Working Papers of the Cornell Phonetics Laboratory

No. 2 • April 1988

Research in Laboratory Phonology

The Role of the Sonority Cycle in Core Syllabification
G. N. Clements 1

The Delta Programming Language: an Integrated Approach to Nonlinear Phonology, Phonetics, and Speech Synthesis
Susan R. Hertz 69

Articulatory Binding
John Kingston 123

Shorter Notices

Research Facilities in the Cornell Phonetics Laboratory
G.N. Clements, S.R. Hertz, J. Kingston 173
Preface

The Working Papers of the Cornell Phonetics Laboratory (WPCPL) are issued on an occasional basis to report on current research by the lab's members and other matters of interest to the phonetics and linguistics community. Work appears here in pre-publication form and may appear in different form elsewhere.

The major reports in this issue were originally presented at the First Conference on Laboratory Phonology, held at the Ohio State University on June 5-7, 1987. They appear here in considerably revised form and will also appear in Papers in Laboratory Phonology I: Between the Grammar and the Physics of Speech, edited by John Kingston and Mary Beckman (to be published by the Cambridge University Press). Work in laboratory phonology emphasizes experimental and computational approaches, i.e. the methods of the laboratory, to phonological questions, and attempts to bring phonologists, phoneticians, and researchers in other disciplines to a sense of sharing a common enterprise: the study of speech sounds and how they pattern in language. As exhibited in these papers and at the conference, laboratory phonology currently ranges widely, from phonological approaches enlightened by relevant research in phonetics to phonetic approaches informed by an awareness of linguistic theory and methodology. We expect that in time, these two now-distinct ways of working will be replaced by entirely integrated approaches.

Because we strongly support the attempt to integrate the methodology and findings of phonology and phonetics, papers in future numbers of the Working Papers will include as far as possible papers that represent the integration we call "laboratory phonology," although papers representing more traditional phonetic and phonological approaches will appear here as well. We also expect that some issues will be devoted entirely to monograph-length studies, and that others will include shorter reports on ongoing research or technical issues.

The Working Papers are distributed free of charge to other phonetics, phonology, and speech laboratories interested in exchanging their own reports with us, and are otherwise available at cost. To set up an exchange or obtain a copy, write to: Working Papers, Phonetics Laboratory, Department of Modern Languages and Linguistics, Morrill Hall, Cornell University, Ithaca, N.Y. 14853-4701.

Nick Clements
Susan R. Hertz
John Kingston
Phonetics Laboratory
Department of Modern Languages and Linguistics
Morrill Hall
Cornell University
Ithaca, N.Y. 14853-4701
tel. (607) 255-7394

Research Personnel (listed alphabetically):
Theresa Carroll, Administrative Aide
Nick Clements, Professor of Linguistics and Cognitive Studies
John Drury, Graduate Student in Linguistics
Roger Garnett, Electronics Technician
Susan R. Hertz, Senior Research Associate in Linguistics
Rukayyah Herzallah, Graduate Student in Linguistics
John Kingston, Assistant Professor of Linguistics and Cognitive Studies
Yetunde Laniran, Graduate Student in Linguistics
Stuart Milliken, Graduate Student in Linguistics
Clifford Olling, Systems Programmer/Analyst
Annie Riaolland, Visiting Scholar (C.N.R.S., Paris)
Susana Sainz, Graduate Student in Linguistics
Chilin Shih, Assistant Professor of Linguistics and Chinese
James Stewart, Graduate Student in Computer Science
Renée Zakia, Graduate Student in Linguistics

This issue of the Working Papers of the Cornell Phonetics Laboratory was produced by Nick Clements and self-edited by the authors. Copies of these reports are distributed free of charge to phonetics, phonology and speech laboratories interested in exchanging their own reports with us, and are otherwise available for the at-cost price of $8.00 per issue.
The Role of the Sonority Cycle in Core Syllabification

G. N. Clements

1. Introduction

One of the major concerns of laboratory phonology is that of determining the nature of the transition between discrete phonological structure (conventionally, "phonology") and its expression in terms of nondiscrete physical or psychoacoustic parameters (conventionally, "phonetics"). A considerable amount of research has been devoted to determining where this transition lies, and to what extent the rule types and representational systems needed to characterize the two levels may differ (see Keating (1985) for an overview). For instance, it is an empirical question to what extent the assignment of phonetic parameters to strings of segments (phonemes, tones, etc.) depends upon increasingly rich representational structures of the sort provided by autosegmental and metrical phonology, or upon real-time realization rules -- or indeed upon some combination of the two, as many are coming to believe. We are only beginning to assess the types of evidence that can decide questions of this sort, and a complete and fully adequate theory of the phonetics/phonology interface remains to be worked out. A new synthesis of the methodology of phonology and phonetics, integrating results from the physical, biological and cognitive sciences, is required if we are to make significant progress in this area.

The present study examines one question of traditional interest to both phoneticians and phonologists, with roots that go deep into modern linguistic theory. Many linguists have noted the existence of cross-linguistic preferences for certain types of syllable structures and syllable sequences. These have been the subject of descriptive studies and surveys such as that of Greenberg (1978), which have brought to light a number of generalizations suggesting that certain syllable types are less complex or less marked than others across languages. We must accordingly ask how, and at what level these tendencies are expressed and explained within a theory of language. ¹

From the late nineteenth century onwards, linguists have proposed to treat generalizations of this sort in terms of a Sonority Sequencing Principle governing the preferred order of segments within the syllable. According to this principle, segments can be ranked along a "sonority scale" in such a way that segments ranking higher in sonority
stand closer to the center of the syllable and segments ranking lower in sonority stand
closer to the margin. While this principle has exceptions and raises questions of
interpretation, it expresses a strong crosslinguistic tendency, and represents one of the
highest-order explanatory principles of modern phonological theory.

A theory incorporating such a principle must give an adequate account of what
“sonority” is, and how it defines the shape of the optimal or most-preferred syllable type.
Up to now there has been little agreement on these questions, and phoneticians and
phonologists have characteristically taken different approaches to answering them.
Phoneticians have generally elected to focus their attention on the search for physical or
perceptual definitions of sonority, while phonologists have looked for formal explanations,
sometimes claiming that sonority has little if any basis in physical reality. It seems
appropriate to reconsider these questions at this time, especially in view of the advances
that have been made elsewhere in understanding syllable structure and its consequences for
the level of phonetic realization.

My purpose here will be to examine the status of sonority within phonological theory.
I will propose that an adequate account of sonority must be based on a principle termed the
Sonority Cycle, according to which the sonority profile of the preferred syllable type rises
maximally at the beginning and drops minimally at the end (the term cycle will be used
exclusively in this study to refer to this quasiperiodic rise and fall). We will see that this
principle is capable of providing a uniform explanation not only for crosslinguistic
generalizations of segment sequencing of the sort mentioned above, but also for an
impressive number of additional observations which have not been related to each other up
to the present time. Regarding its substantive nature, I will suggest that sonority is not a
single, multidimensional property of segments, but is derived from more basic binary
categories, identical to the major class features of standard phonological theory (Chomsky
and Halle 1968) supplemented with the feature “approximant.”

2. The Sonority Sequencing Principle: a Historical Overview

The notion that speech sounds can be ranked in terms of relative stricture or
sonority can be found in work as early as that of Whitney (1865). However, the first com-
prehensive attempts to use such a ranking to explain recurrent patterns of syllable structure
are due to Sievers (1881), Jespersen (1904), Saussure (1916), and Grammont (1933).
Sievers observed that certain syllable types were commonly found in languages, while others differing from them only in the order of their elements were rare or nonexistent. For example, he noted that *mla, mra, alm, arm* were relatively frequent in languages, while *lma, rma, aml, amr* were not. On this basis he assigned the liquids a higher degree of sonority than the nasals. Proceeding in this fashion, Sievers arrived at a ranking of speech sounds in terms of their inherent sonority. In a syllable consisting of several sounds, the one with the greatest sonority is termed the peak, or sonant, and the others the marginal members, or consonants. According to Sievers' sonority principle, the nearer a consonant stands to the sonant, the greater must its sonority be.²

In Jespersen's version of the theory, which is the more familiar today, the sonority principle was stated as follows: "In jeder Lautgruppe gibt es ebensoviele Silben als es deutliche relative Höhepunkte in der Schallfülle gibt" ("In every group of sounds there are just as many syllables as there are clear relative peaks of sonority") (p. 188). Jespersen’s version of the sonority scale is given in (1):³

(1)
1. (a) voiceless stops, (b) voiceless fricatives
2. voiced stops
3. voiced fricatives
4. (a) voiced nasals, (b) voiced laterals
5. voiced r-sounds
6. voiced high vowels
7. voiced mid vowels
8. voiced low vowels

Drawing on the work of Sievers and Jespersen, we may state a provisional version of the *Sonority Sequencing Principle* as follows:⁴

(2) *Sonority Sequencing Principle:*

Between any member of a syllable and the syllable peak, only sounds of higher sonority rank are permitted.

Under this principle, given the sonority scale in (1), syllables of the type *tra, dva, sma, mra* are permitted, while syllables like *rta, vda, msa, mla* are excluded. Crosslinguistic comparison supports the view that clusters conforming to the Sonority Sequencing
Principle are the most commonly occurring, and are often the only cluster types permitted in a given language. Clusters violating this principle do occur, as we will see, but they are relatively infrequent, and usually occur only in addition to clusters conforming to it.

The theory of sonority just characterized was first developed in the late nineteenth century, when the notion of a synchronic grammar as a system of categories and representations defined at various degrees of abstraction from the physical data was yet to emerge in a clear way. The present revival of the sonority principle in phonological theory has taken place in a very different context, that of the initial period of response to Chomsky and Halle's *Sound Pattern of English*. Chomsky and Halle proposed, among other things, a major revision of the distinctive feature system of Jakobson, Fant and Halle (1952) which retained its binary character but reorganized the way sounds were classified by features. One aspect of the new system was a characterization of the traditional notion *degree of stricture* in terms of a set of binary *major class features*. These features -- first identified as [sonorant, consonantal, vocalic], with [vocalic] later replaced by [syllabic] -- were grouped together on the basis of their similar function in accounting for the basic alternation of opening and closing gestures in speech (Chomsky and Halle 1968, 301-2). In terms of their function within the overall feature system, these features played a role analogous to that of sonority in prestructuralist phonology. By excluding any additional feature of sonority from their system, Chomsky and Halle made the implicit claim that such a feature was unnecessary.

The notion of *scalar* or multivalued features was first introduced into generative phonology by Foley (1970, 1972) as an alternative to binary feature systems. Foley’s approach was intended as a radical alternative to the approach of Jakobson, Halle, and Chomsky. His main proposals were that (i) all binary features should be replaced by a set of scalar features, and that (ii) these scales do not refer to phonetic properties of segments, but are justified only by recurrent cross-linguistic aspects of segment behavior, as evidenced particularly in sound change. Foley’s scalar feature of resonance (1972) is given in (3):

(3) 1. oral stops
2. fricatives
3. nasals
4. liquids
5. glides
6. vowels

Through its influence on Zwicky (1972), Hankamer and Aissen (1974), and Hooper (1976), all of whom cite Foley’s work, this view of resonance gained wide currency and in its later adaptations came to have a substantial influence on the subsequent development of syllable theory within generative phonology.

More recently, the Sonority Sequencing Principle in something close to its original version has had a general revival in the context of syllable phonology (major references include Hooper 1976, Kiparsky 1979, Steriade 1982, and Selkirk 1984). In a significant further development, Hooper (1976) proposed a principle according to which the sonority of a syllable-final consonant must exceed that of a following syllable-initial consonant (equivalently, the second must exceed the first in “strength”). This principle, originally proposed for Spanish, has been found to hold in other languages, though usually as a tendency rather than an exceptionless law (Devine and Stephens 1977, Christdas 1988), and has come to be known as the Syllable Contact Law (Murray and Vennemann 1983). We may state it as follows, using $ to designate a syllable boundary:

(4) *The Syllable Contact Law*:

In any sequence $C_A$ $C_B$, there is a preference for $C_A$ to exceed $C_B$ in sonority.

3. Current Issues in Sonority Theory

In spite of its importance, sonority remains an ill-defined, if not mysterious concept in many respects, hence the urgency of reexamining its theoretical and empirical bases. Among the questions we would like to be able to answer are the following: how, exactly, is sonority defined in phonological theory? Is it a primitive feature, or is it defined in terms of other features? What are its phonetic properties? Assuming that some version of the Sonority Sequencing Principle is correct, at what linguistic level does it hold? Over what morphological or prosodic domain are sonority constraints most appropriately defined? Can we define a single sonority scale valid for all languages, or must we recognize a significant degree of cross-linguistic variation? These are some of the questions that will be addressed in the remainder of this study.
3.1. At What Level Does the Sonority Sequencing Principle Hold?

One important issue concerns the level at which the SSP holds. A surface-oriented version of the SSP might claim that it holds without exception of surface syllabification in all languages. In such a view, the SSP would project a unique and exhaustive syllabification over any arbitrary string of phonemes in surface representation, containing no prior annotations showing where the syllable peaks lie. In agreement with much of the more recent literature, I will suggest that such an interpretation of the SSP is incorrect, and that the SSP holds at a more abstract level than surface representation.

Consider the representative examples in (5). The cases in (5a) represent sonority “plateaus”, in which two adjacent consonants at the beginning or end of a word have the same sonority rank. In (5b) we find representative cases of sonority “reversals”, in which the sonority profile first rises, then drops again as we proceed from the edge of the word inward. As these examples show, cases can be cited involving all major segment classes: fricatives, liquids, nasals and glides. In (5c) we find cases in which the syllable peaks are not sonority peaks (at least not prior to the assignment of syllabicity), but are adjacent to elements of higher sonority. For example, the syllabic peak of English yearn is the liquid [r], which is adjacent to the glide [y] and thus does not constitute a sonority peak. Contrasting examples such as pedaller, pedlar show that syllable peaks are not fully predictable in all languages on the basis of the surface context. The level of representation assumed here is approximately that of systematic phonetic representation prior to the application of (automatic or language-particular) phonetic realization rules, though as transcription practices vary from one writer to another and are often inexplicit, this assumption may not accurately reflect the writer's intention in all cases.

(5) a. **Consonant sequences with sonority plateaus:**

   English: apt, act, sphere
   Russian: mnu ‘I crumple’, tkut ‘they weave’, kto ‘who’, gd’e ‘where’
   Mohawk: tkataweya’t ‘I enter’, kka:weš ‘I paddle’
   Marshallese: qqin ‘to be extinguished’, kken ‘to be invented’, lliw ‘angry’

b. **Consonant sequences with sonority reversals:**

   English: spy, sty, sky, axe, apse, adze
German: Spiel ‘game’, Stein ‘stone’, Obst ‘fruit’, letzt ‘last’
Russian: rta ‘mouth (gen.)’, 1ba ‘forehead (gen.)’, mgla ‘mist’
Cambodian: psa: ‘market’, spiy ‘cabbage’
Pashto: wro ‘slowly’, wlar ‘he went’, lmar ‘sun’
Ewe: yra ‘to bless’, wlu ‘dig’
Klamath: ms’s ‘prairie dog’, ltewa ‘eats tules’, toq’lga ‘stops’
Mohawk: kskohárya’ks ‘I cut dead wood’
Ladakhi: lpaxs ‘skin’, rtn-pa’heel’, rgyal’a’road’
Kota: anźrčgɛg’vdik ‘because will cause to frighten’
Abaza: yg’yzmlr stxd ‘they couldn’t make him give it back to her’
Tocharian A: yneš ‘apparent’, ysar ‘blood’
Yateé Zapotec: wbeγ ‘hoe’, wše-zí-le ‘morning’, wza-’a ‘I ran’

c. Syllables whose peaks are not sonority peaks:

English: yeam [yrm], radio, pedaller [ped-].-r] vs. pedlar [ped-lər]
German: kōnnen [kön-n] ‘to be able’, wollen [vol-n] ‘to be willing’ vs. Köln
[kölN] ‘Cologne’
French: rel(e)ver [rä]-ve] ‘to enhance’, troua [tru-a] ‘(s)he dug a hole’ vs.
trois [trwa] ‘three’, haï [a-i] ‘hated’ vs. ail [ay] ‘garlic’
[fwi-mos] ‘we went’ vs. huimos [u-i-mos] ‘we fled’; piara [pya-ra] ‘herd’ vs.
piara [pi-a-ra] ‘chirp’ (past subj.)
PIE: *wλkos ‘wolf’ (cited by Saussure as evidence against the sonority theory)
Turkish: dağa [da-a] ‘mountain (dat.)’ vs. dağ [da:] ‘mountain’ (nom.)
Berber: ti-wn-tas ‘you climbed on him’, ra-ymm-γi ‘he will grow’, ra-tk-ti ‘she
will remember’
Swahili: wa-li-m-pa ‘they gave him’, wa-i-te ‘call them’, ku-a ‘to grow’ vs.
kwa ‘of, at’
Bella Coola: nmnmak ‘both hands’, mnmnts ‘children’, sk’lxlxc ‘I’m getting
cold’

The facts are not at issue in most of these cases, and in many similar cases that could be
cited. Some writers’ accounts are detailed enough to allow us to accept the transcriptions
with a very high degree of confidence. For example, Jaeger and Van Valin (1982) report that the glide \([w]\) occurs regularly before obstruents in word-initial clusters in Yateé Zapotec (see examples in (5b) above). They argue that \([w]\) cannot be reanalyzed as or derived from the vowel \([u]\) for a number of reasons: (i) \([w]\) has the acoustic properties of a glide, namely short duration and a rapidly moving second formant; (ii) \([u]\) does not occur elsewhere in the language, underlyingly or on the surface; (iii) otherwise there are no vowel-initial words in Yateé Zapotec; (iv) \([w]\) carries no tone, whereas all syllables otherwise have phonemic tone; (v) \([w]\) usually functions as an aspect prefix, and all other aspect prefixes are consonants. As Jaeger and Van Valin point out, this latter fact helps to explain the survival of these rare cluster types, since the \([w]\) carries essential morphological information that would be lost if \([w]\) were deleted.

