phonological evidence to suggest an internal classification of the Gbe lects as summarized in (1).

(1) An internal classification of Gbe lects:

Vhe dialects: - *e and *ɛ › e/ɛ/ə
- *t^{h} › t, and *d^{h} › d
- *ʁ^{w} > ʋ, and *χ^{w} > *f*
- *H^{w} › w
(Eʋe, Awlan, Pecí, Wací, Kpándo etc.)

Gen dialects: - *e and *ɛ › e
- *d^{h} › d, and *t^{h} › t
- *t^{s} › s, and *d^{z} › z
- *χ^{w} > p
- *H^{w} › w/ʁ^{w}
(Gẽ-Mina, Anɛxɔ, Agɔi)

Ajá dialects:
- *e and *ɛ > e
- *d^h > d, and *t^h > t
- *t^s > s, and *d^z > z
- *t > tʃ, and *d > dʒ /_u, ũ
- *H^w > w
- *ãi > ĩ ĩ, and *ɔ̃i > ũĩ

(Aja, Stádó, Hwe, Dogbó, etc.)

Fon dialects:
- *t^s > s, and *d^z > z
- *t^h > s, and *d^h > z
- *H^w > ʁ^w
- only two nominal prefixes (at most)

(Fɔn, Gun, Agbóme, Maxí, etc.)

Phla-Pherá d.:
- *t^s > s, and *d^z > z
- *t^h > s, and *d^h > z
- *H^w > ʁ^w
- more than two nominal prefixes

(Tɔli, Tɔfin, Phla, Ayizɔ, etc.)

The major question addressed in this paper is whether, given the synchronic situation in the Gbe chain, one can reconstruct the bilabial fricatives in Proto-Gbe, or whether they have been innovated in the Vhe section, as suggested in (1). Our discussion starts with the relevant data (§2), then considers the alternative hypotheses (§3), and presents additional evidence for our preferred hypothesis (§4) before highlighting the implications of our conclusion for feature geometry (§5).

2. Relevant Data

Let us start with a survey of the labial consonants in Gbe. By labiality, we here refer to the participation of at least one lip in the production of the sound: thus labial consonants include the bilabials, labio-dentals, labial velars and "labialized" consonants attested in stems. An inventory of Gbe labial consonants is thus presented in (2).

(2) Labial consonants in Gbe:

(a) common: b m f v w kp gb
(b) lect-specific: ƒ ʋ (Vhe dialects)
p (Gen dialects)
χ^{w} ʁ^{w} (Ajá, Fon and Phla-Pherá dialects, and partly Gen dialects)

Since the bilabial fricatives occur almost exclusively in the Vhe dialects, we must establish their phonemic status in those lects, without reference to other Gbe lects. Taking standard Eʋe as our reference here, we present in (3) items showing that the bilabial fricatives not only contrast with one another, but also that they contrast with all the other labials attested.

(3) Bilabial fricatives and other labials in Vhe dialects:

1. afá 'divination'
2. aƒá 'outcry'
3. aʋa 'war'
4. ava 'barn'
5. ʋu 'open'
6. wu 'kill'
7. aƒu 'ocean'
8. ʋya 'whip'
9. ƒo 'millet'
10. -kpo 'stick'
11. -ʋɔ 'door'
12 -bɔ 'cricket'
13. ʋã 'move'
14. mã 'slim down'
15. agba 'luggage'
16. ƒle 'buy'

Situating Vhe in the context of Gbe, we illustrate in (4) the correspondents of the Vhe bilabial fricatives in other Gbe lects. The data in (4) are organized in two parts: (4a) illustrates (Vhe) ƒ, ʋ : (Gen) p, ʁ^{w} : (Ajá, Fon, Ph-Ph) χ^{w}, ʁ^{w}, whereas (4b) illustrates (Vhe) ƒ, ʋ : (Gen) p, ʁ : (Ajá, Fon, Ph-Ph) χ, ʁ; this is because in (4a) we have non-rounded vowels after the consonants considered, whereas in (4b) we have rounded vowels, a situation of complementary distribution that we consider very significant.

(4) Vhe bilabial fricatives and their correspondents:
(Vhe) ƒ, ʋ : (Gen) p, ʁ^{w}: (Ajá, Fon, Ph-Ph) χ^{w}, ʁ^{w}

		Vhe	Gen	Ajá	Fon	Ph-Ph	gloss
(a)	1.	aƒá	apá	aχ^{w}á	aχ^{w}á	aχ^{w}á	'outcry'
	2.	aʋa	aʁ^{w}a	aʁ^{w}a	aʁ^{w}a	aʁ^{w}a	'war'
	3.	-ƒe	epe	eχ^{w}e	-χ^{w}e	-χ^{w}e	'year'
	4.	ʋe	ʁ^{w}e	ʁ^{w}e	ʁ^{w}e	ʁ^{w}e	'be small'

(b)	5.	υu	ʁũ	ʁũ	ʁũ	ʁũ	'open'
	6.	*f*ú	pú	χú	χú	χú	'swim'
	7.	*f*o	po	χo	χo	χo	'beat'
	8.	-υɔ	eʁɔ̃	eʁɔ̃	-ʁɔ̃	-ʁɔ̃	'door'

Given the fact that in (4b), Vhe *f*, υ correspond to χ, ʁ in Ajá, Fon and Phla-Pherá dialects, it becomes necessary to show that χ and ʁ are common to all Gbe lects, although not in the items illustrated in (4); this is done in (5), where these sounds occur before rounded (5b) as well as non-rounded (5a) vowels.

(5) Back (velar/uvular) fricatives in all Gbe lects:

		Vhe	Gen	Ajá	Fon	Ph-Ph	gloss
(a)	1.	aχa	aχa	aχa	aχa	aχa	'rib'
	2.	aʁa	aʁã	aʁã	aʁã	aʁã	'drink, n.'
	3.	χe(ví)	χeví	χeví	χɛ	χɛ	'bird'
	4.	ʁe	ʁẽ	ʁẽ	ʁɛ̃	ʁɛ̃	'breed'
(b)	5.	-χɔ	eχɔ	eχɔ	-χɔ	-χɔ	'hut'
	6.	ʁo	ʁo	ʁɔ̃	ʁɔ̃	ʁo	'uproot'
	7.	-χó	eχó	eχó	χó	-χó	'history'
	8.	ʁũ	ʁũ	ʁũ	ʁũ	ʁũ	'mystery'
	9.	aʁɔ̃́	aʁɔ̃́	aʁɔ̃́	aʁɔ̃́	aʁɔ̃́	'navel'

Given the fact that in (4) we established a correspondence series *f*, υ : χ^w, $ʁ^w$, it would be good to establish the monophonemic status of χ^w and $ʁ^w$ in the lects in which they occur; this is all the more necessary since in (4a) we have χ^w, $ʁ^w$, but in (4b) χ, ʁ in the same lects. We do this by shifting our attention to Gun, a Fon dialect, for which we present data in (6), using two arguments: the fact that χ^w and $ʁ^w$ occur in most grammatical categories (6a), and their behavior in reduplication (6b): for details see (Capo 1978).

(6) Back labialized fricatives in Fon, Ajá and Phla-Pherá dialects:
Gun as an example to show their monophonemicity

(a) grammatical categories and the presence of /a, ã/

1. χ^{w}e 'go'
2. ʁ^{w}e 'be small'
3. aχ^{w}á 'outcry'
4. ʁ^{w}ã 'move'
5. χ^{w}lɛ́ 'plane, v.'
6. ʁ^{w}lɛ̃ 'save'
7. χ^{w}ì 'line'
8. -ʁ^{w}ɛ 'jugement'

(b) sample of reduplication

	stem	reduplication	gloss
1.	ʁ^{w}ã	ʁ^{w}ĩʁ^{w}ã	'move'
2.	ʁ^{w}e	ʁ^{w}iʁ^{w}e	'be small'
3.	χ^{w}á	χ^{w}iχ^{w}á	'be half-ripe'
4.	bɛ	bibɛ	'hide'
5.	sa	sisa	'sell'
6.	blá	blablá	'tie'
7.	χyá	χyaχyá	'dry, v.'
8.	gblé	gblegblé	'spoil'
9.	ʁ^{w}lɛ̃	ʁ^{w}lɛ̃ʁ^{w}lɛ̃	'save'

Comparing relevant data in (5) and (6), one realizes that χ^{w} and ʁ^{w} contrast with χ and ʁ; however, there is a distributional gap in that whereas χ and ʁ occur before both rounded and non-rounded vowels (in all lects), χ^{w} and ʁ^{w} occur only before non-rounded vowels. In Capo (1981), we have postulated a synchronic rule in Fon, Ajá and Phla-Pherá lects delabializing /χ^{w}, ʁ^{w}/ before rounded vowels; it appears now that, instead of a P-rule, we must account for this gap by a Morpheme Structure Condition, because a verb like [χo] 'beat', which we interpreted as /χ^{w}o/, reduplicates as [χiχo], not [χ^{w}iχo] as we would have expected, given the fact that the reduplicative vowel *i* is not an appropriate environment for the delabialization rule. We present the suggested MSC informally as (7).

(7) $$^{*}\left\{\begin{matrix}\chi^{w}\\ \text{ʁ}^{w}\end{matrix}\right\}\begin{bmatrix}V\\ +\text{round}\end{bmatrix}$$

3. Interpretations

3. 1. Guiding Principles

We shall now consider three alternative hypotheses as to the sources of the bilabial fricatives being considered in this paper. Our hypotheses are based on a number of principles outlined in (8).

(8) Some guiding principles:
- (a) There is no majority rule;
- (b) Dialect distribution *per se* is not important;
- (c) Changes must be seen in terms of rules/processes;
- (d) Rules should be evaluated on the basis of plausibility, naturalness, and predictability;
- (e) Pattern congruity may enhance the direction of change;
- (f) Proto-segments may coincide with attested phonemes in at least one daughter language, but they may also not surface in any of the daughter languages.

3. 2. First Hypothesis

Suppose that Proto-Gbe had *p and *B, or *'p and 'b (where B stands for an indeterminate voiced bilabial stop, and 'p, 'b stand for lenis bilabial stops), then we would need the diachronic rules in (9) to derive the modern reflexes:

(9) Diachronic rules if the Proto-Gbe phonemes were *p,*B or *'p,*'b

a) *p > *f* or *'p > *f* (Vhe)
*B > ʋ or *'b > ʋ (Vhe)

i.e. [-cont] > [+cont]

b) *p > χ^{w} or *'p > χ^{w} (Aj, Fo, Ph-Ph)
*B > ʁ^{w} or *'b > ʁ^{w} (Ge,Aj,Fo,Ph-Ph)

i.e. [+lab, -cont] > [+rnd, +cont, +back]

According to this hypothesis and its entailed diachronic rules, Gen lects are partly conservative (in that they retained basically the voiceless bilabial stop), while Vhe lects on the one hand, and on the other Ajá, Fon and Phla-Pherá lects are innovative. We may point out, however, that whereas the process depicted in (9a) seems a reasonable innovation (as a lenition process changing stops to fricatives, even though the process also occurs word initially), the one in (9b) seems hard to justify in that we fail to see what would motivate the backness and roundness of the reflexes (in addition to their fricative nature) in (Gen), Ajá, Fon and Phla-Pherá dialects.

3. 3. Second Hypothesis

Suppose that Proto-Gbe had *f and *ʋ , then we would need the diachronic rules in (10) to derive the modern reflexes:

(10) Diachronic rules if the Proto-Gbe segments were *f,*ʋ
a) *f > p (Gen), i.e. [+cont] > [-cont]
b) *f > χ^{w} (Ajá,Fon,Ph-Ph)
*ʋ > ʁ^{w} (Gen,Ajá,Fon,Ph-Ph), i.e. [+lab] > [+back, +rnd]

According to this hypothesis and its entailed diachronic rules, Vhe lects are conservative (in that they have retained the bilabial fricatives, as attested in the parent language), while Gen lects on the one hand, and Ajá, Fon and Phla-Pherá lects on the other are innovative. We may point out, however, that whereas (10a) can be argued for on the ground that the strengthening may be due to the absence of /p/ in the parent language (pattern congruity), (10b) seems hard to justify in that we fail to see what would motivate the backness and roundness of the reflexes in (Gen), Ajá, Fon and Phla-Pherá dialects.

3. 4. Third Hypothesis

Suppose that Proto-Gbe had *χ^{w} and *ʁ^{w}, then we would need the diachronic rules in (11) to derive the modern reflexes:

(11) Diachronic rules if the Proto-Gbe segments were *χ^{w}, *ʁ^{w}
a) *χ^{w} > f (Vhe)
*ʁ^{w} > ʋ

i.e. [+back, +rnd] > [+lab, (-rd)]
b) *χ^{w} > p (Gen)
i.e. [+back, +rnd, +cont] > [+lab, -cont, (-rd)]

According to this hypothesis and its entailed diachronic rules, Ajá, Fon and Phla-Pherá dialects are conservative (in that they have retained the labialized velar/uvular fricatives of the parent language), while Gen lects are partly innovative (p being a creation) and Vhe lects innovative. We would like to point out that both (11a) and (11b) seem plausible in that: (i) the labiality of the output segments was already present in the input segments as roundness; (ii) the backness of the input segments was exchanged for the reinforcement of the labiality in the output segments; (iii) the output in (11b), i.e. the fact that we have a stop, can be argued for on account of pattern congruity (see (10) above).

There is a problem associated with this hypothesis, however, in that it claims that the MSC proposed in (7) for (Gen), Fon, Ajá and Phla-Pherá dialects would not apply to Proto-Gbe, since it clearly reconstructs *ʁ^{w}ũ 'open', *χ^{w}ú 'swim', *χ^{w}o 'beat' and *-ʁ^{w}ɔ̃ 'door' for items 5-8 in (4) above, as opposed to *-χɔ 'hut', *ʁõ 'uproot', *-χó 'history', *-ʁũ 'mystery' and *aʁɔ̃́ 'navel' for items 5-9 in (5) above. That is the only way to arrive at the Vhe reflexes.

4. Additional Evidence for the Third Hypothesis

The above-mentioned problem is not an insurmontable one. In fact, our preferred hypothesis is that the Vhe *f*, ʋ indeed developed from Proto-Gbe */χ^{w}, ʁ^{w}/, and we present below additional evidence to support this stand and clear up possible objections.

The first set of evidence comes from the fact that in a few Vhe words where one would have expected *f* and ʋ, we have χ and ʁ instead; it seems reasonable to view these items as relics to which the Vhe-specific diachronic rules failed to apply. Some of these items are presented in (12) below.

(12) Failure to obtain *ƒ* and ʋ in a few words in some Vhe dialects:

	Awlan	Pecí	Kpándo	Wací	gloss	Fon
1.	ʋɔtrú	ʁɔtrú	ʁɔtrú	ʋɔtrú	door	ʁɔ̃
2.	aʋatʃo	aʁatʃo	aʁatˢo	--	liar	--
3.	aƒélí	aƒélí	aƒélí	aχʷélí	a deity	aχʷélí

(see also Nutsugan 1975 where more compounds are found)

The second set of evidence comes from the fact that in all present-day Gbe lects, there is a coalescence rule that gives rise to labialized consonants, including labialized velar/uvular consonants. There is every reason to believe that such a coalescence rule was already present in Proto-Gbe, and therefore χʷ and ʁʷ could be synchronically derived in Proto-Gbe as phonetic entities; if this is accepted, the fact that they became phonemicised would not be unexpected. Examples of such synchronic labialized consonants are presented in (13) from different dialects.

(13) Derivation of labialized consonants in all Gbe lects:

(a) Wací as an example:

1.	/χɔ + i /	→	[χɔɛ] ~ [χwɛ] ~ [χʷɛ]	'take it'
2.	/tu + i /	→	[tui] ~ [twi] ~ [tʷi]	'grind it'
3.	/ekɔ + a/	→	[ekɔa] ~ [ekwa] ~ [ekʷa]	'the neck'

(b) Agbóme as another example:

4.	/ʁu + i/	→	[ʁui] ~ [ʁwi] ~ [ʁʷi]	'kill it'
5.	/gbo + i/	→	[gboe] ~ [gbwe] ~ [gbʷe]	'cut it'
6.	/ɖṹ + ɔ́/	→	[nṹɔ̃́] ~ [nwɔ̃́] ~ [nʷɔ̃́]	'the thing'

(c) Awlan as yet another example:

7.	/χoχó + i/	→	[χoχúí] ~ [χoχwí] ~ [χoχʷí]	'old'
8.	/kɔkɔ + i/	→	[kɔkoe] ~ [kɔkwe] ~ [kɔkʷe]	'saint'
9.	/tʃí + go + í/	→	[tʃígúí] ~ [tʃígwí] ~ [tʃígʷí]	'tube'

The third set of evidence, and by far the strongest, comes from the curious status of /χʷ/ in Gen dialects. Synchronically, although there is a /ʁʷ/ phoneme in Gen dialects, /χʷ/ occurs in only two stems and would be at best treated as a marginal phoneme. Within the hypothesis adopted, we expect the Gen lects to evolve Proto-Gbe *χʷ to /p/. This is exactly what happens, as illustrated in (4) above. There is, however, a very common word, the one for 'home', that is rendered in Gen by /aχʷé/ and not by /apé/ as expected; what makes the situation interesting is that we do have the expected /apé/ in some compounds, as illustrated in (14) below, together with a near synonym /aχó/ which corresponds to /aχʷé/ in Fon dialects, for example.

(14) The curious status of /χʷ/ in Gen:

	Fon	Vhe	Gen	gloss
1.	aχʷé	aƒé	aχʷé	house/home
2.	χʷétɔ́	aƒétɔ́	aχʷétɔ́	landlord
3.	(mɛɖaχó) X	aƒétɔ́ X	apétɔ́ X	Mr. X
4.	(gã́)	aƒétɔ́	apétɔ́	boss
5.	aχʷé	aƒé	aχó	home/residence
6.	χʷlé	(ƒlé)	χʷlé	offer in sacrifice

The near synonymity of /aχʷé/ and /aχó/ in Gen is illustrated in (15), where Gen utterances are compared with Fon and Vhe utterances of the same meanings.

(15) Gen utterances compared with Fon and Vhe

Gen:	1a:	é yi aχʷé	'he went home, i.e. to his home town'
	1b:	é yi aχó me	'he went home, i.e. to his residence'
	1c.	é yi Kofí bé aχó me	'he went to Kofi's house'
Fon:	2a:	é yi χʷé	'he went home, i.e. to his home town'
	2b:	é yi χʷé gbe	'he went home, i.e. to his residence'
	2c:	é yi Kofí sín χʷé gbe	'he went to Kofi's house'
Vhe:	3a:	é yi aƒé	'he went home, i.e. to his home town'
	3b.	é yi aƒé me	'he went home, i.e. to his residence'
	3c.	é yi Kofí ƒé aƒé me	'he went to Kofi's house'

What can we make out of (14) and (15)? First, it seems reasonable to argue that in Gen /aχʷé/ 'home' is a relic of the Proto-Gbe form /aχʷé/ because whereas the Gen-specific diachronic rule failed to apply, the Vhe-specific diachronic rule did apply. Secondly, the very fact that in most other Gbe lects we have the same phonological form to cover both Gen /aχó/ and /aχ^{w}é/ is an indication that our intuition of near synonymity is a valid one. In fact, it leads us to believe that if the two items are not the same, they are nonetheless definitely related. Now, let us assume that items 1 and 5 in (14) were two different lexical items in early Proto-Gbe, yet related such that 1 was derived from 5 through an *{-i}- suffix (discussed in Capo 1983), i.e. */aχó/ and */aχói/, the latter realized as *[aχóé] ~ *[aχwé] ~ *[aχ^{w}é] (just like in (13) above); this would explain the situations in the various lects in two ways. (i) Gen dialects would typically reflect the situation in Proto-Gbe, and hence the basically same meaning of the two forms but used in different contexts; (ii) in other lects, only the form with the {-i} suffix would have survived and the morphological derivation been consequently blurred so that the [χ^{w}] would have been seen as a unit phoneme instead of a sequence of /χ+o/ followed by another vowel. Our assumption is borne out by independent evidence in Ajá dialects, where a similar, if not identical form, /-χ^{w}é/ used as a suffix (or better still a second element of a compound) in Fon and even Gen dialects, is rendered as /χú/ in Ajá dialects in place-names, as evidenced in (Ajá) Akplaχú : (Fon) Akplaχ^{w}é : (Wací) Akpla*fé* 'name of a settlement founded by Akpla.' Note that current thinking (e.g. Heine and Reh 1984) claims that the locative suffix -χ^{w}ɛ/-*f*ɛ is a bleached noun aχ^{w}ɛ/a*f*ɛ, i.e. item 1 in (14) above. However, based on our knowledge of the synchronic phonology and morphology of the Ajá dialects, we can only postulate the underlying form of [akplaχú] as /akpla+aχó+i/; this is because, not only is the {-i} suffix realized as [-i] after /o/, it also closes /o/ (and /e/) to [u] (resp. [i]), and in most cases, the {-i} suffix itself gets deleted, its presence being recoverable through the vowel closing rule (see also Capo 1985). Note, in this regard, that Ajá dialects also have the [aχó] ~ [aχ^{w}é] pairing in the lexicon (see Tchitchi 1984 for a similar, but partly different account).

The three sets of evidence confirm that we cannot but reconstruct *χ^{w} and *ʁ^{w} in Proto-Gbe, and therefore treat the Vhe bilabial fricatives as innovations. In addition, we have seen how [χ^{w}] and [ʁ^{w}] themselves emerged as coalesced forms of χ and ʁ plus back rounded vowels when followed by another vowel, and there was an indication of how the biphonemic complex became reinterpreted as monophonemic. The monophonemic re-interpretation is crucial to the evolution into *f* and ʋ in the Vhe dialects.

5. Related Issues

This account seems compatible with the feature geometry being worked out by Clements (see references). In particular it supports the idea that the same feature [labial] can occur at the C-place tier as well as the V-place tier. Clements (1989) has specifically proposed a tier promotion rule of the form reproduced in (16).

(16) Tier promotion (Clements 1989):
[labial]: V > C
Complex Segment Simplification: yes (unmarked).

This rule claims that the labiality occurring at the V-place tier on a consonant can be promoted to the C-place tier, i.e. a secondary labial articulation may evolve to a primary labial articulation; in other words a "labialized consonant" can become a true labial consonant. That is exactly how one would derive diachronically the Vhe bilabilal fricatives and, with a strengthening process motivated by pattern congruity, the Gen voiceless bilabial stop within the framework of the hypothesis we argued for above. Such a scenario can be be better understood from the feature matrices outlined in (17).

(17) Sound shifts illustrated with the voiceless series

		χ	χ^w	*f*	p
$^*\chi^w > f > p$	C-place				
$^*ʁ^w >$ ʋ	labial	-	-	+	+
	dorsal	+	+	-	-
	V-place				
	labial		+		

One immediate implication of the tier promotion concept is that Gen /p/ would not be a direct reflex of $^*\chi^w$, but would necessarily transit via *$^*f$*. Does that also imply that there is a common ancestor of Gen and Vhe dialects below the Proto-Gbe node? And if that is the case, would the *ʋ of that ancestor (from Proto-Gbe *ʁ^{w}) revert to /ʁ^{w}/ in Gen dialects? It would be difficult to answer those questions in the affirmative, given the fact that Gen dialects share other important characteristics with Ajá, and even Fon and Phla-Pherá dialects, as can be observed in (1) above. The other alternative would be to claim that *f* and ʋ were not innovated at the same time.

In addition, if the MSC postulated in (7) for (Gen), Fon, Ajá and Phla-Pherá dialects is to be maintained, it must be properly understood as the result of a diachronic P-rule (similar to my initial synchronic P-rule), viz (18):

(18)
$$\begin{Bmatrix} {}^{*}\chi^{w} \\ {}^{*}\text{ʁ}^{w} \end{Bmatrix} > \begin{Bmatrix} \chi \\ \text{ʁ} \end{Bmatrix} \;/\; \_\_ \begin{bmatrix} \text{V} \\ +\text{round} \end{bmatrix}$$

Otherwise it may have two interpretations, being the source of two potential rules in case a sequence of χ^{w}, ʁ^{w} plus a rounded vowel would be expected; in other words one of the following repair strategies will apply:

(19) Two ways of correcting the violation of the MSC:
(a) either χ^{w}, ʁ^{w} surface, but the vowels change to their corresponding nonback nonrounded counterparts; or
(b) χ^{w}, ʁ^{w} change to χ, ʁ and the back rounded vowels surface.

Another related issue concerns the attestation of the bilabial fricatives in other Gbe lects. Our initial fieldwork reported in Capo (1981) showed that these sounds were noticed in Alada and Ayizɔ (Phla-Pherá dialects) and in Dogbó (an Ajá dialect), but only before the high front vowels /i, ĩ/ and the yod /y/; in such a context they are in free variation with /f, v/, and we have derived them from the latter, even though we were wondering what could be the motivation for such a rule occurring in a typical "palatalization environment." More recently, Anago (in prep.) has reported that in Tɔfin (a Phla-Pherá dialect) the bilabial fricatives have indeed phonemic status and that the language does not have underlying labio-dental fricatives (because the latter occur only in the speech of the bilingual town dwellers but not with the rural monolingual speakers). On comparative basis, however, it is clear that hese Tɔfin bilabial fricatives do not have the same sources as those of the

Vhe dialects, because they correspond to f,v in all other Gbe lects (including Vhe dialects), as illustrated in (20).

(20) *f* and ʋ in a few other Gbe dialects (Tɔfin and Ayizɔ́),
based on Anago (in prep.) and Capo (1981)

	Wací	Ayizɔ́	Tɔfin	Agbóme	gloss
1.	afɔ	afɔ	a*f*ɔ	afɔ	foot
2.	fá	fá	*f*á	fá	be cool
3.	fɔ̃́	fɔ̃́	*f*ɔ̃́	fɔ̃́	wake up
4.	efe	ɔfɛ̃	ɛ*f*ɛ̃	fɛ̃	nail
5.	efú	ofṹ	o*f*ṹ	fṹ	hair
6.	afí	a*f*í	a*f*í	afí	ash
7.	vo	vo	ʋo	vo	be free
8.	vɔ	vɔ	ʋɔ	vɔ	finish
9.	vé	vɛ́	ʋɛ́	vɛ́	be bitter
10.	eví	oʋí	oʋí	ví	child
11.	a*f*é	oχʷé	ɔχʷé	χʷé	home
12.	aʋa	aʁʷa	aʁʷa	aʁʷa	war

In my previous works, I have reconstructed the correspondence sets illustrated in (20) as *f, *v in Proto-Gbe. From that point of view, Tɔfin would have the diachronic rule informally stated as (21):

(21) Emergence of /*f*, ʋ/ in Tɔfin:
*f, *v > *f*, ʋ

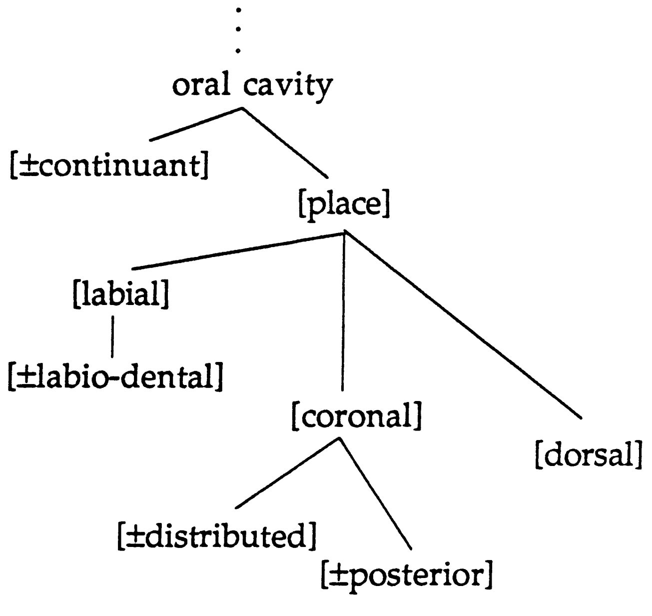
How natural is (21), especially as it is a context-free rule? Note that if one adopts such a rule, it would be nice to argue that it is shared by Alada, Ayizɔ and Dogbó also, except that in those dialects, it would only apply before /i ĩ y/.

Although I would like to continue with this view, perhaps one can also speculate that *f* and ʋ were widely attested in the Gbe chain as a whole and might be reconstructed in Proto-Gbe. With this second alternative, Tɔfin (and Alada, Ayizɔ and Dogbó) would be conservative whereas most of the present-day Gbe dialects would have applied the diachronic rule informally stated in (22).

(22) Emergence of /f,v/ in most present-day Gbe lects:
*_f_, *ʋ > f, v

This alternative seems attractive because typologically many languages have the labio-dental fricatives and not the bilabial fricatives, which means that if a language (such as Proto-Gbe) had the bilabial fricatives but not the labio-dental fricatives, chances are that the bilabials would evolve into the labio-dentals (on cross-language analogical grounds). Should this be the correct prediction, the theory of phonology should incorporate it into its formal apparatus. This is, however, difficult as at now, because in the stop series it is the bilabials that are the most widely attested, and the geometry proposed in Clements (1990) assumes, I believe, that the default value for [labial] is bilabial, since labio-dentals need to be so specified explicitly. We reproduce in (23) the relevant portion of the feature geometry proposed by Clements (1990).

(23) Feature geometry (Clements 1990):



6. Conclusions

We would like to suggest here that perhaps late Pre-Gbe/early Proto-Gbe had both *_f_, *ʋ and *χʷ, *ʁʷ, but not *f, *v. This suggestion implies the following scenarios outlined in (24).

(24) Some sound changes in the history of Gbe:

a) Middle Proto-Gbe had innovated *ƒ, *ʋ > f, v (perhaps on the basis of snobism/language contact); by the time of this change, the Tɔfin speakers were already in the process of migration, hence they maintained the earlier state (without f,v but with ƒ,ʋ) !

b) Proto-Vhe (one of the daughter languages of Proto-Gbe) had innovated *χʷ, *ʁʷ > ƒ, ʋ; apart from its naturalness as argued earlier, this change was facilitated by the fact that Proto-Gbe *ƒ, *ʋ had already shifted to f, v (as in (a)), and so there was no merger, nor confusion.

c) Probably independently, Fon, Ajá and Phla-Pherá dialects have innovated *χʷ, *ʁʷ > χ, ʁ / __ [+round] vowel.

d) Assuming the wave theory model, the Proto-Vhe innovation spread to Gen only for the voiceless fricative where it was taken a step further, hence *χʷ >ƒ > p, while the "delabialization rule" also spread to Gen from the other end in respect of the voiced fricative.