Such problems for surface-oriented versions of the SSP were recognized in the earliest literature. Sievers, for example, noted the existence of anomalous clusters such as those in \(pta, kita, apt, akt\) in which the second member was a dental sound. He suggested that they might be due to the “ease of the articulatory transition” to a dental, but did not offer independent reasons for assuming that transitions to dentals were simpler than transitions to other places of articulation. Sievers had similar trouble dealing with the reversed-sonority clusters in syllables like \(spa, sta, aps, ats\), and attempted to explain them by introducing the notion of the “secondary syllable” (Nebensilbe), a unit which counted as a syllable for the purposes of the SSP but not for linguistic rules, such as stress placement and the like. Jespersen proposed special principles to account for type (5c) violations involving rising sonority ramps (as in Spanish \(río\)), but had nothing to say about falling ramps (as in Spanish \(país\)).

I will suggest here that the SSP holds at deeper levels of representation than surface representation, and in particular that it governs underlying syllabification in the lexical phonology. More exactly, the underlying representations of any language are fully syllabified in accordance with certain principles of core syllabification, which are sensitive to sonority constraints (see further discussion in section 5.1). However, some consonants remain unsyllabified (or extrasyllabic) after the core syllabification rules have applied, and may therefore create violations of sonority restrictions. Such consonants either become syllabified at a later point in the derivation, or are deleted.

Such an analysis is strongly supported when we examine the phonological characteristics of sonority violations more closely, since we often find convincing evidence
that they involve consonants that are not incorporated into syllable structure in the early stages of phonological derivations. In Sanskrit, for example, such consonants regularly fail to participate in reduplication, a fact which can be explained on the assumption that unsyllabified consonants are invisible to reduplication (Steriade 1982). In Turkish, Klamath, and Mohawk, among other languages, such consonants regularly trigger the application of a variety of epentheses (Clements and Keyser 1983, Michelson 1983); we can view these rules as having the function of incorporating unsyllabified consonants into syllable structure, and thus of simplifying representations. The fact that extrasyllabic elements tend to be removed from representations by rules of epentheses is an instance of the more general principle that rules tend to apply in such a way as to replace complex representations with simpler ("euphonic", "eurhythmic") ones. Furthermore, in English and many other languages, violations of the sonority constraints are restricted to the edges of the syllabification domain, reflecting the preference of extrasyllabic elements for this position (Milliken 1988). Thus, for example, the sonority violation found in the final cluster of *apt* can be explained on the assumption that the [t] is extrasyllabic at the point where the sonority constraints apply; notice that syllable-final sequence [pt] is found only at the end of level 1 stems, which form the domain of core syllabification (Borowsky 1986). Finally, there is evidence that such consonants may remain unsyllabified all the way to the surface, in a number of languages which permit long, arbitrary strings of consonants in surface representation (see examples in (5b) and the references just below).

Other, related ways of explaining or eliminating surface exceptions to the SSP have been proposed in the earlier literature. These include the restriction of the principle to the syllable "core" as opposed to "affix" (Fujimura andLovins 1978), the recognition of a syllable "appendix" that lies outside the scope of sonority restrictions (Halle and Vergnaud 1980), the postulation of language-particular rules that take precedence over the principle (Kiparsky 1979, 1981), and the treatment of clusters such as English *sp, st, sk* as single segments rather than clusters for the purposes of syllabification (for example, Selkirk 1982). It is not always possible to find convincing independent evidence for such strategies in all cases, however, and it seems possible that a hard core of irreducible exceptions will remain. Languages exhibiting a high degree of tolerance for long, arbitrary sequences of consonants prominently include (but are not limited to) the Caucasian languages, several Berber dialects, numerous Southeast Asian languages and many languages native to the Northwest Pacific Coast.10
3.2. The Phonetic Basis of Sonority

A further issue in sonority theory involves its phonetic basis. Given the remarkable similarity among sonority constraints found in different and widely separated languages, we might expect that sonority could be directly related to one or more invariant physical or psychoacoustic parameters. However, so far there exists no entirely satisfactory proposal of this sort. The problem is due in part to lack of agreement among phonologists as to exactly what the universal sonority hierarchy consists of. As Selkirk points out (1984), it will not be possible to determine the exact phonetic character of sonority until phonologists have come to some agreement about the identity of the hierarchy. But many proposals have been put forward at one time or another, and at present there are a great number of competing sonority scales to claim the attention of phoneticians.

This problem should disappear as better theories of sonority are developed. But in addition to uncertainty regarding the linguistic definition of the sonority scale, there is some question whether a uniform, independent phonetic parameter corresponding to sonority can be found, even in principle. This because the various major classes of speech sounds, have substantially different properties from nearly every point of view: aerodynamic, auditory, articulatory, and acoustic. It is true that a variety of phonetic definitions of sonority have been offered in the past, from the early proposal to define it in terms of the relative distance at which sounds could be perceived and/or distinguished (O. Wolf, cited by Jespersen 1904,187) to a variety of more sophisticated recent suggestions (see for example Lindblom (1983) and Keating (1983) for discussion of an articulatory-based definition of sonority, and Price (1980) for discussion of an acoustic-based definition). But as Ohala and Kawasaki state (1985,122), "no one has yet come up with any way of measuring sonority" -- not at least a widely agreed-upon method based on a uniform phonetic parameter corresponding to a linguistically-motivated sonority scale.11 The ultimate response to this problem may be to deny that sonority has any regular or consistent phonetic properties at all (Hankamer and Aissen 1974, Hooper 1976, Foley 1977). While this represents a position to which we might be driven out of necessity, we do not choose it out of preference, since in the absence of a consistent, physical basis for characterizing sonority in language-independent terms we are unable to explain the nearly identical nature of sonority constraints across languages.

We should not be overly concerned about the difficulty of finding well-defined phonetic definitions of sonority, however. In the past, attempts to define linguistic constructs in
terms of physical definitions (or operational procedures) have usually proven fruitless, and it is now widely agreed that abstract constructs are justified just to the extent that they are tightly integrated into the logical structure of predictive and explanatory theories. Thus, no adequate phonetic definition has ever been given of the phoneme, or the syllable -- and yet these constructs play a central and well-understood role in modern phonology. Similarly, the notion of sonority is justified in terms of its ability to account for crosslinguistic generalizations involving phoneme patterning, and need not have a direct, invariant expression at the level of physical phonetics. The problem raised by our (at present) incomplete physical understanding of sonority reduces to the problem of accounting for why linguistically-motivated sonority rankings are very much the same across languages. I will suggest that the sonority scale is built into phonological theory as part of universal grammar, and that its categories are definable in terms of the independently-motivated categories of feature theory, as discussed in section 4.

3.3. The Redundancy of the Feature “Sonority”

A further issue concerns the feature characterization of sonority. A multivalued feature of sonority makes good sense as an alternative to the binary major class features of standard theory, but is harder to justify as a supplement to them, since sonority can be adequately defined in terms of these independently-motivated features (see discussion below), and is thus redundant in a theory containing them. One way of eliminating this redundancy is to eliminate the major class features themselves. This approach is considered by Selkirk, who suggests that the work of the major class features can done by (a) a feature representing the phonetic dimension of sonority, (b) the sonority hierarchy, and (c) the assignment of a sonority index to every segment of the language (1984, 111). Hankamer and Aissen are even more categorical, stating “the major class features of the standard feature system do not exist” (1974, 142). Most proponents of the sonority hierarchy have been reluctant to go this far, but there has been little explicit discussion of how we may justify the presence of two types of features with largely overlapping functions in feature theory.

4. The Major Class Features and the Definition of Sonority

On the basis of this review of the issues, let us consider how sonority can be incorporated into a formal theory of phonology and phonetics.
As has previously been noted (Lekach 1979), the sonority scale can be defined in terms of independently-motivated binary features. I will adopt a definition involving the four major class features shown in (6), where O = obstruent, N = nasal, L = liquid, G = glide, and V = vowel (the choice of this particular scale will be justified below). The sonority scale is derived by taking the sum of the plus-specifications for each feature.

(6) \[
\begin{array}{cccc}
O & N & L & G \\
- & - & - & + \\
- & - & + & + \\
- & + & + & + \\
- & + & + & + \\
\end{array}
\]

"syllabic" vocoid approximant sonorant

rank (relative sonority)

Three of these features are familiar from the earlier literature. "Syllabic" can be interpreted as referring to the prosodic distinction between V and C elements of the timing tier, or alternative characterizations of syllable peaks in prosodic terms. For reasons given in Clements and Keyser (1983, 8-11, 136-7), I will take this feature to have no intrinsic physical definition, but to be defined in language-particular terms: "syllabic" segments are those which attract the properties of the syllabic nucleus in any particular language, while "nonsyllabic" segments are those which do not. "Vocoid", a term introduced by Pike (1943), is simply the converse of the traditional feature "consonantal" and is defined accordingly. "Sonorant" has its usual interpretation.

In order to complete our definition of the sonority scale in terms of binary features we require a further feature, grouping liquids, glides and vowels into one class and nasals and obstruents into another. This is exactly the function of the feature “approximant,” proposed by Ladefoged (1982) who defines it as “an articulation in which one articulator is close to another, but without the vocal tract being narrowed to such an extent that a turbulent airstream is produced” (1982,10). In order to clearly exclude nasals (which do not involve a turbulent airstream), I will consider an approximant to be any sound produced with an oral tract stricture open enough so that airflow through it is turbulent only if it is voiceless. 12

The recognition of approximant as a feature is justified by the fact that approximants tend to pattern together in the statement of phonological rules. For example, many
languages allow complex syllable onsets only if the second member is an oral sonorant, i.e. an approximant in our terms. Similarly, nonapproximants often pattern together. In Luganda, only nonapproximants occur as geminates: thus we find geminate /pp, bb, ff, vv, mm/, etc., but not /ww, ll, yy/.

We will treat approximant as a binary feature, like the other major class features. In Ladefoged’s account (see Ladefoged 1982, 10, 38-9, 61-2, 256, 265), approximant is not a feature category, but a value or specification of the feature category stop. This category is a three-valued scalar feature whose values are stop, fricative, approximant. This system makes the prediction that the values stop and approximant are mutually exclusive, and thus do not allow any segment to bear both of these values at once. The crucial data here involve laterals, which in the feature classification given in Halle and Clements (1983) are classified as [-cont], and in the present account must also be [+approximant]. Under the view proposed here, then, /l/ is both a stop and an approximant, and may function as such in one and the same language. This appears to be correct. For example, in English /l/ functions with the other approximants in its ability to occur as the second member of complex syllable onsets: /pl, tr, kw/, etc., while nasals may occur in this position only after /s/. But /l/ also patterns with nasals in the rule of intrusive stop formation, in which an intrusive stop is inserted between a nasal or lateral and the following fricative in words like den[t]se, fal[t]se, ham[p]ster, weal[t]thy. This rule involves a “lag” of the features [-cont] and [place] onto the following segment (see Clements 1987a for discussion).

In an alternative proposal, Van Coetsem (1979) has suggested reintroducing the feature “vocalic” alongside “syllabic.” As he points out, this would correctly allow us to distinguish nasals and liquids by a major class feature, unlike the SPE feature system which must make use of the feature “nasal.” In our proposal (and Ladefoged’s), however, this task is accomplished by the major class feature “approximant”. The major difference between the two proposals is that if we chose “vocalic” instead of “approximant,” glides would be ranked at the same sonority level as liquids by the algorithm given above:
More importantly, a system with the feature "approximant" seems to reflect natural groupings of sounds better than one with "vocalic." Thus, liquids and glides ([+approximant] nonsyllabics) frequently fall together as a class in the statement of rules, while obstruents, nasals, and glides ([vocalic] nonsyllabics) rarely or never do. One of the primary functions of the feature [vocalic] in earlier feature systems was to designate the natural class of liquids, characterized as [+vocalic,-consonantal] sonorants. However, this function is equally well served by a feature system containing [approximant], in which liquids are designated by the features [+approximant,-vocoid]. I conclude, therefore, that the major class features are correctly represented as in (6).

The scale in (6) is incomplete in that it only provides for vowels (i.e., syllabic vocoids) and not for syllabic consonants. In some languages, nasals and even obstruents can function as syllable peaks, in certain circumstances (for syllabic obstruents see Bell's (1978) survey article, Dell and Elmedlaoui (1985) for discussion of a dialect of Berber, Clements (1986) for discussion of syllabic geminates in LuGanda, and Rialland (1986) for discussion of syllabic consonants derived through compensatory lengthening in French). In principle, any segment can occupy the syllable peak, but the ability of a given segment to function as a syllable peak is related to its rank on the sonority scale. Our model predicts the following ranking of syllabic segments:

\[
\begin{align*}
1 &< N < L < V < G < V \\
+ &\quad + &\quad + &\quad + &\quad + \\
+ &\quad - &\quad - &\quad + &\quad + \\
+ &\quad - &\quad + &\quad + &\quad + \\
+ &\quad + &\quad + &\quad + &\quad + \\
1 &\quad 2 &\quad 3 &\quad 4 &\quad \text{rank}
\end{align*}
\]
(Note that a syllabic glide is identical to a vowel, or to put it another way, a glide is simply a nonsyllabic vowel: cf. Pike (1943).) However, this does not quite accord with the facts, since as Bell has noted (1978), syllabic nasals are generally preferred to syllabic liquids in languages that have just one or the other. The notion “relative sonority,” as defined in (8), does not therefore extend unproblematically to syllable peaks, which require separate discussion.  

Combinations of major class features other than those given in (6) and (8) are non-occurring, and are excluded by the following universal redundancy rules:

\[(9) \begin{align*}
\text{a. } & [-\text{sonorant}] \rightarrow [-\text{approximant}] \\
\text{b. } & [-\text{approximant}] \rightarrow [-\text{vocoid}]
\end{align*}\]

These rules entail the following, by contraposition:

\[(c) \begin{align*}
& [+\text{approximant}] \rightarrow [+\text{sonorant}] \\
& [+\text{vocoid}] \rightarrow [+\text{approximant}]
\end{align*}\]

I will assume that these redundancy rules apply to the output of each phonological rule as well-formedness conditions, and readjust the values for approximant, vocoid and sonorant as necessary.

The four major class features together with the redundancy rules given above define twenty-one natural classes (or twenty-nine, if we count single-member classes). These are conveniently suggested in terms of the following 2x4 array of segment types. Terms that can be enclosed in a vertically or horizontally oriented rectangle constitute a natural class. Three examples are given for illustration.

\[(10) \begin{align*}
\text{[+syllabic]:} & \begin{array}{ccc}
\text{O} & \text{N} & \text{L} \\
\text{V}
\end{array} \\
\text{[-syllabic]:} & \begin{array}{ccc}
\text{O} & \text{N} & \text{L} \\
\text{G}
\end{array}
\end{align*}\]
The three boxes represent the class of syllabic consonants, the class of consonantal sonorants, and the class of nonsyllabic approximants, from upper left to lower right. Notice that a single step to the right or up results in a one-degree increase in sonority rank. This array clearly shows the special status of the feature [syllabic] in the sonority scale: it is the only major class feature that crossclassifies all others. This suggests that [syllabic] has a different status in feature representation than the true major class features, perhaps not functioning as a feature at all but as a prosodically defined position within the syllable, as suggested above.

The sonority scale as given in (6) does not include a subdivision of obstruents into stops and fricatives, or into voiceless and voiced obstruents; nor does it recognize a distinction between lateral and central liquids. This is because to date (see further discussion below), the best-motivated crosslinguistic generalizations involving sonority, such as Greenberg’s, do not appear to require any further subdivision of the sonority scale. However, some individual languages have been proposed as motivating further subdivisions, such as one between voiced fricatives and other obstruents, or one between lateral and central liquids. To accommodate such languages, linguists have proposed to recognize more elaborate versions of the sonority scale including additional features such as [continuant], [voiced], [coronal], etc., and have proposed that the features relevant for the definition of sonority can vary from one language to another. This approach will not be adopted here, however. The explanatory value of sonority theory lies in its ability to predict valid crosslinguistic generalizations. As soon as we allow the sonority scale to vary in its identity from one language to another, we seriously undermine its explanatory role by increasing the number of ways in which it will accommodate potential exceptions, thus reducing the number of crosslinguistic generalizations that it accounts for.

I will argue below that there are considerable advantages to maintaining a strong, predictive version of sonority theory. Much of the apparent evidence for language-particular variation in the sonority scale comes from observations which can be explained in other ways. For example, the tendency for voiced fricatives to be excluded as the first member of initial clusters, found in a number of languages including English, can be understood in terms of a principle (to be described in section 5.3) which holds that all else being equal, sequences containing less marked segments are favored over sequences containing more marked segments. Crosslinguistically, voiced fricatives are more marked
than either stops or voiceless fricatives. This principle extends to many other distributional regularities that had previously been thought to require language-particular modifications of the sonority scale, such as the apparently greater sonority of coronal as opposed to noncoronal consonants in some languages. The strongest position consistent with the available evidence is that a single scale, perhaps O < N < L < G < V or a simple variant of this, defines the unmarked order of segments within the syllable across languages, and that apparent deviations from this scale have independent explanations.

4.1. An Alternative View: Sonority as a Multivalued Feature

What has tempted many linguists to consider sonority to be a single, multivalued feature is the fact that it arrays segment classes into a hierarchy. Other binary features in phonology do not seem to have this property. The notion of hierarchy, it can be argued, is most simply and directly expressed in terms of a single multivalued feature, rather than by making use of several binary features constrained by redundancy rules such as those in (9a-d). A multivalued feature of this sort is equally capable of capturing the necessary natural classes, if we allow that rules may refer to any continuous sequence of positions along the hierarchy. For example, given the sonority scale in (6), the natural class of liquids, glides and vowels can be referred to by the expression “[2-4sonority]” (cf. Zwicky 1972, Selkirk 1984 for discussion of such a proposal).

Let us consider the notion of hierarchy in general terms. Given a binary feature system, a hierarchy is defined whenever we have an implicational relation holding between two features. For example, given the two binary features F, G and the implication: [+F]→[+G] (entailing [-G] → [-F] by contraposition), we define a three-term hierarchy over the segment classes A, B, C:

\[
\begin{array}{ccc}
& A & B & C \\
- & - & + & [F] \\
- & + & + & [G] \\
\hline
0 & 1 & 2 & \text{rank}
\end{array}
\]

The number of terms in the hierarchy increases as we add implicational relations. Thus, the four-term hierarchy O < N < L < G results from the presence of the two implicational statements (redundancy rules) (9a-b) given earlier.
Hierarchies of this type are common elsewhere in grammar. For example, in some Bantu languages we find the following nominal hierarchy defined by the accessibility of a given nominal to direct object status: 1st person > 2nd person > 3rd person human > 3rd person animal > 3rd person inanimate (Hyman, Duranti and Morolong 1980). In principle, it would be possible to define this hierarchy in terms of a single multivalued feature whose meaning is roughly 'similarity or closeness to ego'. However, the various positions on this hierarchy do not form a continuum, but a series of discrete steps, most of which are found to play a role elsewhere in grammar: for example, the distinction between first, second and third person animate commonly plays a role in Bantu inflectional morphology. In such cases linguists do not usually assume that a single, multivalued feature or parameter is at work, but rather that the hierarchical scale is built up out of independently-needed linguistic categories linked by implicational relations.

This seems to be the appropriate way to view the sonority hierarchy. There is, moreover, considerable phonetic and perceptual rationale for the definition of the sonority scale given in (6) in terms of the major class features. "Sonority" is a composite property of speech sounds which depends on the way they are specified for each of a certain set of features. Plus-specifications for any of these features have the effect of increasing the perceptibility or salience of a sound with respect to otherwise similar sounds having a minus-specification, for example by increasing its loudness (a function of intensity), or making its formant structure more prominent. By defining the sonority scale in terms of several independent features rather than attempting to define it in terms of a single, uniform phonetic parameter, we take a significant step toward solving the problem of "defining" sonority in phonetic terms. Moreover, we are able to relate the notion "relative sonority" directly to perceptibility, since each of the acoustic attributes associated with a plus-specification for a major class feature enhances the overall perceptibility of the sounds that it characterizes. (See Stevens and Keyser (1987) for recent discussion of the acoustic correlates of distinctive features in somewhat similar terms.)

In sum, we may regard the major class features as defining the relative sonority of the various speech sounds in just this sense. Although the notion of relative sonority cannot be defined in terms of any single, uniform physical or perceptual property, we need not conclude that it is a fictitious or purely subjective matter, as long as we consider it a composite attribute of speech sounds, defined in terms of a set of major class features which themselves have relatively well-defined attributes.
5. The Sonority Cycle

Let us now consider the way sonority-based constraints are to be formulated in core phonology. I will propose a model involving two principles, which I will term the principles of *Core Syllabification* and *Feature Dispersion*. These two principles, taken together, implement the principle of the *Sonority Cycle*.

5.1. The Core Syllabification Principle

I will assume that there is a more or less well-defined portion of the lexical phonology characterized by certain uniform, perseverative properties. For example, in some languages the set of syllabification rules responsible for the syllabification of underlying representations reapplies to the output of each phonological and morphological operation throughout a portion of the lexical phonology; we may call these the *core syllabification rules*, after Clements and Keyser (1981, 1983). Similarly, by a principle of conservation some languages maintain a uniform phoneme inventory throughout much or all of the lexical phonology, an effect of “structure-preservation” which Kiparsky has proposed to account for in terms of *marking conditions* (Kiparsky 1985). Furthermore, in some languages, we observe constraints on segment sequences that hold both of nonderived stems and derived stems, giving rise to “conspiracy” effects that cannot be accounted for by syllabification principles alone (see Christdas (1988) for discussion of Tamil). The segmental and sequential uniformity characterizing these inner layers of the lexical phonology does not generally extend to the postlexical phonology, and does not necessarily even characterize the entire lexical phonology where violations of structure preservation (in the strict sense that precludes the introduction of novel segment types) are found in a number of languages (Clements 1987b). I will refer to the portion of the lexical phonology subject to such perseverative well-formedness conditions as the *core phonology*, and the syllabification rules that operate at this level the *core syllabification rules*.

Let us consider the nature of core syllabification more closely. As has widely been noted, syllables are normally characterized by a rise and fall in sonority which is reflected in the sonority scale values characterizing each of their segments. Sequences of syllables display a quasiperiodic rise and fall in sonority, each repeating portion of which may be termed a *sonority cycle*. It is possible to fit a curve or outline over such representations which reflects this rise and fall, as shown in (12), consisting of two cycles:
The number of cycles whose peaks fall on the top ([syllabic]) line of this diagram will correspond exactly to the number of syllables, except that a plateau along the top line (representing a sequence of vowels) may be parsed as a sequence of syllable peaks, as in many of the examples of (5c).

We may formulate a preliminary version of the Sonority Sequencing Principle in terms of this cyclic organization. It will be stated in terms of three steps or actions which are performed successively on segment strings to create syllables. The first of these searches for [+syllabic] segments as defined by the language in question, and introduces a syllable node over them (cf. Kahn 1976). This step presupposes that syllabic segments are already present in the representation at this point, whether created by rule or underlying (as is required in the case of languages that have unpredictable distinctions between vowels and glides or other segments differing only in syllability, as in French, discussed below). Further segments are syllabified by first adding segments to the left that have successively lower sonority values, and then doing the same for unsyllabified segments on the right. This yields the following principle of unmarked syllabification. I will call it the Core Syllabification Principle (CSP) for reasons that will become clear in the subsequent discussion.

(13) The Core Syllabification Principle (CSP):

a. Associate each [+syllabic] segment to a syllable node.

b. Given P (an unsyllabified segment) preceding Q (a syllabified segment),
   adjoin P to the syllable containing Q iff P has a lower sonority rank than Q.
   (iterative)
c. Given Q (a syllabified segment) followed by R (an unsyllabified segment),
adjoin R to the syllable containing Q iff R has a lower sonority rank than Q. (iterative)

The first iteration of (13b), which creates CV syllables, is not restricted by the sonority condition, since languages allowing syllabic consonants may permit segments of equal or higher sonority to be syllabified to their left: cf. English *yearn*, from underlying */yərn/* in which */t/* is syllabic, and similar examples in other languages (see Steriade (1982) for further observations on the special status of CV syllables). A further necessary qualification is that some languages place upper limits on the length of initial and/or final clusters created by (13), as in Turkish which does not permit syllable-initial consonant clusters in native words.14

The “left-precedence” or “onset-first” principle rendered explicit by the precedence given to (13b) over (13c) is widely observed in languages. Sievers (1881) had already noticed the widespread tendency toward syllabifications of the form V.CV, where C is a single consonant or a “permissible initial cluster”. This observation was generalized by later linguists as the Maximal Onset Principle, which states that intervocalic clusters are normally divided in such a way as to maximize syllable onsets (see Pulgram (1970), Bell (1977), Selkirk (1982), and others). This principle applies as a strong crosslinguistic tendency just as long as the result is consistent with the CSP and with any additional language-particular restrictions on syllable length or syllable composition.15

In accordance with this principle, *template* is syllabified as follows:

(14)

```
   +   +   +   +   +   +   +   +
  +   +   +   +   +   +   +   +
 +   +   +   +   +   +   +   +
 t   e   m   p   l   e   y   t
 σ   σ
```

syllabic
vocoid
approximant
sonorant
As a consequence of the Core Syllabification Principle, intervocalic clusters will be syllabified in such a way as to both maximize the length of syllable onsets and increase the difference in sonority between their first and last members. This follows from (13b), which due to its iterative nature will continue to adjoin consonants to the initial cluster as long as each new one added is lower in sonority than the previous one. A second consequence is that a syllable which is nonfinal in the domain of core syllabification will have a minimal decay in sonority, since less sonorous consonants to its right will normally have been syllabified into the following syllable by the prior application of (13b). Both points are illustrated in the syllabification of template, above, in which the Core Syllabification Principle requires $p$ to syllabify rightward rather than leftward, giving the first syllable a relatively small decay in sonority at its end and the second a relatively sharp rise at its beginning. A third consequence is that syllables which are final in the domain of core syllabification should tend to show a maximal decay in sonority, since they do not compete for consonants with a syllable to the right. The right margins of final syllables should thus tend to resemble the “mirror image” of the initial margins of initial syllables as far as their sonority profiles are concerned.

This prediction, though frequently true, is not exceptionless, however. In many languages the preferred syllable type is open, and closed syllables tend to be characterized by small rather than large drops in sonority, both medially and finally. When languages allow both sonorants and obstruents in final position, the variety of obstruents which can occur here is frequently smaller than the variety of permissible sonorants. It seems that the universally preferred syllable type tends to resemble the simple, open CV syllable as closely as possible; and a syllable approximates this type more closely to the extent that it declines less in sonority at its end. Thus a better characterization of the sonority cycle principle is that the preferred syllable type shows a sonority profile that rises maximally toward the peak and minimally towards the end, proceeding from left to right.

This principle expresses a valid crosslinguistic tendency, but does not exclude the presence of less preferred core syllable types in given languages. For example, many languages tolerate V-initial syllables, which begin with no rise in sonority at all. However, such syllables are normally restricted to word- or morpheme-initial position: in internal position, hiatus across syllable boundaries is very commonly eliminated in the core phonology by such processes as glide formation and vowel deletion. Similarly, many languages tolerate syllable types with abrupt drops in sonority at their end, or indeed
syllables that have fairly complex final clusters that do not obey sonority sequencing restrictions at all. However, such clusters are generally restricted to the edges of morphologically-defined domains. For example, English has a high tolerance for syllables ending in obstruent clusters (Dewey 1923, Roberts 1965). Within roots and level 1 stems, however, they are restricted to final position; internally, the inventory of syllable finals is much more restricted, and strongly favors sonorants over obstruents (Borowsky 1986). The suspension of normal sonority constraints in peripheral position in the domain of syllabification can be formally characterized in terms of the notion of extraprosodicity as governed by the Peripherality Condition (Hayes 1981, Harris 1983), but may have a deeper explanation in the observation that peripheral segments are not subject to competing syllable divisions, and thus cannot give rise to alternative syllable parsings.

Viewed in this way, the sonority cycle provides a rationale for the ordering of (13b) over (13c) in the statement of the Core Syllabification Principle. This ordering, as already noted, reinforces the tendency of syllables to show a gradual decay in sonority toward their end. We will see shortly that a more precise characterization of the notion of the sonority cycle, implemented in terms of the Dispersion Principle, allows a significant simplification of the Core Syllabification Principle.

The Core Syllabification Principle (henceforth, CSP) is defined within the domain of core syllabification, which is fixed on language-particular grounds. This domain is the morphologically-determined portion of a form to which the core syllabification rules apply. Within this domain, the core syllabification rules and principles apply recursively to the output of each phonological or morphological operation. Thus in German, the domain of core syllabification has been identified as the morpheme (Laeufer 1985), while in English, as just noted, it is most likely identical to the stem formed by the level 1 morphology.

As noted, the CSP operates only within the margin of freedom allowed by a particular language. Thus if a language does not allow initial clusters, an intervocalic cluster will usually be heterosyllabic, even if the second member of the cluster is higher in sonority than the first. Examples are Turkish and Klamath, whose syllable-sensitive phonologies always treat the first of two intervocalic consonants as closing the first syllable, regardless of the sonority profile of the cluster (Clements and Keyser 1983). Another type of constraint is illustrated in the Germanic languages, where it is widely observed that a short stressed vowel attracts a following consonant into its syllable, even when the resulting cluster violates the CSP (Murray and Venneman 1983, Laeufer 1985). This principle has a
counterpart in the English rule of Medial Ambisyllabification (Kahn 1976), which applies without regard to the general preferences expressed by the CSP. Further, many languages systematically syllabify vowels and glides together to form diphthongs, even when the following segment is a vowel. Thus in English, the glide \[y\] in *biology* \[bayˈələʤ\] is syllabified with the first syllable, not with the second, as is evidenced by the failure of the first vowel to reduce to schwa by Initial Destressing. These are common ways in which language-particular rules may take precedence over the CSP. These rules themselves, it should be noted, are not arbitrary but reflect independently-observed tendencies, such as the widespread dispreference for tautosyllabic clusters, or the preference for stressed syllables to be heavy.

5.2. The Dispersion Principle

The Core Syllabification Principle (CSP) expresses a generalization about the way sequences of segments are commonly organized into syllables. It classifies syllables into two types, those that conform to the CSP, and those that violate it by presenting sonority plateaus or sonority reversals. Most frequently, if a language has syllables that violate the CSP it also has syllables that conform to it. Accordingly we will call syllables that conform to the CSP “simple” or “unmarked” syllables and those that violate it “complex” or “marked” syllables.