The discussion in this paper has raised at least two interesting issues. First, if indeed the unmarked value for [labial] is [-labio-dental], i.e. bilabial as implied in (23), why is it that /ƒ, ʋ/ are rare in the languages of the world, whereas /p, b/ are quite common (with due recognition that stops are more widely attested than fricatives)? Or is it the case that, as borne out by the typological patternings observed, the unmarked value for [labial] should be correlated with the value of [continuant]? Notice however, that since [labial] is dominated by [place] which is a coordinate branch of [continuant], this formal correlation is difficult to express. Obviously, more research is needed in this area. The second issue raised has to do with data (20), especially with the ƒ, ʋ shown by Ayizɔ. If one adopts my earlier analysis postulating a synchronic rule /f v/ → ƒ ʋ /__ i, y, the question would be: what is the feature shared by both [ƒ, ʋ] and [i, y]? If one adopts the alternative considered in (22), that is a diachronic rule *ƒ, *ʋ > f, v (except before i, y in some dialects), the question would be: what in /i, y/ is responsible for the non-application of the sound shift? Thus in both alternatives, there is a "hidden principle" yet to be uncovered with regard to the theory of distinctive features.

References

Alarcos Llorach, E. 1961. *Fonología española*. Madrid: Gredos.

Anago, P. (in prep). Etude phonologique comparée du tɔligbe et du tɔfingbe, Master's thesis, Univ. Nat. du Bénin.

Bendor-Samuel, J. 1989. *The Niger-Congo Languages*. Lanham: University Press of America.

Capo, H.B.C. 1978. "A propos des corrélations de labialisation et de palatalisation en fon," *Annales de l'Université d'Abidjan* (H: Linguistique) XI,1: 5-19.

___1981. A diachronic phonology of the "Gbe" dialect cluster. Unpublished doctoral dissertation, Univ. of Ghana, Legon.

___ 1983. "I et U en hwe et leur place dans la reconstruction du proto-gbe," *Journal of West African Languages* 13,1: 3-8.

___ 1985. "Palatalisation et fermeture vocalique en hwe," *Africana Marburgensia* 18,1: 24-40.

___ 1988. *Renaissance du gbe*. Hamburg: Buske.

Clements, G.N. 1985. "The Geometry of Phonological Features,"*Phonology Yearbook* 2: 225-252.

___1989 "A Unified Set of Features for Consonants and Vowels" (preliminary version), ms., Cornell University.

___ 1990. "A Model of Feature Organization," ms., Cornell University.

Elugbe, B.O. 1989. *Comparative Edoid: Phonology and Lexicon*. Port Harcourt: Univ. of Port Harcourt Press.

Heine, B. and M. Reh 1984. *Grammaticalization and Reanalysis in African Languages*. Hamburg: Buske.

Nutsugah, C. 1975. A comparative phonological and lexical study of Aŋlɔ, Peki and Kpándo dialects of Ewe, ms., Univ. of Ghana, Legon.

Real Academia Española 1973. *Esbozo de una nueva gramática de la lengua española*. Madrid: Espasa-Calpe.

Tchitchi, T.Y. 1984. *Systématique de l'àjágbe,* unpublished doctoral dissertation, Université de Paris 3.

Williamson, K. 1989a. "Niger-Congo Overview," in Bendor-Samuel 1989, 3-46.

___ 1989b. "Benue-Congo Overview," in Bendor-Samuel 1989, 247-274.

On the Representation of Clicks

Ann R. Bradlow

0. Introduction

At the level of phonetic description, click consonants involve a double articulation, and a timing of articulatory gestures which results in an ingressive, velaric airstream mechanism. Phonological analyses regarding the linguistic representation of clicks vary in the status - major or minor - they afford each of the two place of articulation specifications. These segments are typologically unusual in that they occur only in certain language families of southern and eastern Africa. This paper combines information about the phonetic description and phonological patterning of clicks, with typological universals of doubly articulated segments, to argue that clicks are back consonants with a secondary front closure.

The most complex and extensive inventory of click consonants described in the literature is found in !Xóõ, a Bushman language spoken in Botswana and Namibia. This language has a total of 80 distinctive clicks, being almost double the number of clicks in Zu/ʔhoãsi (also known as !Xu), the next most extensive click language. Due to this complexity, !Xóõ is a particularly interesting language to study, and owing to the extensive fieldwork of Dr. Anthony Traill, data on this language is readily available. For these reasons, this paper will focus primarily on data from !Xóõ.

The remainder of this paper is structured as follows. In section 1, I provide a brief phonetic description of clicks. Section 2 reviews the main points of the various phonological feature analyses of clicks. Section 3 discusses the notion of primary and secondary articulations and how it bears on the representation of clicks. This section also proposes an analysis of clicks in which they are represented as back consonants with a secondary front articulation. Section 4 discusses the clicks of the Khoisan language, !Xóõ, and their relationship to the non-click consonants of the language. The comparison of the click and non-click systems of this language will be shown to provide further support for the analysis proposed in section 3. Finally, section 5 provides a summary and conclusions of the discussion.

1. Phonetic description of clicks.

In the articulation of clicks, there are two points of closure: one towards the front of the oral cavity and one farther back. The familiar and well described (Beach 1938, Ladefoged

1975, Traill 1985) mechanism by which a click is produced involves the influx of air into the oral cavity upon the release of the front closure. This occurs as a result of the decrease in air pressure in the space between the lowered middle part of the tongue and the roof of the mouth. The lowering of the tongue is a suction movement which causes the rarefaction of the air between the two closures. The release of the back closure is achieved with the efflux of a pulmonic or glottalic airstream. A requirement for the articulation of clicks is that both closures are in place before the lowering of the tongue body, and that the front closure is released prior to the release of the back closure. It is precisely this ordering of events, in conjunction with the suction movement in the oral cavity, which produces the influx of air which we hear as a "click."

Palatographic studies of the clicks of !Xóõ presented in Traill (1985) show the five influx places of articulation. These can be described as bilabial, dental, post-dental/alveolar, palatal and lateral. Traill avoids the term "alveolar" since a common physiological feature of the San population is the absence of a conventional alveolar ridge; in this population the palate tends to slope back smoothly. Examples of each of the five places of articulation for the front closure of the clicks in !Xóõ are given below, along with their IPA symbols.

Table 1. Examples of the 5 influxes in !Xóõ (Traill p 124)

	symbols	example	gloss
bilabial	ʘ	ʘôõ	'dream'
dental	ǀ	ǀâa	'move off'
post-dental	ǂ	ǂàã	'knock'
palatal	!	âã	'wait'
lateral	ǁ	ǁáã	'poison'

An articulatory distinction between the dental and post-dental clicks on the one hand and the palatal and lateral clicks on the other, is that the former are laminal whereas the latter are apical. Another phonetic detail which sets the labial, dental and lateral clicks apart from the post-dental and palatal clicks is that the former are released with friction whereas the latter are released instantaneously. These phonetic distinctions are summarized in table 2 below. Traill observes 16 distinctions along the place and manner dimensions for the articulation of the back closure in the clicks of !Xóõ. There is a contrast between a velar and uvular place of articulation and the efflux may be achieved with varying degrees of closure, oral or nasal voicing, pulmonic or glottalic airstreams and various degrees of aspiration, which can precede, follow or be simultaneous with the release. Combined with the five possible influxes this yields a total of 5x16=80 clicks in this language.

Table 2. Phonetic features of the influxes.

	apical	laminal	fricated release
labial	-	-	+
dental	-	+	+
post-dental	-	+	-
palatal	+	-	-
lateral	+	-	+

A crucial phonetic detail in the production of clicks is the relative timing of the closures and releases of the front and back articulations. Maddieson and Ladefoged (1989) discuss the possible timing relations of multiply articulated segments such that both articulations are audible and neither is subsumed by the other. Figure 1 below provides a schematic diagram of the possible timing relations.

(1) (a) dorsal |----------| (b) dorsal |------------------|
coronal |--------| coronal |----------|

In the case of non-click, multiply articulated segments the relative timing of the two gestures is necessarily as shown in figure (1a), where the solid lines indicate the duration of the coronal and dorsal closures respectively, in undefined units of time. This timing is required in order for there to be robust cues to the existence of two separate, articulatory gestures. In the case of clicks, the timing of the two gestures is as shown in (1b). The back closure must be in effect throughout the articulation of the front closure in order to facilitate the characteristic velaric airstream mechanism. It is the click sound which results from the velaric airstreams which provides the phonetic cue to the presence of a multiple articulation. In the simplest case, the release of the back closure will take place immediately after the release of the front closure. In this case the onset of the following vowel occurs immediately after the front release. This is the case of the "basic" click with the unaspirated, voiceless velar stop accompaniment. Similarly, the nasal and voiced clicks, display an essentially immediate onset of the following vowel. The rest of the click accompaniments involve the production of phonetic material which is audible between the suction, clicking sound and the onset of the following vowel.

2. The phonological feature analysis of clicks.

In "Preliminaries to Speech Analysis," Jakobson, Fant and Halle (1963) provide very brief comments on the feature specifications for clicks. In their system, clicks, and other doubly articulated consonants are "...but special forms of consonant clusters. They are extreme cases of co-articulation" (page 23). Thus in this system there is no need for a special feature [suction] for it is the timing of the releases and their extreme contraction in time which produces the influx. Similarly, the issue of primary and secondary articulations is not relevant in this feature system since clicks are treated as consonant clusters.

Trubetzkoy (1969) treats clicks as a correlation that arises as a result of a secondary series of localization entering into an opposition with its corresponding basic series. The click correlation is distinct from the correlations of palatalization, velarization and labialization in that the latter are all correlations of timbre with specific, vocalic-like "colorings," whereas the suction property of clicks is accounted for as a phonetic detail. The supplemental velar closure which is present in all click consonants is the feature which is of phonological importance, however it is distinct from the correlation of velarization in that the closure is consonantal, rather than vocalic, in nature. Trubetzkoy points out that a velar closure is present in the gutturalized and labiovelarized consonants of other languages, "though perhaps not in quite as energetic a form." (p. 138) This view of clicks as consonants with a special type of secondary articulation is the precursor to the view taken by Chomsky and Halle in S.P.E. And, the observation that a parallel can be drawn between the inventory of click and non-click consonants within a language is echoed in Traill's (1985) study of !Xóõ.

In SPE, Chomsky and Halle introduce the feature [suction]. They characterize clicks as ".. noncontinuants with extreme velarization" (p. 319). Thus in this feature system, clicks are treated as consonants with secondary velarization. The feature [suction] is classified as a manner of articulation feature which involves a supplementary movement. A specification of [+suction] is the mark of a movement within the vocal tract which results in a decrease in pressure at the time when both closures have been achieved. It is in opposition to the feature specification [+pressure] which indicates a movement in the vocal tract which results in an increase in supraglottal pressure. The introduction of this feature into the SPE system provides a means of distinguishing click consonants from other multiply articulated consonants. For example, they note that the Guang languages have consonants which, according to Ladefoged, combine a labial and velar articulation with no suction. The Yoruba labio-velars are intermediates between the Guang type of doubly articulated consonants and the true clicks in that they involve a movement of the back articulator which decreases the oral cavity pressure and results in a suction feature of the front closure. However, this is achieved while there is still an outward flow of air from the lungs. In other words these doubly-articulated segments are produced with a combination of ingressive and egressive airflow. The sequencing of closures and releases for the true clicks is such that the suction is achieved while the back closure is still in effect and the flow of air upon release of the front closure is ingressive. Thus in this system the feature [suction] distinguishes different types of multiple articulations, and clicks are treated as [+suction] consonants with secondary velarization.

In his study of the !Xóõ language, Traill (1985) also adopts the feature [suction] as a means for distinguishing the click consonants from the non-click consonants. He concludes his study with the proposal that clicks be considered as clusters of independent consonants. His primary argument in favor of the cluster analysis is that all of the click effluxes exist in the language as independent consonants, some of which combine with other consonants to form non-click clusters. By analyzing clicks as clusters of consonants, the consonant inventory of !Xóõ is dramatically reduced in size, bringing it more in line with the majority of the world's languages. Note however, that the feature specification [+suction] is still necessary under the cluster analysis in order to distinguish the voiceless unaspirated, voiced and nasal clicks, that is the clicks which are articulated without any distinctive markings on the efflux, from their homologous non-click counterparts. In the examples in (2) the feature [suction] distinguishes the first segments of each pair.

(2)	(a) voiceless, unaspirated:	|aa	'move off'	taa	'person'
	(b) voiced:	|gaa	'work'	dam	'hunger'
	(c) nasalized:	|naa	'see you'	n	'I'

Although Traill is correct in his claim that the consonant inventory of !Xóõ is drastically reduced under the cluster analysis, this analysis is not without shortcomings. Firstly, a result of this analysis is that this language has extremely complex clusters which is a typological rarity in itself. Furthermore, in this analysis, as it is presented by Traill, the back articulation of clicks is treated either as a phonetic detail, as in the case of the voiceless unaspirated, voiced and nasal clicks, or as an independent consonant which follows the click. However below I present evidence from the phonological patterning of clicks with other back consonants, that the back articulation for all clicks is phonologically relevant.

In this language there is a constraint which applies to vowels following back consonants. The underlying vowel inventory of !Xóõ consists of five basic vowel contrasts, namely /i, e, a, o, u/, however following a back consonant all vowels will appear on the surface as [a, o, or u]. This "Back Vowel Constraint" can be formulated as in (3).

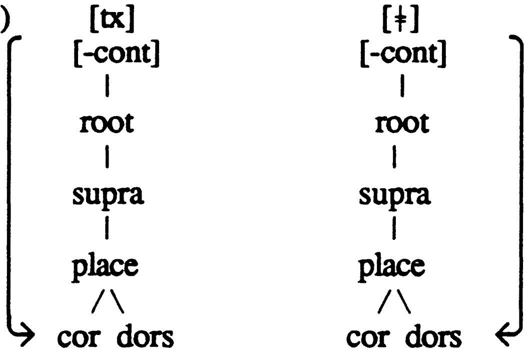
With few exceptions all the click consonants exert this effect on following vowels, indicating that clicks and back consonants form a natural class which can be characterized by the feature specification [+back].

(3) Back Vowel Constraint
If C_1 V_1
[+back]
then C_1 V_1
[+back] [+back] (Traill '85, p90)

In Traill's analysis of clicks as consonant clusters, he draws a distinction between the "basic" clicks, that is the voiceless unaspirated, voiced and nasal clicks, and all other clicks. Basic clicks are those which lack prominent features on the release of the back closure and which are thus not analyzed as clusters of independent consonants. All other clicks have audible release features on the back closure which, under the cluster analysis, are considered to be independent consonants which follow the "basic" click. This analysis fails to capture the linguistic significance of the back closure of the basic clicks with respect to a phonological process in the language, the Back Vowel Constraint, since it treats the closure at the velum for these clicks as a phonetic detail which is required for the production of the influx. Thus, contrary to the phonological feature specifications we expect from Traill's analysis, the specification [+back] must be added to the feature characterization of the simple clicks if we still wish to maintain a cluster analysis of clicks and their accompaniments. In other words, we have demonstrated the necessity of specifying a double articulation for even the basic clicks. However the status of the back articulation as a primary or secondary articulation still remains to be resolved.

In the analysis proposed by Sagey (1986), clicks are complex segments which have the back closure as the major articulator and the front closure as the minor articulator. In Sagey's representation, the distinguishing mark of a click is a velar closure as the major articulation in the complex segment. The mechanism by which the major and minor articulations in a complex segment are specified is directly related to the phonological degree of closure features of the segment. The major articulation will receive the degree of closure associated with the segment and therefore is able to enter into oppositions based on these features. The minor articulation, on the other hand, will have predictable stricture features associated with it. So, for example in !Xu, Sagey distinguishes clicks from all other corono-dorsals by having the dorsal articulation as major. For [tx] the dorsal articulation has predictable (i.e., always fricated) degree of closure, whereas for clicks, the coronal articulation has predictable degree of closure. (As in !Xóõ, the degree of closure of the front articulation is predictable from the place specification, as shown in table 2). Thus we have the representations shown in (4) to distinguish [ǂ] from [tx] in !Xu. The crucial difference is in the articulation being pointed to by the pointer which identifies the major articulator. In the case of [tx] the coronal node is pointed to, whereas for the click, [ǂ], the feature dorsal is the major place of articulation. This representational system uniquely specifies clicks within the language by having the dorsal articulation as the major articulator. Furthermore, in this way the feature [suction] is rendered redundant.

(4)



An important characteristic of Sagey's theory is that she draws a distinction between major and minor articulations, and avoids the terms "primary" and "secondary". In her feature system a major articulator is defined as "... an articulator to which the phonological degree of closure features of the segment apply" (p. 203). A minor articulator will always have a degree of closure which is predictable within the particular language. This characterization of major and minor articulators allows for the possibility that minor articulators are of an equal, or even greater, degree of closure. Indeed, there are cases of multiply-articulated segments for which the articulator with distinctive degree of closure shows a lesser degree of closure than the other, minor articulator. For example, Sagey cites the corono-labial segments in Margi, /ps/, where the non-continuant, labial articulator has a non-distinctive degree of closure and the continuant, coronal articulator contrasts for degree of closure. When applied to clicks, Sagey's feature theory provides a straightforward means of capturing the predictable degree of stricture of the front closure. And, since this theory allows for the possibility that the minor articulator shows a consonantal type of closure, the problem of a non-vocalic secondary articulation is avoided.

Sagey also presents an argument from the point of view of phonological processes in favor of her proposal. In her theory, complex segments involve two articulations which are represented under one x-slot and thus occupy only one unit of phonological timing. Unlike contour segments, the two articulatory gestures of a complex segment are phonologically unordered relative to each other. The segment as a whole is subject to linguistic rules which affect either of the features represented at the subsegmental level. Moreover, a complex segment with two, unordered place features, say F and G, may behave as an F segment with respect to its right edge for the purposes of one rule but as a G segment with respect to that same edge for the purposes of another rule. Clusters, on the other hand, are strictly ordered and will behave as such with regard to phonological processes.

Sagey claims (p. 126) that the !Xóõ clicks display exactly the type of behavior expected of complex segments with respect to phonological processes. In this language there is an assimilatory process which raises and fronts the vowel /a/ when it is followed either immediately or after an intervening consonant by /i/, or by /n/, and is preceded by a dental consonant. For the purposes of this rule, the click, /ǂ/, patterns with other dental consonants such as /t/ and /l/. This rule with examples are given in (5), (Sagey p. 127).

(5) Dental Assimilation
a -> ɜ, ɨ / dental ____ i, n
/tan/ [tɜn] 'to it' (Traill p.73)
/ǂali/ [ǂɨli] 'fold ' (Trail p.70)

Thus the click /ǂ/ behaves phonologically according the place features of its front closure at its right edge. However, for the purposes of the !Xóõ "Back Vowel Constraint" this same segment behaves according to the feature specifications of the back closure at its right edge. So, the patterning of /ǂ/ with other dental consonants with regard to the dental assimilation rule and its classification as a back consonant from the point of view of the "Back Vowel Constraint", is presented by Sagey as evidence that this click must have two, phonologically unordered places of articulations, making it a singular, complex segment.

A weakness of this analysis of clicks as complex segments is with regard to the interaction of the place specifications and the following vowel as evidence for unordered multiple articulations. In a feature geometry which has different place features specified on different tiers, any feature of a given segment is accessible to non-adjacent segments on the segmental tier for rule interaction so long as there is no intervening specification on the tier in question. For example, in a $C_1C_2V_1$ sequence in which C_1 has a specification [+F] for some feature F, and C_2 is specified [+G] for some other feature G, both of these feature specifications are accessible to V_1 without violation of the prohibition against crossing association lines. The adjacency of C_2 and V_1 on the segmental tier makes this observation trivial for the interaction of the feature G with the following vowel. The lack of a [-F] specification for C_2 which would block the interaction of C_1 and V_1 makes the observation true for the interaction of the feature F and V_1. The case of clicks in the Khoisan languages is exactly analogous to the example discussed above whether the accompaniment is taken to be an independent consonant or if clicks are analyzed as unitary segments. Thus Sagey's argument based on phonological patterning in favor of clicks as complex segments is not completely compelling, although it has certain strong points.

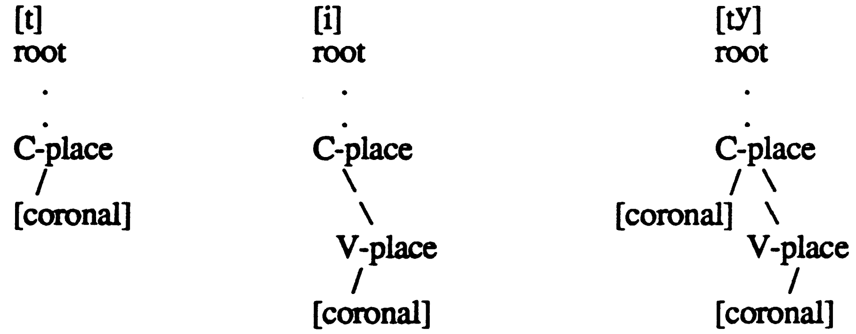
In contrast to the analyses in which the two articulators of clicks are treated as either two independent segments (e.g. Jakobson, Fant and Halle , Traill), or one major (primary) and one minor (secondary) articulation (e.g. Trubetzkoy , Chomsky and Halle, Sagey), Maddieson and Ladefoged (1989) claim that neither of the two click articulators is less major, or less primary, than the other. They therefore treat clicks as multiply articulated segments with two primary (or major) articulators. However, their arguments are mainly in favor of viewing the back articulator as primary rather than as secondary, as proposed in SPE. They do not discuss the possibility of viewing the front articulator as secondary to the primary back articulator.

3. Clicks as back consonants with anterior secondary articulations revisited

This section presents a modified version of Sagey's feature representation for clicks, in which the front articulation is secondary to the back articulation. We then examine further some of the characteristics of primary and secondary (or, major and minor) articulations as they are characterized by the proposed structure, and show how they bear on the status of the click articulations within the sound structure of !Xóõ.

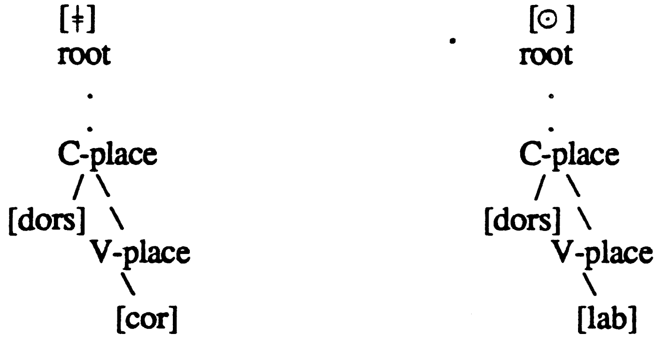
Clements (1991) proposes a feature geometry in which the place node of a segment splits into two lower level nodes. The higher level node is termed the consonant, or C-place node, while the lower level node is referred to as the vocalic, or V-place node. A segment may bear features on just the C-place node (i.e. plain consonants), on just the V-place node (i.e. vowels), or on both the C-place and the V-place nodes (i.e. consonants with a secondary articulation). The abbreviated tree structures in (6), based on Clements (1991) show the representations for [t], [i] and [tʸ].

(6)



As discussed elsewhere (Clements 1991, Goodman 1990), motivation for this structure comes from the relatively free interaction of vowels across consonants, as opposed to the blocking effect of vowels on consonantal spreading. The above representation accounts for this fact by placing the V-place node under the C-place node. Furthermore, this representation provides a clear account of interactions between vowels and secondary articulations of consonants. For example, in Ponapean, labial consonants with a labiovelar secondary articulation, such as $/p^w/$, pattern together with back rounded vowels (Goodman 1990). Sagey's insight with regard to the non-contrastive degree of stricture of minor articulations can be captured in this system by assigning the contrastive degree of stricture to the place specifications under the C-place node in multiply-articulated segments. In this way there is no longer a need for the pointer, and the contrastive degree of closure of primary articulators versus non-contrastive degree of closure of secondary articulators is captured in an equally natural way. The representation for clicks in this theory is illustrated in (7) by the partial feature trees for [ǂ], and [ʘ].

(7)



A prediction of this theory, which represents primary articulations under the C-place node and secondary articulations under the V-place node, is that interactions between parallel tiers (that is, interactions between two adjacent C- or V- place tiers) will be "stronger" than interactions between cross tiers (that is interactions between adjacent C- and

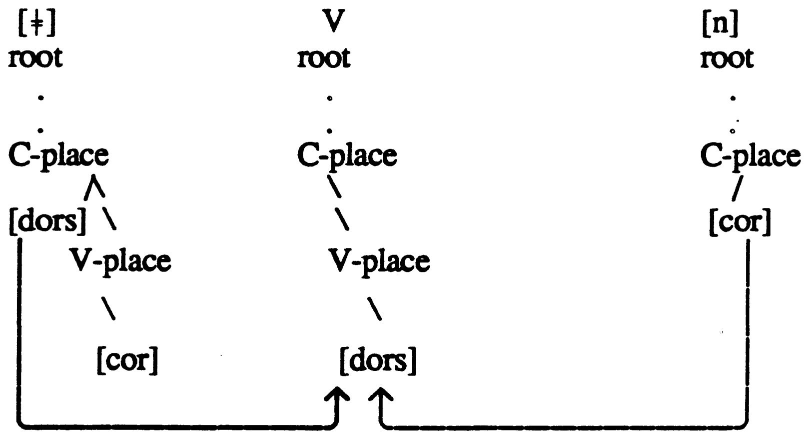
V-place tiers). As we will see below, by representing clicks as back consonants with labial or coronal secondary articulations, we can account for the difference between a partial and a total assimilatory process in a straightforward manner.

As discussed above, !Xóõ has a constraint which backs vowels following back consonants, including clicks. Thus after all back consonants the vowels will be taken from the set, [a, o, u]. This is a strong constraint in the language and will also apply to loan words, such as the Afrikaans word [dɔŋki] 'donkey' yielding [tonti] in !Xóõ. In this case, the final CV sequence, of which the consonant is [+back] and the vowel is [+front], is nativized by !Xóõ to yield two front segments. There is one class of exceptions to the Back Vowel Constraint which shows a regular assimilation of the vowel following a dental or post-dental stem-initial click, to the Class 1 noun suffix [-i]. This process results in an alternation of the stem vowel, as shown by the examples in (8). Apparently, the stem vowel in the singular form assimilates completely to the suffix vowel via a rule such as a→ i /{ǀ, ǂ} _ i (Traill, p. 91).

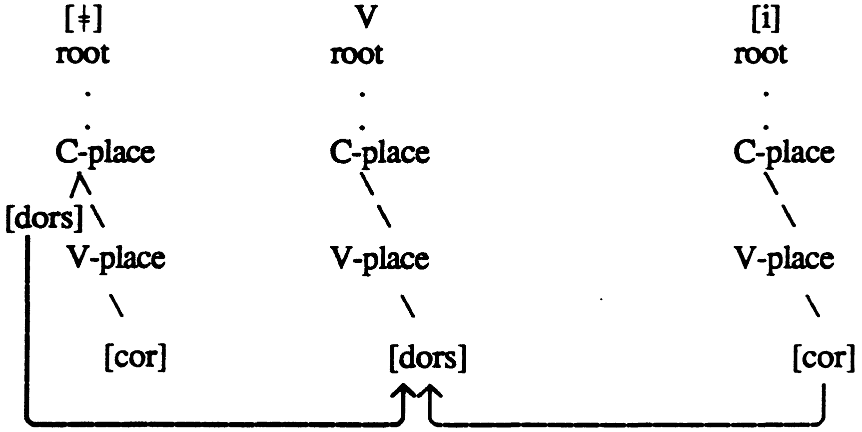
(8)	ǀʔi.i	'lover'	ǀʔa.ba+te	pl.
	ǀi.i	'steenbuck'	ǂa.ba+te	pl.
	ǂqhe.e	'bush'	ǂqha.m	

This language also has a co-occurrence constraint against a sequence of two vowels of which the first member is a coronal vowel, that is /i/ or /e/. Thus we have well-formed sequences such as /ai, ae, ui, ue, oi, oe/, whereas */ia, ea, iu, eu, io, eo/ are not permissible. This can be stated as a restriction against sequences of coronal and dorsal feature specifications on the V-place tier (*[coronal dorsal]$_{\text{V-place}}$). Now, in the case of the dental and post-dental clicks which have the coronal articulator represented under the V-place node, the Back Vowel Constraint would appear to create sequences which violate this restriction. In fact, these are precisely the cases where we find a relaxation of the Back Vowel Constraint in the form of the dental assimilation rule shown above in (5), and in the systematic exceptions to the Back Vowel Constraint in (8). Thus it appears that in these cases the *[coronal dorsal]$_{\text{V-place}}$ restriction receives reinforcement by the presence of an adjacent coronal segment to its right, which will spread its coronality to the vowel. Feature tree representations for these two interactions are given below in (9). The bold arrows connect the two tiers over which the Back Vowel Constraint operates. The unbolded arrows connect the two tiers which enter into the assimilatory processes.

(9)(a) Dental Assimilation



(b) Back Vowel Constraint exceptions



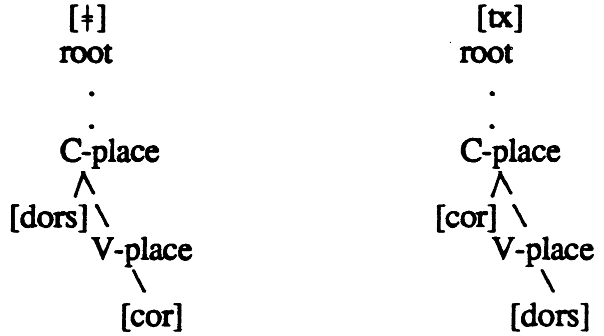
Recall that the assimilation of the dental assimilation rule is a rule of partial assimilation which causes the vowel to be centralized, whereas the examples in (8) show a total assimilation and the vowel surfaces as a front vowel. This difference in strength of the assimilatory process can be accounted for in this system of representation by appeal to the difference between cross- and parallel-tier interactions. The interaction in the dental assimilation rule when conditioned by a following [n], as shown above in (9a), occurs across tiers; the C-place coronal specification of the [n] interacts with the V-place specification of the vowel. In the case represented by (9b), the V-place - V-place interaction overrides the C-place - V-place interaction of the Back Vowel Constraint, resulting in apparent exceptions to this constraint. If the coronal specification of the click were represented as the primary articulator under the C-place node and the dorsal click specification were represented under the V-place node, then we would expect that the two interactions exert equal influence over the vowel in the examples in (8). However, this expectation is not borne out and it appears that this interaction is indeed a V-place - V-place interaction.

In the case of the dental assimilation rule conditioned by a following [n], the interaction of this rule and of the Back Vowel Constraint are both characterized by this representation as cross-tier interactions. We would therefore expect them to be of equal "strength," and this expectation is borne out by the fact that the assimilation is partial rather than complete as in the parallel-tier, V-place - V-place interactions of (9b). In fact, Traill notes that in certain cases where the dental assimilation rule is conditioned by a following vowel, [i], the target vowel may assimilate fully yielding a long [i:] (Traill p70). This is a result that we would expect given the proposed representation since this assimilation operates over parallel tiers. Furthermore, there is evidence to show that the assimilation can be suspended if there is a back vowel following the coronal consonant. For example, we find [ǂɨli] where the assimilation has applied and [ǂana] were the assimilation is blocked by the final [a]. Thus there is apparently a competition of assimilatory processes which results in alternating surface vowels. If the click consonant were represented with the back specification under the V-place node, we would expect the Back Vowel Constraint to override any other C-place - V-place interaction. In this case, the Back Vowel Constraint would be a "strong" V-place - V-place interaction and should show no interference from any C-place - V-place interaction.