Apart from the two-way distinction between unmarked and marked syllable types, the CSP does not have anything to say about the relative complexity of syllables. This topic is treated in this section. Our basic claim will be that syllables are simple just to the extent that they conform to the optimal syllable as defined by the sonority cycle. Thus, the simplest syllable is one with the maximal and most evenly-distributed rise in sonority at the beginning and the minimal drop in sonority (in the limit case, none at all) at the end. Syllables are increasingly complex to the extent that they depart from this preferred profile.

In order to characterize “degree of distance from the optimal syllable” in this sense, we will first define a measure of dispersion in sonority, and then formulate the Dispersion Principle in terms of it. This is the principle that will serve as the basis for ranking syllable types in terms of relative complexity. As stated here, it is defined only upon syllables that accord with the CSP, that is, those that show a steady rise in sonority from the margin to the peak; other (“marked”) types of syllables must be ranked by a separate method of evaluation involving an extension of the complexity metric to be given below.
In order to state the Dispersion Principle in the most revealing form, it proves convenient to make use of the *demisyllable*, a notion drawn from the work of Fujimura and his collaborators. I will begin by defining this term as it is used here. A syllable is divided into two overlapping parts in which the syllable peak belongs to both; each of these parts is termed a demisyllable. In the case of syllables beginning or ending in short vowels, one demisyllable is the short vowel itself. Thus, for example, the syllable [kran] consists of the demisyllables [kra, an], the syllable [spawl] of [spa, awl], the syllable [pa] of [pa, a], the syllable [ap] of [a, ap], and so forth.

The demisyllable can be defined more formally as follows:

(15) A *demisyllable* is a maximal sequence of tautosyllabic segments of the form \( C_m ... C_n \) or \( VC_m ... C_n \), where \( n \geq m \geq 0 \).

The idea underlying the use of the demisyllable is that the sonority profile of the first part of the syllable is independent of the sonority profile of the second part. That is, there are no dependencies holding between the two parts of the syllable as far as sonority is concerned. Thus the attribute “dispersion in sonority” is most appropriately defined over the demisyllable.

If we now restate the principle of the sonority cycle in terms of demisyllables, and consider only “unmarked” demisyllables (those that conform to the CSP), we will say that the *initial* demisyllable maximizes the contrast in sonority among its members, while the *final* demisyllable minimizes it. The contrast in sonority between any two segments in a demisyllable can be stated, as a first approximation, as an integer \( d \) designating the distance in sonority rank between them. For example, given the sonority scale \( O < N < L < G < V \) the distance in sonority rank between \( N \) and \( V \) is 3, regardless of their relative position in a demisyllable.

The notion “dispersion in sonority” can be stated in terms of a measure of dispersion, \( D \), of the distances in sonority rank \( d \) between the various pairs of segments within a demisyllable. \( D \) characterizes demisyllables in terms of the extent to which the sonority distances between each pair of segments is maximized: the value for \( D \) is lower to the extent that sonority distances are maximal and evenly distributed, and higher to the extent that they are less maximal or less evenly distributed. It can be defined by the following equation,
which is used in physics in the computation of forces in potential fields, and which has also been used by Liljencrants and Lindblom (1972) to characterize the perceptual distance between vowels in a vowel system.

\[
D = \sum_{i=1}^{m} \frac{1}{d_{ij}^2}
\]

Here, \(d\) is the distance in sonority rank between each \(i\)th pair of segments in the demisyllable (including all nonadjacent pairs), and \(m\) is the number of pairs in the demisyllable, equal to \(n(n - 1)/2\), where \(n\) is the number of segments. It states that the \(D\), the dispersion in sonority within a demisyllable, varies according to the sum of the inverse of the squared values of the sonority distances between the members of each pair of segments within it.

Assuming the sonority scale in (6), this gives the following values of \(D\) for simple CV and VC demisyllables:

\[
\begin{align*}
(17) \quad \text{OV, VO} & = 0.06 \\
\text{NV, VN} & = 0.11 \\
\text{LV, VL} & = 0.25 \\
\text{GV, VG} & = 1.00
\end{align*}
\]

For CCV and VCC demisyllables, we have the following:

\[
\begin{align*}
(18) \quad \text{OLV, VLO} & = 0.56 \\
\text{ONV, VGO, OGV, VNO} & = 1.17 \\
\text{NLV, VGN, NGV, VLN} & = 1.36 \\
\text{LGV, VGL} & = 2.25
\end{align*}
\]

We observe that initial demisyllables with low values for \(D\) are those that show an optimal sonority profile, i.e. a sharp and steady rise in sonority, while in the case of final demisyllables those with high values for \(D\) show the best profile, i.e. a gradual drop in sonority. We may accordingly state the Dispersion Principle as follows:
(19) *Dispersion Principle*:
   a. The preferred initial demisyllable minimizes $D$
   b. The preferred final demisyllable maximizes $D$

It can be noted in passing that other ways of defining the value of $D$ are possible in principle. For example, it might be more appropriate to restate (16) over the sum of sonority distances for adjacent pairs of segments only. As it happens, this version of (16) gives only slightly different values of $D$, since the value of $d^2$ is always very small for nonadjacent pairs, and proves to yield no differences in actual demisyllable rankings. Other possible versions, involving some simple summation of the distance between members instead of the inverse of the square, prove not to yield the desired complexity rankings, and need not be discussed here.

We may now define a Complexity Metric making use of the Dispersion Principle as stated in (19). This metric defines complexity rankings in terms of values of $D$, and states separate conditions for initial and final demisyllables.

(20) *Complexity Metric*: For any demisyllable of length $l$,
   a. the complexity ranking, $C$, of an initial demisyllable increases as its ranking in terms of $D$ increases
   b. the complexity ranking, $C$, of a final demisyllable increases as its ranking in terms of $D$ decreases

In the case of initial demisyllables of a given length, this metric will assign the rank 1 to the demisyllable with the lowest value of $D$, the rank 2 to the next highest, and so forth. The demisyllable OV, for example, has the lowest value for $D$, and therefore the lowest complexity rank (1); NV has the second highest value for $D$, and thus the second lowest complexity rank (2); and so forth.

Two-member demisyllables fall into four degrees of complexity, as do three-member demisyllables. Complexity rankings for two-and three-member demisyllables are shown in Table 1. It should be noticed that $C$ is not proportional to $D$ itself, but rather to the *ranking* defined by $D$.

(insert Table 1 about here)
(20) does not assign a value for C to one-member demisyllables (V), which nevertheless vary in complexity according to whether they constitute initial or final demisyllables just as longer ones do. We will therefore extend our measure of complexity in a natural way to account for these. An initial one-member demisyllable V must be regarded as highly complex as it fails to show any rise in sonority whatsoever. It must therefore be regarded as more complex than the most complex two-member initial demisyllable GV, which shows a slight (i.e. one step) rise in sonority. Since GV has a complexity rank of 4, we will assign the initial demisyllable V a complexity rank of 5. A final one-member demisyllable V, on the other hand, must be regarded as relatively simple since it conforms exactly to the pattern of the optimal CV syllable, showing no decline in sonority at all. We will give this demisyllable the complexity rank of 0, one step lower than the next most favored final demisyllable, VG. Thus we have the additional rankings in Table 2, which rank one-member demisyllables with respect to two-member demisyllables.

(insert Table 2 about here)

Four-member demisyllables fall into one of three complexity ranks, as shown in Table 3. The longest demisyllables that can be evaluated by this procedure, assuming the scale \( O < N < L < G < V \), are the singleton five-member sets, ONLGV and VGLNO, for which \( D = 5.03 \).

(insert Table 3 about here)

The same system extends to demisyllables with syllabic consonants as peaks. Recall that all syllabic consonants have a sonority ranking of 1 more than their nonsyllabic counterparts, as was shown in (8). Thus for the case of demisyllables of length 2, for instance, we have the rankings in Table 4. Notice that in this case, corresponding initial and final demisyllables must have the same relative rankings; the Complexity Metric (20b) must accordingly be restricted to demisyllables whose peaks are of the same type.\textsuperscript{20}

(insert Table 4 about here)
As noted earlier, this result is not well supported in one respect: syllables with syllabic nasals are more frequently reported across languages than syllables with syllabic liquids. We leave the question open whether this unexpected frequency should be regarded as reflecting on the relative complexity (markedness) of these two types, or some other factor.

The complexity rankings in Tables 1-4 define a hierarchy over demisyllables. We may now state the following implications for core phonology, which hold at the level resulting from initial syllabification, which I will call L(IS). The implications are stated only over demisyllables of the same type, where the type of a demisyllable depends on (i) whether it is an initial or final demisyllable, (ii) whether its peak is a vowel or consonant. In addition, the Complexity Hierarchy is stated only over demisyllables of the same length (where V-demisyllables count as if they were of length 2, for the reasons explained just above). A separate Length Hierarchy is stated over demisyllables of different lengths. These two statements together form the Complexity and Length Hierarchies, stated in (21):

(21) a. The Complexity Hierarchy:
For any given type t and length l, the presence in L(IS) of a demisyllable of complexity rank n implies the presence of a demisyllable of complexity rank n-1.

b. The Length Hierarchy:
For any given type t, the presence in L(IS) of a demisyllable of length l (l > 2) implies the presence of a demisyllable of length l-1.

By the Length Hierarchy (21b), for example, the presence of a CCV demisyllable in L(IS) implies the presence of CV, and so forth for longer demisyllables. The Length Hierarchy does not project a ranking for V-demisyllables, since as just mentioned these count as representing length 2; instead, V-demisyllables are ranked with respect to others by (21a), which treats a final V-demisyllable as simpler than any VC demisyllable, and an initial V-demisyllable as more complex than any CV demisyllable, by Table 2. Notice that it is only by placing V-demisyllables under the scope of the Complexity Hierarchy in this way that we are to offer a principled account of the fact that V-demisyllables are more complex than CV initial demisyllables, but simpler than VC final demisyllables (rather than, for example, the contrary); if we were to rank them under the Length Hierarchy instead, this asymmetry in behavior would have to be accounted for in terms of an arbitrary stipulation.
These principles allows us to characterize a language as more or less complex according to the following properties of demisyllables occurring at L(IS):

(22) a. the maximal value of $n$ in (21a);
   b. the maximal value of $l$ in (21b);
   c. the presence of “marked” demisyllables (those violating the CSP).

The Complexity/Length Hierarchy (21) represents a claim about the organization of phonological systems at the level of core syllabification. It maintains that core syllabification rules do not create complex syllable types unless they create the more simple syllable types. Surface exceptions to (21) arise as a result of segmental rules creating new cluster types and later syllabification rules applying after the level of core phonology. Both types of surface exception can be illustrated from French.

In French, we find surface syllables of several types. In the first place, we find the unmarked demisyllable types OLV (drap ‘sheet’, vrai ‘true’), OGV (dieu ‘god’, chouette ‘owl’), NGV (mieux ‘best’, nuage ‘cloud’), and LGV (rien ‘nothing’, lieu ‘place’, rouan ‘roan’, lui ‘him’), as well as a full range of CV demisyllables. In the second place, we find demisyllable types such as OOV (style ‘style’, sphère ‘sphere’, psychose ‘psychosis’) and more rarely ONV, NNV (pneu ‘tire’, mnémonique ‘mnemonic’); in addition we find a few s-initial CCCV demisyllables, such as spleen ‘spleen’, strict ‘strict’. The second group can be identified as nonbasic syllable types due to the fact that they are restricted to initial position in the syllabification domain: thus we do not find tautosyllabic sc clusters internally in morphemes and simple stems (Lowenstamm 1981). We need consider only the first set, therefore, all of which may occur word-internally as well as word-initially and which are accordingly good candidates for core syllables at the level L(IS).

Among unmarked demisyllables of length 3, then, we find OLV and three types of CGV syllables: OGV, NGV, and LGV. Missing are ONV and NLV. Since the presence of LGV (of complexity rank 4) implies the presence of ONV and NLV (of complexity ranks 2 and 3, respectively), we have an apparent violation of (21). It remains to determine, however, whether CGV demisyllables are actually present in L(IS).

Glides and vowels are underlyingly contrastive in French, but this contrast is restricted to words like abbaye [abei] ‘abbey’, abeille [abey] ‘bee’, where the glide is final; we find no comparable contrasts in prevocalic position. For example, we find cahot [kao] ‘jolt’ and
caillot [kaio] 'clot' but no contrastive word [kaio]. Surface GV syllables ordinarily derive from underlying VV sequences, since such syllables behave as vowel-initial with respect to rules that distinguish consonants and vowels. Thus we find les [lez] amis 'the friends' contrasting with les [le] copains 'the pals', illustrating the fact that the final [z] of les is deleted before consonants; [z] is retained, however, before the surface glide [y] in les yeux [lez yø] showing that this must be a vowel at the time z-deletion applies (see Clements and Keyser 1983,96-99 for fuller discussion). We conclude that GV syllables do not occur at the level L(IS) and therefore that initial demisyllables of length 2 are restricted to a maximum complexity of 3 at this level. (Note, however, that a small number of loanwords allow initial underlying glides, such as yod 'yod', whisky 'whisky'; cf. les [le] yods, les [le] whisky).

By the principle of resolvability (Greenberg 1978, 250; Clements and Keyser 1983, 47-8), the presence of a tautosyllabic cluster ABC implies the independent occurrence of tautosyllabic AB and BC. If CGV demisyllables were created by core syllable rules at L(IS), we would have a violation of this widely-observed principle, since as just shown GV does not occur independently at this level. As CGV syllables do not contrast with CVV syllables, however, we may eliminate them from the level of initial syllabification L(IS) and derive them from the CVV demisyllables by a rule of Glide Formation, which turns high vowels into glides before vowels. This rule accounts not only for the presence of (C)GV in monosyllabic roots, but for alternations such as manie [mani] 'I handle' vs. manier [manye] 'to handle', or avoue [avu] 'he admits' vs. avouer [avwe] 'to admit' (see e.g. Dell 1980, Noske 1982).

This leads us to the following analysis of core demisyllables in French. The maximal complexity for initial demisyllables of length 2 is 3 (with a few word-initial exceptions in nonnative words, as mentioned) and the maximal complexity for demisyllables of length 3 is 1, the default value for this case. Thus initial syllabification creates only OV, NV, LV, and OLV, consistently with (21). CGV demisyllables arise through the rule of Glide Formation, which applies obligatorily in initial and postvocalic position and optionally postconsonantally. For some, but not all speakers it is also obligatory when defined entirely within a single morpheme (for such speakers lieu 'place' is always [lyø], never [liø]). The output of Glide Formation is fully syllabified, but respects the length constraints which continue to operate through the core phonology: thus it cannot create CCGV demisyllables, and is blocked in words like plier 'bend', crier 'cry', and grief
'grievance', which remain bisyllabic. Interestingly, for some speakers Glide Formation can apply in s-initial words like skier [skye] 'to ski'. We may assume that for these speakers s-initial clusters are created by a post-core rule syllabifying initial s with a following consonant; at this point the core syllable constraints are no longer operative. For other speakers, this rule belongs to the core phonology.

We see, then, that surface exceptions to (21) may not be exceptions at the level of initial syllabification, at which (21) is defined. In French, surface exceptions arise in two ways: through the creation of new sequence types by the operation of Glide Formation in the core phonology, which are resyllabified subject to the length restrictions, and through the creation of new syllable types (such as s-initial clusters) by syllabification rules applying subsequently in the derivation, perhaps in the post-core phonology. This analysis directly captures the generalization that the length condition on the output of Glide Formation is identical to the length condition holding on underlying syllables. More generally, it supports our claim that sonority constraints are most suitably defined in core phonology, rather than in surface structure (section 3.1).

Let us turn finally to the status of ‘marked’ demisyllables containing violations of the sonority cycle, in the form of sonority ‘plateaus’ or ‘reversals’ such as OOV, NOV, or LOV. Such structures are not uncommon at the surface-phonetic level in languages, as we saw in section 3.1, and may arise through core or post-core syllabification processes as we have just seen in French. The essential observation here is that such sonority violations are usually restricted to the periphery of the syllabification domain, where they do not give rise to problems of syllable division. A language that exhibits LOV demisyllables word-initially, for instance, does not usually tolerate them word-internally, just as languages that permit VOO demisyllables word-finally do not usually allow them in non-final syllables.

This observation does not require any new principles, since it follows directly from the CSP. Word-initially, the CSP syllabifies a sequence like LOV as L-OV, where the L remains extrasyllabic. Some languages may then have special rules allowing this segment to be incorporated into the syllable, while others may require it to be deleted. Word-internally, however, the sequence VLOV will be syllabified VL-OV by the CSP; this will be true even if the language in question has a rule creating LOV demisyllables word-initially, under the assumption that such special rules apply after rules that implement the CSP. Therefore LOV demisyllables will not be created word-externally, except in the
highly unusual case in which a language has a rule overriding the CSP in just this context, for example by carrying out a resyllabification.

There is a straightforward way to determine the relative complexity of demisyllables containing sonority "plateaus" and "reversals", which have not so far been integrated into the evaluation system. The basic observation is that the deviance of "marked" demisyllables is proportional to their distance from "unmarked" demisyllables. Thus sonority reversals are more complex than sonority plateaus, and the complexity of sonority reversals increases in proportion to the extent of the reversal: e.g. NOV is more complex than OOV, LOV than NOV, GOV than LOV, etc. We may assume an appropriate extension of the Complexity Metric to cover these cases.

This section has proposed a formal procedure for determining the relative complexity of demisyllables of various types -- and hence (though derivatively) of syllables. We need not attribute such computation to the explicit knowledge of native speakers in any sense. Rather, the relationships we have sought to bring out are properties of the representations as such, and can presumably be apprehended by speakers without carrying out conscious mathematical calculations - just as we can detect whether billiard balls are evenly dispersed on a billiard table without doing computations on a pocket calculator. The procedures do no more than elicit relationships that are already inherent in linguistic representations themselves, and render them accessible to study.

5.3. The Sequential Markedness Principle

Certain sequencing constraints holding within syllables cannot be accounted for by the theory developed so far. Let us consider cases in which place of articulation seems to play a role.

Greenberg (1978) observes what he terms the "law of the final dental-alveolar", which he formulates as follows: "every language [in the sample] with final clusters contains at least one cluster with a final obstruent in the dental-alveolar region" (p. 268). That is, if a language allows VCC demisyllables to occur in final position, at least one of these is of the form VCT, where T represents a dental or alveolar obstruent. Examples given by Greenberg include Classical Greek, with the three final clusters ps, ks and nks, Latin whose final clusters all end in s or t, Balti with ks, rs, ns, and Maasai with only rn, rt, and rd. A similar implication holds of initial demisyllables: as Greenberg notes, "every
language [in the sample] with initial clusters contains at least one cluster with an initial consonant in the dental-alveolar region” (269). As an example he cites Chiricahua Apache, in which the only initial clusters are st and sd.

Some linguists have suggested, on the basis of observations similar to these, that coronal segments should be assigned a special rank of their own on the sonority scale. This would allow coronals to be formally treated as different in sonority from segments formed at other places of articulation. Closer consideration shows, however, that this approach weakens the notion of sonority to an undesirable degree, and does not fully explain the special status of anterior coronals (dentals and alveolars) as opposed to posterior coronals (palato-alveolars).

One reason not to assign coronals a special place of their own on the sonority scale is that the distinction between coronal and noncoronal segments of the same major class does not correspond to any consistent difference in perceptibility, unlike the major class features which define the scale given in (6). For example, [s] is a more salient segment than [f] or [θ] in terms of intensity and loudness, and is thus presumably more sonorous in this view, but nevertheless occurs peripherally to [p] and [k] in initial clusters like English spit, skit and in final clusters like lapse, tax where the theory requires it to be less sonorous. Nor, in particular, does such an approach help explain why [s] and [z] normally occur peripherally to fricatives at other places of articulation, as in English sphere, Jeeves or Dutch school [sxoːl], aardigst [...xst] ‘nicest’. If we are to maintain that coronals are less sonorant than noncoronals on the basis of the patterning of [s], we must abandon the claim that sonority is related to increased perceptibility, which seems otherwise correct.

Moreover, it is difficult to find any general position for coronals that would give a correct general account of their exceptional freedom of occurrence. On the one hand, to handle initial clusters like sp, sf, sk, sx or final clusters like pt, kt, ps, fs, ks, xs we would have to assign the coronals a lower sonority rank than noncoronals, as we have just seen. But there are considerations arguing for just the opposite analysis. As Steriade (1982) observes, we may account for the common exclusion of the initial clusters tl, dl in languages otherwise permitting OL clusters freely by the minimal distance principle if we claim that t, d have a higher sonority rank than p, b, k, g: under this assumption, t,d are “closer” in sonority to l than are the noncoronal stops, and hence can be excluded by a minimal distance constraint. And as Selkirk notes (1984), we must assign coronals a higher rank than noncoronals to account for languages such as Spanish and Italian in which
only sonorants and $s$ may close the syllable, in order to designate this set of segments as a natural class on the sonority scale. But such inconsistency in the place of coronals argues strongly against this approach.

Furthermore, it seems that one and the same language may treat coronals inconsistently. In English, as we have seen, coronals typically pattern peripherally to noncoronals in initial and final obstruent clusters, a fact which suggests that they have a lower rank on the sonority scale. This is supported by Stuart Milliken’s observation (personal communication) that in single morphemes, an obstruent may follow an oral stop only if it is a coronal, regardless of whether the cluster is intervocalic or final: thus we have intervocalic clusters such as those in chapter, capsule, abdomen, pretzel, factor, and pixel beside final clusters such as those in rapt, lapse, ritz, fact, tax, but no words with stop-initial clusters ending in noncoronals like chaper, rap, etc. 21 If we regard coronals as ranking lower in sonority than noncoronals, this will follow from the CSP and the Syllable Contact Law (section 2). On the other hand, other facts in English argue that coronals are higher ranking than noncoronals. For example, $[s]$ is the only fricative that can precede a noncoronal oral stop in a morpheme: whisper, whisker, lisp, risk but *whisper, whiskee, lifp, rifk; this can be explained under the Syllable Contact Law and the Minimal Distance Principle only if it is higher in sonority than the noncoronal fricative $f$. Moreover, English shares the property of many languages mentioned above, according to which $t, d$ are excluded in initial clusters before $l$; as pointed out, this also follows from the Minimal Distance Principle if coronals are higher ranking than noncoronals.

For these reasons it seems undesirable to introduce further subdivisions of the sonority scale to accommodate distinctions in place of articulation. Examining the relevant facts more closely, it seems that another principle may be better able to provide an explanation. The observations made so far show that in initial and final clusters, anterior coronals have a freer privilege of occurrence than other consonants do; they are often the only segments able to occur as the first or second member of clusters. It would seem reasonable to relate this fact to an independent property of anterior coronals, which is that they are formed at the least marked place of articulation, by most markedness criteria (see Stevens and Keyser 1987 for recent discussion). The complexity of any sequence of segments can be considered a function both of its length and of the individual segments that compose it. Thus although a two-member cluster is more complex than a one-member cluster, it is less complex if contains an anterior coronal than if it contains some other consonant, all else
being equal. Normal markedness principles, therefore, lead us to expect exactly the pattern of preference for anterior coronals that we have observed.

We can make this observation explicit in terms the following principle, which presumably does not need to be stated as an axiom in grammatical theory as it should follow from an adequate, completely elaborated theory of markedness:

\[(23) \textit{Sequential Markedness Principle:}\]

For any two segments A and B and any given context X__Y, if A is simpler than B, then XAY is simpler than XBY.

Thus \(pt\) is simpler than \(pk\) by virtue of the fact that \(t\) is simpler than \(k\), and so forth. This principle extends to most of the observations we have made above. Clearly, however, its scope is much broader. Beside explaining the preference for \(sk\) over \(fk\), for example, it also explains the general preference for clusters containing \(s\) as opposed to all other coronal fricatives: \(s\) is the least marked fricative. This explains why initial clusters in English include \(sn\), \(sm\) but not \(fn\), \(fm\), \(θm\), \(θn\), and only marginally \(šn\), \(šm\). Since voiceless fricatives are less marked than voiced fricatives, it also explains the presence of initial \(fr\), \(fl\), \(sn\) beside the absence of \(vr\), \(vl\), \(zn\), etc. Thus we need a principle like (23) in any case. But once we have it, it accounts for the preference for dentals and alveolars without the need for further elaboration of the theory of sonority.\(^{22}\)

6. Theoretical Results

Let us consider some of the general results and crosslinguistic predictions of our approach to sonority. These are taken up under five headings: 6.1 Sonority Sequencing Restrictions; 6.2 the Maximal Onset Principle; 6.3 Minimal Distance Constraints; 6.4 the Syllable Contact Law; and 6.5. Core Syllable Typology.

6.1. Sonority Sequencing Restrictions

As has already been pointed out, the account of sonority given above is based in the first instance on crosslinguistic generalizations of the sort noted by Greenberg (1978). These generalizations strongly support the sonority scale \(O < N < L < G < V\). I summarize Greenberg's main results in (24)-(26), below.\(^{23}\) These examples consist of
implicational statements of the general form, "if a language has property A, it also has property B". We can symbolize such statements by means of expressions of the form "A → B," or "A implies B." Statements of this type are often understood as providing an indication of the relative markedness of the two properties in question, the unmarked (or less marked) value appearing to the right of the arrow. I present Greenberg's results under three general headings, which subsume Greenberg's implicational statements. (These headings are my own, not Greenberg's.)

Under (24) I have grouped a number of implications supporting the proposition that the unmarked order of segment types within an initial demisyllable is ONLGV, and within a final demisyllable VGLNO. This proposition follows from the the sonority scale O < N < L < G < V and from the CSP (13), which plays the important role of distinguishing between "unmarked" and "marked" demisyllable types. For example, since "marked" LOV demisyllables are not formed by the CSP, they require the complexity of an extra syllabification rule, and are furthermore ranked as several degrees more complex than "unmarked" demisyllables such as OLV created by the CSP, by the extension of the complexity metric suggested at the end of section 5.2. As (24) shows, Greenberg's results strongly support this proposition, in the sense that nine of his statements are entailed by it and none are inconsistent with it. (I retain Greenberg's numbering.)

(24) The unmarked order of segment types in a demisyllable is ONLGV or VGLNO:

(17) LOV → OLV
(18) VOL → VLO
(19) GOV → GOV; VOG → VGO
(24) LNV → NLV
(25) VNL, VNN, VLL → VLN
(36') VNN → VNO

Under (25) I have grouped further statements relating to the proposition that segments within the initial demisyllable tend to be equally and maximally distributed in sonority. Two implicational statements are entailed by this proposition and none contradict it. This proposition does not follow from the Core Syllabification Principle itself: in particular, the three demisyllable types mentioned in these statements are equally consistent with this principle, which establishes no ranking among them. This proposition does, on the other hand, follow from the Dispersion Principle (19), as we have already seen.
(25) Segments within the initial demisyllable tend to be equally and maximally distributed in sonority:

(33) NLV → OLV

(37) ONV → OLV

The converse of this proposition, that segments in final demisyllables tend to be minimally or unequally distributed in sonority, is contradicted by one statement, (34), according to which VLN → VLO. How is this discrepancy to be explained? We have already observed the common operation of rules that append extrasyllabic segments to the ends of syllabification domains. Such rules commonly create highly marked clusters, as in English, German, and many of the languages surveyed by Greenberg in which initial and final obstruent clusters occur with fairly high frequency. Segments appended by these rules are often coronal obstruents, a fact which reflects the unmarked status of these segments (cf. the Sequential Markedness Principle), as well as the fact that obstruents are just those segments that will least often create violations of the CSP when appended to the beginning or end of a domain. Since these properties follow from our other principles, however, it is not necessary to introduce new principles to deal with them. We therefore consider the preference for VLO demisyllables over VLN demisyllables in word-final position as reflecting the operation of rules that append extrasyllabic segments, preferentially obstruents, to the ends of domains. We expect this preference not to obtain in nonfinal syllables.

Finally, in (26) I give a number of statements supporting the view that fricatives and stops should be considered equal in rank, just as is claimed by the scale O < N < L < G < V. In these statements, “S” is to be read “stop” and “F” as “fricative”. The general proposition supported by Greenberg’s results here is that sequences differing in their specification for [continuant] are preferred to sequences agreeing in this specification.

(26) Contrast in continuancy is favored over its absence:

(7) SSV → FSV, SFV

(8) VSS → VFS, VSF

(9) FFV → FSV, SFV

(10) VFF → VFS, VSF
This same principle may be able to account for the widely-observed preference for
demisyllables of the form [rV] or [drV] over demisyllables of the form [t1V] or [dlV]. In
the first of these, the consonant cluster contrasts in terms of the feature [continuant], while
in the second it does not, assuming the correctness of our earlier assumption that laterals
are [-continuant].

Interestingly, Greenberg's results do not support the common view that voiced
obstruents outrank voiceless obstruents in sonority. The reason for this is that obstruent
clusters show a strong tendency to share all laryngeal features, including voicing.

We see, then, that the principles developed here account correctly for the crosslinguistic
generalizations noted by Greenberg, including certain ones (25) that did not follow from
earlier versions of sonority theory. In addition, they make further predictions regarding
preferred segment order that cannot be directly confirmed on the basis of Greenberg's
study (which did not attempt to evaluate all possible orderings of the segment classes
O,N,L,G,V), and which must be the subject of future research.

6.2. The Maximal Onset Principle

This approach also allows us to derive a further generalization, the Maximal Onset
Principle. In section 5.1, the Maximal Onset Principle was stipulated as part of the
statement of the Core Syllabification Principle, by giving statement (13b) precedence over
(13c). We observed that this order of precedence was in accordance with the properties of
the sonority cycle, but we were unable at that point to derive it from any higher-level
principle.

Suppose now we generalize the Complexity Metric given earlier in (20) so that it
compares demisyllables of different lengths as well as those of the same length (that is, we
delete the phrase "for any demisyllable of length l" in (20)). If we do this, the system of
evaluation developed in section 5.2 will select syllabifications that are in accordance with
the Maximal Onset Principle over syllabifications that violate it. As a result, the principle
will now follow directly from the Complexity Metric.

We may see this best by considering some examples. The simplest case of the Maximal
Onset Principle requires that VCV sequences should be syllabified as V-CV instead of VC-
V. This already follows from the Complexity Metric, since for any value of C, V is a
simpler final demisyllable than VC, and CV is a simpler initial demisyllable than V (see Tables 1 and 2). The generalized version of the Complexity Metric will derive the Maximal Onset Principle in the case of prototypical VCCV clusters. For example, V-OLV is selected over VO-LV, since V is a simpler final demisyllable than VO, and OLV is a simpler initial demisyllable than LV. Accordingly, it is no longer necessary to give (13b) explicit precedence over (13c). The CSP can be restated as follows:

(27) Core Syllabification Principle (revised):
   a. Associate each [+syllabic] segment to a syllable node
   b. Given P (an unsyllabified segment) adjacent to Q (a syllabified segment), if P is lower in sonority rank than Q, adjoin it to the syllable containing Q. (iterative)

We may allow syllabification to take place simultaneously rather than directionally, with the Complexity Metric deciding between otherwise well-formed alternatives.

This theory makes slightly different predictions from the earlier one in certain respects. For one, it does not select unambiguously between certain syllabifications, such as VN-GV versus V-NGV, both of which have an aggregate complexity of 4 by Table 1. Notice, however, that for a language to provide a test case it must allow both VN and NGV demisyllables in the first place; but languages of this type are not common, and it is not clear whether there is a crosslinguistic preference for either of the alternative syllabifications. The theory extends unambiguously (and apparently correctly) to the prototypical cases of VCCV syllabification, however, and we need not regard its failure to make predictions in marginally attested cases as a shortcoming.

More significantly, the present theory differs from standard versions of the Maximal Onset Principle in being defined in terms of the universal sonority scale rather than in terms of language-particular syllabification rules. This means that in cases of alternative syllabifications, the correct one will normally conform to the universal sonority scale, under the present account. Again, evidence bearing on this claim is hard to find, but some recent studies suggest that it may be correct. Hayes (1981) finds that intervocalic sequences of [s] + oral stop in English tend to syllabify as VC-CV in spite of the fact that we find [s] + oral stop clusters word-initially (see however Davidsen-Neilsen (1974) for contrary results), and Lowenstamm (1981) reports similar observations for French.
6.3. Minimal Distance Constraints

A further result concerns what we may term "minimal distance constraints." It has been noticed by a number of linguists that not all syllables that are well-formed in terms of the Sonority Sequencing Principle actually constitute well-formed syllables in given languages. Some languages show a strong preference for syllables in which adjacent elements are not too close to each other in sonority rank. For example, Harris (1983) observes that in Spanish, initial clusters of the form ON and NL are systematically excluded, while those of the form OL are allowed. He suggests that this is not an arbitrary property of Spanish, but reflects a tendency for languages to prefer syllables in which adjacent elements are separated by a specificable minimal distance on the sonority scale. As he further points out, if the sonority scale for Spanish is taken to be \( O < N < L < G < V \), then we may say that Spanish requires adjacent consonants in the same syllable to be nonadjacent, i.e. to observe a minimal distance of 2, on the sonority scale.