To summarize this section so far, we have seen that a representation of clicks as back consonants with a secondary front articulation in a feature geometry which splits the place node into a C-place node with a lower level V-place node, provides a natural account for the predictable degree of stricture of the front articulator. Also, by having the front articulator represented as a secondary articulation under the V-place node, we can explain the differences in degree of assimilation between vowels and following consonantal and vocalic coronal segments. Thus we have evidence from phonological rules of !Xóõ in favor of an analysis in which clicks are back consonants with a secondary front articulation.

At this point it is interesting to consider various other properties of the proposed structure for multiply articulated segments. Given this feature geometry, we expect to find cases of complex segments which differ only in which articulator is selected as major and which is selected as minor. Indeed, as initially observed by Sagey for !Xu, the closely related !Xóõ contrasts the non-click complex segment /tx/ and the click /ǂ/. Both of these segments involve both coronal and dorsal articulations. The difference between them is precisely a difference in terms of which articulator is treated as primary. For the click it is the coronal articulator which shows predictable degree of stricture. For the non-click the situation is exactly reversed. In this case the back articulation is always fricated. Thus, for the non-click complex segments, the coronal articulator is represented under the C-place node and the dorsal articulator is represented under the V-place node. For clicks it is the dorsal articulator which appears under the C-place node. These opposing representations are given below in (10). Note how this representation allows us to avoid using the pointer of Sagey's theory.

(10)



As noted above, Sagey's feature geometry allowed for the possibility that minor articulators show equal, or even greater, closure than major articulators. The defining feature of a minor articulator is that it does not bear distinctive degree of stricture features. In the theory proposed by Clements, this property of minor articulators is expressed by placing a secondary articulation under the V-place node, and by having the distinctive degree of stricture features of the segment apply to the place specification under the C-place node. The phonetic interpretation of the closure for the secondary, V-place features is determined in a language specific way. Note that in neither Sagey's nor Clement's feature system is the feature [suction] needed for the representation of clicks, since it is the distinctions between primary (or, major) and secondary (or, minor) articulations which defines the click segments phonologically. Thus the specific timing of the articulator closures and releases which results in the ingressive airstream mechanism is treated as part of the segment's phonetic interpretation.

A prediction of this analysis in both Sagey's and Clement's theories, is that no language will contrast two segments which differ only in either timing of the articulatory gestures or in degree of stricture of the minor articulator. This is because these properties of the segments are treated as part of their phonetic interpretations. For example, we expect that no language will contrast [kY] and [ǂ]. Both of the segments are back consonants with secondary coronalization and thus have identical phonological representations. They differ in the degree of stricture of the secondary articulation; for [kʸ] the secondary articulation has a vocalic degree of stricture whereas for [ǂ] the coronal articulation is consonantal. Indeed, I know of no click language which has such a contrast.

Thus we have seen that a representation for clicks such that the back closure is represented under the C-place node and the front closure is under the V-place node, finds support from both linguistic processes and representational generalizations. The major advantage of this system over Sagey's system of feature representation is that it provides an explanation for certain asymmetries that are observed between consonant-vowel interactions and vowel-vowel interactions.

4. A closer look at the click and non-click systems of !Xóõ

A comparison of the click and non-click consonants of !Xóõ reveals some interesting properties of the general consonant system of this language. Traill observes that there is a parallel between the click accompaniments and the non-click consonants of the language, which is seen by the existence of almost all the click accompaniments as independent consonants. This fact is Traill's primary argument in favor of a cluster analysis of clicks. In this section, we examine this parallel between the clicks and non-clicks, and show that the analysis of clicks proposed in the previous section provides a natural system of classification for the !Xóõ consonant system which is consistent with cross-linguistic trends.

The clicks of !Xóõ show five influx places of articulation; bilabial, dental, post-dental, palatal and lateral. Amongst the non-click consonants we find bilabial, dental, and post-dental places of articulation, as well as a marginal lateral. However, there are no non-click palatal consonants. For the click back articulations we find a velar and uvular place of articulation. Both of these places of articulation are also found amongst the non-clicks. For both the clicks and non-clicks we find all four logically possible combinations of voicing and aspiration. In all cases voicing is realized as voice-lead and aspiration is realized as post-release aspiration. Table (3) shows these sets of consonants. Table (3a) shows non-click consonants which contrast along the voicing and aspiration dimensions. Table (3b) shows the clicks with a velar back closure and their variants with respect to voicing and aspiration. Table (3c) shows the corresponding series of clicks with a uvular back closure. Parentheses indicate that the segment is marginally present in the language and all orthographic conventions adopted by Traill are retained. Thus, *dth* represents a dental segment with prevoicing and post-aspiration.

Table (3)

(a) Non-click consonants:

	labial		dental		post-dental		velar		uvular	
	unasp	asp.	unasp.	asp.	unasp.	asp.	unasp.	asp.	unasp.	asp.
vcl	(p)	(ph)	t	th	ts	tsh	k	kh	q	qh
vcd	b		d	dth	dz	dtsh	g		(N)G	
lat.						(l)				

(b) Click consonants with a velar back closure:

	labial		dental		post-dental		palatal		lateral	
	unasp.	asp.	unasp.	asp.	unasp.	asp.	unasp.	asp.	unasp.	asp.
vcl	ʘ	ʘh	ǀ	ǀh	ǂ	ǂh	!h	!	ǁ	ǁh
vcd	ʘg	gʘh	ǀg	gǀh	ǂg	gǂh	!g	g!h	ǁg	gǁh

(c) Click consonants with a uvular back closure:

	labial		dental		post-dental		palatal		lateral	
	unasp.	asp.	unasp.	asp.	unasp.	asp.	unasp.	asp.	unasp.	asp.
vcl	ʘq	ʘqh	ǀq	ǀqh	ǂq	ǂh	!q	!qh	ǁq	ǁqh
vcd	mʘ G		nǀG		nǂG		n! G		n ǁG	

Within the proposed representation for clicks, the non-click consonants in table (3a) are related to the clicks in tables (3b) and (3c) in that they are non-complex consonantal variants of the click double articulations. Thus the labial, dental and post-dental non-clicks in table (3a) are consonantal versions of the corresponding click secondary articulations. The velar non-click consonants in table (3a) are simply the click primary articulations in (3b) with no secondary articulations. Similarly, the uvular non-click consonants have their click counterparts which are shown in table (3c).

The next set of parallels between the click and non-click consonants is exemplified by those pairs of doubly articulated click and non-click consonants which differ in which is the primary and which is the secondary articulator. These consonants are shown in table (4).

Table (4)

(a) non-click segments with a secondary uvular fricative:

	dental	post-dental
vcl	tx	tsx
vcd	dtx	dtsx

(b) click segments with a uvular fricative accompaniment:

	labial	dental	post-dental	palatal	lateral
vcl	⊙x	ǀx	ǂx	!x	ǁx
vcd	g⊙x	gǀx	gǂx	g!x	gǁx

Amongst the non-clicks we find that only the coronal segments bear the secondary articulation. The clicks, however, show both coronal and labial secondary articulations. Note that in all of the segments in table (4), following Traill, the uvular fricative is transcribed as [x] rather than [χ]. This reflects a traditional means of transcription and is therefore retained.

Table (5)

(a) ejected non-click segments:

	post-dental	velar	uvular
vcl	tsʔ	kxʔ	(qʔ)
vcd	dtsʔ	gkxʔ	

(b) click segments with ejected accompaniments:

	labial	dental	post-dental	palatal	lateral
	⊙qʔ	ǀqʔ	ǂqʔ	!qʔ	ǁqʔ
vcl	⊙kxʔ	ǀkxʔ	ǂkxʔ	!kxʔ	ǁkxʔ
vcd	g⊙kxʔ	gǀkxʔ	gǂkxʔ	g!kxʔ	gǁxʔ

(c) ejected non-click segments with ejected secondary articulations:

	dental	post-dental
vcl	tʔkxʔ	tsʔkxʔ
vcd	dtʔkxʔ	dtsʔkxʔ

The non-click system shows both voiced and voiceless ejected uvular affricates (transcribed as [kxʔ] and [gkxʔ]), a marginal ejected uvular stop and ejected post-dental segments. In the click system we find all of the above back segments with all five of the click front, secondary articulations. These series of clicks and non-clicks are shown in tables (5a-c).

Like the non-clicks in table (3a) which are the non-complex counterparts of the clicks in (3b), the velar and uvular consonants in table (5a) are the non-complex versions of the clicks in table (5b). This language also has ejected dental and post-dental non-clicks with ejected velar secondary articulations (table 5c). Thus, like the dental and post-dental plain stops with the secondary uvular fricative in table (4a) which contrast with the clicks in (4b), we find an opposition between the clicks in table (5b) and non-clicks in table (5c) based on which articulator is selected as primary. In the case of an ejected secondary articulation, (5c), we find that the primary coronal is ejected too.

For the non-click nasals we find plain and preglottalized bilabial and dental nasals. For the nasalized clicks we find plain, preglottalized and voiceless variants. Thus, for the nasal series, we find that the non-clicks correspond to the click secondary articulations. However the non-click system is less extensive than the click system since it lacks the voiceless variants. Table (6) shows these segments.

Table (6)

(a) nasal non-click segments:

	labial	dental
plain	m	n
preglott.	ʔm	ʔn

(b) click segments with nasal accompaniments:

	labial	dental	post-dental	palatal	lateral
plain	⊙ n	ǀn	ǂn	! n	ǁn
preglott.	ʔ⊙ n	ʔǀn	ʔǂn	ʔ ! n	ʔǁn
voiceless	⊙ n̥	ǀn̥	ǂn̥	! n̥	ǁn̥

The only remaining click accompaniment is the glottal stop, which has no reflex in the non-click system.

From the comparison of the non-click and click systems of !Xóõ, we see that the non-clicks can be related to the clicks in two different ways. Firstly, there is a set of non-clicks which exist as the plain (as opposed to complex) versions of the click primary and secondary articulations. Thus we have four of the five click influxes (secondary articulations) which exist as independent non-click consonants. And we find plain non-click velar and uvular variants of all of the clicks. The nasal click and non-click segments are also related in this way. That is, the non-click nasals are related to the nasal clicks in that they are non-complex versions of the click secondary articulations. Similarly, the ejected click and non-click consonants are related in this way to the non-click [kxʔ], [gkxʔ], and [qʔ], which are non-complex versions of their click counterparts.

The second way in which the non-click and click consonants are related is with regard to which articulator is selected as the major articulator. Thus we have click consonants in which the back articulator is primary and non-click consonants in which the front articulator is primary. This type of relationship is limited to the dentals and post-dentals.

Based on the data in the UPSID sample (Maddieson 1984), we find a general tendency for complex segments to co-occur with their non-complex counterparts. Thus we find a universal tendency for languages which have /k^{ʔw}/ to also have /kʔ/ (94.4% of cases), and /k^{w}/ (88.8% of cases). Similarly, all languages in the sample with /k^{w}/, and 84% of languages with /kʔ/, also have /k/. Thus we have a very strong implicational hierarchy which can be stated as follows; /k^{ʔw}/ → /kʔ/, /k^{w}/ → /k/. A similar hierarchy holds for uvular stops so that we can formulate an implicational hierarchy such as /q^{ʔw}/ → /qʔ/, /q^{w}/ → /q/. Under the proposed analysis, we find that this tendency is realized in !Xóõ for all the clicks. As shown above, we find that all of the click primary (back) articulations exist in the language as non-click consonants with no additional secondary articulations.

The UPSID sample also shows that there is a weak tendency for /k^{w}/ to co-occur with /w/. This suggests that there is an implication that consonant secondary articulations will also exist in the language as independent segments. In !Xóõ, as expected, we found that this implication holds, although not quite as strictly as the implication discussed above. Most of the secondary (front) click articulations exist in the language as independent consonants. However, the language has no palatal non-clicks whilst it has the full complement of palatal clicks. Thus, we have seen that with an analysis of clicks in which the back articulation is primary and the front articulation is secondary, the relationship between the clicks and non-clicks follows the expected pattern based on cross-linguistic tendencies.

5. Summary and conclusions

This paper presents arguments in favor of an analysis of clicks in which the back articulator is represented under the primary, C-place node and the front articulator is represented under the secondary, V-place node. This analysis captures generalizations of three kinds. Firstly, as initially observed by Sagey, the front articulation of clicks does not enter into any oppositions based on degree of stricture. Within the theory adopted here, the degree of stricture features of the segment apply to the C-place node and the articulation represented under the V-place node will not contrast for degree of closure. Secondly, it is has been shown that this representation accounts for certain asymmetries between consonant and vowel interactions. Finally, under this analysis, we find that the general consonant inventory of !Xóõ, a typologically unusual language, can be shown to conform to universal tendencies regarding the co-occurrence of complex and plain consonants.

References

Beach, D. M (1938) "The phonetics of the Hottentot language," Cambridge: W. Heffer and Sons

Chomsky, Noam and Morris Halle (1968) The Sound Pattern of English. New York: Harper and Row.

Clements, G.N. (1989) "A Unified Set of Features for Consonants and Vowels." Cornell University, ms.

Clements, G. N. (1991) "Place of articulation in consonants and vowels: A unified view," to appear in B. Laks and A. Rialland, eds., L'architecture et la géométrie des répresentations phonologiques Editions du C.N.R.S., Paris

Goodman, B. (1990) "Ponapean labiovelarized labials: Evidence for internal segment structure," to appear in *NELS 20* , Montreal, Canada

Jakobson, R., Fant, G. and M. Halle (1963) Preliminaries to Speech Analysis. Cambridge, MA; MIT Press. Fourth printing, first published 1951.

Ladefoged, Peter and Anthony Traill. (1984). "Linguistic phonetic description of clicks". *Language*. 60.1: 1-20

Ladefoged, Peter. (1975). A course in phonetics. New York: Harcourt Brace Jovanovich

Maddieson, Ian and Peter Ladefoged. (1989). "Multiply Articulated Segments and the Feature Hierarchy". *UCLA-Working Papers in Phonetics 72*.

Marchal, Alain. (1987) "Des clics en français?" *Phonetica* 44:30-37

Sagey, Elizabeth. (1986). The Representation of Features and Relations in Nonlinear Phonology, Ph.D dissertation, MIT, Cambridge, Ma.

Traill, Anthony. (1978). "Another click accompaniment in !Xóõ". *Khoisan Linguistic Studies* 5.22-9.

Traill, Anthony. (1985). Phonetic and phonological studies of !Xóõ Bushman. Hamburg: Helmut Buske Verlag

Trubetzkoy, N. S. (1969) Principles of Phonology. Berkeley; University of California Press. Translated by Christiane Baltaxe.

The American English Flapping Rule and the Effect of Stress on Stop Consonant Durations

Alice Turk

1. Introduction

The experiment described in this paper concerns the American English Flapping Rule, whereby non-word-initial intervocalic /t/ and /d/ are flapped when preceding an unstressed vowel, as in the words 'méṭal' and 'pyrámiḍal'; in contrast the /t/ preceding a stressed vowel in 'meṭállic' is not flapped. Acoustically, a major difference between flapped and non-flapped alveolars is durational: flaps are considerably shorter than their non-flapped counterparts. In this study we measured the durations of intervocalic stops in both the non-flapping and flapping environments as well as the durations of vowels that preceded the alveolars in order to answer the following questions:

1. Does a general timing mechanism underly the rule? It may be the case that the extreme shortening found in /t/ and /d/ is an instantiation of a more general timing principle. We discuss possible mechanisms for the length differences between poststress and prestress stops as well as implications for the formulation of the American English flapping rule.

2. Are vowels preceding flaps that were underlying /d/'s longer than vowels preceding flaps that were underlying /t/'s? It is well-known that vowels preceding voiced stops tend to be longer than vowels preceding voiceless stops (e.g. Chen, 1970). Longer vowels before flapped /d/'s than before flapped /t/'s would suggest that lengthening before voiced segments is due to the underlying phonological distinction ([+voice] vs. [-voice]) rather than to the phonetic realization of the consonant (which, in the case of most flaps, is as a voiced segment). In fact, one study (Fox and Terbeek, 1977) found that vowels preceding flapped /d/'s were longer than vowels preceding flapped /t/'s.

3. Does a longer phonological phrase affect segment durations?

2. The Phonology of Flaps

Flaps are allophones of /t/ and /d/ formed by a rapid movement of the tongue-tip making contact with the alveolar ridge, followed by immediate release. They have characteristically short durations of 10-40 ms. (Zue and Laferriere, 1979). In American English, /t/ and /d/ are realized as flaps in the following environments:

1. After a stressed vowel and before an unstressed vowel:
   ´v_v For example: mé/t/al -> me[D]al ([D] is used as the symbol for a flap).
2. Optionally between two unstressed vowels:
   v_v For example: provoca/t/ive -> provoca[D]ive or provoca[t]ive
3. Optionally word-finally between two vowels (stressed or unstressed):
   v_#v For example: no/t/ a/t/ all -> no[D] a[D] all or no[t] a[t] all.
   priva/t/e airplane -> priva[D] airplane

Word-initial alveolars are never flapped.

The experiment described in this paper is less concerned with the phonological features involved in the flapping rule than with a phonetic analysis of the phenomenon. However, the phonological accounts of the flapping phenomenon are relevant because they all refer to prosodic structure in the formulation of the rule, suggesting that all consonants occurring in the flapping environments should exhibit the same phonetic consequences of being in that particular environment. These accounts vary in the segmental environment they claim is relevant, the distinctive features with which they characterize the flap, the domain of the application of the rule, and in the prosodic environment of flapping. From a prosodic point of view, the accounts can be divided into three principle views: 1. Flaps are ambisyllabic (Kahn, 1976; Gussenhoven, 1986); 2. They are syllable-final (Selkirk, 1982; Inouye, 1989); 3. They are non-foot-initial (Kiparsky, 1979). These views are described below.

2.1 Flaps are ambisyllabic

Kahn (1976) and Gussenhoven (1986) both claim that the occurrence of alveolar flaps can be predicted by referring to the syllabic structure notion of ambisyllabicity. It is described as follows for word-medial consonants (Kahn, 1976:55,56):

Ambisyllabicity

In [-cons] C C0 V
|.·' \|/ | Associate C and S1
S1 S2 -stress

Kahn describes four allophones of the voiceless alveolar stop:

1. Voiceless aspirated alveolar stop: [th] ex: creativity
2. Unreleased voiceless alveolar/glottal stop: [t'] ex: hat
3. Released voiceless alveolar stop: [t] ex: stem
4. Voiced alveolar flap: [D] ex: item

Their distribution can be predicted from the syllable structure described above:

Stops which are both syllable-initial and non-syllable-final are aspirated.
Alveolar stops which are syllable-final and non-syllable-initial are glottalized.
Stops which are non-syllable-initial and non-syllable-final are released and not aspirated.
Alveolars which are both syllable-initial and syllable-final (ambisyllabic alveolars) are flapped.

The rule for flapping (ordered after the rule for aspiration) can be written as follows:

$$/t/ \rightarrow [D] \;/\; [-\text{cons}] \;\_\_\_\_\; a\langle\#(\#)\rangle \quad \begin{array}{l} +\text{syllabic} \\ b\,\langle-\text{stress}\rangle \end{array}$$

Conditions:

1. ~a -> b (if the alveolar stop is word-medial, the following vowel must be unstressed for flapping to occur).
2. if a, within connected phrases only.
3. Does not apply in artificially slow speech.

Kahn (1976) assigns [-cons] to prevent the alveolar stops in "after", "faster", and "kept it" from flapping.

Kahn hesitates as to which feature to assign to the flap itself: in order to capture the voicing characteristic of many flaps, he assigns [-stiff v.c]; and to be consistent with feature descriptions of trilled /r/'s which are considered taps, he assigns the flap the feature specification of [+sonorant].

2.2 Flaps are Syllable-final

Selkirk (1982) differs from Kahn (1976) principally in that in her account, the consonants Kahn describes as ambisyllabic are resyllabified in order to be the coda of the preceding syllable. She cites the following evidence for resyllabifying the consonants including those which are flapped: the lengthening of vowels before voiced consonants, the nasalization of preceding vowels, and l-velarization, based on the premise that these are syllable-conditioned processes. Selkirk claims that resyllabified consonants behave as though they were underlyingly syllable final: it is the property of syllable finality that accounts for the flapping phenomenon.

Selkirk's syllabification rules have different implications for the prediction of alveolar stop allophones in American English. Instead of having to specify that aspiration applies to stops which are syllable initial and not syllable final, her rule is simplified in that all syllable-initial stops are aspirated. To distinguish syllable-final alveolars which are to be realized as flaps from those which are to be realized as glottalized alveolars, she posits the feature Release as the determining parameter. Alveolars which are [-release] are glottalized; alveolars which are [+release] are flapped:

From Selkirk (1982):

Tap (= Flap)

t,d -> D / s(...[-cons] ____)s
|
+release

Glottalization

p,t,k -> p', t', k' / ____
|
-release

Selkirk claims that [+release] is the unmarked value for the feature and describes at length the environments for [-release], although, as Inouye (1989) points out, she does not unify the environments for [-release] in any principled way. Although Selkirk states that the feature Release is not known to be phonemically contrastive in any language (and that perhaps it is not a distinctive feature, she notes that different languages have different realizations of Release for their stop consonants (e.g. French stops are always released except when followed by a homorganic consonant; in Vietnamese, prepausal consonants are not released), and that Release must be specified at some level of representation.

Another account which treats flaps as syllable-final is that of Inouye (1989). Her paper deals mainly with the issue of the phonological specification of the flap, however, which has been virtually ignored in the accounts of Kahn (1976) and Selkirk (1982). She argues that the flap is phonologically dynamic, and proposes that the flap is actually a contour segment (cf. Sagey (1986)), based on evidence that the flap exhibits edge effects (the lefthand environment ≠ the righthand environment). On the left edge, flapping is reported to be blocked by [+cons] segments (that have a constriction in the oral cavity); on the right edge, flapping seems to be blocked by segments which require the tongue tip to be above a certain threshold. Her argument rests on the commonly held observation that flapping does not occur in words such as "after", "faster", and in phrases such as "kept it" (the left edge is [+cons]). It also does not occur when the right edge is [+coronal], as in "beaten" She argues from such evidence that the dynamic nature of the flapping gesture has not been captured either in the feature specification of Kahn (1976) ([+son], [-stiff v.c], or in Selkirk's requirement that only alveolar stops which are [+release] be flapped. Inouye proposes a multivalued feature Aperture with which to characterize the approach, contact and release stages of the flap, and describes the flapping rule as one of spreading of the aperture node from adjacent segments onto the coronal segment.

2.3 Flaps are Non-foot-initial

Kiparsky (1979) differs from the rest in that his analysis does not refer to syllable structure in describing the flapping environment. Kiparsky claims that the rules for alveolars are foot-sensitive: only non-foot-initial alveolars which follow a [-cons] segment and precede a [+syllabic segment] may flap.

For example:

[no]S [[t]o]Σ [ma[D]oes]Σ, where Σ denotes a foot.

Basically, within a foot, alveolar stops that precede unstressed vowels can be flapped:

re[pé[D]i[D]ive]Σ

To summarize, scholars are divided as to what is the prosodic environment of flapping. Kahn and Gussenhoven claim that flaps are ambisyllabic; Selkirk and Inouye claim flaps are syllable-final; and Kiparsky argues that they are non-foot-initial. The results of this experiment will be discussed in terms of which prosodic account they support.

3. The Phonetics of Flapping

The study draws on the observation of Zue and Laferriere (1979) that characteristically flaps are very short (10-40 ms); the experiment described here seeks to determine whether this characteristic short duration is a product of the stress on surrounding segments: i.e., whether it is a phonetic consequence of being in that prosodic environment.

Several studies have dealt with the effect of stress on the durations of intervocalic consonants.

Zue and Laferriere (1979) in a study of /t/ and /d/ in American English found that word-medial alveolars are realized as extremely short flaps when immediately following a stressed vowel or when between two unstressed vowels. They report that flaps between two unstressed vowels are significantly longer than flaps immediately following a stressed vowel.

As for other places of articulation, Umeda (1977) and Stathopoulos and Weismer (1983) found that closure durations of both voiced and voiceless stops immediately preceding a stressed vowel were longer than that of stops immediately following a stressed vowel. Umeda's study consisted of one speaker reading a twenty minute essay. Umeda notes that velars show the smallest range of difference, although no tests of statistical significance are reported. /b/ was not included in the prestress environment, nor were total stop durations reported for the voiceless stops. Stathopoulos and Weismer (1983) report similar results listed in the following table:

Mean Durations of stop closures in 2 positions:

	1.prestress	2.poststress
b	92 s.d. 10	66 s.d.16
p	96 s.d. 12	87 s.d 11
d	76 s.d 13	41 s.d.15
t	82 s.d. 13	44 s.d.11
g	68 s.d. 10	56 s.d. 7
k	72 s.d. 11	71 s.d. 13

Closure durations for /b,p,d,t, and g/ seem to be significantly different in the two stress environments, (although they do not report statistical significance). Interestingly, for the non-alveolar segments, differences are larger in the voiced cases (/b,g/).

In another study, Davis and Summers (1989) report that in word-medial VC sequences, total durations (closure + aspiration) were longer in a prestress environment than in a poststress environment. However, the test tokens used for velars were not well-paired in terms of segmental environment--the post-stress velars were prevocalic, whereas the prestress velars occurred before /r/ (*sagging* and *sacking* vs. *degrees* and *decrees*). They did not show any statistical tests for the effect of stress on consonant duration.

In general, all of the above studies suggest that consonants preceding stress within a foot are longer than consonants that follow stressed vowels. However, a clearer picture of the effects of stress on both parts of intervocalic consonant durations (closure + aspiration) is needed. The experiment described in this paper was designed to provide a complete description of the effects of stress on word-medial stop consonants.

Of relevance to the second part of the experiment is the study by Fox and Terbeek (1977) on the durations of vowels preceding flaps. Fox and Terbeek conducted a study of 20 disyllabic word pairs (ex: *writer*, *rider*) spoken by 21 speakers. They discovered that stressed vowels preceding flapped /d/'s were significantly longer than stressed vowels preceding flapped /t/'s, suggesting that the lengthening before voiced stops rule makes reference to the phonological representation of segments, not to their phonetic realization.

However, this seems to be the case only for stressed vowels. Davis and Summers (1989) confirmed Fox and Terbeek's results for stressed vowels, but they found that unstressed vowels preceding voiced prestress stops were longer, but not significantly so, than those preceding voiceless stops. In the experiment described below, the durations of both stressed and unstressed vowels preceding alveolar stops are reported.

4. The Experiment

The first part of this experiment was designed to determine systematically whether the stress effect on stop durations suggested by previous research can in fact be generalized to word-medial stop consonants of all places of articulation. Measurements of closure, aspiration and total (closure +aspiration) duration were taken from stops of all places of articulation in real words. The division of stops into closure and aspiration portions may show which part of the stop is shortened or lengthened. Two different sentence lengths were used to see if a longer phonological phrase had any effect on the duration of the test consonant. None was expected: Klatt (1976) observed that there is little evidence that speakers adjust segment durations in order to satisfy a global rhythmic constraint. However, an affect of sentence length on segment duration would indicate that segmental timing is dictated at least in part at a prosodic level higher than the word.

In the second part of the experiment, the duration of vowels preceding the alveolars in the different stress environments was measured to see if lengthening before stop voicing is observed for both types of flaps.

4.1 Methods

4.1.1 Stimuli

Closure, aspiration, and total durations of /p,t,k,b,d,g/ and the durations of vowels preceding alveolars were investigated when the test consonant occurred in three word-medial environments:

1) Before a stresssed syllable (the non-flapping environment).

2) After a stressed syllable and before an unstressed syllable (a flapping environment).

3) Between unstressed syllables (Optional flaps: Zue and Laferriere found that flaps in this environment were longer than the other flaps)

The test words were chosen such that for a given word-medial consonant, the segments surrounding the stops were as similar as possible:

	v_V	V_v	v_v
/p/	repaír	léper	cáliper
/t/	metállic	métal	diámeter (speakers RG, AM and LE)
/k/	locále	lócal	fóllicle
/b/	rebél	rábble	párable
/d/	medállion	médal	pyrámidal (speakers RG, AM and LE)
/g/	cigár	vígor	vínegar

The frame sentences (listed in Appendix A) were selected so that the test words would be in similar prosodic positions. Sentence length was increased by adding a modifier before each test word.

In all, there were 18 test words x 2 sentence lengths x 4 speakers x 5 repetitions. Alveolar data was collected for three out of the four speakers.

Test sentences were presented to the speakers along with an equal number of decoy sentences. The sentences were typed on notecards and were shuffled before each repetition of the set. They were read at a normal speaking rate, with natural intonation patterns in an IAC sound-treated booth.

4.1.2 Speakers

Two female and two male native speakers of American English served as speakers for the experiment. All were in their early twenties and speak a standard dialect of American English. Speaker AM (male) is a native of Ithaca, NY; Speaker AW (male) grew up in New York City; Speaker LE (female) is from California; Speaker RG (female) is a native of Ann Arbor, Michigan.

4.1.3 Measurements

The speech was digitized at 8 kHz, lowpass filtered at 4 kHz and stored on a SUN 3/160 workstation. Measurements were taken from waveforms and spectrograms displayed using AT&T Bell Labs WAVES software. The beginning of the stop was judged to be the point of F1 offset, and likewise, the end of the stop was the F1 onset. When F1 could not be seen clearly, measurements were taken at points of obvious change in the waveform where there was no higher formant energy in the spectrogram.

4.2 Results

A repeated measures ANOVA with factors of stress environment (Stress Position), underlying phonological voicing (Voicing), place of articulation (Place), sentence length (±Modifier) and the repeated measures factor of speaker (Speaker) shows main effects of Stress Position, Place, and Voicing on closure, aspiration and total consonant durations:
Effect of Stress on

Closure: $F(2,144) = 20.660$; $p < .0001$
Aspiration: $F(2,144) = 435.858$; $p < .0001$
Total duration: $F(2,144) = 421.103$; $p < .0001$

Effect of Place of Articulation on

Closure: F(2,144) = 315.866; p < .0001
Aspiration: F(2,144) = 148.960; p < .0001
Total duration: F(2,144) = 252.543; p < .0001

Effect of Voicing on

Closure: F(1,144) = 18.47; p < .0001
Aspiration: F(1,144) = 1858.35; p < .0001
Total duration: F(1,144) = 1443.26; p < .0001

Effect of Speaker (Repeated Measures Factor) on

Closure: F(3,432) = 227.03; p < .0001
Aspiration: F(3,432) = 135.51; p < .0001
Total duration: F(3,432) = 334.62; p < .0001

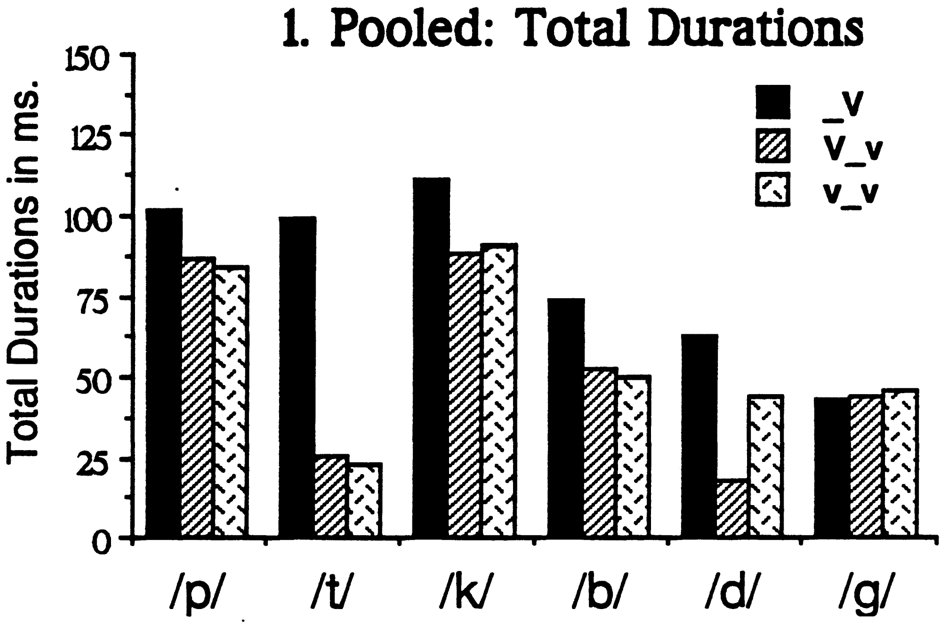
Sentence length had no significant effect on either the closure duration, the aspiration duration, or the total duration of the stops (p > .05). Interactions of the repeated measures factor of Speaker with all combinations of Stress Position, Voicing, and Place are significant at the p<.05 level, indicating that the different speakers behave differently. All interactions which include the factors of Stress Position, Voicing, and Place, but not ±Modifier are significant at the p<.05 level, which indicates that the effect of different stress environments varies according to the voicing and place specification of the stop.

Since the stress effects are confounded by the place of articulation and underlying voicing of each stop, repeated measures ANOVAs were conducted which tested the effect of Stress Position and ±Modifier on stop closure durations, aspiration durations and total durations for each type of stop. Likewise, single factor analyses of variance testing the effects of Stress Position on each type of measurement for each speaker were done to expose the nature of inter-speaker variability. Results broken down in this way are given below.

4.2.1 The Effect of Stress Position

4.2.1.a On Total Stop Durations

A repeated measures ANOVA including the factors of Stress Position and ±Modifier with the repeated measures factor of Speaker shows a main effect of Stress Position on the total durations (closure + aspiration) of all stops except /g/. F and p values are listed in Appendix A. As can be seen from Figure 1, the difference in duration in different stress environments lies primarily between the non-flapping and the flapping environments. Repeated measures ANOVAs with factors of ±Modifier, Stress which included only the two flapping environments, and the repeated measures factor of Speaker showed that total durations in the two flapping environments were significantly different only in the case of /d/ (F (1,16) = 126.89; p < .0001) (Figure 1).

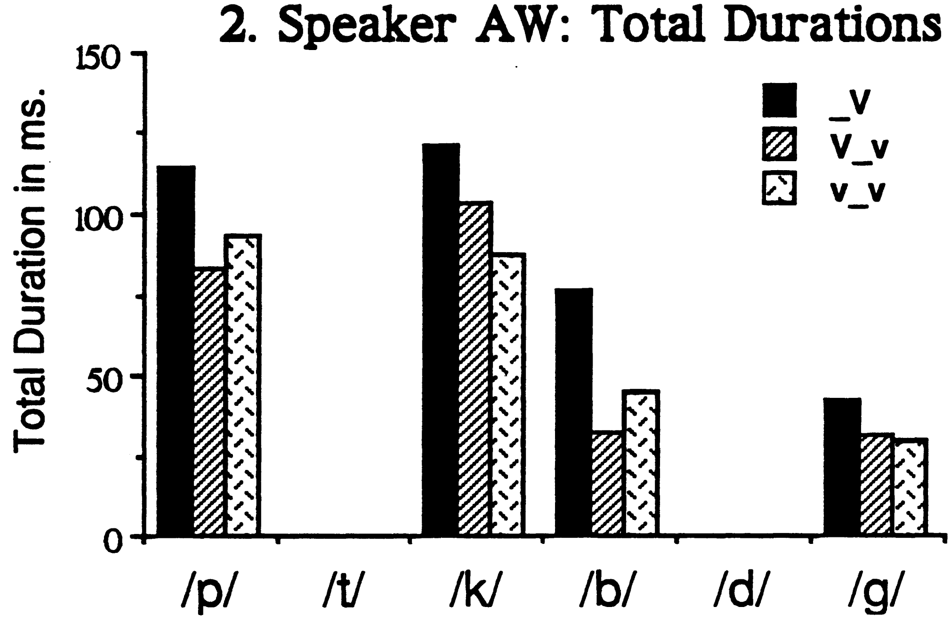


Results from ANOVAs and Scheffe F-tests for each speaker follow. F and p values are given in Appendix C.

Speaker AW (Figure 2):

Bilabial stops and /k/ are significantly longer in the prestress (non-flapping) environment than in the two flapping environments. The effect of Stress on /g/ was in the predicted direction, although differences were not significant.

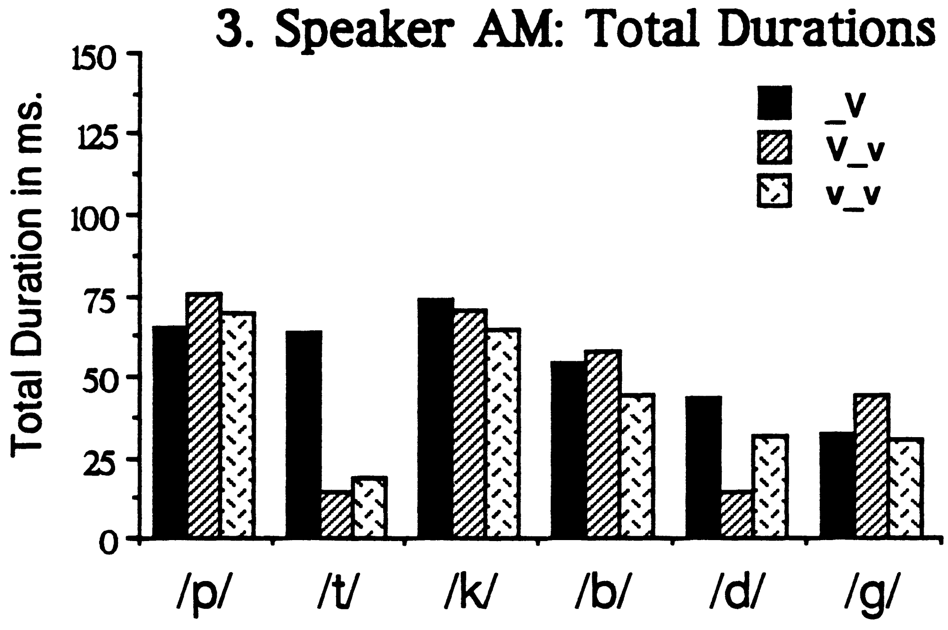
There is no significant difference between the closure durations in the two flapping environments.


2. Speaker AW: Total Durations

Speaker AM (Figure 3):

Total durations are significantly longer in the prestress environment only in the case of the alveolars. /k/ exhibits differences in the expected direction, although differences are not significant.

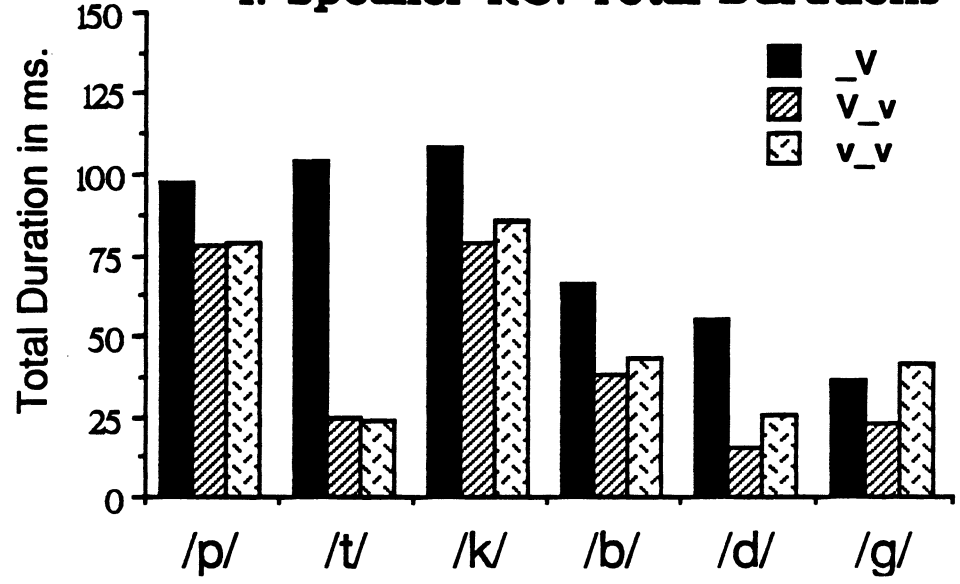
The durations of stops in the two non-flapping environments are significantly different only in case of /b,d, and g/. Zue and Laferriere (1979) had found that flaps which occur between two unstressed vowels are longer than those in the environment V_v. This is the case for /d/ in the present study, but /b/ and /g/ have the opposite pattern.


3. Speaker AM: Total Durations

Speaker RG (Figure 4):

Total durations are longer in the prestress environment for all stops except /g/. There is no significant difference between the two non-flapping environments for any stop, although the trend in all cases is for the stop between two unstressed vowels to be longer.

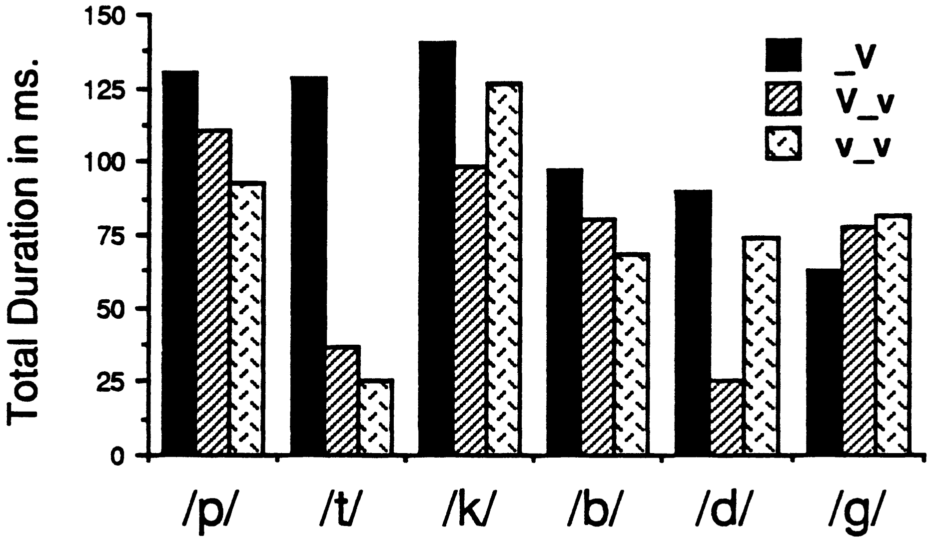
4. Speaker RG: Total Durations



Speaker LE (Figure 5):

Total durations are longer in the prestress environment for all stops except /g/. Significant differences between the two flapping environments are found for /p,k,b and d/. /d and k/ are longer in the v_v flapping environment than in the V_v flapping environment.

5. Speaker LE: Total Durations



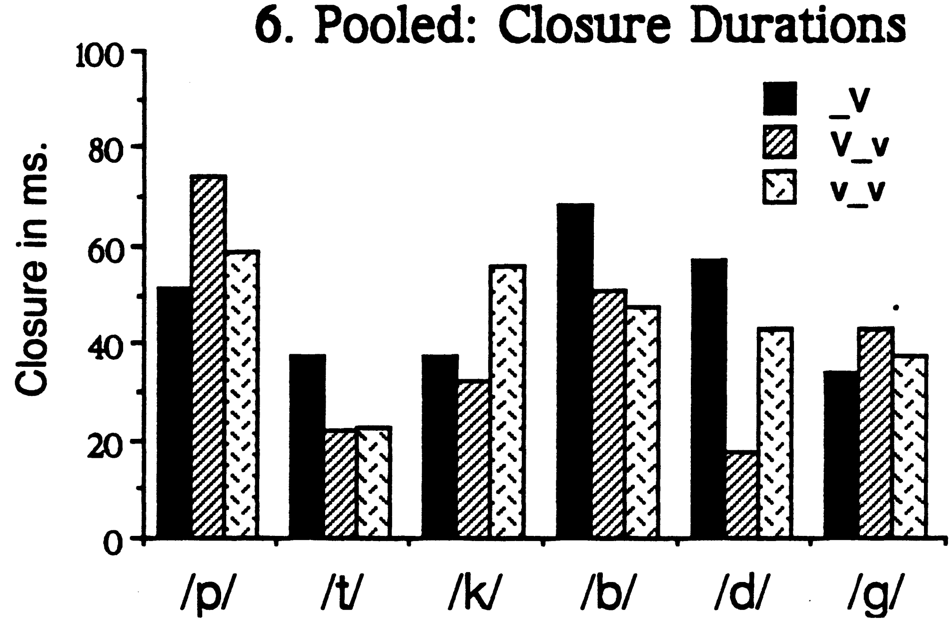
To summarize, total durations of all stops except for /g/ are longer in the non-flapping environment than in the flapping environments for three out of four speakers.

The durational differences that Zue and Laferriere (1979) observed for alveolars in the two flapping environments (/t/ and /d/ tend to be longer in the environment v_v than in the environment V_v) were not observed for stops of other places of articulation, nor did the alveolars the present experiment consistently follow the pattern Zue and Laferriere observed.

In order to determine which component of each consonant (closure or aspiration, or both) is responsible for observed total duration differences, the effect of stress on both closure and aspiration are shown in the two sections which follow.

4.2.1.b On Closure Durations

In a repeated measures ANOVA including Stress Position, ±Modifier and the repeated measures factor of Speaker, Stress Position had a significant effect on the closure durations of /t,b, and d/. A significant effect in the opposite direction was found for /p/ -- its closure duration in the prestress environment is shorter than its closure duration in the flapping environments.

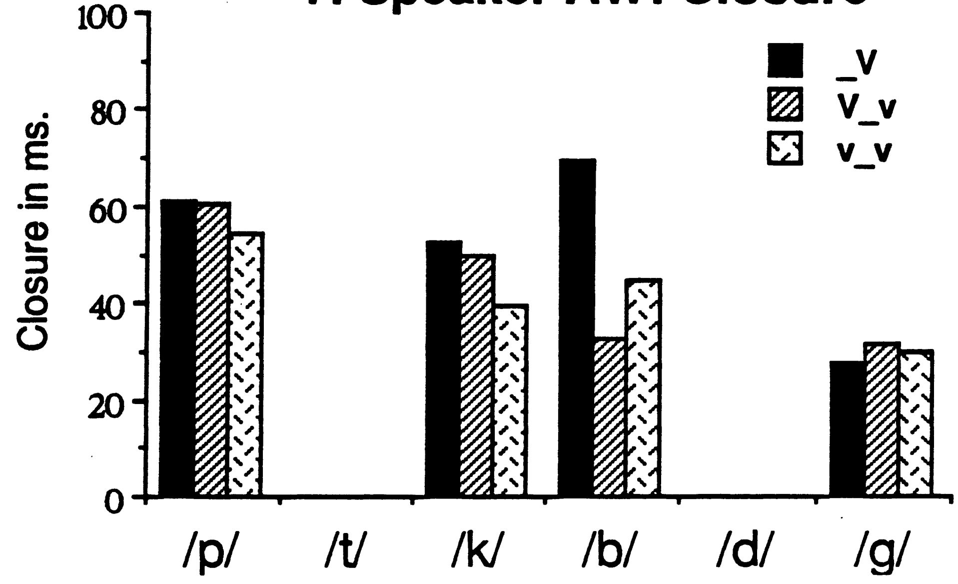

6. Pooled: Closure Durations

Speaker AW (Figure 7):

Closure durations for /b/ are significantly longer in the non-flapping environment than in both of the flapping environments. The closure durations for /k/ are significantly longer in the non-flapping environment than in the v_v environment, but the difference in the non-flapping environment and the V_v environment is not significant, although in the expected direction. The differences shown for /p and g/ between the non-flapping and flapping environments are not significant.

As for durational differences between stops in the two flapping environments, /b/ is the only consonant whose closures are significantly longer in the v_v environment than in the V_v flapping environment.

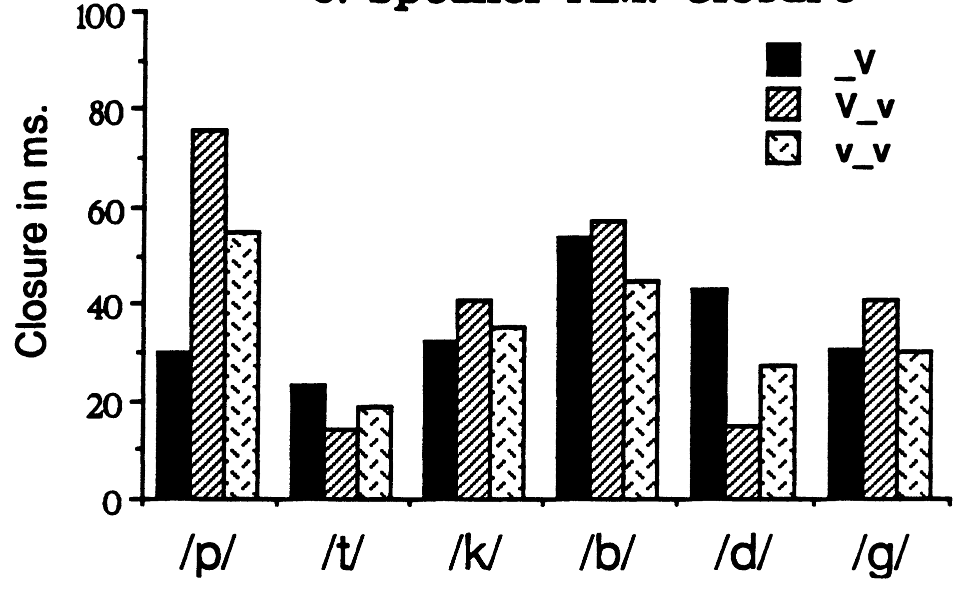
7. Speaker AW: Closure



<u>Speaker AM</u> (Figure 8):

Only the closure durations of the alveolars (/t and d/) are significantly longer in the prestress environment than in both of the flapping environments. The rest of the consonants (/p,k,b,g/) are longest in the V_v environment.

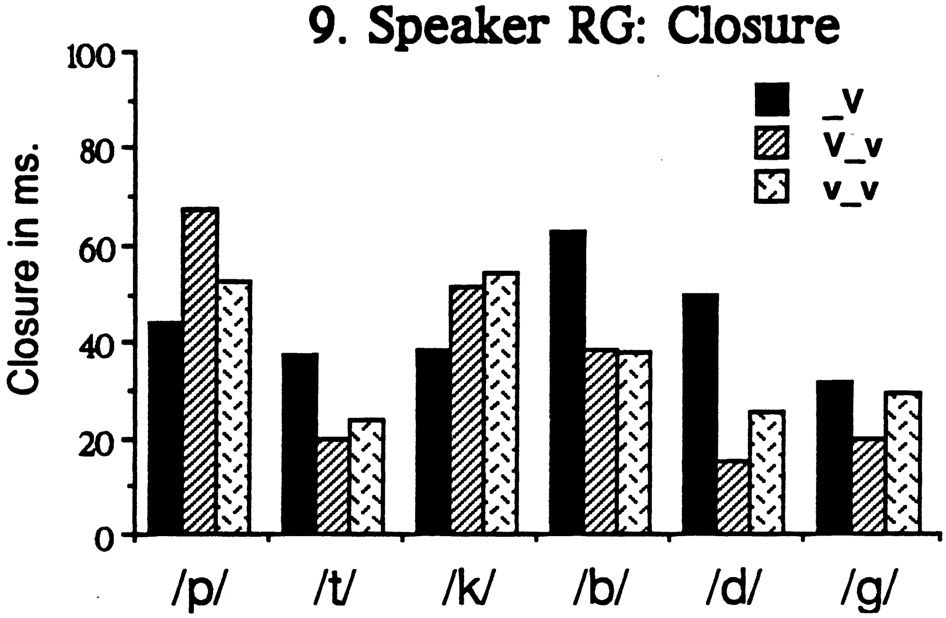
8. Speaker AM: Closure



Speaker RG (Figure 9):

Closure durations for /t,b, and d/ are significantly longer in the prestress environment than in both flapping environments.

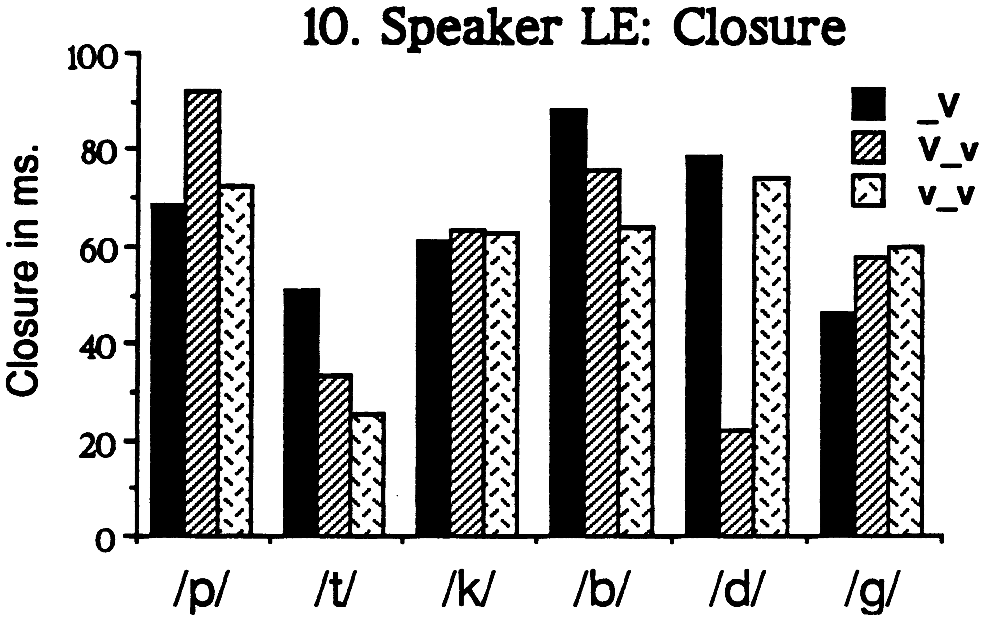
A significant difference in the expected direction between the two flapping environments was found only for /d/.



Speaker LE (Figure 10):

Closure durations for /t,b, and d/ are longer in the prestress environment than in both of the flapping environments, although the difference between the closure duration of /d/ in the v_V and v_v environments was not significant. The closure durations of /p,k and g/ are shorter in the prestress environment.

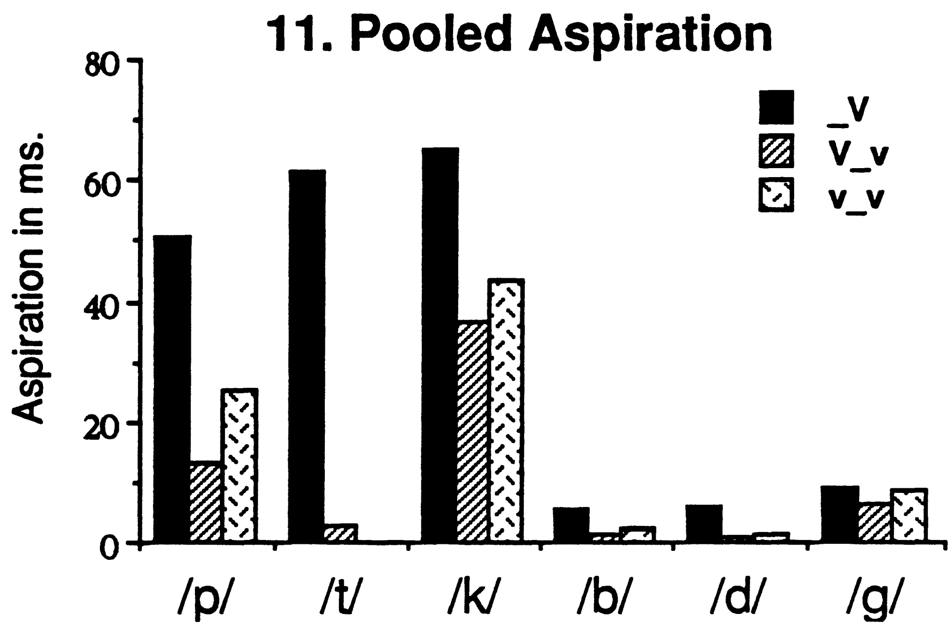
Only /d/ shows a significant difference in the expected direction between the two flapping environments.



To summarize, the closure durations of /t/ and /d/ are consistently longer in the prestress environment than in the flapping environments. As for consonants of other places of articulation, /b/ is the only one which is longer in the prestress environment for three out of four speakers.

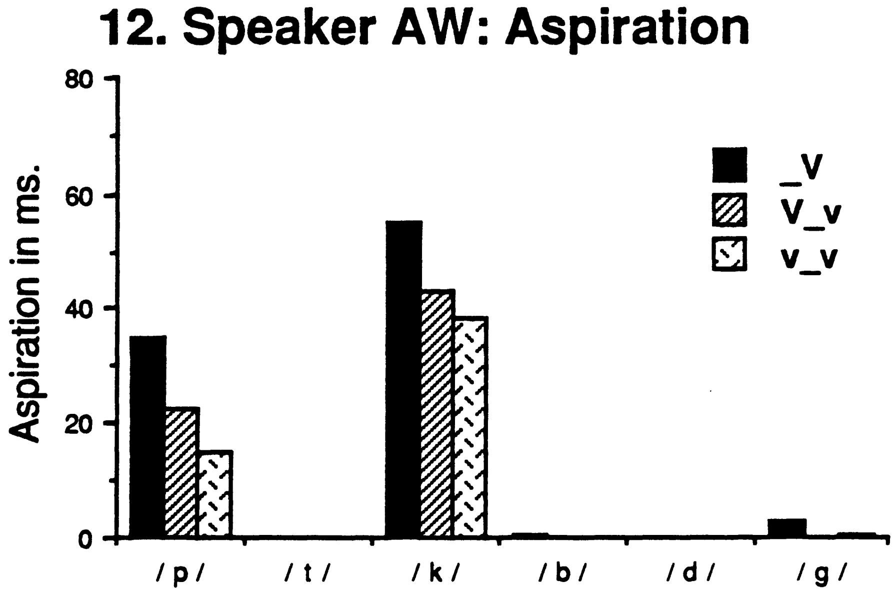
4.2.1.c. On Aspiration Durations:

Stress Position had a significant effect on the aspiration durations of all stops except /g/ (Figure 11). The aspiration durations for /p/ were significantly different in the expected direction in the two flapping environments (F(1,16) = 23.97; p < .0002).



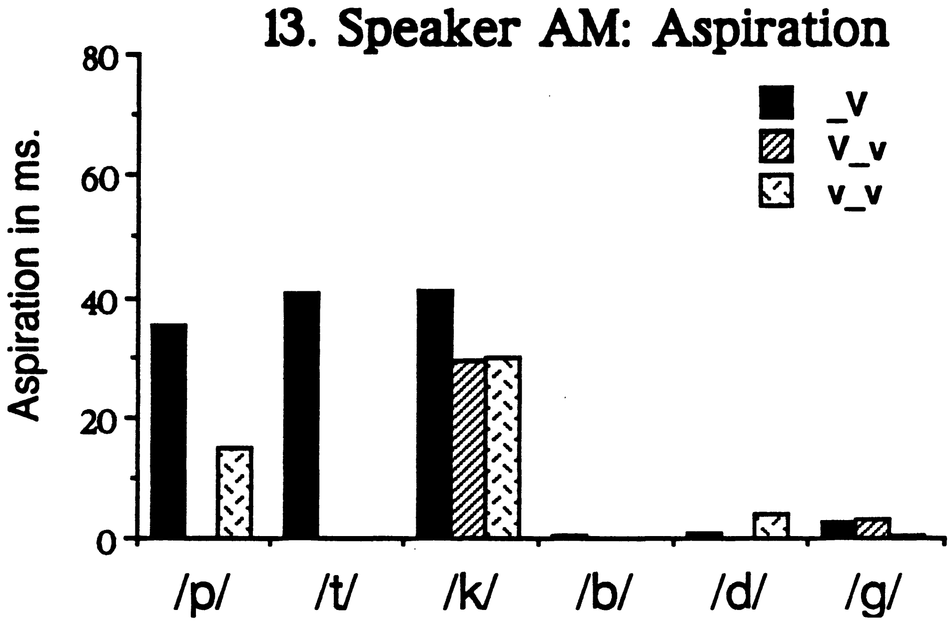
Speaker AW (Figure 12):

The aspiration durations are significantly longer in the prestress environment for /p, k,b, and g/. Only /p/ shows a significant difference between the two flapping environments.



Speaker AM (Figure 13):

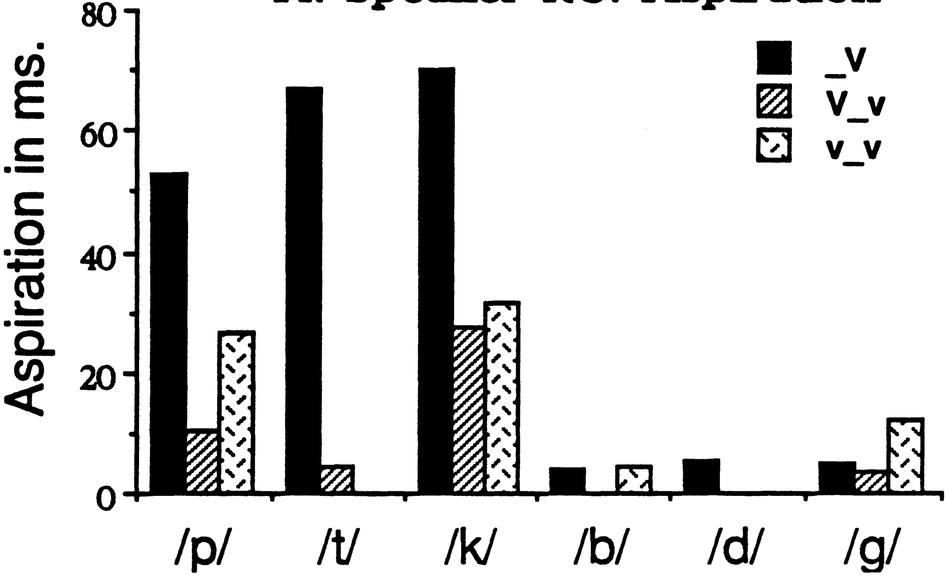
The aspiration durations are significantly longer in the prestress environment than in the other two environments for /p,t,k/. /p/ aspiration shows a significant difference between the two flapping environments.