To the extent that statements of this sort prove to be simple and uniform across languages, they can be taken as providing further confirmation for the essential correctness of sonority theory, without which these statements cannot be easily expressed. However, a number of observations suggest that this principle needs some qualification. First, minimal distance constraints only seem to apply in the initial demisyllable; they typically do not govern final demisyllables, where segments tend to be close to each other in sonority, as we have seen. Furthermore, to the extent that it has been applied to a wider set of languages, this principle turns out to require increasingly idiosyncratic, language-particular versions of the sonority hierarchy in order to be made to work (see Steriade (1982), Harris (1983), Selkirk (1984), van der Hulst (1984) and Borowsky (1986) for discussion of minimal distance constraints in a variety of languages involving several different sonority scales). While the notion that segments within the syllable should not be too similar in terms of their sonority rank undoubtedly offers valid insights into syllable structure, its formalization in terms of minimal distance constraints may not be the most satisfactory way of capturing these intuitions.

The approach given here derives the main effects of minimal distance constraints without raising these problems. To see this, let us consider an example. Spanish, as noted, requires that consonants in initial clusters observe a minimal distance of 2 on the sonority scale. In a theory formally incorporating minimal distance constraints, such
statements are part of the grammar, and the minimal distance value governing syllabification in each language must correspondingly be discovered by each language learner. Under such an account, the simplest possible language would be one with no minimal distance constraints at all. But this seems incorrect: minimal distance constraints appear to be quite widely observed across languages, and seem to represent the unmarked option. If this is so, we would prefer an account in which such constraints need not be stated explicitly in the grammar, but can be derived from independent principles.

Under the present theory, such an account is possible. We may describe a language such as Spanish by saying that its initial CCV demisyllables have a maximal complexity of 1. Thus the only permitted initial demisyllables are of the form OLV (see Table 1). If we assume that 1 is the default value in universal grammar, this value does not have to be learned. We can account for more complex cases (those of languages which tend not to observe minimal distance constraints) by assuming that the learner abandons the default hypothesis only in the face of clear evidence to the contrary. For example, if a language allows ONV or OGV demisyllables in addition to OLV demisyllables, the value of the demisyllable rises from 1 to 2, and the learner abandons the null hypothesis. This result follows as a consequence of the principles given earlier, and provides a straightforward account of the valid empirical core of the notion of “minimal distance,” while accounting for the skewing between initial and final demisyllables. A further result is that by stating the Dispersion Principle over demisyllables rather than over consonant clusters (as in earlier approaches), we are able to bring the syllable peak into the domain of our statements, and account for the general preference for LV demisyllables over GV demisyllables.

6.4. The Syllable Contact Law

The theory presented above not only derives the effects of the Sonority Sequencing Principle intrasyllabically, it also derives the Syllable Contact Law transsyllabically. This principle, it will be recalled, holds that the preferred contact between two consecutive syllables is one in which the end of the first syllable is higher in sonority than the beginning of the second. In an extended version of this principle, Murray and Vennemann (1983) propose that the optimality of two adjacent, heterosyllabic segments increases in proportion to the extent that the first outranks the second in sonority. In this view, a sequence such as am.la, for example, constitutes a lesser violation than a sequence such as at.ya. Their version of the principle is paraphrased in (28):
(28) *The Extended Syllable Contact Law* (after Murray and Vennemann 1983, 520):

The preference for a syllabic structure $A \ S B$, where $A$ and $B$ are segments and $a$ and $b$ are the sonority values of $A$ and $B$ respectively, increases with the value of $a$ minus $b$.

This statement extends the Syllable Contact Law to syllable contacts of all types, including V$S$C. The consequence is that sequences like *at-a* exemplify the worst possible syllable contact and *a-ta* the best. This fully general version of the principle gives us the following implicational ranking of syllable contacts, in which the contact types improve as we proceed upward and rightward across the table:

(29)

<table>
<thead>
<tr>
<th></th>
<th>V</th>
<th>G</th>
<th>L</th>
<th>N</th>
<th>O</th>
</tr>
</thead>
<tbody>
<tr>
<td>V</td>
<td>V.V</td>
<td>V.G</td>
<td>V.L</td>
<td>V.N</td>
<td>V.O</td>
</tr>
<tr>
<td>G</td>
<td>G.V</td>
<td>G.G</td>
<td>G.L</td>
<td>G.N</td>
<td>G.O</td>
</tr>
<tr>
<td>L</td>
<td>L.V</td>
<td>L.G</td>
<td>L.L</td>
<td>L.N</td>
<td>L.O</td>
</tr>
<tr>
<td>N</td>
<td>N.V</td>
<td>N.G</td>
<td>N.L</td>
<td>N.N</td>
<td>N.O</td>
</tr>
<tr>
<td>O</td>
<td>O.V</td>
<td>O.G</td>
<td>O.L</td>
<td>O.N</td>
<td>O.O</td>
</tr>
</tbody>
</table>

In the present theory, neither the Syllable Contact Law nor the Extended Syllable Contact Law need be stated separately, but follow from the principle of the Sonority Cycle as characterized in the earlier discussion. Suppose we view the complexity of any given syllable contact as a linear function of the complexity of each of its component demisyllables, taken individually. The ranking in (29) then follows straightforwardly from the complexity metric for individual demisyllables proposed in Tables 1-4.

To see this, let us assign an aggregate complexity score to each of the contact types in (29) calculated as a sum of the complexity values of each of the demisyllables that constitute it. We need consider only sequences in which neither demisyllable has more than two members, since this is the prototypical case. Thus the contact N.G (representing the demisyllable sequence VN.GV) is assigned a score of 7, since the first demisyllable has a complexity value $C$ of 3 and the second a complexity value $C$ of 4 (see Table 1). Proceeding in this way, we may construct a matrix from the table given in (29) by entering the appropriate scores for each contact type. We see that the optimality of a given contact type is a simple function of its aggregate complexity:
(30)  
<table>
<thead>
<tr>
<th></th>
<th>V</th>
<th>G</th>
<th>L</th>
<th>N</th>
<th>O</th>
</tr>
</thead>
<tbody>
<tr>
<td>V</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>G</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>L</td>
<td>7</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>N</td>
<td>8</td>
<td>7</td>
<td>6</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>O</td>
<td>9</td>
<td>8</td>
<td>7</td>
<td>6</td>
<td>5</td>
</tr>
</tbody>
</table>

6.5. Core Syllable Typology

In Clements and Keyser (1983), it was pointed out that the inventory of core syllable types is subject to certain widely observed constraints. The following types of core syllable inventories are commonly found across languages (where each C and V can represent a potential cluster):

(31)  
Type I: CV  
Type II: CV, V  
Type III: CV, CVC  
Type IV: CV, V, CVC, VC

On the other hand, other logically possible types of core syllable inventories are apparently lacking:

(32)  
a. V, VC  
b. CVC, VC  
c. CV, V, VC  
d. CV, CVC, VC  
e. CV, V, CVC  
f. CV, VC  
g. V, CVC  
h. V, VC, CVC

Thus we find languages whose core syllable types fall into the set of categories in (31), but none whose core syllable types correspond to any of the sets in (32). Clements and Keyser point out that the attested sets in (31) are characterized by two logical implications: a closed
syllable type implies an open syllable type, and a vowel-initial syllable type implies a consonant-initial type. The CV syllable type is universal, as it is implied by all the others.

These relations follow from the principles of markedness presented above. By the Complexity Hierarchy stated in (21a), closed syllables imply open syllables because final VC demisyllables imply final V demisyllables; and similarly, V-initial syllables imply C-initial syllables because V-initial demisyllables imply CV-initial demisyllables. Skewed inventories such as the one in (32c) are excluded by the fact that core syllabification rules are defined on demisyllables, not syllables: thus the presence of a CV initial demisyllable and a VC final demisyllable in the core phonology is sufficient to determine the presence of a CVC core syllable type.

7. Residual Problems

This section examines three residual problems: the question of consonant strength hierarchies, the treatment of geminates and other linked sequences, and the place of the major class features in the feature hierarchy.

7.1. Consonant Strength Hierarchies

Many linguists have noted a tendency for phonological “strengthening” processes to apply preferentially at the beginning of the syllable and “weakening” processes at the end. For example, Foley (1977) identifies an Inertial Development Principle according to which (i) strong elements strengthen first and most extensively and preferentially in strong environments, and (ii) weak elements weaken first and most extensively and preferentially in weak environments. Strong environments include syllable-initial position, and weak environments syllable-final position.

Some linguists have attempted to explain this tendency by treating consonant strength as the converse of sonority. In this view, we may simply invert the arrows in the sonority scale to define a scale of strength: O > N > L > G > V. If strength is interchangeable with sonority in this way, we may restate Foley’s generalization to say that sonority tends to decrease toward the beginning of the syllable and increase toward the end. But this is just to say that demisyllables tend to get simpler; and this is in turn a further instance of the more general principle, stated in section 3.1, that rules tend to apply in such a way as to
replace complex representations with simpler ones. Thus under these assumptions, the "Inertial Development Principle" would follow as a consequence of sonority theory.

A caveat is in order, however. The Inertial Development Principle as stated by Foley has many exceptions in both directions: as several linguists have pointed out, many processes predicted to occur don't, and many processes predicted not to occur do. Further, many proposed examples of consonant strengthening processes, such as fortition or gemination, cannot be adequately stated in terms of differences in sonority unless we assume a version of the sonority scale making much finer distinctions that are justified by crosslinguistic generalizations of segment sequencing. These problems undermine the attempt to derive a theory of sound change from sonority theory rather seriously.

Indeed, one might expect, if historical change were to follow a tendency to simplify all demisyllables, that all languages would eventually wind up with simple CV syllables and no others. This raises a problem inherent in all theories of markedness: why don't all languages eliminate marked structure in favor of unmarked structure everywhere? The answer seems to be that what is simple by one criterion is complex by another, and that all criteria cannot be satisfied at all times. Rules which simplify production may complicate perception, and rules that simplify perception may reduce the efficiency of the speech code. These competing tendencies (and other problems) make it very difficult to construct strongly predictive theories in the area of sound change. It seems best, therefore, to treat the Inertial Development Principle and similar theories with a measure of caution, and to hope that more adequate theories of sound change will be developed which take such complexities into account.

7.2. The Special Status of Linked Sequences

We have said nothing as yet about a significant set of exceptions to the principles of syllable contact discussed in section 6.4. Many languages allow just a small set of intervocalic consonant clusters, typically including geminates and homorganic NC (nasal + consonant) clusters. Indeed, some languages, including Japanese, Southern Paiute, and Luganda, allow only these. As Prince points out (1984, 242-3), the generalization seems to be that languages otherwise eschewing heterosyllabic consonant clusters may allow them just in case they involve linked sequences: sequences sharing a single set of features. More precisely, what seems to be required is that the adjacent consonants share the place of articulation node. This is exactly what geminates and homorganic NC clusters have in
common, as is shown by the following, simplified diagrams of the sequences [tt] and [nt], respectively (for this notation see Clements 1985):

\[
\begin{array}{c}
\text{(33)} \\
\text{root:} \\
\text{supralaryngeal:} \\
\text{place:} \\
\text{[coronal]:}
\end{array}
\]

\[
\begin{array}{ll}
\text{a.} & \text{C} \quad \text{C} \\
\text{b.} & \text{C} \quad \text{C}
\end{array}
\]

Intuitively, what makes these sequences simple is the fact that they involve only a single specification for place of articulation; indeed there is some evidence that the NC clusters may be gesturally equivalent to a single consonant at the same place of articulation in some languages (see Browman and Goldstein (1986) for English and Chaga, but cf. Fujimura and Lovins 1978, note 43 for contrary results for English). We conclude from these observations that intersyllabic articulations involving a single place specification are simpler than those involving two (or more) place specifications. This principle must clearly take precedence over the sonority principles stated earlier. (On the other hand, sonority considerations may help to explain the fact that the C in NC clusters is almost always an obstruent.)

Why are geminates heterosyllabic instead of tautosyllabic? In other words, why is a word like \textit{totta} universally syllabified \textit{tot-ta} rather than \textit{to-tta}? The answer to this lies in the CSP. As it scans the skeletal tier, the CSP syllabifies leftward as far as possible, first adjoining the second half of the geminate with the final vowel. It cannot syllabify the first half of the geminate with that vowel, since both skeletal C-elements dominate a single segment and thus have the same sonority rank. Consequently the first half of the geminate syllabifies with the preceding vowel. That it syllabifies into the preceding demisyllable at all, rather than e.g. being deleted due to its low sonority rank, reflects a general principle, overriding sonority considerations, to the effect that linked material is syllabified whenever possible (Christdas 1988).26
7.3 The Status of the Major Class Features in the Feature Hierarchy

A further question concerns the status of the major class features in the feature hierarchy. In the view presented in Clements (1985), major class features are placed under the domination of the supralaryngeal node. By assigning the major class features to the supralaryngeal node rather than to the root node, we predict that laryngeal "glides" — segments which have only laryngeal specifications — are not ranked in any position on the sonority scale, and are not characterized for any major class features. This seems correct from a crosslinguistic perspective. Laryngeals tend to behave arbitrarily in terms of the way they class with other sounds, avoiding positions in syllable structure that are available to true glides, and patterning now with obstruents, now with sonorants in a way often better explained by their historical origin in any given language than by their inherent phonological properties.

In assigning these features to separate tiers, however, we predict that they should be able to engage in assimilatory spreading. As pointed out independently by Schein and Steriade (1986) and Bruce Hayes (personal communication), the support for this prediction is at present quite thin. An alternative view is that they have the status of "annotations" on the supralaryngeal class node, in the sense that they are features characterizing this node, but which are not linked to it by association lines. This assumption would entail that major class features spread if and only if the supralaryngeal node spreads.27

8. General Discussion

Let us review the answers we have proposed to some of the questions raised at the outset of this study:

1. *How is sonority defined in phonological theory? Is it a primitive, or is it defined in terms of other, more basic features?* We have proposed that sonority is not a primitive phonological feature, but a derived phonological property of representations definable in terms of the major class categories [syllabic, vocoid, approximant, sonorant].

2. *What are its phonetic properties?* We have suggested that the phonetic correlates of sonority are just those of the major class features which define it, which share a
“family resemblance” in the sense that all of them contribute to the overall perceptibility of the classes of sounds they characterize.

3. At what linguistic level do sonority sequencing constraints hold? We have proposed that the constraints are primarily defined in core phonology (more specifically, at the level of initial syllabification L(IS)), where syllabification obeys the Complexity/Length Hierarchy. Later rules, especially those applying at the periphery of the syllabification domain, may introduce new, more complex syllable types which create surface exceptions to the sequencing constraints.

4. Over what units are sonority constraints defined? It has been shown that sonority constraints are defined over demisyllables, rather than over syllables or other subsyllabic units. The demisyllable is necessary and sufficient to the statement of sonority constraints.

5. Can languages vary in their choice of sonority scales? We have argued that a single sonority scale, that given in (6) or a simple variant of it, characterizes sonority in all languages. Apparent language-particular variation may reflect the effect of the Sequential Markedness Principle, which holds that if only a subset of a particular major class is allowed in some position, it will be the least marked subset.

6. Can syllable types be ranked along a scale of complexity? It has been argued that demisyllables (and derivatively, syllables) can be ranked along a scale of complexity according to the principle of the Sonority Cycle, which holds that the preferred syllable shows a sharp rise in sonority and a gradual fall. This principle is supported by the range of evidence discussed in section 6.

This approach is further supported by the simplification it allows in the description of particular languages. The core syllable inventory of a given language is largely determined by a small number of alternative choices and parameters:

(34)  i. the domain of core syllabification (nonderived stem, derived stem at level n, word, etc.);
ii. type of permitted syllable peaks (V, L, N, ...);
iii. maximum length of each demisyllable type;
iv. maximum degree of complexity of each type of demisyllable (predicts the presence of all the less complex demisyllables of that type).