Speaker RG (Figure 14):

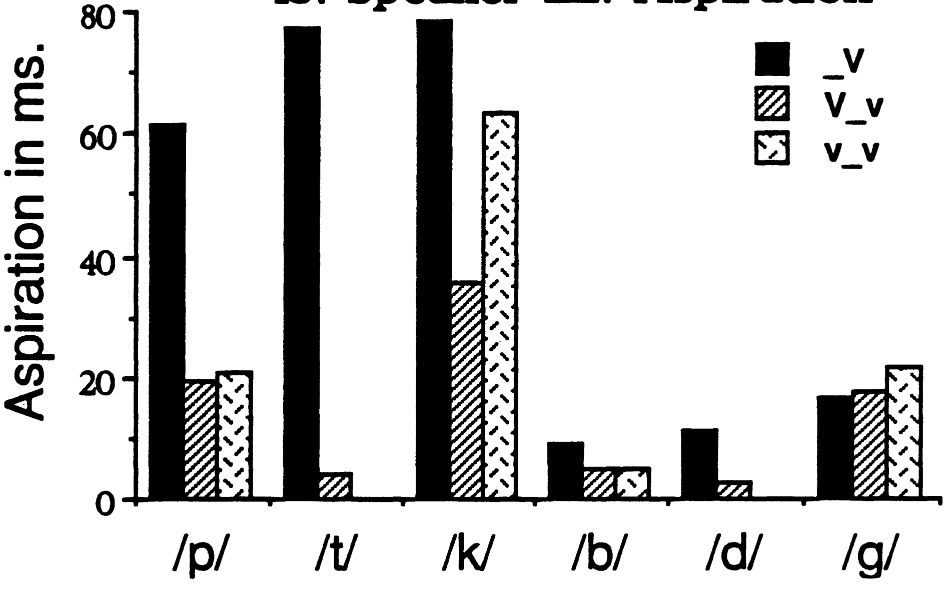
The aspiration durations are significantly longer in the prestress environment for /p,t,k/.

Aspiration durations for /p and g/ are significantly longer in the v_v environment than in the V_v environment.


14. Speaker RG: Aspiration

<u>Speaker LE</u> (Figure 15):

The aspiration durations are significantly longer in the prestress environment for /p,t,k, and d/. The aspiration durations for /k/ are significantly longer in the v_v environment than in the V_v environment.


15. Speaker LE: Aspiration

To summarize, all speakers show greater aspiration durations in the prestress environment than in the other two environments for all voiceless stops.

The aspiration durations of /p/ are longer in the v_v flapping environment than in the V_v flapping environment for three out of four speakers.

4.2.2. The Effect of a Longer Phonological Phrase

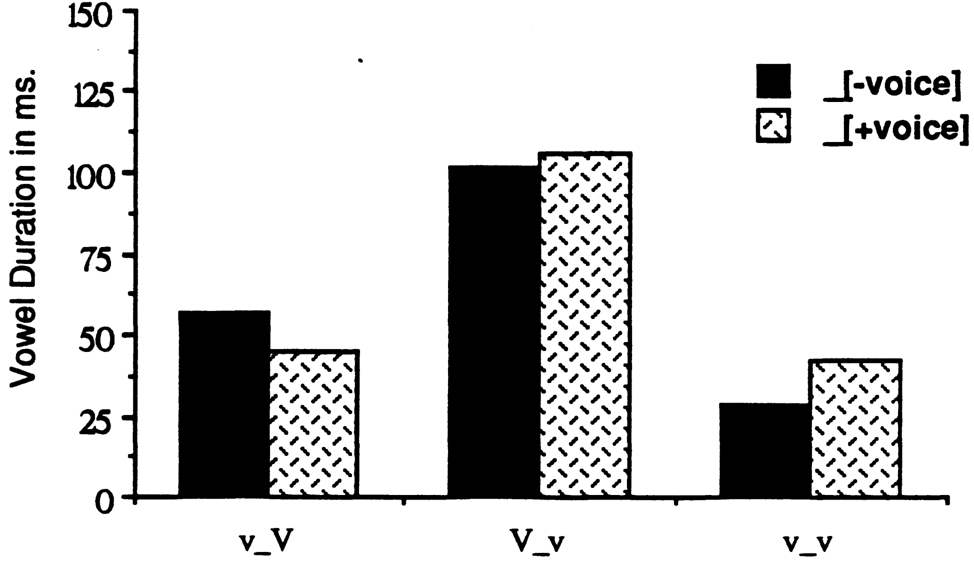
For the pooled data, the main effect of ±Modifier is significant only in the cases of the aspiration and total durations of /d/ ($F(1,24) = 12.80$; $p < .0015$--aspiration; $F(1,24) = 7.74$; $p < .0103$) and in the aspiration durations of /k/ ($F(1,24) = 6.205$; $p < .02$). These results confirm the observation of Klatt (1976) that there is little evidence that speakers adjust segment durations in order to satisfy a global rhythmic constraint.

4.2.3. Durations of Vowels Preceding /t/'s and /d/'s

Data from three speakers were included in a repeated measures ANOVA with factors of Stress Position,Voicing, ±Modified, and the repeated measures factor of Speaker. It shows a significant main effect of consonantal Stress Position ($F(2,48) = 585.97$; $p < .0001$) on the duration of the vowel before /t/ or /d/, which confirms that stressed vowels are longer than schwa (Stressed vowels had a mean duration of 104 ms; unstressed vowels had a mean duration of 43 ms). There is a significant interaction of Speaker with Stress Position indicating that stress position affects the durations of vowels differently for each speaker.

There is no main effect of Voicing. However, there is a significant interaction of Stress Position and Voicing on the duration of vowels before /t/ and /d/ ($F(2,48) = 19.42$; $p < .0001$), which indicates that the effect of Voicing on the length of the preceding vowel varies with the position of stress (Figure 16).

16. Pooled: Vowel Durations before /t/, /d/



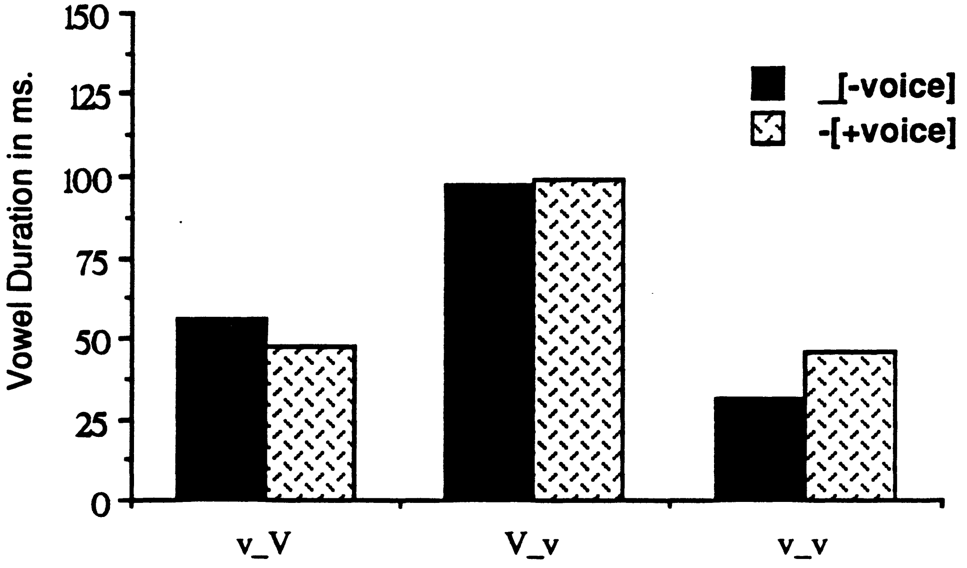
To explore the interaction of Stress Position and Voicing further, repeated measures ANOVAs were conducted to determine the effect of Voicing on the durations of vowels preceding /t/ and /d/ for each consonantal stress environment. In the prestress environment, vowels preceding /t/ were significantly longer than vowels preceding /d/ (F(1,18) = 20.36; $p < .0003$), and there was a significant main effect of Speaker (F(2,36) = 7.09; $p < .0025$). In the V_v consonantal environment, vowels preceding /d/ were longer than vowels preceding /t/, but not significantly (F(1,18) = 1.38; $p > .2561$). There was a significant interaction with Speaker (F(2,36) = 25.23; $p < .0001$). In the v_v environment, vowels preceding /d/ were significantly longer than vowels preceding /t/ (F(1,18) = 18.67; $p < .0004$), and again, there was a significant interaction with Speaker (F(2,36) = 3.85; $p < .0305$).

Mean durations of the pooled data for vowels preceding /t/ and /d/ in each stress environment are given below:

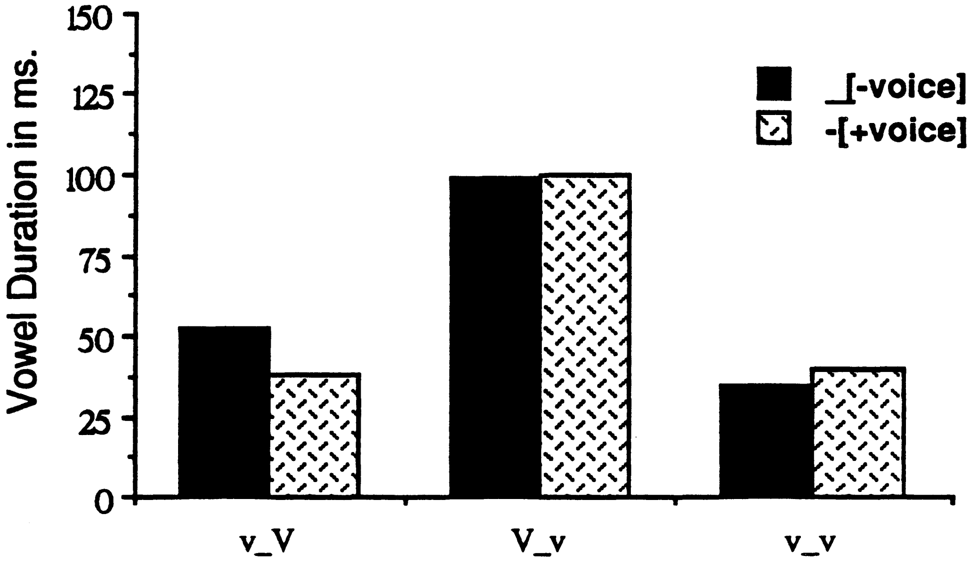
	before /d/ ([+voice])	before /t/ ([-voice])
vCV	n = 30 45 ms.	n = 30 56 ms.
VCv	n = 30 106 ms.	n = 30 101 ms.
vCv	n = 30 43 ms.	n = 30 29 ms.

Graphs of the results from each speaker follow; they mimic the patterns found in the pooled data.

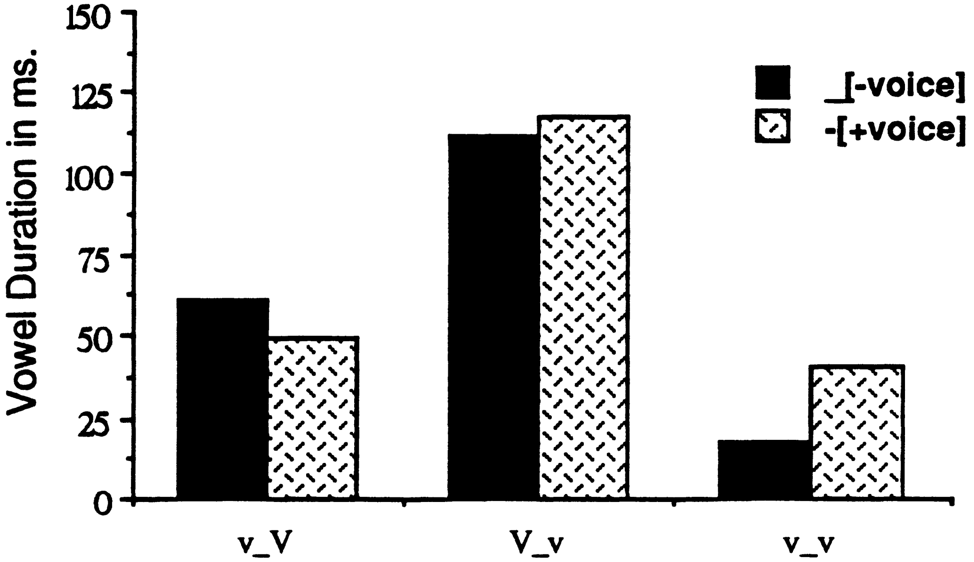
17. Speaker AM: Vowel Duration before /t/, /d/



18. Speaker RG: Vowel Duration before /t/, /d/



19. Speaker LE: Vowel Duration before /t/, /d/



4.3. Discussion

Generalizations that can be made from the results are discussed below:

4.3.1 The Effect of Stress

All stops (except /g/) are longer in total duration in prestress position than they are in either of the flapping environments. This finding suggests that relative consonant length is an acoustic correlate of the prosodic structure required by formulations of the flapping rule: the extreme shortening found in the alveolar cases is an instantiation of a more general timing phenomenon. The failure of /g/ to conform to the pattern could be a result of the relative inertia of the tongue dorsum as compared to other articulators.

The general pattern of the results bears upon the formulation of the prosodic domain of the American English flapping rule. There is considerable disagreement in the literature over what the prosodic environment for flapping is: Kahn (1976) and Gussenhoven (1986) claim that flaps are ambisyllabic; Selkirk (1982) and Inouye (1989) claim that flaps are syllable-final, and Kiparsky (1979) claims that flapping is not syllable-conditioned, but occurs when alveolars are non-foot-initial.

The present experiment shows that total duration is an acoustic correlate of the prosodic structure common to all stops in each stress environment. If we assume that flaps are ambisyllabic, following Kahn (1976), we must therefore say that ambisyllabic consonants are shorter than those that belong to only one syllable. Chierchia (1983/1986) argues that geminate consonants (presumably long) are ambisyllabic. The two accounts do not necessarily conflict, if it is the case that the mapping from prosodic phonology to the acoustics is language-specific. However, it seems counterintuitive to have such extremely different mappings in different languages.

In Selkirk's (1982) theory, word-medial consonants following stressed vowels are resyllabified and become the codas of the preceding syllable. Results of the present study show that in an algorithm mapping prosodic structure to acoustic values, syllable-final consonants would be assigned relatively short durations, while syllable-initial consonants would be assigned relatively long durations. The results indicate that there would have to be no difference in the durations assigned to syllable-final consonants in stressed vs. unstressed syllables. However, Stathopoulos and Weismer (1983) found that in fact, closure durations of word-final consonants in CVCVC nonsense disyllables were shorter when the final syllable was stressed than when it was unstressed. Their results show that the durations of syllable-final stops vary according to the stress assigned to the syllable. The fact that the present experiment shows no difference in stop duration in the two non-flapping environments (V_v) and (v_v), does not support Selkirk's theory, where the stops would be codas of syllables that differ in stress, and would be expected to have different durations.

In Kiparsky's (1979) formulation of the flapping rule, no resyllabification is required; non-foot-initial alveolars are flapped. The shortening of stops in the flapping environment could be explained in two ways: (1) The shorter stops are shorter because they are compensating for the length of the stressed vowel that occurs earlier in the foot. Such an explanation would be consistent with theories of isochrony, which seek to explain the phenomenon of decreasing syllable duration with an increasing number of syllables in a word (Kozhevnikov and Chistovich (1965), Port (1980), Allen (1976) and Lehiste (1976). The unit of isochrony is unclear, but Lehiste (1970) feels that it is either the foot or the word. It is also unclear exactly how the compensation takes place. It is possible that shortening takes place within syllabic units; alternatively, compensation could occur on a segmental level without reference to syllable structure; (2) Assuming duration is assigned in syllabic units, the shorter stops could be shorter simply because they are onsets in unstressed syllables (V_v and v_v), while the longer stops are long because they are onsets in stressed syllables (v_V). The results of Stathopoulos and Weismer (1983) for word-initial stop closures support this view. In nonsense disyllables of the form CVCVC, closure durations of word-initial stops were shorter when the first syllable was unstressed than when it was stressed.

Of the three prosodic domains posited for the flapping rule, the foot of Kiparsky (1979) seems to have the most support from results of the experiment described here.

Although the total durations show a clear pattern of being longer in the prestress environment and shorter in the flapping environment, there is a dichotomy in closure duration differences for voiced and voiceless stops. Closure durations of voiced stops are significantly longer in prestress position than in poststress position, whereas the closure durations of voiceless stops are not significantly longer in prestress position. This is consistent with the results of Stathopoulos & Weismer (1983) who found that closure durations for non-alveolar voiceless stops did not exhibit much of a difference between prestress and poststress position durations, whereas closure durations for voiced stops were very different in prestress position vs. poststress position. However, Stathopoulos & Weismer (1983) found that, despite the dichotomy between voiced and voiceless stops, the closure durations for all stops were longer in prestress position, which is not the case here. This experiment shows that for the voiceless stops, the shortening/lengthening occurs primarily in the aspiration portion, whereas for the voiced stops, shortening/lengthening occurs primarily in the closure portion. Aspiration thus appears to be more compressible/stretchable than closure. Kozhevnikov and Chistovich (1965) and Klatt (1976) have noted that vowels are more strechable/compressible than consonants. Our experiment supports their findings, and also suggests that aspiration could be added to the stretchability/compressibility hierarchy: Vowels are the most stretchable/compressible, followed by aspiration; closure durations are the least stretchable/compressible.

4.3.2. Vowels Preceding /t/ and /d/

Vowels preceding flapped /d/'s are significantly longer than vowels preceding flapped /t/'s only when the vowel before the flap is unstressed. When the vowel is stressed, length differences are in the expected direction, but are not significant. When the alveolar is in the non-flapping environment, the unstressed vowels preceding /d/ are actually shorter than vowels preceding /t/. The results differ slightly from those of both Fox & Terbeek (1977) and Davis and Summers (1989). Fox and Terbeek (1977) found that stressed vowels before flapped /d/ were reliably longer than stressed vowels before /t/. They concluded that the durations of preceding vowels were conditioned by the phonological specification of voicing for the consonant. Our results, while consistent with their conclusion, do indicate that the magnitude of lengthening before [+voiced] flaps varies with the stress environment of the flap. Furthermore, the slightly different results suggest that dialectal/idiolectal differences play a role in lengthening before voicing phenomena.

Davis and Summers (1989) found that vowels preceding prestress voiced stops tended to be longer (but not significantly) than vowels preceding prestress voiceless stops. The results in this experiment are in the opposite direction. Davis and Summers (1989) used their non-significant result to support the view of Maddieson (1985) that lengthening before voicing is not a syllable-conditioned process (the lengthened vowel is not tautosyllabic with the following consonant in the vCV case). The results of this experiment tend to support the view expressed in Selkirk (1982) that lengthening before voicing is a syllable-conditioned process. However, the most that can be tentatively concluded from the variety of results obtained here and across the literature is that there are dialectal/idiolectal differences with respect to lengthening before voicing phenomena. Alternatively, it is possible that in some studies, observed effects are artifacts of the way the test sentences are presented.

4.3.3. The Effect of a Longer Phonological Phrase

The presence of a modifier before the test word has virtually no effect on the durations in the test consonants for all speakers. These results corroborate Klatt's (1976) observation that speakers tend not to adjust segment durations in order to satisfy global rhythmic constraints.

4.3.4. Interspeaker Variability

The stress effect described above as well as the lengthening before [+voiced] flaps varies according to speaker; speaker AM in this study showed significant length differences between the flapping and non-flapping environments only for the alveolars and /k/, while in other speakers the effect is more widespread.

5. Summary

The experiment has shown the following:

1) For all stops except /g/, a phonetic consequence of occurring in flapping environments is reduced length: Total stop durations word-medially in all places of articulation tend to be longer in prestress position. This suggests that the American English Flapping Rule is driven by an underlying timing mechanism. The results are slightly different than those of Stathopoulos and Weismer (1983) who found that stop closure durations tend to be longer in prestress position.

2) The results suggest that the American English Flapping Rule is a foot-conditioned process; they support the account of Kiparsky (1979) in which only non-foot-initial alveolar stops may flap.

3) Voiced stop shortening/lengthening takes place in the closure position, whereas voiceless stop lengthening/shortening takes place mostly in the aspiration portion. Comparing the closure, aspiration, and vowel duration differences in different stress environments shows a compressibility/stretchability hierarchy where vowels are most stretchable/compressible, aspiration is less compressible/stretchable, and closure is the least compressible/stretchable.

4) Unlike the results of Fox and Terbeek (1977), significant differences were found between measurements of vowels preceding flapped /t/'s and measurements of vowels preceding flapped /d/'s only when the alveolar occurred between two unstressed vowels.

Notes

1. In this study, closure durations of voiced stops are shorter than the closure durations of voiceless stops for all places of articulation in the poststress environment (V_v), and closure durations of voiced stops are shorter than the closure durations of voiceless stops for non-alveolars in the between unstressed environment (v_v). In contrast, in the prestress environment (v_V), voiced stop closures are longer than voiceless stop closures for labials and alveolars. In the velar case, voiced stop closures are shorter than voiceless stop closures. These findings support the findings of Davis and Summers

(1989) that the direction of length differences between closures of voiced and voiceless stops in the prestress position is variable. Also, the findings support their claim that closure duration is a reliable cue to voicing after a stressed vowel.

References

Allen, G.D. (1975) "Speech Rhythm: Its Relation to Performance Universals and Articulatory Timing," Journal of Phonetics 3, 75-86.

Chen, M. (1970) "Vowel Length as a Function of the Voicing of the Consonant Environment," Phonetica 22, 129-159.

Chierchia, G. (1983/1986) "Length, Syllabification, and the Phonological Cycle in Italian," Journal of Italian Linguistics 8:5-34.

Davis, S. and W.V. Summers (1989) "Vowel length and closure duration in word-medial VC sequences," Journal of Phonetics 17, 339-353.

Fox, R.A. and D. Terbeek (1977) "Dental flaps, vowel duration and rule ordering in American English"Journal of Phonetics 5, 27-34.

Gussenhoven, C. (1986) "English Plosive Allophones and Ambisyllabicity," Gramma 10.2, Instituut Nederlands, Un. of Nijmegen, 119-141.

Inouye, S.B. (1989) "The Flap as a contour segment," UCLA Working Papers 72, 39-81.

Kahn, D. (1976) Syllable-based generalizations in English phonology, Doctoral Dissertation, MIT, distributed by Indiana University Linguistics Club, Indiana University, Bloomington.

Kiparsky, P. (1979) "Metrical structure assignment is cyclic," Linguistic Inquiry 10, 3, 421-441.

Klatt, D.H. (1974) "The duration of [s] in English words," Journal of Speech and Hearing Research 17, 51-63.

Klatt, D.H. (1976) "Linguistic uses of segmental duration in English: acoustic and perceptual evidence," Journal of the Acoustical Society of America 59, 1208-1221.

Kozhevnikov, V.A., and L.A. Chistovich (1965) Speech: Articulation and Language, translated by Clearinghouse for Federal and Scientific Information (Washington D.C.).

Lehiste, I. (1970) Suprasegmentals, Cambridge (Massachussetts) and London: MIT Press.

Lehiste, I. (1977) "Isochrony Reconsidered," Journal of Phonetics 5, 253-263.

Maddieson, I. (1985) "Phonetic Cues to Syllabification," in V. Fromkin, ed., Phonetic Linguistics Essays in Honor of Peter Ladefoged, Academic Press, Inc., pp. 203-221.

Port, R.F. (1981) "Linguistic Timing Factors in Combination," Journal of the Acoustical Society of America 69, 262-274.

Selkirk, L. (1982) "The Syllable," in H. van der Hulst and N. Smith, (eds.) The Structure of Phonological Representations (Part 2), Foris Publications 337-383.

Stathopoulos, E. and G.Weismer (1983) "Closure duration of stop consonants," Journal of Phonetics 11, 395-400.

Umeda, N. (1977) "Consonant duration in American English," Journal of the Acoustical Society of America 61, 846-858.

Zue, V. and M. Laferriere (1979) "Acoustic study of medial /t,d/ in American English,"Journal of the Acoustical Society of America 66: 1039-1050.

Appendix A

He avoided the leper for health reasons.
He avoided the dirty leper for health reasons.
She preferred to repair the green car.
She preferred to thoroughly repair the green car.
He misplaced the caliper on his workbench.
He misplaced the outside caliper on his workbench.
He found the metal in the ground. (Speakers RG, AM and LE)
He found the copper metal in the ground. (Speakers RG, AM and LE)
Her dress looks metallic from a distance. (Speakers RG, AM and LE)
Her dress looks vaguely metallic from a distance. (Speakers RG, AM and LE)
She measured the diameter with a ruler. (Speakers RG, AM and LE)
She measured the indicated diameter with a ruler. (Speakers RG, AM and LE)
We expect the local on track fifteen.
We expect the incoming local on track fifteen.
She found the locale of the murder.
She found the mysterious locale of the murder.
He observed the follicle under the microscope.
He observed the little follicle under the microscope.
He watched the rabble from the tower.
He watched the cheering rabble from the tower.
She preferred to rebel against her captors.
She preferred to publicly rebel against her captors.
Seh studied the parable in religion class.
She studied the second parable in religion class.
She won the medal at the Olympics. (Speakers RG, AM and LE)
She won the silver medal at the Olympics. (Speakers RG, AM and LE)
He found the medallion on the floor. (Speakers RG, AM and LE)
He found the copper medallion on the floor. (Speakers RG, AM and LE)
The object looked pyramidal in shape. (Speakers RG, AM and LE)
The object looked perfectly pyramidal in shape. (Speakers RG, AM and LE)

He admired the vigor of the athletes.
He admired the steroid-induced vigor of the athletes.
He smoked the cigar in the waiting room.
He smoked the Cuban cigar in the waiting room.
He added the vinegar to the salad.
He added the cider vinegar to the salad.

Appendix B

1. The Effect of Stress on Total Durations

F and p values of the effect of Stress on total durations of /p,t,d,b,d,g/ from a repeated measures ANOVA with dependent variables of stress environment (Stress Position), sentence length (±Modifier), and the repeated measures factor of Speaker.

/p/: F (2,24) = 17.37; p < .0001
/t/: F (2,24) = 370.80; p < .0001
/k/: F (2,24) = 57.17; p < .0001
/b/: F (2,24) = 80.76; p < .0001
/d/: F (2,24) = 176.73; p < .0001
/g/: F (2,24) = .56; p < .5784

2. The Effect of Stress on Closure Durations

F and p values of the effect of Stress on closures durations of /p,t,d,b,d,g/ from a repeated measures ANOVA with dependent variables of stress environment (Stress Position), sentence length (±Modifier), and the repeated measures factor of Speaker.

/p/: F (2,24) = 54.27; p < .0001
/t/: F (2,24) = 17.14; p < .0001
/k/: F (2,24) = 2.423; p < .0001
/b/: F (2,24) = 61.49; p < .0001
/d/: F (2,24) = 167.31; p < .0001
/g/: F (2,24) = 1.47; p < .5784

3. The Effect of Stress on Aspiration Durations

F and p values of the effect of Stress on aspiration durations of /p,t,d,b,d,g/ from a repeated measures ANOVA with dependent variables of stress environment (Stress Position), sentence length (±Modifier), and the repeated measures factor of Speaker.

/p/: F (2,24) = 78.24; p < .0001
/t/: F (2,24) = 405.40; p < .0001
/k/: F (2,24) = 67.40; p < .0001
/b/: F (2,24) = 15.74; p < .0001
/d/: F (2,24) = 13.33; p < .0001
/g/: F (2,24) = 3.21; p < .5784

Appendix C

F and p values from one factor ANOVAs for each consonant for each speaker:

A. Stress on Total Consonant Duration

Speaker AW:	Speaker AM:
/p/: F(2,27) = 10.88; p < .0003	/p/: F(2,27) = 4.18; p < .0003
/t/: not collected	/t/: F(2,27) = 106.62; p < .0001

/k/: $F(2,27) = 16.93$; $p < .0001$
/b/: $F(2,27) = 69.04$; $p < .0001$
/d/: not collected
/g/: $F(2,27) = 2.3877$; $p > .1109$

/k/: $F(2,29) = 3.17$; $p > .0578$
/b/: $F(2,27) = 11.20$; $p < .0001$
/d/: $F(2,27) = 36.03$; $p < .0001$
/g/: $F(2,27) = 9.98$; $p < .0006$

Speaker RG:
/p/: $F(2,27) = 1.68$; $p > .2055$
/t/: $F(2,27) = 84.26$; $p < .0001$
/k/: $F(2,27) = 12.72$; $p < .0001$
/b/: $F(2,27) = 15.43$; $p < .0001$
/d/: $F(2,27) = 26.09$; $p < .0001$
/g/: $F(2,27) = 3.19$; $p > .0569$

Speaker LE:
/p/: $F(2,27) = 16.47$; $p < .0001$
/t/: $F(2,27) = 89.32$; $p < .0001$
/k/: $F(2,27) = 40.90$; $p < .0001$
/b/: $F(2,27) = 27.76$; $p < .0001$
/d/: $F(2,27) = 61.65$; $p < .0001$
/g/: $F(2,27) = 4.22$; $p < .0254$

Results from a one factor ANOVA which compares the total durations of stops in the two flapping environments (V_v vs. v_v):

Speaker AW:
/p/: $F(2,18) = 3.31$; $p > .0853$
/t/: not collected
/k/: $F(2,18) = 12.05$; $p < .0027$
/b/: $F(2,18) = 10.09$; $p < .0052$
/d/: not collected
/g/: $F(2,18) = .11$; $p > .7416$

Speaker AM:
/p/: $F(2,18) = 3.14$; $p > .0934$
/t/: $F(2,18) = 2.10$; $p < .1647$
/k/: $F(2,18) = 4.20$; $p > .0554$
/b/: $F(2,18) = 22.41$; $p < .0002$
/d/: $F(2,18) = 26.24$; $p < .0001$
/g/: $F(2,18) = 18.27$; $p < .0005$

Speaker RG:
/p/: $F(2,18) = .13$; $p > .7178$
/t/: $F(2,18) = .02$; $p > .8989$
/k/: $F(2,18) = 1.19$; $p > .2906$
/b/: $F(2,18) = .60$; $p > .4498$
/d/: $F(2,18) = 9.64$; $p < .0061$
/g/: $F(2,18) = 6.07$; $p < .024$

Speaker LE:
/p/: $F(2,18) = 7.31$; $p < .0145$
/t/: $F(2,18) = 1.39$; $p > .2545$
/k/: $F(2,18) = 27.46$; $p < .0001$
/b/: $F(2,18) = 9.85$; $p < .0057$
/d/: $F(2,18) = 50.33$; $p < .0001$
/g/: $F(2,18) = .32$; $p > .5769$

B. On Closure Durations:

Speaker AW:
/p/: $F(2,27) = .96$; $p > .3942$
/t/: not collected
/k/: $F(2,27) = 6.53$; $p < .0049$
/b/: $F(2,27) = 44.99$; $p < .0001$
/d/: not collected
/g/: $F(2,27) = .31$; $p > .7333$

Speaker AM:
/p/: $F(2,27) = 102.39$; $p < .0001$
/t/: $F(2,27) = 4.22$; $p < .0254$
/k/: $F(2,27) = 4.91$; $p < .0152$
/b/: $F(2,27) = 10.46$; $p < .0004$
/d/: $F(2,27) = 48.93$; $p < .0001$
/g/: $F(2,27) = 7.89$; $p < .002$

Speaker RG:
/p/: $F(2,27) = 15.54$; $p < .0001$
/t/: $F(2,27) = 11.41$; $p < .0003$
/k/: $F(2,27) = 5.89$; $p < .0075$
/b/: $F(2,27) = 18.32$; $p < .0001$
/d/: $F(2,27) = 28.55$; $p < .0001$
/g/: $F(2,27) = 2.43$; $p > .1068$

Speaker LE:
/p/: $F(2,27) = 16.03$; $p < .0001$
/t/: $F(2,27) = 5.04$; $p < .0138$
/k/: $F(2,27) = .21$; $p > .8136$
/b/: $F(2,27) = 16.59$; $p < .0001$
/d/: $F(2,27) = 65.14$; $p < .0001$
/g/: $F(2,27) = 4.24$; $p < .0251$

Results from a one factor ANOVA which compares the closure durations of stops in the two flapping environments (V_v vs. v_v):

Speaker AW:
/p/: F(2,18) = 1.42; p > .2483
/t/: not collected
/k/: F(2,18) = 6.35; p < .0214
/b/: F(2,18) = 10.09; p < .0052
/d/: not collected
/g/: F(2,18) = .11; p > .7416

Speaker AM:
/p/: F(2,18) = 43.51; p < .0001
/t/: F(2,18) = 2.10; p > .1647
/k/: F(2,18) = 4.93; p < .0393
/b/: F(2,18) = 22.41; p < .0002
/d/: F(2,18) = 29.10; p < .0001
/g/: F(2,18) = 14.04; p < .0015

Speaker RG:
/p/: F(2,18) = 15.00; p < .0011
/t/: F(2,18) = 1.90; p > .1852
/k/: F(2,18) = .41; p > .5304
/b/: F(2,18) = .01; p > .9345
/d/: F(2,18) = 16.93; p < .0007
/g/: F(2,18) = 2.72; p > .1162

Speaker LE:
/p/: F(2,18) = 16.59; p < .0007
/t/: F(2,18) = .62; p > .4398
/k/: F(2,18) = .01; p > .9305
/b/: F(2,18) = 8.29; p < .01
/d/: F(2,18) = 65.72; p < .0001
/g/: F(2,18) = .23; p > .6354

C. On Aspiration Duration

Speaker AW:
/p/: F(2,27) = 23.30; p < .0001
/t/: not collected
/k/: F(2,27) = 8.09; p < .0018
/b/: F(2,27) = 14.58; p < .0001
/d/: not collected
/g/: F(2,27) = 8.32; p < .0015

Speaker AM:
/p/: F(2,27) = 80.89; p < .0001
/t/: F(2,27) = 192.68; p < .0001
/k/: F(2,27) = 19.44; p < .0001
/b/: F(2,27) = 1; p > .3811
/d/: F(2,27) = 2.29; p < .1207
/g/: F(2,27) = .99; p > .38

Speaker RG:
/p/: F(2,27) = 15.02; p < .0001
/t/: F(2,27) = 86.43; p < .0001
/k/: F(2,27) = 39.23; p < .0001
/b/: F(2,27) = 2.74; p > .0823
/d/: F(2,27) = 3.82; p < .0347
/g/: F(2,27) = 3.36; p < .0496

Speaker LE:
/p/: F(2,27) = 56.83; p < .0001
/t/: F(2,27) = 251.91; p < .0001
/k/: F(2,27) = 29.93; p < .0001
/b/: F(2,27) = 1.50; p > .24
/d/: F(2,27) = 13.28; p < .0001
/g/: F(2,27) = .60; p > .555

Results from a one factor ANOVA which compares the aspiration durations of stops in the two flapping environments (V_v vs. v_v):

Speaker AW:
/p/: F(2,18) = 22.26; p = .0002
/t/: not collected
/k/: F(2,18) = 1.45; p > .2442
/b/: 0 ms aspiration for all tokens.
/d/: not collected
/g/: 0 ms aspiration for all tokens.

Speaker AM:
/p/: F(2,18) = 69.78; p < .0001
/t/: 0 ms aspiration for all tokens.
/k/: F(2,18) = .10; p > .7541
/b/: 0 ms aspiration for all tokens.
/d/: F(2,18) = 2.95; p > .103
/g/: F(2,18) = 2.26; p > .1499

Speaker RG:
/p/: F(2,18) = 13.30; p < .0018
/t/: F(2,18) = 2.22; p > .1539
/k/: F(2,18) = .6515; p > .4301
/b/: F(2,18) = 3.78; p > .0676
/d/: F(2,18) = 1.00; p > .3306
/g/: F(2,18) = 5.66; p < .0286

Speaker LE:
/p/: F(2,18) = .07; p > .7902
/t/: F(2,18) = 3.78; p > .0677
/k/: F(2,18) = 19.19; p < .0004
/b/: F(2,18) = .01; p > .921
/d/: F(2,18) = 2.23; p > .1529
/g/: F(2,18) = .67; p > .4247

The Timing of Phones and Transitions: Toward a Nucleus-Based Model of English Duration

Susan R. Hertz

This paper presents a brief overview of a new duration model for General American English, which I am currently developing in connection with work being done at both Cornell University and Eloquent Technology, Inc. in speech synthesis by rule. In addition to pursuing the standard goals of synthesis rule development, such as high intelligibility and naturalness, we strive in this work to develop a linguistically realistic system with which we can gain insights into the timing structure of English and other languages.

The paper expands upon Hertz (1991), which describes and motivates a model of speech timing called the *phone-and-transition model.* While the phone-and-transition model provides an appropriate framework for synthesis of any language or dialect (see Hertz, 1990a), this paper will focus on its application to General American English (henceforth "GA"), presenting the basic rule algorithm for determining the internal timing structure of stressed syllable nuclei in GA.

The first section gives a brief outline of the phone-and-transition model. The second section describes the methodology being used to develop the duration model for GA. The third section discusses the duration model, focussing on the rules that compute the durations of the sub-units of stressed syllable nuclei. A final section provides some concluding remarks.

1. The Phone-and-Transition Model

The duration rules build on a phonetic model in which *phones* and *transitions* are explicit units that can be manipulated independently by rules (Hertz, 1991). Consider, for example, the following representation of the word *tot* in GA, as uttered in the frame *Say ____for me*:

(1) *tot:*

```
phone:       |t    |     |a    |     |t    |
F2:          |1800 |     |1300 |     |1700 |
transition:  |     |trans|     |trans|     |
millisecond: |95   |70   |85   |50   |85   |
```

This structure, which is produced by our current speech synthesis rules for GA, is called a *delta*, because it consists of multiple interconnected *streams* (e.g., `phone`, `F2`, `transition`, and `millisecond`) like a river delta. (In this and other sample deltas below, only the streams relevant to the example are shown. Deltas produced by our synthesis rules contain many other streams as well.)

In this delta, each unit in the `phone` stream (i.e., each phone) is synchronized by the vertical bars (*sync marks*) with the formant values (*targets*) that are a direct result of the articulation of the phone itself.[1] Intervening transitions represent the movement of the formants from the target of one phone to the target of the next. The formant values during the transitions between phones are computed by interpolating between the formant targets of adjacent phones over the duration specified for the transition (e.g., by interpolating between 1800 Hz and 1300 Hz over 70 ms for the transition between the phones `[t]` and `[a]`). (Although not shown in this example, a phone sometimes has two targets with different values at the beginning and end of the phone, in which case there is interpolation within the phone as well.)

A model based on explicit phones and transitions leads to more straightforward timing rules than conventional models in which the transitions are not treated as independently manipulable units. Consider, for example, the spectrograms of *tot* `[tat]` and *dot* `[dat]` in Figure 1. In these spectrograms, we see that the aspiration after the `[t]` of *tot* is superimposed on the transition between the `[t]` and the following `[a]`. In *dot*, on the other hand, the transition from `[d]` to `[a]` is voiced. Within the phone-and-transition model, we can easily capture the generalization that stop aspiration aligns with the transition following the stop, as shown below for *tot*:

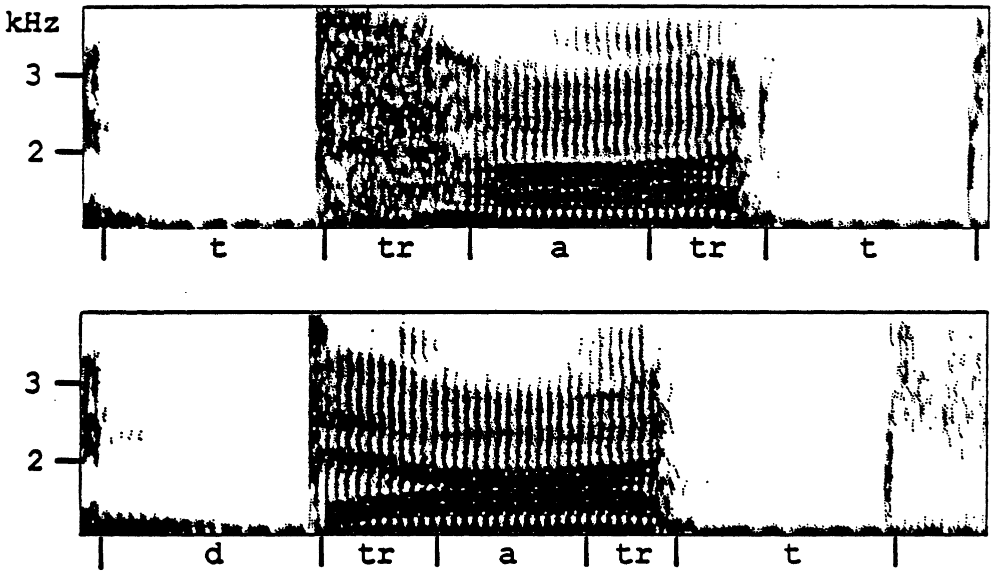
(2) *tot:*

```
phone:       |t   |     |a   |     |t   |
F2:          |1800|     |1300|     |1700|
aspiration:  |    |70   |               |
voicing:     |0         |60        |0   |
transition:  |    |trans|    |trans|    |
millisecond: |95  |70   |85  |50   |85  |
```

All vertical bars in the same column represent the same sync mark. Thus, 70 decibels of aspiration amplitude is synchronized with the transition after the phone `[t]`, and voicing (i.e., 60 dB of voicing amplitude) starts at the beginning of the phone `[a]`. This structure

---

[1] This example only shows the second formant targets. Targets for all of the formants are generally aligned with each other. In positioning the targets, we have abstracted away from small timing differences between formants in natural speech, where the formants for a phone do not always reach their targets at precisely the same point (Kewley-Port, 1982). In our experience, these differences are not important for intelligibility, though some of them may result in improved naturalness. We believe that such refinements can be made by low-level adjustments to the positioning of the relevant targets after all the basic duration rules have applied. Any such adjustments that might be needed are a minor concern compared to the considerable advantages for rule development of using a uniform structure for the timing of formant targets.



`tr` = transition; 1 inch = 105 ms

Figure 1. Spectrograms of *tot* and *dot.*

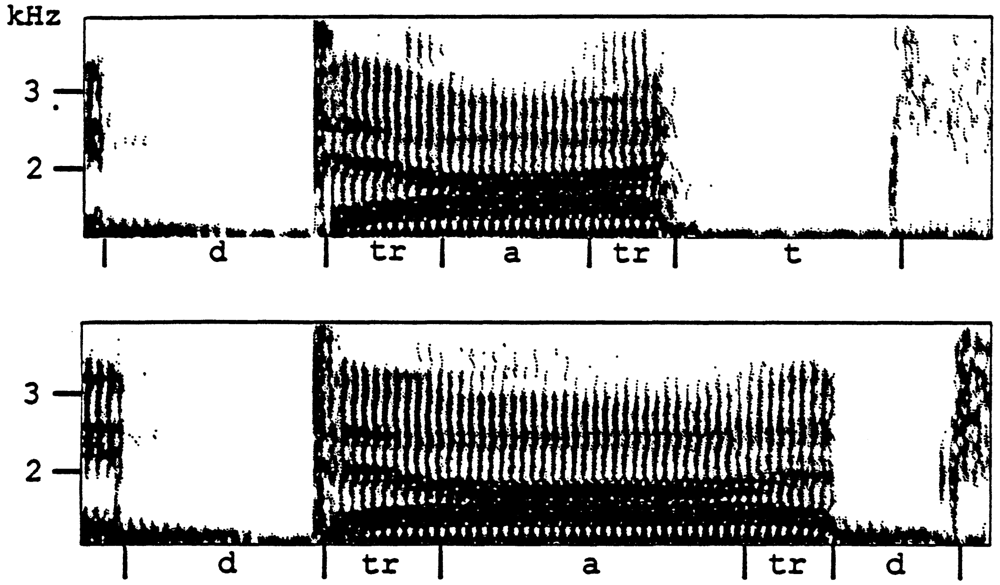
may be contrasted with that of *dot*, in which voicing starts at the beginning of the transition following [d], and there is no aspiration:

(3) *dot:*

```
phone:        |d   |     |a   |     |t    |
F2:           |1800|     |1300|     |1700|
aspiration:   |0                          |
voicing:      |0   |60              |0    |
transition:   |    |trans|    |trans|     |
millisecond:  |85  |50   |55  |50   |95   |
```

With independent phones and transitions, then, we can express the timing behavior of aspiration and voicing straightforwardly. More conventional models, without independent transitions (e.g., Klatt, 1979; Hertz, 1982), have led to arbitrary treatments of aspiration and to unnecessarily complicated rules, as discussed in Hertz (1991).

Explicit phones and transitions also lead to a straightforward account of vowel lengthening. For example, when a vowel lengthens before a tautosyllabic voiced stop, the lengthening occurs almost exclusively in the phone, with the adjacent transitions lengthening relatively little or not at all, as shown by the spectrograms of *dot* [dat] and *Dodd* [dad] in Figure 2. It is interesting to note that although the utterances represented in these spectrograms were articulated very carefully, with the lengthening for [dad] even slightly exaggerated, the lengthening is nonetheless restricted primarily to the phone [a], with the initial transitions in both utterances having virtually identical durations.



tr = transition; 1 inch = 105 ms

Figure 2. Spectrograms of *dot* and *Dodd.*

While the final transitions have different durations, we attribute the difference to early cessation of voicing during the movement from [a] to [t], rather than to lengthening before [d], as discussed further in Section 3.

The following deltas of *dot* and *Dodd* clearly indicate that the lengthening before [d] occurs in the phone:

(4). *dot:*

phone:	d		a		t
transition:		trans		trans	
millisecond:	85	50	55	50	95

Dodd:

phone:	d		a		d
transition:		trans		trans	
millisecond:	85	50	110	50	50

The durations in the deltas are the values produced by our rules, not the precise values in the example spectrograms. As discussed in Section 2, our rules are based on general patterns observed across many tokens, not properties of any specific utterance.

Based on observations about timing patterns such as those for aspiration and vowel lengthening, we have adopted independent phones and transitions as the basis for the duration rules, which account for a wide range of timing patterns in GA. We have found similar considerations to motivate independent phones and transitions for other American English dialects and for other languages. For example, we have also observed relatively stable consonant-vowel transitions in Japanese and Hindi syllables with phonemically (and phonetically) long and short vowels. Hertz (1990a) illustrates how independent phones and transitions allow us to make generalizations across dialects and languages.

2. Methodology

The present duration rules for GA are based on an extensive study of the speech of a female speaker of GA, although the basic principles underlying the model have all been observed for other speakers of the dialect. The primary methodology being used to develop the model is to formulate and test hypotheses cyclically, alternating between analysis and synthesis. In the analysis phase of each cycle, we study natural speech data, primarily by examining spectrograms, and develop hypotheses concerning the underlying phonetic structure of the utterances. To help develop these hypotheses, we segment spectrograms into phones and transitions (see below), and measure and compare their durations. (Spectrograms are made with the Kay real-time DSP Sona-Graph (model 5500) and Entropic System's Waves/ESPS speech analysis software running on a Sun Workstation.)

The synthesis phase of the rule development methodology is carried out with Eloquent Technology's Delta System, a software system for phonology and phonetics that is centered around the multi-stream delta data structure for representing utterances (Hertz, 1988a, 1990b; Charif, Hertz, & Weber, 1992). The Delta System includes an interactive environment with which deltas can be built "by hand," and a programming language for writing formal rules. In the synthesis phase, we use both components of the system to write specific rules and build utterance representations that embody the hypotheses formed during the analysis phase. Speech is then synthesized on the basis of the deltas, and the synthetic output is evaluated. Among other things, evaluation includes listening to the synthetic speech back-to-back with natural speech, and visually comparing spectrograms of synthetic and natural speech. We also administer periodic intelligibility tests.

To expedite rule development in the future, we are currently implementing a multi-dialect linguistic/acoustic database, which we can query to extract generalizations for a number of American English dialects and a number of languages. The database contains both linguistic information (syllable structure, phonemic structure, degrees of stress, etc.) and acoustic values (durations of phones and transitions; formant values at the edges of phones; periods of voicing, aspiration, noise, and nasalization; and so on) for a large number of utterances. We are using the Waves/ESPS speech analysis software to segment spectrograms into phones and transitions, to mark other relevant information, and to automatically extract formant values. We then enter the durations of the phones and transitions and the other information obtained with Waves/ESPS into the database, using a program that we wrote. The database is currently implemented with Borland's Paradox relational database software on an IBM PC networked to the Sun Workstation on which the initial speech analysis is done.

3. The Duration Model

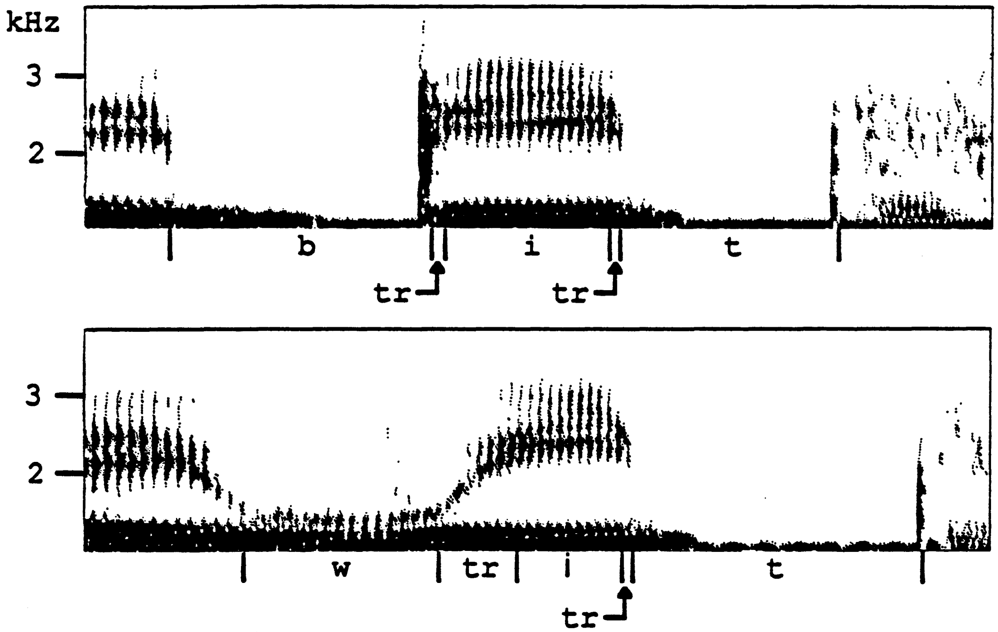
The duration rules are based on the premise that a unit that we tentatively call the *acoustic nucleus*, which contains specific phones and transitions as sub-units (see below), serves a basic organizational role within the durational system. For example, within the acoustic nucleus, there is a trading relationship among the phone and transition durations (which we believe holds cross-linguistically) such that the total of the phone and transition durations will yield a relatively constant duration for a nucleus of a given type. Similarly, certain processes that change duration in English, such as lengthening before voiced obstruents, operate on the component phones (and in some cases also the transitions) in the acoustic nucleus in such a way that the duration of the nucleus as a whole is modified by the appropriate amount.

This section presents an overview of the rules that generate durations for stressed syllable nuclei in our current synthesis program for GA. The first subsection motivates the acoustic nucleus unit on which the rules are based; the second subsection discusses the actual rules.

3.1. The Acoustic Nucleus

As a point of departure, consider the duration of vowel phones and transitions in monosyllables with the structure C_1VC_2, where C_1 is a voiced consonant, V a vowel, and C_2 a voiceless stop (henceforth "final voiceless stop monosyllables"). In such syllables, we observe that the longer the adjacent transitions are, the shorter the vowel phone is, with the total duration of the transitions and the phone remaining relatively constant in a given context, as can be seen in the spectrograms of *beat* and *wheat* in Figure 3.

In both *beat* and *wheat*, the total duration of the phone `[i]` and the transitions on either side is about 100 ms. This duration, however, is distributed differently among the transitions and the `[i]` in the two cases. In *beat*, the transition from `[b]` to `[i]` is about 10 ms long, and `[i]` itself is 85 ms. In *wheat*, on the other hand, the transition from `[w]` to `[i]` is about 40 ms long, and `[i]` itself is 55 ms. The final transitions into `[t]` are about 5



tr = transition; 1 inch = 105 ms

Figure 3. Spectrograms of *beat* and *wheat.*

ms long in each case. Thus, in *beat*, the total transition duration is 30 ms longer than in *wheat*, and the [i] is 30 ms shorter.

Since the transitions together with the vowel phone have a relatively constant duration, it is convenient to group the transitions plus the phone into a higher level unit of duration, which, as mentioned above, we call the *acoustic nucleus*, or simply *nucleus* (although we recognize that our use of this term differs from its various senses in phonology). Consider, for example, the following deltas for *beat* and *wheat*:

(5) *beat*:

```
nucleus:      |  |nuc               |  |
phone:        |b |      |i  |       |t |
transition:   |  |trans |   |trans  |  |
millisecond:  |  |10    |85 |5      |  |
                 \_________________/
                         100
```

wheat:

```
nucleus:      |  |nuc               |  |
phone:        |w |      |i  |       |t |
transition:   |  |trans |   |trans  |  |
millisecond:  |  |40    |55 |5      |  |
                 \_________________/
                         100
```

Note the different distribution of the total nucleus duration among the components of the nucleus in the two cases.

With an explicit nucleus unit, then, we can express the trading relationship between phones and transitions straightforwardly. In addition, we can easily write rules that modify the durations of phones and transitions differentially. For example, we can model lengthening before voiced obstruents with a rule that stretches the phone in the nucleus, but not the neighboring transitions, when the nucleus precedes a voiced obstruent (see Section 1 above).

It is not only vowel phones, however, that lengthen in syllables ending in a voiced obstruent, but also any following tautosyllabic glides and liquids, as discussed in Hertz (1991; see also Chen, 1970). For example, in words like *wide* [`wayd`], *willed* [`wIld`], and *wild* [`wayld`], the vowel *and* the following [`y`] and/or [`l`] lengthen.[2] We thus include in the nucleus not only the vowel of the syllable, but also any following glides and liquids (along with the relevant transitions), as illustrated by the following deltas for the words *white* and *wilt*:

(6) *white*:

```
nucleus:      |  |nuc                    |  |
phone:        |w |     |a |     |y |     |t |
transition:   |  |trans|  |trans|  |trans|  |
```

wilt:

```
nucleus:      |  |nuc                    |  |
phone:        |w |     |I |     |l |     |t |
transition:   |  |trans|  |trans|  |trans|  |
```

Given such structures, the lengthening rule can be expressed quite simply as a rule that lengthens all phones within a nucleus that precedes a voiced obstruent. Additional motivation for the nucleus comes from the fact that the precise degree of lengthening is constrained by the overall nucleus duration, as discussed in Section 3.2.

3.2. The Rules

The duration rules begin by assigning to each acoustic nucleus a duration typical of the duration of the nucleus in a final voiceless stop monosyllable uttered in the frame *Say ____for me*. (Syllables occurring in this frame will henceforth be referred to as "phrase-medial.") The rules then modify the starting durations according to the actual context. Since we have observed consistent differences among the durations of nuclei before different voiceless stops (e.g., nuclei before [`p`] tend to be shorter than nuclei before [`t`]

---

[2] The transitions seem to be more stable in some sonorant sequences than others, but the general principle that most of the lengthening is in the phones holds across the various sequences. We are still examining the factors that determine whether transitions will lengthen and by how much.

or [k]), we have arbitrarily selected monosyllables ending in [t] for the starting durations.[3]

The starting duration of the nucleus (transitions plus phones) depends on the particular phones in the nucleus. For example, it is well-known for single-phone nuclei that a nucleus containing a mid tense vowel ([e] or [o]) is longer than a nucleus containing a high tense vowel ([i] or [u]), all other factors being equal. Similarly, a nucleus containing a tense vowel is longer than a nucleus containing the corresponding non-tense vowel (e.g., a nucleus with [e] is longer than one with [E]), and nuclei containing low vowels are longest, everything else being equal (Peterson & Lehiste, 1960).

The rules distribute the nucleus duration among the component phones and transitions as follows. First, durations are assigned to the transitions and to any non-vowel phones (i.e., glides and/or liquids) in the nucleus, with rules that depend on the features of the constituent phones. Then the vowel of the nucleus is assigned a duration by subtracting the total duration of the transitions and non-vowel phones (henceforth the "total non-vowel duration") from the total nucleus duration.

Consider, for example, the nucleus of *white*, which contains the two phones [a] and [y]. First, the rules assign to the nucleus a duration of 145 ms, the default starting duration assigned to two-phone nuclei. (A few two-phone nuclei—particularly those consisting of tense vowels followed by [l]—are assigned longer durations.) Next, the rules assign specific durations to the transitions and to the phone [y], as shown below:

(7) *white:*

```
nucleus:      |  |nuc                     |  |
phone:        |w |     |a |     |y  |     |t |
transition:   |  |trans|  |trans|   |trans|  |
millisecond:  |  |30    |  |80    |15 |10    |  |.
```

The rules then give the phone [a] the duration needed to bring the total nucleus duration to 145 ms—that is, 10 ms:

---

[3] While the choice of the starting environment is somewhat arbitrary, there is preliminary evidence from our dialect studies that using the durations in voiceless stop monosyllables will allow for the simplest description of durational differences among dialects. In particular, we have observed in preliminary studies of Black English and some Southern dialects, that the vowel durations in final voiceless stop monosyllables are similar to those in GA, while the vowel durations in other contexts (e.g., before voiced stops) are very different. If this preliminary result turns out to hold across dialects, it will justify considering the durations of acoustic nuclei in voiceless stop monosyllables to be basic in some sense, and hence, to be the appropriate starting durations. The dialects would then be considered to differ primarily in degrees of lengthening of the nuclei, rather than shortening.

(8) *white:*

```
nucleus:     |   |nuc                      |  |
phone:       |w  |     |a  |     |y  |     |t |
transition:  |   |trans|   |trans|   |trans|  |
millisecond: |   |30   |10 |80   |15 |10   |  |
                            145
```

Now consider the word *wilt*. As in *white*, the rules first assign a total nucleus duration of 145 ms. Then they assign durations to the transitions and to the [l]. In this case, however, the total non-vowel duration is 165 ms, which happens to be 20 ms greater than the starting nucleus duration. Thus, when the non-vowel duration is subtracted from the total nucleus duration, the vowel receives a "negative duration" of –20 ms, as shown below:

(9) *wilt:*

```
nucleus:     |    |nuc                        |    |
phone:       |w   |     |I   |     |l   |     |t   |
F2:          |600 |     |1640|     |760 |     |1700|
transition:  |    |trans|    |trans|    |trans|    |
millisecond: |    |40   |-20 |60   |50  |15   |    |
                                 145
```

Negative durations are overridden with the appropriate positive durations in most lengthening contexts. When a vowel phone still has a negative duration after all the duration rules have applied, the vowel is given a duration of 0 ms, and the rules (still tentative) shorten selected transitions in the nucleus by the appropriate amounts. In the case of *wilt*, for example, no further duration rules apply to the nucleus, so the transitions between [w] and [I] and between [I] and [l] are each shortened by 10 ms, yielding:

(10) *wilt:*

```
nucleus:     |    |nuc                        |    |
phone:       |w   |     |I   |     |l   |     |t   |
F2:          |600 |     |1640|     |760 |     |1700|
transition:  |    |trans|    |trans|    |trans|    |
millisecond: |    |30   |0   |50   |50  |15   |    |
                                 145
```

Note that since [I] has a duration of 0 ms, its second formant value of 1640 Hz functions as a durationless target. The second formant pattern moves from the 600 Hz target of the [w] over 30 ms to the 1640 Hz target of [I]. From there it moves immediately over 50 ms to the 760 Hz target of the [l], which is held for 50 ms before the pattern moves on to the 1700 Hz target of the [t].

In all of the examples so far, the transitions have been treated as part of the nucleus.[4] This, however, has been a slight oversimplification, since it is actually only voiced portions of transitions that contribute duration to the nucleus, and are therefore included in the computation of the total non-vowel duration. The transition into the final [t] of *wilt* is only voiced for a small fraction of its duration. (The early cessation of voicing partway through transitions into voiceless obstruents is discussed in Hertz (1991; see also Klatt, 1976). However, since this transition is so short to begin with, this detail does not significantly affect the duration computations presented for *wilt* above.

Let us now compare the words *tot* and *dot*. For each of these words, the rules assign a total nucleus duration of 140 ms, but the total non-vowel duration differs considerably. In *dot*, the initial 50 ms transition is voiced, but only the first 35 ms of the final transition is voiced. The total non-vowel duration is thus 85 ms. Now consider *tot*, which is like *dot* in all respects except that the initial transition is aspirated and not voiced, and therefore does not contribute duration to the nucleus. Thus the total non-vowel duration for *tot* is only 35 ms (the duration of the voiced portion of the final transition), and the vowel receives a starting duration of 105 ms (140 – 35). The different starting durations of the vowels of *tot* and *dot* are illustrated in the following deltas, which also show the cessation of voicing 35 ms into the final transition:

(11) *tot:*

```
nucleus:      |        |nuc     |      |
phone:        |t |     |a   |       |t |
aspiration:   |0 |70   |0              |
voicing:      |0       |60      |0     |
transition:   |  |trans|    |trans  |  |
millisecond:  |  |70   |105 |35 |15 |  |

                         140
```

dot:

```
nucleus:      |  |nuc         |      |
phone:        |d |     |a  |       |t |
aspiration:   |0                      |
voicing:      |0 |60           |0     |
transition:   |  |trans|   |trans  |  |
millisecond:  |  |50   |55 |35 |15 |  |

                      140
```

In reality, we find that the vowel of *tot* is actually slightly shorter than 105 ms (as shown in Example (2) above), and consequently, the final total nucleus duration in *tot* is

---

[4] We are still investigating how best to treat transitions between vowels of words like *neon*, in which the transition between the [i] and [a] could be grouped with the first or second nucleus, or a portion of the transition could be included with each.

slightly shorter than in *dot.* We attribute this difference to an independent rule that shortens all vowels after aspiration by 20 ms, whether the aspiration results from a voiceless stop or from [h] (cf. Peterson & Lehiste, 1960).