In addition, languages may have core syllabification rules defining well-formed "marked" demisyllable types; filters specifying systematic gaps in the set of well-formed demisyllables; and perhaps rules defining the occurrence of permissable extrasyllabic elements ("appendices", "affixes") in domain-peripheral position. It is likely, however, that such rules play a less important role in core syllabification than has previously been thought.

These results have consequences for questions regarding the formalization of syllable representation. There are many current views concerning the nature of subsyllabic constituency, and the nature of the evidence supporting one or another of these views is not always as clear or straightforward as we might like. We might take as the null hypothesis the view that there is no subsyllabic constituency at all (Kahn (1976), Clements and Keyser (1983)). Most syllable theoreticians, however, have proposed more elaborate views. Some have proposed to recognize a ternary division of the syllable into an onset, peak, and coda (Hockett 1955, Pike 1967, Davis 1985). Others recognize a basic subdivision of syllables into two immediate constituents consisting of the onset on the one hand and a unit consisting of the peak and coda (sometimes termed the rhyme) on the other (see Kuryłowicz 1948, Fudge 1969, Selkirk 1982, among others). Still others group the peak and the onset into a unit forming an immediate constituent with the coda, calling both constituents "moras" or "weight units" (Yoshida 1983, Hyman 1984, McCarthy and Prince 1986). These positions by no means exhaust the range of proposals that have been made at one time or another. Disagreement reflects not only the different properties of the particular languages taken into account by any given author, but also the choice of what criteria to take as primary. For example, some linguists treat "external" evidence from sources such as language games and speech errors as crucial, while others regard such evidence as secondary to "internal" evidence; some draw their evidence from relatively abstract phonological analysis, while others place major emphasis on phonetic processes. Beyond these differences, however, there seems to be a certain amount of inherent ambiguity in phonological systems themselves, many of them lending themselves equally well to more than one analysis (van der Hulst 1984).
The study of sonority-based constraints bears on this discussion quite directly, since most linguists have recognized sequential constraints within the syllable as providing crucial evidence for syllable constituency. We have found no direct support from this quarter for any of the above theories *in toto*. Rather we have seen that sequencing constraints are most simply and effectively stated in terms of the demisyllable. Note that if this view is correct, it predicts that we should not expect to find sonority-based dependencies holding across demisyllables. Thus (without further qualification of the theory) we would not anticipate finding sonority-based dependencies holding between initial and final demisyllables, such that an initial demisyllable having a sonority profile of type A fails to combine with a final demisyllable whose sonority profile is of type B. Nor should we expect to find sonority-based dependencies holding across syllable boundaries. These predictions are correct, as far as I know. Thus, for example, we have found that apparent "syllable contact" dependencies are derivable from an independently-motivated metric needed to express the relative complexity of individual demisyllables, and require no separate statement. 28

These results, if correct, provide support for certain aspects of the theories mentioned above, but not for others. The initial demisyllable is identical to the initial mora (or weight unit) of moriac theories, and the final demisyllable is similar to the rhyme of onset/rhyme theories. Thus some of the predictions of these theories follow equally from a theory incorporating demisyllables. Other predictions, however, do not. Thus there is no formal counterpart in demisyllable representations for the onset or the final mora, and at least the latter construct appears to play a crucial role in theories incorporating it, for example by allowing us to distinguish between heavy and light syllables. Note that it is crucial that the final demisyllable (unlike the final mora) include V, since otherwise we would have no way of expressing the preference for e.g. VG demisyllables over VO demisyllables, and would not be able to derive such statements as the Complexity Hierarchy or the Syllable Contact Law. A full evaluation of the consequences of these results for the theory of syllable-internal constituency would take us well beyond the goals of the present study, however, and therefore we must leave it for further study at a later time.

9. Conclusion

The notion of the Sonority Cycle as developed above provides us with a basis for explaining the striking and significant regularities in syllable structure that we find across
languages, and for integrating these observations into a formal theory of syllable representation, allowing us to capture many generalizations that have up to now been inadequately understood or explained.

Our results suggest that a significant crosslinguistic regularity of phonological structure (the Sonority Cycle) may be most clearly revealed at a level of representation considerably removed from surface representation (or acoustic reality), but that this principle has a regular expression at the phonetic level through the mediation of the major class features which provide its vocabulary. Such a conclusion should not be surprising in view of the fact that what is perceptually real for native speakers may bear little direct resemblance to the speech signal itself; indeed this has been the lesson of phonological studies since the emergence of the modern concept of the phoneme in the work of Sapir, Trubetzkoy, Jakobson and others in the early 1930s. This result by no means implies a divorce between linguistics and phonetics, but rather takes us a step further toward solving the long-standing enigma of how abstract linguistic form is communicated through the medium of the speech waveform: significant patterning relations may be encoded at a certain degree of abstraction from the physical data, but must have a clear and constant manifestation in the speech signal if they are to be successfully conveyed from speaker to hearer.
Table 1

Complexity rankings for demisyllables of two and three members, based on the sonority scale $O < N < L < G < V$.

<table>
<thead>
<tr>
<th></th>
<th>$D$</th>
<th>$C$</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Two-member demisyllables:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>i. initial:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OV</td>
<td>.06</td>
<td>1</td>
</tr>
<tr>
<td>NV</td>
<td>.11</td>
<td>2</td>
</tr>
<tr>
<td>LV</td>
<td>.25</td>
<td>3</td>
</tr>
<tr>
<td>GV</td>
<td>1.00</td>
<td>4</td>
</tr>
<tr>
<td>ii. final:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VO</td>
<td>.06</td>
<td>4</td>
</tr>
<tr>
<td>VN</td>
<td>.11</td>
<td>3</td>
</tr>
<tr>
<td>VL</td>
<td>.25</td>
<td>2</td>
</tr>
<tr>
<td>VG</td>
<td>1.00</td>
<td>1</td>
</tr>
<tr>
<td>b. Three-member demisyllables:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>i. initial:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OLV</td>
<td>.56</td>
<td>1</td>
</tr>
<tr>
<td>ONV, OGV</td>
<td>1.17</td>
<td>2</td>
</tr>
<tr>
<td>NLV, NGV</td>
<td>1.36</td>
<td>3</td>
</tr>
<tr>
<td>LGV</td>
<td>2.25</td>
<td>4</td>
</tr>
<tr>
<td>ii. final:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VLO</td>
<td>.56</td>
<td>4</td>
</tr>
<tr>
<td>VGO, VNO</td>
<td>1.17</td>
<td>3</td>
</tr>
<tr>
<td>VLN, VGN</td>
<td>1.36</td>
<td>2</td>
</tr>
<tr>
<td>VGL</td>
<td>2.25</td>
<td>1</td>
</tr>
</tbody>
</table>
Table 2

Complexity rankings for one-member demisyllables (compared to two-member demisyllables):

<table>
<thead>
<tr>
<th></th>
<th>(D)</th>
<th>(C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. initial:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>V</td>
<td>undefined</td>
<td>5</td>
</tr>
<tr>
<td>b. final:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>V</td>
<td>undefined</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 3

Complexity rankings for four-member demisyllables:

<table>
<thead>
<tr>
<th></th>
<th>(D)</th>
<th>(C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. initial demisyllables:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ONGV</td>
<td>2.53</td>
<td>1</td>
</tr>
<tr>
<td>OLGV, ONLV</td>
<td>2.67</td>
<td>2</td>
</tr>
<tr>
<td>NLGV</td>
<td>3.61</td>
<td>3</td>
</tr>
<tr>
<td>b. final demisyllables:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VGNO</td>
<td>2.53</td>
<td>3</td>
</tr>
<tr>
<td>VGLO, VLNO</td>
<td>2.67</td>
<td>2</td>
</tr>
<tr>
<td>VGLN</td>
<td>3.61</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 4

Complexity rankings for demisyllables with syllabic consonants as peaks:

<table>
<thead>
<tr>
<th></th>
<th>(D)</th>
<th>(C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>OV/VO</td>
<td>.06</td>
<td>1</td>
</tr>
<tr>
<td>OL/LO</td>
<td>.11</td>
<td>2</td>
</tr>
<tr>
<td>ON/NO</td>
<td>.25</td>
<td>3</td>
</tr>
<tr>
<td>OQ/QO</td>
<td>1.00</td>
<td>4</td>
</tr>
</tbody>
</table>
Acknowledgements

I would like to thank Harry van der Hulst, John McCarthy, Stuart Milliken, Donca Steriade, and participants in a seminar on syllable phonology given on three occasions during 1985-1986 at Cornell University, the University of Washington, and the Summer Linguistics Institute at the University of Salzburg, for their valuable critical reactions to various presentations of the ideas in this paper. I am further grateful to Annie Rialland for discussion of the French data, and to Mary Beckman, Osamu Fujimura, and John Kingston for their written commentary on earlier drafts. All of these have contributed in some way to improvements in style and substance, although they do not necessarily agree with its conclusions. Earlier versions of this paper were presented at Yale University in November, 1985, at the Annual Meeting of the Linguistic Society of America in Seattle, Washington, in December 1985 and at the Workshop on Features, Wassenaar, The Netherlands in June, 1986.

Notes

1. Cross-linguistic generalizations such as these, at varying degrees of abstraction from the primary data, provide the explicanda of theory construction in linguistics, and form the basis of the hypotheses and models that eventually come to constitute a formal linguistic theory, i.e. a theory of possible grammars and of optimal grammars.

2. Sievers distinguished between the Drucksilbe, conceived of as an articulatory-defined syllable produced with a single independent expiratory pulse, and the Schallsilbe, an auditorily-defined syllable determined by the relative audibility or sonority (Schallfülle) of its members. These two criteria do not always coincide, as is evidenced by the German (or English) word hammer which constitutes one Drucksilbe but two Schallsilben. Of these, the Schallsilbe is most relevant to sonority theory. See Bloomfield (1914) for a brief summary of Sievers' ideas, in which the two syllable types are termed "natural syllable" and "stress syllable".

3. Jespersen's scale differed from Sievers' in ranking all voiceless sounds before all voiced, in not attributing a separate rank to voiceless stops and fricatives, and in assigning nasals and liquids the same rank.

4. The version of the Sonority Sequencing Principle given here follows Sievers rather than Jespersen, as this is the version that is most widely followed today, cf. e.g. Kiparsky (1979) and Lowenstamm (1981). Jespersen allowed elements of equal sonority to be adjacent within the
syllable. His reluctance to adopt the more restrictive version may have been motivated by the common occurrence of initial clusters like st and final clusters like ts, which constitute anomalies under Sievers' formulation, but not under Jespersen's, where s and t are of equal rank and may thus occur adjacent to each other.

5. See Pike (1942, 137-148) for a presentation of this notion, as well as Catford (1977) for more recent discussion.

6. There has been relatively little critical discussion of the notion of sonority in the recent literature; a notable exception is Bell and Saka (1983).

7. Early proponents of the theory, such as Sievers and Jespersen, did not distinguish between underlying and surface representation, and consequently assumed a surface-oriented version of the principle. Discussion in the context of generative phonology has generally recognized that the SSP interacts with other rules and principles which may give rise to surface-level exceptions. For example, Kiparsky (1979, 1981) notes that the SSP may be overridden by language-particular rules, while Fujimura and Lovins (1977) allow exceptions within syllable "affixes" that lie outside the "core".

8. This statement must be qualified by the observation that the identity of the sonority scale varies in detail from one linguist to another. What is a sonority reversal for one writer may be a sonority plateau for another, and what is a sonority plateau for one may constitute an ascending or descending ramp for another. This qualification extends to the further discussion below. Note also that the cases in (5a) represent violations of Sievers' version of the Sonority Sequencing Principle as given in (2), but not of Jespersen's, which tolerates clusters of equal sonority within the syllable.

9. Data sources for the less familiar languages are as follows: Mohawk (Michelson, 1983), Cambodian (Huffman 1972), Marshallalese (Bender 1976), Ewe (author's field notes, standard dictionaries), Pashto (Bell and Saka 1983), Klamath (Barker 1963), Ladakhi (Koshal 1979), Kota (Emenau 1944), Abaza (Allen 1956), Tocharian A (Coppieters 1975, J. Jasanoff, p.c.), Yatóe Zapotec (Jaeger and Van Valin 1982), Turkish (Clements and Keyser 1983), Berber (Dell and Elmedlaoui 1985), Luganda (Tucker 1962), Bella Coola (Nater 1984).

10. A few representative references follow: Allen 1956 (Abaza), Dell and Elmedlaoui 1985 (Berber), Huffman 1972 (Cambodian), Nater 1984 (Bella Coola). See also Bell and Saka (1983) for a detailed examination of Pashto. (Notice that while Dell and Elmedlaoui argue that Berber largely conforms to sonority sequencing restrictions, they also recognize language-particular configurations in which these requirements are suspended.)
11. Heffner (1950,74) states that “sonority may be equated more or less correctly with acoustic energy and its quantities determined accurately by electronic means”, citing Fletcher (1929) in support. It is true that Fletcher’s methods of measuring the “phonetic power” of segments give us a ranking grossly similar to familiar sonority scales, with vowels at one end and obstruents at the other. But Fletcher’s results do not support the finer distinctions usually thought to be required for linguistic purposes. Thus by one of his measures (the “threshold” method), the nonanterior sibilants represented by orthographic *ch, sh* ranked higher in power (roughly equivalent to sonority) than nasals and all other obstruents, the voiceless stop [k] ranked higher than fricatives or voiced stops, while the glides and nonsyllabic [r] were not measured at all. Moreover, Fletcher observed a high degree of interspeaker variation, suggesting that crucial details of such phonetic measures might vary substantially from speaker to speaker.

12. This definition, which follows Catford (1977,119-127), includes voiceless sonorants, which are normally produced with audible turbulence. Unlike Catford, however, I consider all vowels to be approximants. The sonority ranking of voiceless approximants is not well established, and requires further examination. The term “approximant” was first introduced in Ladefoged (1964), and replaces the older term “frictionless continuant.”

13. Bell notes: “among the languages with only syllabic nasals, very few are subject to vowel reduction; of those with syllabic liquids, all but a handful do have some form of vowel reduction. The formation of syllabic liquids may be strongly disfavored where nonreduced vowel syncope is the process of origin, but not disfavored under reduced-vowel syncope” (171).

14. The CSP differs from a similar algorithm given for English syllabification by Kahn (1976) in being universal rather than language-particular. It does not syllabify in terms of language-particular initial and final clusters, as does Kahn’s rule, but in terms of the universal sonority scale. In this view, language-particular differences in core syllabification are attributed to further parameters of core syllabification, such as length constraints of the sort just mentioned, or to further rules of core syllabification that apply independently of sonority restrictions, as discussed below.

15. Versions of the Maximal Onset Principle were known to the ancient Sanskrit and Greek grammarians (Varma 1929, Allen 1951). It is usually considered to have had exceptions in Indo-European, however; see Hermann (1923), Borgstrøm (1937), Schwyzer (1939), and Lejeune (1972) for relevant discussion.

16. This assumes either simultaneous or right-to-left application of the core syllabification rules. As we will see below, our final statement of the Core Syllabification Principle in (27) will be consistent with both of these modes of application.
17. This skewing may explain the asymmetries between initial and final clusters noted by Reilly (1986).

18. The term *demisyllable* as used here is inspired by Fujimura's account, but differs from it in significant respects. Fujimura has used it to designate a phonetic sequence used for the purposes of speech synthesis and automatic speech recognition (Fujimura et al. 1977), and has characterized it as follows:

> We have tentatively decided on an operational rule for “cutting” each syllable in two, producing initial and final demisyllables ... The cutting rule may be stated: ‘Cut 60 msec after release, or if there is no release, 60 msec after the onset of the vocalic resonance.’ This is usually a point shortly after the beginning of the so-called steady state of the vowel, that is, after the consonant-vowel transition.

In this usage, the demisyllable is an acoustic unit. Fujimura also conceives of the demisyllable as a phonological unit, one of the two halves into which syllables “cores” are divided (Fujimura and Lovins 1977, Fujimura 1979, 1981). This unit has not previously been used in the statement of phonological rules and constraints to my knowledge, although it has been identified with the onset/rhyme distinction inside the syllable core (Fujimura 1981,79). In my usage, for reasons to be made clear, demisyllables are not identified with onsets and rhymes; see especially section 8 for discussion.