Unlike the lengthening before tautosyllabic voiced consonants, the shortening after aspiration is local to the first phone of the nucleus, regardless of the phone or the overall structure of the nucleus. There are other rules that modify the durations of specific phones in the nucleus, without regard to the overall nucleus structure. For example, word-final vowels, liquids, and glides are lengthened.

Let us now consider in more detail how syllable nuclei lengthen before tautosyllabic voiced stops. In general, the amount of lengthening of the nucleus depends on its structure. Nuclei that have relatively short starting durations (e.g., those consisting of a single non-low vowel) tend to be about 1.5 times as long as they are before a tautosyllabic voiceless stop, while long nuclei (e.g., those consisting of three phones or a tense vowel followed by [l]) tend to be about 1.3 times as long. Rather than handling these cases with different lengthening rules, we posit maximum durations for different nuclei before voiced stops, and stretch the nuclei to 1.5 times their starting duration or to their maximum duration, whichever is less. While the linguistic evidence does not help us choose between these two approaches, a maximum duration is more manageable from a programming point of view, since we can put durations measured directly from spectrograms into the rule program as the maximum durations, rather than having to compute the appropriate percentages of lengthening for the various individual cases. Maximum durations also allow for easier comparisons of duration rules across dialects.

It is important to note that a maximum duration is specific to a given context. Thus, for example, a three-phone nucleus before a tautosyllabic voiced stop stretches to a maximum duration that is less than 1.5 times its starting duration, but before a tautosyllabic voiced fricative, it stretches to more than 1.5 times its starting duration.

In general, a nucleus is stretched by lengthening the phones in it to attain the appropriate nucleus duration. Consider, for example, the nuclei of *wilt* and *willed.* Initially, both words are assigned the durations shown below for *wilt*, in accordance with the strategy discussed above (see Example (9)).

(12) *wilt:*

```
nucleus:      |  |nuc                         |  |
phone:        |w |     |I   |     |l    |     |t |
transition:   |  |trans|    |trans|     |trans|  |
millisecond:  |  |40   |-20 |60   |50   |15   |  |
                  \_____________  ___________/
                                \/
                               145
```

In *willed*, the desired nucleus duration is 210 ms, the maximum duration for phrase-medial nuclei consisting of lax vowels followed by [l] before a tautosyllabic voiced stop. This represents an increase in duration of 65 ms over the 145 ms duration of the same nucleus before a tautosyllabic voiceless stop. Since, for all intents and purposes, all of the lengthening occurs in the phones, the total duration of the phones [I] and [l] must be increased from the initial total duration of 30 ms (see the above delta) to 95 ms (30 + 65) to achieve the appropriate nucleus duration, as shown below:

(13) *willed:*

```
nucleus:      |  |nuc                      |  |
phone:        |w |     |I  |     |l  |     |d |
transition:   |  |trans|   |trans|   |trans|  |
millisecond:  |  |40   |25 |60   |70 |15   |  |
                            210
```

We have written a preliminary set of rules that distributes the total phone duration among the individual phones in a complex nucleus in a lengthening environment. Since these rules are still undergoing revision as we explore more data, they are not discussed further in this paper.[5]

So far we have seen examples of phrase-medial syllables, in which only the phones of the nucleus are lengthened. In phrase-final position, in contrast, not only the phones, but also the transitions tend to be lengthened, though transitions are not lengthened nearly as much as phones. As a first approximation, the rules currently lengthen all transitions in a phrase-final nucleus by 40%. They then lengthen the nucleus to 2.5 times its present duration (i.e., its duration after any other lengthening rules have applied, such as the rule for voiced stops), though since nuclei with particular structures are given maximum durations, the effective lengthening is sometimes less. While in other contexts, only a few maximum durations need to be posited, and the cases are quite general, phrase-finally there are several very specific nucleus structures that seem to require different maximum durations. For example, a phrase-final nucleus consisting of the vowel [I] followed by a voiceless stop has a maximum duration of 120 ms. That is, although it is very short to begin with, it nevertheless is not lengthened beyond a certain

---

[5] In the model described in Hertz (1991), the phones of complex nuclei (which were never assigned negative durations) were lengthened before voiced stops by different percentages, depending on their linguistic features, without regard to the overall nucleus duration. The result of this strategy was that the final durations of nuclei such as those of *bead* and *weed* were often quite different. In our current model, the nuclei of such words would have the same duration, which much more closely accords with the facts. The average duration of the nuclei may actually differ by about 5 ms, but certainly not by as much as our earlier model would have predicted. As we put more GA data in the database, we will refine our present model to correctly predict any systematic small differences that we may have missed from our more informal analyses of spectrograms.

point. There are several other cases that are similar, often involving high vowels and vowels before voiceless stops.

While we have focussed on rules for stressed monosyllables, there are a number of rules that adjust nucleus durations in other contexts. Among these are rules that shorten the nuclei of stressed monosyllables that are not word-final, and rules that shorten unstressed syllables and function words.

4. Final Remarks

This report has presented the basic premises of a new model of duration for General American English. This model is centered around the notion of an acoustic nucleus organized into phones and transitions. The model is leading to much simpler synthesis rules than the more conventional model that we employed in our previous synthesis rules for English (Hertz, 1982, 1988b). It is also serving as a fruitful basis for our cross-dialect and cross-language investigations, in which we are trying to separate universal rules from language-specific ones (Hertz, 1990a).

The development of the specific duration rules for General American English is still in progress, and future analysis and synthesis cycles are likely to lead to new or revised rules, so specific rules and algorithms discussed in this paper should be regarded as tentative. We do not, however, expect to make changes in the basic premises of the model.

Acknowledgments

Many thanks to Marie Huffman, Abby Cohn, Allard Jongman, and David Lewis for helpful comments on drafts of this paper. The development of the duration model has been supported in part by (1) the U.S. Dept. of Education, under Contracts RS89071002 and RS90087003 to Eloquent Technology, Inc., the author's company; (2) the NIH under Grant 1 R43 DC00758-01 to Eloquent Technology, Inc; and (3) the New York State Science and Technology Foundation, under grant RDG 89174 to Cornell University. The content of this publication does not necessarily reflect the views or policies of these agencies.

References

Charif, R., Hertz, S. R., and Weber, T. (1992). *The Delta User's Manual*, Ithaca, New York: Eloquent Technology, Inc.

Chen, M. (1970) Vowel Length variation as a function of the voicing of the consonant environment, *Phonetica* **22**, 129-159, 1970.

Hertz, S. R. (1982) From text to speech with SRS, *Journal of the Acoustical Society of America*, **72**, 1155-1170.

Hertz, S. R. (1988a) Delta: flexible solutions to tough problems in speech synthesis by rule, *The Official Proceedings of Speech Tech 88*, New York: Media Dimensions Inc.

Hertz, S. R. (1988b) *SRS Phoneme-to-Speech Rules for English*, Ithaca, New York: Eloquent Technology, Inc.

Hertz, S. R. (1990a) A modular approach to multi-language and multi-dialect speech synthesis, Proceedings of the ESCA Conference on Speech Synthesis.

Hertz, S. R. (1990b) The Delta programming language: an integrated approach to non-linear phonology, phonetics, and speech synthesis. In *Papers in Laboratory Phonology I: Between the Grammar and the Physics of Speech* (J. Kingston & M. Beckman, editors), Cambridge: Cambridge University Press.

Hertz, S. R. (1991) Streams, phones and transitions: toward a new phonological and phonetic model of formant timing, *Journal of Phonetics* **19**, 91-109.

Kewly-Port, D. (1982) Measurement of formant transitions in naturally produced stop consonant-vowel syllables, *Journal of the Acoustical Society of America*, **72**, 379-389.

Klatt, D. H. (1976) Linguistic uses of segmental duration in English: acoustic and perceptual evidence. *Journal of the Acoustical Society of America*, **59**(5), 1208-1221.

Klatt, D. H. (1979) Synthesis by rule of segmental durations in English sentences. In *Frontiers of Speech Communication Research* (B. Lindblom & S. Ohman, editors), New York: Academic Press.

Peterson, G. and Lehiste, I. (1960) Duration of syllable nuclei in English, *Journal of the Acoustical Society of America*, **32**, 693-703.

Phonetic priming effects in auditory word recognition

Joan A. Sereno and Allard Jongman

1. Introduction

The present research is concerned with the representation and access of lexical form during spoken language comprehension. We address the nature of the input representation, focusing on how lower-level acoustic-phonetic information is used to access lexical representations in spoken word recognition.

In general, the representation of a given lexical form may be a specific acoustic-phonetic realization of the input, or it may be a more abstract representation which is independent of various sources of acoustic variation such as the sex of the speaker or the rate at which the utterance was produced. Consider, for example, the phoneme /t/ in the words *top* and *stop*. Both words obviously contain a /t/ segment. Nevertheless, /t/ in *top* and /t/ in *stop* are quite distinct acoustically. The acoustic properties or features that make up these segments are different. The /t/ in 'top' is an unvoiced aspirated alveolar stop consonant and is transcribed in a narrow phonetic transcription as [t^h], whereas a /t/ which is not syllable-initial, as in 'stop', is an unvoiced unaspirated alveolar stop consonant and is transcribed as [t].

The question arises whether listeners in the process of spoken language comprehension make use of the difference in acoustic information between these two /t/ segments. Two opposing viewpoints are possible. Either the listener engages all available acoustic-phonetic information to segment words, or else the listener resorts to a segmental representation and uses an abstract representation in accessing the lexicon. When listeners hear an aspirated [t^h], is it the case that they retain the information that this segment is aspirated and that it therefore could only have occurred in syllable-initial position, or do listeners simply categorize this segment as a /t/ phoneme?

The present set of experiments addresses this issue using a priming paradigm. In priming experiments, the relationship between prime and target can be systematically varied. Ordinarily, there is a relationship of meaning (e.g., Meyer and Schvaneveldt, 1971), in which recognition of a target word preceded by an associatively related prime item is facilitated.

Relevant for the present research is the case when prime and target are related phonologically. The most extensively studied topic in this domain is rhyme priming (e.g., Meyer, Schvaneveldt, and Ruddy, 1974; Shulman, Hornak, and Sanders, 1978; Hillinger, 1980). In a visual lexical decision task, both Shulman et al. (1978) and Hillinger (1980) found that reaction time to a target item (e.g., *tribe*) is faster when preceded by a graphemically and phonologically related prime (e.g., *bribe*) compared to a control prime. This result suggests that both orthographic and phonological similarity facilitate responses to target items.

Additionally, it has been shown that this effect is not exclusively due to a purely visual effect of overlapping letters. Meyer et al. (1974) and Shulman et al. (1978) contrasted prime-target pairs that were both graphemically and phonologically related (e.g., *bribe-tribe*) to pairs that were only graphemically related (e.g., *freak-break*). Comparing these two conditions to their respective controls, they showed that reaction times to stimuli sharing only graphemic information were slower than those to stimuli sharing both graphemic and phonological information, suggesting that phonological as well as graphemic information is effective in priming.

In an attempt to separate the individual contributions of graphemic and phonological information to the priming effect, Hillinger (1980) used prime-target pairs that shared phonological but not graphemic information (e.g., *eight-mate*). For such phonologically similar but graphemically dissimilar pairs, Hillinger obtained as much facilitation as when prime and target shared both phonological and graphemic information, suggesting that phonological rather than graphemic overlap primarily contributes to the rhyme priming effect. Recently, however, Martin and Jensen (1988) failed to replicate these results and concluded that visual lexical decisions are made without phonological mediation.

Given such conflicting evidence for the role of graphemic and phonological information in visual rhyme priming, it is surprising that so few studies have investigated rhyme priming effects within an auditory paradigm. In a rare example, Burton (1989) obtained significant rhyme priming effects, relative to controls, for both word prime-target pairs (*bat-cat*) and nonword prime-target pairs (*gat-cat*) using an auditory lexical decision task. In these experiments, rhymes were kept orthographically similar since Seidenberg and Tanenhaus (1979) had previously found that orthographic differences did affect rhyme detection when stimuli were presented auditorily.

In general, rhyming can be defined in terms of a preponderant overlap of phonological information between prime and target, as in the rhyming pair *bribe-tribe*. It also involves coincidence at the *end* of words. This is intriguing in light of recent theories emphasizing the importance of information at word *onset*. A prominent theory of spoken word recognition is the Cohort Theory (Marslen-Wilson and Welsh, 1978; Marslen-

Wilson and Tyler, 1980; Marslen-Wilson, 1987). In the Cohort model, there is a continuous mapping of sensory input onto representations of lexical form, starting at word onset. A cohort of all word candidates beginning with the same acoustic-phonetic sequence is activated. Members of this cohort are progressively deactivated on the basis of incoming lower-level (acoustic-phonetic) and higher-level (semantic and syntactic) information until only the word to be recognized remains activated. According to the Cohort model, then, words are apprehended in a serial fashion beginning at their onset.

To test the claims of the Cohort model, Slowiaczek, Nusbaum, and Pisoni (1987) examined 'phonological priming' effects by varying the number of phonemes that prime and target shared. In contrast to previous studies investigating rhyme priming, the overlap in phonological information occurred starting from word *onset*. Slowiaczek et al. (1987) used an identification task in which subjects were to identify target words embedded in noise at various signal-to-noise ratios. These target words were preceded by prime words in the clear which shared zero, one, two, three, or all phonemes with the target, beginning at word onset. For example, for a given target (*dread*), the prime had either: 1) no phonemes in common (*scream*), 2) one phoneme in common (*dove*), 3) two phonemes in common (*drill*), 4) three phonemes in common (*dress*), or 5) all phonemes in common (*dread*). According to the Cohort model, one might expect an increase in facilitation to target items as the phonological overlap between prime and target increased. Overall, phonological priming did track phonological overlap between prime and target. However, no significant differences were obtained between unrelated primes and primes with one phoneme overlap, nor between primes with two and three phonemes in common. Thus, these results only tepidly support Cohort theory since priming was not observed when prime and target had one phoneme in common, and priming did not increase from two to three shared phonemes.

In a follow-up study using similar experimental conditions, Slowiaczek and Pisoni (1986)[1] found that facilitation occurred only in the identity condition in which prime and target shared all phonemes. In all other conditions, no facilitation was observed relative to the baseline condition (no phonemes in common). Instead of an identification task, Slowiaczek and Pisoni (1986) used an auditory lexical decision task. Based on their finding that partial phonological overlap did not facilitate responses to target items, Slowiaczek and Pisoni (1986) concluded that the Cohort model was not supported. Similar results from both lexical decision and shadowing experiments have also recently been reported (Radeau, Morais, and Dewier, 1989).

However, as Slowiaczek and Pisoni (1986) suggest, the lack of phonological priming (with the exception of the identity condition) might be due to the use of whole-word primes which created an inhibition effect, thereby obscuring priming that may have been present. Slowiaczek and Pisoni (1986) also invoked the notion of task demands to resolve the difference in results between their study and that of Slowiaczek et al. (1987). In order to recognize a word in the perceptual identification task, subjects were compelled to use the phonological information in the signal. Furthermore, the use of white noise in the perceptual identification task degraded stimulus information such that subjects attended inordinately to phonological cues. In the lexical decision task, this phonological information may already have been replaced by a more abstract lexical representation. Slowiaczek and Pisoni (1986) conclude, for these reasons, that the perceptual identification task may be more sensitive to phonological information than the lexical decision task.

In order to investigate partial phonological priming, we devised a technique which we will call 'phonetic' priming. Phonetic priming employs a priming paradigm in which subjects are to make a lexical decision to a target item when preceded by either a neutral or related prime item. Both prime and target are presented auditorily. However, in contrast to the traditional auditory priming paradigm, the prime consists of a single phonetic segment rather than a word or nonword. The phonetic priming paradigm offers several advantages. As with previous methodologies, this paradigm allows variation in overlap (e.g., initial, medial, or final) between prime and target. Importantly, however, using a single phonetic segment strips the prime of lexical status. Priming effects will therefore not be inhibited by the use of whole-word primes, as may have been the case in Slowiaczek and Pisoni (1986). Finally, presentation of a single phonetic segment for a limited duration substantially reduces the temporal interval between prime and target, enhancing sensitivity to early processes of auditory word recognition.

The next sections describe two experiments using the phonetic priming paradigm to test whether listeners are sensitive to isolated phonemic information during word recognition. The first experiment involves vowels in medial position while the second experiment focuses on fricatives in initial position.

2. Experiment 1 - Vowel Phonetic Priming

The purpose of this experiment was to determine whether a phonetic segment prime facilitates the response to a stimulus item containing that segment. In this experiment, an isolated vowel segment [e,a,o] preceded target items containing that vowel sound in medial position. Both word and nonword target items were employed and these stimuli were presented to subjects in an auditory lexical decision task.

2.1 Method

2.1.1 Subjects. Thirty university students from the subject pool at the Max Planck Institute for Psycholinguistics were paid to participate in the experiment. All were native speakers of Dutch and reported no history of speech or hearing disorders.

2.1.2 Stimuli. Seventy-two stimuli (36 words and 36 nonwords) were used in this experiment. The 36 word stimuli were selected from the Dutch written word frequency norms collected by Uit den Boogaart (1975). All words were monosyllabic nouns, with 12 words containing the stressed vowel [e], 12 containing the stressed vowel [a], and 12 containing the stressed vowel [o].

The vowels [e,a,o] were chosen since they are monophthongal and are phonemically long (Cohen, Ebeling, Fokkema, and Van Holk, 1972; Trommelen and Zonneveld, 1979). In Dutch, only long vowels occur in syllable-final position while short vowels occur only in closed syllables. It is, therefore, possible for long vowels to occur in isolation while short vowels cannot. For this reason, only long vowels were used as primes in this experiment.

The vowels [e,a,o] can be categorized on the basis of features that make up these segments (Chomsky and Halle, 1968). Dutch vowels are voiced and they are typically defined in terms of the position of the articulators (the tongue and lips), with the dimensions height, backness, and roundness being most relevant. In this way, [e] can be described as [+high], [-back], [-round]; [a] as [-high], [+back], [-round]; and [o] as [+high], [+back], [+round].

Words in all vowel subgroups were matched for frequency (Uit den Boogaart, 1975). The mean frequency of occurrence per million for words containing [e], [a], [o] was 44, 37, and 47, respectively.

In addition, a set of 36 monosyllabic nonwords was constructed, one third ofwhich contained the stressed vowel [e], one third [a], and one third [o]. All nonwords obeyed the phonotactic constraints of Dutch (see Bakker, 1971).

All word and nonword stimuli were consonant-initial, with a (C)CVC(C) structure, and they were matched for mean number of phonemes.

All subgroups were of comparable duration. For the word stimuli, durations for words containing [e], [a], and [o] were 477 ms, 476 ms, and 479 ms, respectively, while for the nonwords mean durations were 484 ms, 484 ms, and 485 ms, respectively.

Durations of the isolated prime vowels [e], [a], and [o] were 219 ms, 217 ms, and 220 ms, respectively.

2.1.3 Stimulus Preparation. All stimuli were recorded by a male speaker on a Revox B77 MK II tape recorder in a sound-proof booth (Philips amplisilence) using a Sennheiser MD211N microphone. The words and nonwords were read in a list and the prime segments were produced in isolation. The stimuli were digitized on a VAX750 computer at a sampling rate of 20 kHz with a 10 kHz low-pass filter setting. The stimuli were then excised using both auditory and visual criteria.

Three test conditions were constructed so that every subject heard all target items but no subject heard a target item more than once. Thus, in each condition there were 12 target words and 12 target nonwords each containing the vowel [e], [a], and [o]. All target words and nonwords were preceded by one of the priming vowel segments [e,a,o]. One third of these target items was preceded by a matching vowel prime and two thirds by a non-matching vowel prime.

Three experimental tapes (one for each condition) were prepared. On the second channel of each tape, a timing pulse was set concurrently with the acoustic onset of the target stimuli. This pulse served to start the clock of a MicroMax computer which stored responses and reaction times.

The stimuli were presented at a fixed rate with a prime stimulus followed by a target item. There was a 200 ms ISI between offset of the prime and onset of the target. Reaction times were measured from the onset of the target until a key press was made. Following the offset of the target, there was a 3 second silent interval until the next trial started. This sequence was repeated for each stimulus item.

2.1.4 Procedure. All subjects were tested individually in an auditory lexical decision task. The test tapes were played to subjects on a Revox B77 tape recorder using Sennheiser HD424 headphones. Subjects were told that on each trial they would hear a vowel sound (either [e], [a], or [o]) followed by a target stimulus, and that they were to identify the target stimulus as either a word or a nonword. Subjects were instructed to respond as quickly and accurately as possible to each target item. All responses were made by pressing one of two clearly marked buttons on a response box placed in front of the subject. Each trial was completed when subjects used the index finger of their preferred hand to press one of two equidistantly-placed response buttons. Following instructions, subjects were given a set of 10 practice items to introduce them to the procedure. These practice items were not used in the experiment. The entire experiment lasted approximately 30 minutes.

3. Results

Mean lexical decision latencies and error rates for the different experimental conditions of Experiment 1 are given in Table 1. The data in all experiments were submitted to an analysis of variance (ANOVA) with both subjects (F1) and items (F2) as random variables. All means presented are taken from the subject analyses. There was a total of 72 errors in this experiment, representing 3.3% of all responses. All trials in which errors occurred were excluded from the reaction time analyses.

A three-way ANOVA (Lexical Status X Target Vowel X Prime Vowel) was conducted. Three effects were significant only in the subject analysis: a main effect for Lexical Status [F1(1,29)=8.20, MSe=8034, p<.008; F2(1,66)=1.18, MSe=20982, p>.25]; a main effect of Target Vowel [F1(2,58)=15.52, MSe=3465, p<.001; F2(2,66)=1.25, MSe=20982, p>.25]; and a Lexical Status X Target Vowel interaction [F1(2,58)=14.42, MSe=4216, p<.001; F2(2,66)=1.17, MSe=20982, p>.30]. However, none of these effects was significant in the item analysis and therefore they are not generalizable. There was also a significant Lexical Status X Target Vowel X Prime Vowel interaction in both subject and item analyses [F1(4,116)=3.68, MSe=4989, p<.007; F2(4,132)=3.15, MSe=1720, p<.02]. Mean reaction times for word and nonword stimuli are given in Table 1 for each target vowel preceded by each prime vowel. In general, identical vowels in prime and target words facilitated reaction time relative to non-identical controls while a contrasting pattern occurred for identical and non-identical vowels in nonwords.

Separate analyses were conducted on the words and nonwords. For the words, a two-way ANOVA (Target Vowel X Prime Vowel) revealed a main effect of Target Vowel only in the subject analysis [F1(2,58)=24.52, MSe=3443, p<.001; F2(2,33)=1.96, MSe=18490, p>.15]. There was a significant interaction between Target Vowel and Prime Vowel [F1(4,116)=3.70, MSe=6562, p<.007; F2(4,66)=3.99, MSe=1839, p<.006]. A posthoc Scheffe comparison revealed that reaction times to word stimuli (840 ms) preceded by a matching prime vowel were significantly faster than to word stimuli (870 ms) preceded by a non-matching prime vowel (p<.05).

For the nonwords, a two-way ANOVA (Target Vowel X Prime Vowel) revealed a main effect of Target Vowel only in the subject analysis [F1(2,58)=7.10, MSe=4237, p<.002; F2(2,33)=.62, MSe=23473, p>.50]. There were no other significant main

WORDS

target vowel	prime [e]	prime [a]	prime [o]
[e]	854 (5.0)	861 (3.3)	875 (5.0)
[a]	917 (2.5)	860 (5.8)	889 (5.8)
[o]	819 (7.5)	857 (6.7)	807 (7.5)

NONWORDS

target vowel	prime [e]	prime [a]	prime [o]
[e]	856 (0.8)	860 (0.0)	867 (4.2)
[a]	877 (2.5)	898 (0.8)	905 (0.0)
[o]	890 (0.8)	886 (0.0)	900 (1.7)

Table 1. Mean response latencies (ms) and error rates (percent) for word targets (top) and nonword targets (bottom) obtained in the vowel phonetic priming experiment (Experiment 1).

effects or interactions.

Further analyses were conducted in which the vowels were analyzed in terms of their distinctive features ([high], [back], and [round]). Both prime and target vowels were classified in this fashion. In these analyses, means were first calculated by collapsing over the relevant feature. Three separate ANOVAs, one for each feature contrast ([high], [back], and [round]), were then conducted. For example, for the feature [round], the vowels [e] and [a] (both unrounded vowels) were contrasted to the rounded vowel [o]. It should be noted that these are not optimal contrasts due to language constraints of Dutch. The two contrasting groups differ not only in terms of the vowel feature under consideration ([round]) but also in terms of the other defining vowel features ([high], [back]). None of the feature analyses revealed any significant differences. Reaction times to target stimuli preceded by a prime vowel that matched on one of these features were not significantly different from reaction times to target stimuli that did not match in terms of that same feature.

An analysis of the error data was also conducted. A three-way ANOVA (Lexical Status X Target Vowel X Prime Vowel) revealed a main effect of Lexical Status ([F1(1, 29)=28.62, MSe=.14, p<.001; F2(1, 66)=4.67, MSe=2.10, p<.03]). There were significantly more errors for word targets (59 errors) compared to nonword targets (13 errors). There were no other significant main effects or interactions.

4. Discussion

Experiment 1 showed that prior presentation of a vowel segment in isolation influenced the processing of a following target stimulus containing that vowel. This significant interaction between the prime vowel and the target vowel depended on the lexical status of the target. Reaction times to word targets with matching prime and target vowels were facilitated relative to word targets with non-matching prime and target vowels. This facilitation for target words with matching prime and target vowels depended on the exact identity of the vowel segment in the prime and target stimuli occurring in target-medial position. For the nonword stimuli, response times to nonword stimuli with matching prime and target vowels were not significantly different from responses to nonword stimuli with non-matching prime and target vowels.

Having obtained vowel phonetic priming, it was of interest to determine whether this effect could be generalized to other positions in the target stimulus. A second experiment investigating priming of segments in word-initial position was therefore conducted.

5. Experiment 2 - Fricative Phonetic Priming

The purpose of this experiment was to establish whether a phonetic segment prime facilitates the response to a stimulus item that contains that segment. In this experiment, a fricative segment [f,v,s,z] preceded target items containing that fricative sound in initial position. Both word and nonword target items were employed and these stimuli were presented to subjects in an auditory lexical decision task.

5.1 Method

5.1.1 Subjects. Forty university students from the subject pool at the Max Planck Institute for Psycholinguistics were paid to participate in the experiment. All were native speakers of Dutch and reported no history of speech or hearing disorders. None of these subjects had participated in the previous experiment.

5.1.2 Stimuli. One hundred twenty-eight stimulus items (64 words and 64 nonwords) were used in this experiment. The 64 word stimuli were selected from the Dutch written frequency norms of Uit den Boogaart (1975). One half of the words were monosyllabic nouns and the remaining half were bisyllabic nouns. All bisyllabic words had first-syllable stress. The word stimuli were equally divided into four subgroups in terms of their onset consonant, with 16 words each having an initial [f], [v], [s], or [z] fricative segment. None of the stimuli began with a consonant cluster and no stimulus contained any of the fricative segments [f,v,s,z] in non-initial position.

The consonants [f,v,s,z] can also be categorized on the basis of features that make up these segments (Chomsky and Halle, 1968). The segments [f,v] and [s,z] can be distinguished in terms of their place of articulation. [f] and [v] are characterized by the feature [-coronal] and are classified as labiodental fricatives whereas [s] and [z] share the feature [+coronal] and are classified as alveolar fricatives. In a similar fashion, the segments [f,s] and [v,z] are contrasted in terms of voicing. [f] and [s] are characterized as [-voice] and are classified as unvoiced fricatives whereas [v] and [z] share the feature [+voice] and are classified as voiced fricatives.

All subgroups were matched for word frequency on the basis of the frequency norms for Dutch (Uit den Boogaart, 1975). The mean frequency of occurrence per million for words beginning with [f], [v], [s], and [z] was 19, 23, 17, and 31, respectively.

Additionally, a set of 64 nonwords was constructed. One half of the stimuli was monosyllabic and one half bisyllabic. All bisyllabic nonwords were to be pronounced with first-syllable stress. These items were also equally divided in terms of their onset consonant, with 16 nonwords each having an initial [f], [v], [s], or [z] fricative segment. All items obeyed the phonotactic constraints of Dutch (see Bakker, 1971).

Word and nonword stimuli were matched for mean number of phonemes. All subgroups were also of comparable duration. For the word stimuli, durations for words containing [f], [v], [s], and [z] were 611 ms, 556 ms, 607 ms, and 575 ms, respectively, while for the nonwords, mean durations were 619 ms, 599 ms, 628 ms, and 599 ms, respectively. Durations of the isolated prime fricative segments [f], [v], [s], and [z] were 196 ms, 195 ms, 198 ms, and 196 ms, respectively. It has previously been shown that listeners identify fricatives very accurately when presented with the frication noise in isolation (Jongman,1989).

5.1.3 Stimulus Preparation. The stimuli were recorded, digitized, and excised as described in Experiment 1. For this experiment, four test conditions were constructed so that every subject heard all target items but no subject heard a target item more than once. Thus, in each condition there were 16 target words and 16 target nonwords each beginning with an initial [f], an initial [v], an initial [s], and an initial [z] fricative segment. All target words and nonwords were preceded by one of the fricative primes [f,v,s,z]. One fourth of these target items was preceded by a matching fricative prime and three fourths by a non-matching fricative prime.

Four experimental tapes (one for each condition) were then prepared. All stimulus presentation and response procedures were the same as those described in Experiment 1.

5.1.4 Procedure. All subjects were tested individually in an auditory lexical decision task. Subjects were told that on each trial they would hear an isolated fricative sound (either [f], [v], [s], or [z]) followed by a target, and that they were to identify each target as either a word or a nonword. All other experimental procedures were identical to those in Experiment 1. The entire experiment lasted approximately 30 minutes.

6. Results

Mean lexical decision latencies and error rates for the different conditions of Experiment 2 are given in Table 2. The total number of errors was 219, representing 4.3% of all responses. All trials in which errors occurred were excluded from the reaction time analyses.

A three-way ANOVA (Lexical Status X Target Fricative X Prime Fricative) was conducted. There was a significant main effect of Lexical Status [F1(1,39)=28.72, MSe=32939, p<.001; F2(1,120)=25.16, MSe=19629, p<.001]. Responses to words (969 ms) were significantly faster than those to nonwords (1023 ms). A main effect was also obtained for Target Fricative [F1(3,117)=36.63, MSe=7991, p<.001; F2(3,120)=6.10, MSe=19629, p<.001]. A Newman-Keuls posthoc test showed that targets beginning with the fricative [f] (1034 ms) were significantly slower than targets beginning with the fricatives [v] (968 ms) (p<.01), [s] (1006 ms) (p<.05), and [z] (976 ms) (p<.01) and that targets beginning with the fricative [s] (1006 ms) were significantly slower than targets beginning with the fricatives [z] (976 ms) (p<.05) and [v] (968 ms) (p<.05). There were no other significant main effects or interactions.

Additional analyses of the data were conducted examining the contribution of phonetic features as contrasted to phonetic segments. The phonetic feature analysis of the data involved the following factors: Lexical Status (target is either a word or a nonword); Place (fricatives are categorized in terms of place of articulation, with [f,v] as labiodental and [s,z] as alveolar; and Voicing ([f,s] as voiceless and [v,z] as voiced)(e.g., Ladefoged, 1975). The latter two factors (Place and Voicing) were appropriate for classification of both primes and targets.

A five-way ANOVA (Lexical Status X Target Place X Target Voicing X Prime Place X Prime Voicing) yielded the following results. There was a main effect for Lexical Status [F1(1,39)=28.72, MSe=32939, p<.001; F2(1,120)=25.16, MSe=19629, p<.001]. As stated earlier, responses to words were significantly faster than to nonwords. A main effect also obtained for Target Voicing [F1(1,39)=65.22, MSe=11369, p<.001; F2(1,120)=15.78, MSe=19629, p<.001]. Responses to targets beginning with the voiced fricatives [v,z] (972 ms) were significantly faster than responses to targets beginning with the voiceless fricatives [f,s] (1020 ms). The only other significant finding was a Lexical Status X Target Place X Prime Place interaction [F1(1,39)=13.27, MSe=6091, p<.001; F2(1,120)=8.55, MSe=2424, p<.004]. As shown in Table 3, responses to word targets were facilitated when prime and target fricative shared the same place of articulation (962 ms) compared to a different place of articulation (976 ms), while an opposite pattern obtained for nonwords. Responses to nonword targets were inhibited when prime and target fricative matched in terms of place of articulation (1033 ms) relative to when they mismatched (1014 ms). There were no other significant main effects or interactions.

Separate analyses were then conducted on the word and nonword data. For the word stimuli, a two-way ANOVA (Target Fricative X Prime Fricative) was conducted. There was a significant main effect of Target Fricative [F1(3,117)=26.66, MSe=6659,

WORDS

target onset	prime [f]	 [v]	 [s]	 [z]
[f]	1019 (5.6)	1001 (6.9)	1021 (8.1)	1016 (6.3)
[v]	949 (6.3)	931 (9.4)	967 (10.6)	958 (7.5)
[s]	973 (5.0)	974 (4.4)	972 (4.4)	966 (2.5)
[z]	940 (4.4)	954 (3.8)	939 (3.8)	919 (5.0)

NONWORDS

target onset	prime [f]	 [v]	 [s]	 [z]
[f]	1067 (4.4)	1056 (0.6)	1038 (0.6)	1054 (2.5)
[v]	1003 (1.9)	982 (2.5)	973 (0.0)	981 (1.3)
[s]	1022 (5.6)	1042 (6.3)	1052 (3.8)	1046 (3.8)
[z]	984 (2.5)	1015 (2.5)	1020 (1.9)	1034 (3.1)

Table 2. Mean response latencies (ms) and error rates (percent) for word targets (top) and nonword targets (bottom) obtained in the fricative phonetic priming experiment (Experiment 2).

WORDS

target onset	prime labiodental [f,v]	 alveolar [s,z]
labiodental [f,v]	975 (7.0)	991 (8.1)
alveolar [s,z]	960 (4.4)	949 (3.9)

NONWORDS

target onset	prime labiodental [f,v]	 alveolar [s,z]
labiodental [f,v]	1027 (2.3)	1012 (1.1)
alveolar [s,z]	1016 (4.2)	1038 (3.1)

Table 3. Mean response latencies (ms) and error rates (percent) for word targets (top) and nonword targets (bottom) obtained in the fricative phonetic priming experiment (Experiment 2) for the feature analysis.

p<.001; F2(3,60)=3.10, MSe=23878, p<.033]. A Newman-Keuls posthoc test revealed that word targets beginning with the fricative [f] (1014 ms) were significantly slower than word targets beginning with the fricatives [v] (951 ms) (p<.01), [s] (971 ms) (p<.05), and [z] (938 ms) (p<.01).

The word data were also analyzed in terms of phonetic features. A four-way ANOVA (Prime Place X Prime Voicing X Target Place X Target Voicing) yielded the following results. There was a main effect of Target Voicing [F1(1,39)=64.04, MSe=5777, p<.001; F2(1,60)=6.52, MSe=23878, p<.02]. Responses to word targets beginning with the voiced fricatives [v,z] (945 ms) were significantly faster than those to word targets beginning with the voiceless fricatives [f,s] (993 ms). There was a significant interaction of Target Place X Prime Place [F1(1,39)=4.10, MSe=6625, p<.05; F2(1,60)=3.69, MSe=2775, p<.05]. For words, reaction times to targets beginning with the labiodental fricatives [f,v] were facilitated when preceded by primes beginning with the labiodental fricatives [f,v] (975 ms) as compared to primes beginning with the alveolar fricatives [s,z] (991 ms), and reaction times to targets beginning with the alveolar fricatives [s,z] were facilitated when preceded by primes beginning with the alveolar fricatives [s,z] (949 ms) as compared to primes beginning with the labiodental fricatives [f,v] (960 ms).

For the nonwords, a two-way ANOVA (Target Fricative X Prime Fricative) revealed a main effect of Target Fricative [F1(3,117)=19.16, MSe=7830, p<.001; F2(3,60)=3.79, MSe=15381, p<.02]. A Newman-Keuls posthoc test revealed that nonword targets beginning with the fricative [v] (985 ms) were significantly faster than nonword targets beginning with the fricatives [f] (1054 ms) (p<.01) and [s] (1041 ms) (p<.05).

For the feature analysis of the nonword data, a four-way ANOVA (Target Place X Target Voicing X Prime Place X Prime Voicing) was conducted. A main effect was found for Target Voicing [F1(1,39)=33.14, MSe=11211, p<.001; F2(1,60)=10.02, MSe=15381, p<.002]. Reaction times to nonword targets with voiced fricatives [v,z] (999 ms) were significantly faster than to nonword targets with voiceless fricatives [f,s] (1048 ms). There was also a significant interaction of Target Place X Prime Place [F1(1,39)=6.29, MSe=8947, p<.02; F2(1,60)=5.06, MSe=2073, p<.03]. For nonwords, reaction times to targets beginning with the labiodental fricatives [f,v] were inhibited when preceded by primes beginning with the labiodental fricatives [f,v] (1027 ms) as compared to primes with the alveolar fricatives [s,z] (1012 ms), and reaction times to targets beginning with the alveolar fricatives [s,z] were inhibited when preceded by primes beginning with the alveolar fricatives [s,z] (1038 ms) as compared to primes with labiodental fricatives (1016 ms).

An analysis of the error data was also conducted. A three-way ANOVA (Lexical Status X Target Fricative X Prime Fricative) revealed a main effect of Lexical Status ([F1(1,39)=20.40, MSe=.28, p<.001; F2(1,120)=5.58, MSe=2.30, p<.02]). There were significantly more errors for word targets (150 errors) compared to nonword targets (69 errors). There were no other significant main effects or interactions.

The error data were also analyzed in terms of phonetic features. A five-way ANOVA (Lexical Status X Target Place X Target Voicing X Prime Place X Prime Voicing) revealed a significant main effect for Lexical Status, identical to the segment analysis reported above. There were no other significant main effects or interactions.

7. Discussion

Experiment 2 produced the expected effect of lexical status wherein words were recognized faster than nonwords. In addition, reaction times to both words and nonwords beginning with voiced fricatives [v,z] were faster compared to words and nonwords beginning with voiceless fricatives [f,s]. While this latter effect might seem puzzling, a simple explanation may be found in terms of stimulus duration. The mean duration of target items beginning with [v,z] (582 ms) was significantly shorter than that of targets beginning with [f,s] (617 ms) [t=2.37, df=126, p<.02]. Since there was a significant correlation between stimulus duration and reaction time (Pearson correlation, r=.5043, p<.001), the difference in reaction time found between voiced and voiceless target items may be due to the 35 ms difference in stimulus duration.

More importantly, however, the fricative priming experiment showed that prior presentation of a fricative segment in isolation influenced the processing of the following target stimulus. For word targets, reaction times to target words with matching prime and target fricative segments were *facilitated* compared to word targets with non-matching prime and target fricatives. Conversely, reaction times to target nonwords with matching prime and target fricative segments were *inhibited* compared to nonword targets with non-matching prime and target fricatives.

However, this facilitation for target words and inhibition for nonwords with matching prime and target fricatives did not depend on the exact identity of the fricative segment in the prime and target stimuli. Instead, it was based on a match in terms of place of articulation, regardless of voicing. When the initial consonant of prime and target shared the same place of articulation, responses to target words were facilitated and responses to target nonwords were inhibited relative to a non-matching baseline.

8. General Discussion

The present study describes two auditory lexical decision experiments using a phonetic priming paradigm. Both experiments showed that prior presentation of an isolated phonetic segment affected responses to targets containing an identical or similar segment. Facilitation was observed for word targets with matching prime and target phonetic segments, and an opposite effect obtained for nonword targets with matching prime and target phonetic segments. The results for the word stimuli will first be considered, followed by the nonword data.

In Experiment 1, a vowel segment [e,a,o] facilitated responses to target words containing an identical vowel segment as compared to responses to target words containing a non-identical vowel segment. This facilitation was based on the exact identity between prime and target vowel. This result clearly shows the effectiveness of phonetic priming as a novel procedure to test phonetic and phonological aspects of word recognition processes. First, primes consisted of brief, vowel segments, approximately 200 ms in duration, thereby allowing a substantially reduced temporal interval between prime and target. By restricting the stimulus onset asynchrony (SOA) in such a manner, a relatively early stage of word recognition could be investigated (see, e.g., Sereno, 1991). But the more important promise of such a methodology concerns the nature of the prime. Since the prime is a phonetic segment, it is not encumbered with lexical status. In fact, phonetic priming was observed even in the 'less-phonetically-sensitive' lexical decision task. The present results suggest that the difference in results between Slowiaczek et al. (1987) and Slowiaczek and Pisoni (1986) may not be due exclusively to task differences. In contrast to Slowiaczek and Pisoni (1986), the present results indicate that the lexical decision task can be used to tap into a level at which phonological information has not yet been replaced by a more abstract lexical representation. Since, in the present study, phonetic priming was observed using isolated phonetic segments, the lack of partial phonological priming as reported by Slowiaczek and Pisoni (1986) may have been primarily caused by inhibition due to the use of whole word or nonword primes, as the authors themselves have suggested.

In Experiment 2, fricative primes [f,v,s,z] speeded reaction times to word targets containing a similar fricative in initial position. This priming effect, however, was based on a partial overlap of features rather than an exact identity between prime segments and target stimuli. This featural priming appeared only when prime and target fricative shared the same place of articulation and did not depend on a match in terms of voicing. That is, labiodental fricatives primed target words beginning with labiodental fricative segments and alveolar fricatives primed target words beginning with alveolar fricative segments.

These results suggest that subphonemic or featural rather than phonemic or segmental information may play the more important role in lexical access. This finding is supported by the recent research of Marslen-Wilson and Warren (1990) and Lahiri and Marslen-Wilson (1991). Marslen-Wilson and Warren (1990), for example, contrasted a 'segmental' and a 'featural' hypothesis. The segmental hypothesis assumes that phonetic cues are integrated pre-lexically. Lexical access is then mediated by resulting segmental labels. In the featural hypothesis, phonetic cues are integrated at the lexical level. That is, featural information available in the speech signal is directly projected onto the lexicon. Using a lexical decision task in which subjects responded to cross-spliced stimuli with conflicting phonetic cues to place of articulation, Marslen-Wilson and Warren (1990) obtained evidence in favor of the featural hypothesis. They concluded that features are extracted at a pre-lexical level and mapped directly onto lexical representations. The present fricative priming results also suggest that priming is based on the extraction of featural information.

Some recent speech production experiments (Meyer and Gordon, 1985) provide data consistent with the hypothesis that phonetic features are involved in the mechanisms that underlie speech processing. In these experiments, subjects either produced two CV syllables (e.g., [ba], [da]) in sequential order (the primary response) or they produced these two CV syllables in the reverse order (the secondary response). Meyer and Gordon (1985) found longer latencies and more errors when the consonants of the secondary response pair were related (that is, when they shared either a place of articulation feature or a voicing feature) relative to an unrelated control condition. These results suggest that phonetic features play a significant role in the programming and execution of speech sounds. Yaniv, Meyer, Gordon, Huff, and Sevald (1990) also provide additional supportive data for a phonetic feature analysis by examining vocal responses to syllable pairs which contrasted in terms of their medial vowels.

However, Gordon and Meyer (1984) and Meyer and Gordon (1984), using a slightly different experimental procedure, found that shared voicing features influenced response latencies more than shared place of articulation features, suggesting that some features may be more salient than others (see also Miller and Nicely, 1955). Unfortunately, no concrete explanation was given for the differences among experiments. Moreover, other researchers (e.g., Peters, 1963; Shepard, 1972; MacNeilage and Ladefoged, 1976; Pisoni, 1978) have reported that place and voicing features participate equally or even that place of articulation features may be more salient than voicing features in their contribution to speech perception. It is clear that more research is needed to obtain converging evidence for the status of the features of voicing and place of articulation in the perception and production of speech.

If fricative priming is based on an overlap of features, a question that remains is why no identity priming is observed. When prime and target fricative are identical, they share information about both place and voicing. With identity priming, an increase rather than a decrease in facilitation could be expected. The lack of identity priming in Experiment 2 may be compatible with certain recent developments in phonological theory. One such development is a phonological theory of underspecification which argues that segments are underspecified in their underlying phonological representation (e.g., Archangeli, 1984). That is, instead of all featural information being specified for a particular segment, only the *marked, distinctive* information (i.e., the minimal information necessary to distinguish the segment from all other segments in the language) for that segment is listed in its underlying representation. Based on data from a wide variety of languages, underspecification theory claims that for the feature [voice], for example, only the marked value (i.e., the presence of voicing), represented as [+voice], is specified (Mester and Ito, 1989). In the underlying representation, then, both [v] and [z] are specified as [+voice] whereas [f] and [s] are unspecified and, as such, are underlyingly ambiguous with respect to voicing.

In the context of the present experiment, such an account implies that segments may not necessarily be better primes than a subset of their features. When hearing [f] or [s], for example, the listener cannot be sure about the voicing status of these segments since they are unspecified. In contrast, place of articulation always provides an unambiguous cue to the segment's identity. Since [f,v,s,z] are all underlyingly specified in terms of place of articulation but only [v,z] are underlyingly specified in terms of voicing, priming may be observed based on a match in terms of place of articulation rather than a match in both place and voicing.

The fact that identity priming rather than feature priming obtained for the vowels can also be explained within a theory of phonological underspecification. In this theory, there is an asymmetry in the feature specification of [f,v,s,z] such that all these fricatives are underlyingly specified in terms of place of articulation but all are not specified in terms of voicing. However, for the vowels [e,a,o], there can be no such asymmetry. The features [high], [back] and [round] are all underlyingly specified, since they are necessary to distinguish among these three vowels. For the vowels, then, since all features are underlyingly specified, priming is expected to be observed when all features overlap (identity priming), rather than when a subset of these features overlap.

Finally, priming was observed when the matching phonetic segment occurred in either word-medial (Experiment 1) or word-initial (Experiment 2) position. The existence of both medial and initial priming suggests that matching segments at either location facilitate lexical access processes. These findings may be compatible with results of Cutler (1976) and Cutler and Norris (1988) who suggest that a lexical search is initiated on the basis of the strong syllable in a word. Specifically, Cutler and Norris (1988) showed that word detection is delayed when the word consists of two strong syllables, but not when it consists of a strong followed by a weak syllable. They view their data as supporting a model in which only strong syllables trigger a lexical search.

In the present two experiments, the primed phonetic segments always occurred in the strong syllable of the target stimulus. In Experiment 1, only monosyllabic stimuli were used, and in Experiment 2, all bisyllabic stimuli were of the strong-weak type. Thus, both the medial vowel and the initial fricative priming results are compatible with a model in which lexical access is initiated by the strong syllable in a word. However, to explicitly evaluate Cutler's strong syllable proposal, additional experiments would have to be conducted using polysyllabic target stimuli in which the primed segment occurs in either the strong or the weak syllable.

Both the vowel and the fricative phonetic priming experiments showed a contrasting pattern of results in the word targets as compared to the nonword targets with regard to matching and non-matching phonetic features. These results can be interpreted in terms of current interactive-activation models of word recognition (Elman and McClelland, 1984; McClelland and Elman, 1986). In such models, processing units or nodes have excitatory connections between levels and inhibitory connections within levels. Depending on the input, these connections either raise or lower activation levels. Thus, phoneme nodes, for example, have excitatory connections to word level nodes, with the strength of the connections depending on whether the phoneme is present or absent in the word. Phoneme nodes also have inhibitory connections to each other.

The present phonetic priming results can be interpreted within such a framework. Upon hearing the fricative phoneme [s], for example, activation is transmitted from the appropriate [s] phoneme node to higher level word nodes containing that phoneme (e.g., sin, salt, sofa, saint, sad, etc.). Activation of these word nodes is thus raised toward their threshold, thereby increasing the probability that a word node containing the fricative phoneme [s] will exceed its threshold and be activated. Word nodes with non-matching consonants do not have raised activation levels, thus requiring more confirmatory input evidence to reach threshold. Therefore, reaction times to words with matching phonemes will be facilitated relative to reaction times to words with non-matching phonemes.

The story for nonword stimuli is initially similar to that for nonwords but, ultimately, different effects occur since no word nodes are activated. As stated earlier, presentation of the [s] phoneme will produce activation at the word level for words that contain the [s] phoneme. However, when a nonword stimulus is then presented, although, initially, there are raised activation levels for word nodes containing the [s] phoneme, these raised activation levels at the word level do not facilitate a nonword response to nonwords containing the [s] phoneme compared to nonwords which do not contain the [s] phoneme. In fact, raised activation levels for word nodes have generally been found to inhibit nonword responses. For example, both Coltheart, Davelaar, Jonasson, and Besner (1977) and Luce (1986) demonstrated than the number and nature of the words activated by a nonword stimulus influenced reaction time. Specifically, nonwords in high density and high frequency word neighborhoods were responded to more slowly than nonwords in low density and low frequency word neighborhoods. These results suggest that for nonword stimuli, activation of related words does not make responses faster or easier. Rather, activation of word neighbors at the word level appears to inhibit nonword responses. In the context of the present set of experiments, for nonword stimuli, the initial facilitation (that is, a lowered threshold) resulting from a matching phoneme between prime and target is nullified by raised activation of word neighbors which ultimately inhibits nonword responses. In fact, following the presentation of the [s] phoneme, reaction times to nonwords containing the [s] phoneme are inhibited relative to reaction times to nonwords that do not contain the [s] phoneme. Thus, over time, response latencies to word stimuli with a matching phonetic segment are facilitated as a result of activation of word neighbors while nonwords with a matching phonetic segment are inhibited.

Our results, however, suggest that this processing takes place at the featural level rather than at the phonemic level. In vowels, clusters of features equal phonemes but with the fricative consonants, due to underspecification of phonemes in terms of their features, there is not a complete mapping of feature bundles and phonemes. Similar to Stevens' (1986) model of speech perception in which features are directly mapped onto lexical representations, the present results suggest that no intermediate segmental analysis is made prior to access of the phonological representation in the lexicon. Instead, featural information is extracted from the speech signal and immediately transmitted to the lexicon, where it is mapped onto the appropriate featural phonological representation (see, for example, Lahiri and Jongman, 1990).

Interestingly, for the initial fricative phonetic priming experiment, facilitation of matching prime-target fricative segments obtained for words and an inhibitory effect of matching prime-target fricative segments was observed for nonwords. However, for the medial vowel phonetic priming experiment, facilitation of matching prime-target vowel segments was observed in the word stimuli but no effect of matching prime-target vowel segments was found for the nonword stimuli. The differences found in the behavior of the matching and non-matching nonwords for the fricative and vowel phonetic priming experiments may reflect the importance of onset information in lexical access processes (Marslen-Wilson and Welsh, 1978; Marslen-Wilson, 1984; Marslen-Wilson, 1987). When *initial* phonetic information precedes target presentation, there is significant inhibition of nonword targets with matching phonetic segments due to the substantial activation of word neighbors (that is, word candidates with matching initial phonetic segments). However, since medial phonetic information may not be as effective as onset cues in analyzing the sensory input due to the inherent temporal nature of speech, prior presentation of medial phonetic information may also be not as effective in inhibiting nonword targets with matching phonetic segments. The different behavior of the nonword stimuli in the fricative and vowel phonetic priming experiments may provide some support for the notion that in auditory word recognition, onset information has privileged status and seems to have priority in directing word recognition processes.

9. Conclusions

The vowel phonetic priming experiment and the fricative phonetic priming experiment showed that priming based on phonetic overlap can be observed in an auditory lexical decision task. The phonetic priming paradigm provides a tool to investigate how lower-level acoustic-phonetic information is used to access lexical representations in spoken word recognition. The present results suggest that phonetic features rather than segments are active in lexical access. In future experiments, we hope to test this claim in more detail by systematically varying the acoustic-phonetic overlap between prime and target segments. Such manipulations will enable us to investigate the role of fine-grained, acoustic-phonetic information (for example, coarticulatory cues or allophonic information) in lexical access processes.

10. Acknowledgements

The authors would like to thank Dave Balota, Ino Flores d'Arcais, Howard Kurtzman, and Aditi Lahiri for helpful discussions and Cisca Custers, Henning Reetz, Toine Thissen, Jeroen van de Weijer, and Annemie Witjes for technical assistance.

11. Footnotes

1. Slowiaczek and Pisoni (1986) discuss the results of Slowiaczek, Nusbaum, and Pisoni as reported in an unpublished manuscript. This manuscript was later published as Slowiaczek et al. (1987).

12. References

Archangeli, D. (1984). *Underspecification in Yawelmani phonology*. Doctoral dissertation, Cambridge: MIT.

Bakker, J.J.M. (1971). *Constant en variabel. De fonematische structuur van de Nederlandse woordvorm.* Asten: Schriks Drukkerij.

Burton, M.W. (1989). Associative, mediated, and rhyme priming. Unpublished Doctoral Dissertation, Brown University.

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

Cohen, A., Ebeling, C.L., Fokkema, K., & Van Holk, A.G.F. (1972). *Fonologie van het Nederlands en het Fries*. 's-Gravenhage: Martinus Nijhoff.

Coltheart, M., Davelaar, E., Jonasson, J.T., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and Performance VI*. Hillsdale, N.J.: Erlbaum.

Cutler, A. (1976). Phoneme-monitoring reaction time as a function of preceding intonation contour. *Perception & Psychophysics*, 20, 55-60.

Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, 14, 113-121.

Elman, J.L., & McClelland, J.L. (1984). The interactive activation model of speech perception. In N. Lass (Ed.), *Language and Speech.* New York: Academic Press.

Gordon, P.C. & Meyer, D.E. (1984). Perceptual-motor processing of phonetic features in speech. *Journal of Experimental Psychology: Human Perception and Performance*, 10, 153-178.

Hillinger, M.L. (1980). Priming effects with phonemically similar words: The encoding-bias hypothesis reconsidered. *Memory & Cognition*, 8, 115-123.

Jongman, A. (1989). Duration of frication noise required for identification of English fricatives. *Journal of the Acoustical Society of America*, 85, 1718-1725.

Lahiri, A., & Jongman, A. (1990). Intermediate level of analysis: features or segments? *Journal of Phonetics*, 18, 435-443.

Lahiri, A., & Marslen-Wilson, W.D. (1991). The mental representation of lexical form: A phonological approach to the recognition lexicon. *Cognition*, 38, 245-294.

Ladefoged. P. (1975). *A course in phonetics*. New York: Harcourt Brace Jovanovich.

Luce, P.A. (1986). Neighborhoods of words in the mental lexicon. *Research on Speech Perception, Technical Report No. 6*. Bloomington: Speech Research Laboratory.

MacNeilage, P.F., & Ladefoged. P. (1976). The production of speech and language. In E.C. Carterette & M.P. Friedman (Eds.), *Handbook of Perception*, Vol. 7. New York: Academic Press.

Marslen-Wilson, W.D. (1987). Functional parallelism in spoken word-recognition. In U. Frauenfelder & L.K. Tyler (Eds.), *Spoken word recognition*. Cambridge: MIT Press.

Marslen-Wilson, W.D., & Tyler, L.K. (1980). The temporal structure of spoken language understanding. *Cognition*, 8, 1-71.

Marslen-Wilson, W.D., & Warren, P. (1990). Lexical integration of phonetic cues. Submitted for publication.

Marslen-Wilson, W.D., & Welsh, A. (1978). Processing interactions during word-recognition in continuous speech. *Cognitive Psychology*, 10, 29-63.

Martin, R.C., & Jensen, C.R. (1988). Phonological priming in the lexical decision task: A failure to replicate. *Memory & Cognition*, 16, 505-521.

McClelland, J.L., & Elman, J.L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18, 1-86.

Mester, R.A., & Ito, J. (1989). Feature predictability and underspecification: Palatal prosody in Japanese mimetics. *Language*, 2, 258-293.

Mey, J.L. (1968). A case of assimilation in Modern Dutch. *Acta Linguistica Hafniensa*, 11, 123-145.

Meyer, D.E., & Gordon, P.C. (1984). Dependencies between rapid speech perception and production : Evidence for a shared sensory-motor voicing mechanism. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention and Performance X*. Hillsdale, N.J.: Erlbaum.

Meyer, D.E., & Gordon, P.C. (1985). Speech production : Motor programming of phonetic features. *Journal of Memory and Language*, 24, 3-26.

Meyer, D.E., & Schvaneveldt, R.W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, 90, 227-234.

Meyer, D.E., Schvaneveldt, R.W., & Ruddy, M. (1974). Functions of graphemic and phonemic codes in visual word recognition. *Memory & Cognition*, 2, 309-321.

Miller, G.A., & Nicely, P.E. (1955). An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America*, 27, 338-352.

Peters, R.W. (1963). Dimensions of perception of consonants. *Journal of the Acoustical Society of America*, 35, 1985-1989.

Pisoni, D.B. (1978). Speech perception. In W.K. Estes (Ed.), *Handbook of learning and cognitive processes*. Hillsdale: Erlbaum.

Radeau, M., Morais, J., & Dewier, A. (1989). Phonological priming in spoken word recognition: Task effects. *Memory & Cognition*, 17, 525-535.

Seidenberg, M. & Tanenhaus, M. (1979). Orthographic effects on rhyme monitoring. *Journal of Experimental Psychology: Human Learning and Memory*, 5, 546-554.

Sereno, J.A. (1991). Graphemic, associative, and syntactic priming effects at a brief stimulus onset asynchrony in lexical decision and naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17, 459-477.

Shepard, R.N. (1972). Psychological representation of speech sounds. In E.E. David & P.B. Denes (Eds.), *Human communication : A unified view*. New York: McGraw-Hill.

Shulman, H.G., Hornak, R., & Sanders, E. (1978). The effects of graphemic, phonetic, and semantic relationships on access to lexical structures. *Memory & Cognition*, 6, 115-123.

Slowiaczek, L.M., & Pisoni, D.B. (1986). Effects of phonological similarity on priming in auditory lexical decision. *Memory & Cognition*, 14, 230-237.

Slowiaczek, L.M., Nusbaum, H.C., & Pisoni, D.B. (1987). Phonological priming in auditory word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13, 64-75.

Stevens, K.N. (1986). Models of phonetic recognition II: An approach to feature-based recognition. In P. Mermelstein (Ed.), *Proceedings of the Montreal satellite symposium on speech recognition*. Montreal: ICA.

Trommelen, M., & Zonneveld, W. (1979). *Inleiding in de generatieve fonologie.* Muiderberg; Coutinho.

Uit den Boogaart, P.C. (1975). *Woordfrequenties in geschreven en gesproken Nederlands.* Utrecht: Oosthoek, Scheltema, & Holkema.

Yaniv, I., Meyer, D.E., Gordon, P.C., Huff, C.A., & Sevald, C.A. (1990). Vowel similarity, connectionnist models, and syllable structure in motor programming of speech. *Journal of Memory and Language*, 29, 1-27.

DISSERTATION ABSTRACTS*

PHONETIC AND PHONOLOGICAL CONTRASTS IN THE ACQUISITION OF VOICING: A LINGUISTIC AND DEVELOPMENTAL STUDY OF VOICE ONSET TIME PRODUCTION IN HINDI AND ENGLISH

Katharine Davis, Ph.D., Cornell University 1991

The main goal of this dissertation is to investigate cross-linguistic patterns in the acquistion of the phonemic voicing contrast. Experimental acquisition data from Hindi and English, two languages with differing patterns of phonological /voice/, were compared. Acoustic recordings of word-initial velar stops were taken from twenty children aged two to six and from ten adults in each language. The voice onset times of these productions were measured, and statistical comparisons across age groups were performed.

The results of this study show that the English-learning children produce the English /k/ and /g/ phonemes contrastively by the age of two, replicating previous studies. The Hindi-learning children acquire some of the contrasts of their language earlier than others. For example, the children produce /k/ and /k^h/ contrastively by age two, but do not produce /k/ and /g/ contrastively until after age four.

These results confirm a hypothesis that acoustic differences between phonemes predict the developmental patterns of the Hindi learners better than the feature specifications of the phonemes. Such acoustic differences also account for the relatively early contrast produced by the English learners. Not previously noted in the literature is the result that, in adults, the lag times of the Hindi voiced aspirates are significantly shorter than the lag times of the voiceless aspirates.

It is argued that the results of the developmental study can be explained by a phonetic analysis of the /voice/ contrast which redefines the acoustic cue of voice onset time in order to account for the lag times of prevoiced stops. This voice onset time analysis is shown to be compatible with a phonological analysis employing the features [spread] and [voiced]. It is suggested that in the Hindi aspirates, the presence of breathy voice is an optional redundant cue to a [+ voiced] specification.

Based on the findings of the present study, a new model of phoneme acquisition is proposed.

---

* This dissertation is available for purchase through the Department of Modern Languages and Linguistics. Inquiries should be addressed to DMLL Publications, Morrill Hall, Cornell University, Ithaca, NY 14853-4701.