19. It follows from the definition in (15) that in syllables containing long vowels $V_1V_2$ the first demisyllable ends in $V_1$ and the second begins with $V_2$, while in syllables containing diphthongs VC the first ends in V and the second begins with the same V. In languages whose long vowels are characteristically nondiphthongal and therefore of the type VV, the distribution of long vowels tends to be equivalent to that of short vowels (see Vago 1985 for Hungarian). In contrast, in languages having diphthongs and long vowels of the type VC, such as German and English, the distribution of long vowels tends to be equivalent to that of short vowels followed by consonants (Moulton 1956; Selkirk 1982,351).

20. This limitation follows from what has been said before. The preferred initial demisyllable has a high-sonority peak, since lower-sonority peaks minimize the sonority distance between the peak and the margin. But the peak is shared between the initial and final demisyllable; therefore all else being equal, syllables will be simpler to the extent that their peaks are higher on the sonority scale.

21. There are no exceptions to this statement in morpheme-final position. Morpheme-internally, the only common exceptions are *napkin, pumpkin, breakfast, magpie, tadpole, aardvark,*
Afghanistan, and frankfurter. Proper names show frequent violations but may usually be analyzed into a stem and name-forming suffix, as in Bradford/Bedford, Cambridge/Sturbridge, Lindberg/Sandberg, Bradbury/Woodbury, Tompkins/Watkins, Hatfield/Westfield.

22. A similar account of the exceptional status of coronals has been proposed by Devine and Stevens (1977) in the context of their discussion of Latin syllabification (I thank John McCarthy for calling this work to my attention). There are rarer cases of languages exhibiting a preference for noncoronals in certain positions, for which an alternative explanation will be required. One such case involves the occurrence of clusters like $kt$, $pt$, $mn$ in Attic Greek to the exclusion of clusters like $tk$, $tp$, $nm$; Steriade (1982, ch. 4) argues that the initial members of such clusters are extrasyllabic throughout the lexical phonology.

23. A few qualifications are in order. First, Greenberg’s generalizations concerned initial and final position in the word, not the syllable, and therefore do not necessarily translate directly into syllable structure. We have already noted that initial and final clusters in the syllabification domain (typically, the word) often deviate somewhat from initial and final clusters in internal syllables, especially in permitting extrasyllabic sequences or “appendices”. As such sequences often reflect the operation of syllabification rules that override the usual sonority constraints, we would expect Greenberg’s data to be less supportive of the theory developed here than generalizations based exclusively on syllabification data. Second, Greenberg’s survey was based on a study of the descriptive literature, and inherits the analytical weaknesses and inadequacies of its sources. As Greenberg notes, several arbitrary choices had to be made, particularly concerning the decision whether to regard stop-fricative sequences as clusters or affricates. Third, Greenberg’s implicational universals are probably best regarded as statistical rather than categorical in nature. Several implications that were true of the sample have since proven to have exceptions in other languages: thus, Ladakhi has LOV syllables but not OLV syllables (Koshal 1979), and Yatec Zapotec has the rare GOV syllable type, as noted in section 3.1. The counterpart to this is that many statements that were not categorically true of Greenberg’s sample may turn out to be significant when a wider sample of languages is considered.

24. These results do not depend on the identity of the sonority scale we choose; more complex scales recognizing a larger number of points will yield the same relationship between minimal distance and degree of complexity. For example, given the hypothetical seven-point sonority scale $O < Z < N < L < R < G < V$ discussed earlier, the most equally distributed three-member demisyllable will be OLV. As we successively minimize the difference between the medial member of the demisyllable and either endpoint we increase the value for $D$ and thus increase the complexity value
C. For example, OLV has the value .25 for $D$, ONV has the value .34, and OZV has the value 1.07.

25. There is a further difference between this account and accounts making use of the notion "minimal distance". Given the sonority scale $O < N < L < G < V$, our account predicts that we might find languages containing only one of two demisyllables of a given degree of complexity. This is because the Complexity/Length Hierarchy (21) only requires that given the presence of demisyllables of some degree of complexity $n$, demisyllables with lower degrees of complexity must also be present. For example, we should find languages with initial demisyllables of the form OGV but not ONV (or vice-versa), both of which have a complexity rank of 2. A theory in which ONV is excluded by a minimal distance constraint would necessarily exclude OGV at the same time.

26. Homorganic NC sequences are tautosyllabic in many languages, such as Bantu which allows NCV syllables both initially and word-externally. In these cases it is often plausible to analyze the NC sequence as a single prenasalized stop (Clements 1986), so that the demisyllable type is actually CV.

27. See, however, Milliken (1988) for an account of Flap Formation in English (and similar rules in other languages) in terms of the spreading of subsets of major class features.

28. In some languages, however, we find constraints holding across pairs of syllables such that the sonority rank of the onset of the second syllable must be equal to or greater than the sonority rank of the onset of the first. Williamson (1978), in her discussion of such a phenomenon in Proto-jjɔ, observes that it often arises historically through processes of consonant weakening in noninitial syllables, such as intervocalic voicing or spirantization. This phenomenon does not seem to reflect sonority considerations exclusively, since the dependencies in question often involve features such as voicing and continuance and may be equally well viewed as involving assimilation to the intervocalic context. Clearly, however, this is an important potential type of exception to our statement that deserves fuller and more systematic investigation.
References


Heffner, R.-M.S. (1950) *General Phonetics*, University of Wisconsin Press, Madison, Wi.


The Delta Programming Language: An Integrated Approach to Non-Linear Phonology, Phonetics, and Speech Synthesis

Susan R. Hertz

1 Brief Overview

The Delta programming language is designed to let linguists easily formalize and test phonological and phonetic theories. Its central data structure lets rule-writers represent utterances as multiple "streams" of synchronized units of their choice, giving them considerable flexibility in expressing the relationship between phonological and phonetic units. This paper presents the Delta language, showing how it can be applied to two linguistic models, one for Bambara tone and fundamental frequency patterns and one for English formant patterns. While Delta is a powerful, special-purpose language that alone should serve the needs of most phonologists, phoneticians, and linguistics students who wish to test their rules, the Delta System also provides the flexibility of a general-purpose language by letting users intermingle C programming language statements with Delta statements.

2 Introduction

Despite their common interest in studying the sounds of human language, the fields of phonology and phonetics have developed largely independently in recent years. One of the contributing factors to this unfortunate division has been the lack of linguistic rule development systems. Such systems are needed to let linguists easily express utterance representations and rules, and facilitate the computational implementation and testing of phonological and phonetic models.

SRS (Hertz, 1982) is a rule development system that was designed, starting in 1974, for just this purpose—to let linguists easily test phonological and phonetic rules, and explore the interface between phonology and phonetics through speech synthesis. SRS, however, was influenced quite heavily by the theory of generative phonology that was prevalent at the time, a theory that posited linear utterance representations consisting
of a sequence of phoneme-sized segments represented as bundles of features (Chomsky and Halle, 1968). Although at the phonetic level, SRS uses different "streams" for different synthesizer parameters, the parameter values and segment durations must all be set in relation to the phoneme-sized segments at the linear phonological level.

Thus, while SRS lets users express rules in a well-known linguistic rule notation, and easily change the rules, it forces them to work within a particular framework. Because SRS was biased toward a particular theory of sound systems, we became equally biased in our approach to data analysis and rule formulation. For example, we took for granted that phonemes (more precisely, phoneme-sized units) were the appropriate units for the assignment of durations and formant patterns, a basic assumption that blinded us for years to the possibility of alternative models. As alternatives finally emerged, however, the need for a more flexible system for expressing and testing phonological and phonetic rules became apparent.

The clearest requirements for a more flexible rule development tool were a multi-level (or multi-tiered) data structure that could make explicit the relationship between phonological and phonetic units, and a precise and flexible rule formalism for manipulating this structure. In response to these needs, in July 1983 I began the development of a new synthesis system, the Delta System (Hertz, Kadin, and Karplus, 1985; Hertz, 1986), in consultation with two computer scientists, Jim Kadin and Kevin Karplus. The Delta System provides a high-level programming language specifically designed to manipulate multi-level utterance representations of the sorts suggested by our rule-writing experience with SRS (Hertz, 1980; Hertz, 1981; Hertz, 1982; Hertz and Beckman, 1983; Beckman, Hertz, and Fujimura, 1983). This language lets users write and test rules that operate on multi-level utterance representations without having to take care of the programming details that would be required in an ordinary programming language like C. A one-line delta statement might easily take a page to accomplish in C. The ease of expressing and reading rules in Delta enables rule-writers to test alternative strategies freely and conveniently.

While the move from the linear utterance representations central to SRS to the multi-level representations central to Delta parallels the move by phonologists from linear to non-linear representations, the Delta System is a direct consequence of our SRS experience, and, unlike SRS, was developed independently of the phonological theories in vogue at the time. The Delta System is flexible enough to let phonologists and
phoneticians of different persuasions express and test their ideas, constraining their representations and rules in the ways they, rather than the system, see fit. The system assumes as little as possible about the phonological and phonetic relationships that rule-writers may wish to represent in their utterance representations, and the manner in which their rules should apply, allowing them to make dependencies between rules explicit and giving them full control, for example, over whether their rules should apply cyclically or non-cyclically, sequentially or simultaneously, left-to-right or right-to-left, morph by morph or syllable by syllable, to the entire utterance or only a portion thereof, and so on.

In addition to a powerful programming language for building and manipulating multi-level utterance representations, the Delta System provides a flexible interactive debugger. The debugger lets users issue commands to interact with a program while it is executing. It lets users trace their rules during program execution, stop the execution of their program at selected points (e.g., each time the utterance representation or a particular variable changes), display the utterance representation and other data structures, modify the utterance representation “on the fly” (e.g., to hear the result of a longer duration for a particular unit in a program designed for synthesis), and so on. The debugger, like the system in general, is designed for speed and flexibility in the development of phonological and phonetic rules, letting rule-writers test and modify their hypotheses quickly and easily. The debugger is an essential part of the system, but a description of it is outside the scope of this paper. Hertz et al. (1985) describes an early version of the debugger. The debugger has been enhanced substantially since that paper was written. It is now a complete source-level debugger with many more capabilities than those shown in that paper.

The Delta System has been designed to be as portable as possible, to give linguists the widest possible access to it. A Delta program is compiled by the Delta compiler into a C program, and can be run on any computer with a standard C compiler and at least 512K of memory, such as an IBM PC-AT or a Macintosh. Compiling into C has the additional advantage that it lets users integrate C programs with Delta programs at will, even intermingling Delta and C code in a single procedure.\(^1\) The system is also made

---

1. Earlier versions of the system compiled Delta programs into pseudo-machine instructions. The current approach (compiling into C) sacrifices the extremely compact storage of rules achieved by the pseudo-machine approach in favor of much faster rule execution and flexibility in interfacing C routines.
accessible through the comprehensive Delta User's Manual, which contains an overview of the system, a tutorial, extensive reference sections, and sample programs.

The next section of this paper (Section 3) presents selected features of the Delta language, introducing many of the concepts needed to understand the sample programs in subsequent sections. It uses examples from Bambara, a Mande tone language spoken in Mali. These examples anticipate the programs in Section 4, which illustrate how tone patterns and corresponding fundamental frequency values might be assigned in Bambara. Bambara is chosen because it exhibits many of the properties of tone languages that have motivated multi-level representations in phonology (e.g., tone spreading and floating tones), while at the same time providing good examples of the function of phonological units in determining actual phonetic values (e.g., the role of the tones in determining fundamental frequency values). Section 5 presents a model of English formant timing that further illustrates Delta's flexibility in accommodating a wide range of theories about the interface between phonology and phonetics. Finally, Section 6 presents conclusions, gives a brief overview of the features of Delta not described in the paper, and discusses our plans for enhancing Delta and complementing it with another system in the future.

3 Selected Features of Delta

The Delta programming language is a high-level language designed to create, test, and manipulate a data structure called a delta for representing utterances. A delta consists of a set of user-defined streams of tokens that are synchronized with each other at strategic points. The tokens can represent anything the rule-writer wishes—phrases, morphs, syllables, tones, phonemes, sub-phonemic units, acoustic parameters, articulators, durations, classes of features, and so on. This section describes the structure of deltas, focusing first on the kinds of relationships that can exist between tokens in different streams, and then on the language for determining and testing these relationships.

For most of its sample deltas, this section uses the Bambara phrase muso jaabi 'answering the woman', which can be transcribed as [mûsō ! jāːbî]. In this transcription, the grave accent ` represents a low tone and the acute accent ´ a high tone. Thus the first syllable has low tone, and all the other syllables have high tone. The exclamation point represents tonal downstep, the lowering in pitch of the following high tones.
lowering occurs in Bambara after a definite noun. To account for the tonal downstep, linguists, following Bird (1966), posit as a definite marker a “floating low tone” that occurs after the noun and is not associated with any syllable.

Following is a sample delta that a program couched in the framework of autosegmental CV phonology (Clements and Keyser, 1983) might build for the phrase *muso jaabi*:

<table>
<thead>
<tr>
<th>(1)</th>
<th>phrase:</th>
<th>NP</th>
<th>word:</th>
<th>noun</th>
<th>morph:</th>
<th>root</th>
<th>verb</th>
<th>root</th>
</tr>
</thead>
<tbody>
<tr>
<td>phoneme:</td>
<td>m u s o j a a b i</td>
<td>CV:</td>
<td>C V C V C V C V</td>
<td>nucleus:</td>
<td>nuc nuc nuc</td>
<td>syllable:</td>
<td>syl syl syl</td>
<td>tone:</td>
</tr>
</tbody>
</table>

This delta consists of eight streams: phrase, word, morph, phoneme, CV, nucleus, syllable, and tone. The phrase stream has two tokens, NP (noun phrase) and VP (verb phrase); the word stream has two tokens, noun and verb; and so on. The tokens in the CV stream represent abstract timing units, in accordance with CV theory. The long phoneme a is synchronized with two V tokens in the CV stream, while the short vowels are synchronized with a single V. The nucleus stream marks each vowel, regardless of length, as the nucleus of the syllable. The tone stream has four tokens, two L (low) tokens and two H (high) tokens, reflecting the tone pattern given in the transcription above.

The vertical bars in each stream are called sync marks. Our sample delta has eleven sync marks, numbered at the bottom for ease of reference. Sync marks are used to synchronize tokens across streams. For example, sync marks 1 and 3 synchronize all of the tokens that constitute the first syllable. Sync marks 5 and 6 surround a L tone that is not synchronized with any tokens in any other stream. It represents a floating low tone that marks the noun phrase as definite, as discussed above. Sync marks 1 and 6 synchronize this floating tone along with the preceding L and H tones of the root with the NP token in the phrase stream. Following the analysis of Rialland and Sangare (1985), sync marks 6 and 11 synchronize a single H tone with two syllables and one root, rather than a separate H tone with each syllable.

This delta could be expressed in more familiar autosegmental terms as follows:
Note that while the delta representation makes the appropriate associations among tokens using sync marks alone, the autosegmental representation must use brackets in addition to association lines in order to show that the floating low tone is part of the noun phrase.

Furthermore, while the tiers in autosegmental representations are critically ordered with respect to each other in the sense that “tokens” on one tier can only be explicitly linked to tokens on particular other tiers (see, for example, Clements, 1985) the streams of a delta can always occur in any order with respect to each other. The same relationships exist between the tokens in different streams regardless of the order in which the streams are listed. For example, the phoneme, syllable, and CV streams in delta (1) above could also be displayed in the following order:

As in this example, examples of deltas from here on will show only those streams relevant to the discussion at hand.

In addition to syntactic, morphological, and phonological streams, a delta can also have phonetic streams. For example, a time stream might be added to delta (1) as follows:

Here, a duration (in milliseconds) is synchronized with each phoneme token.

In a delta, unlike in an autosegmental representation, a time stream can be used to time articulatory movements or acoustic patterns with respect to phonological units like
phonemes. In the following delta fragment, an F0 (fundamental frequency) stream is used to time F0 targets with respect to the syllable nuclei:

(5) phoneme: | s | o | j | a |
nucleus: | nuc | nuc |
tone: | H | L | H |
F0: | 150 | 130 |
duration: | 1200 | 150 | 130 |

An F0 target is placed halfway through each nucleus. The targets themselves have no duration, being used only to shape the F0 pattern, which moves from target to target in accordance with the specified durations. A program designed for actual synthesis could interpolate the values (transitions) between the targets and send them, along with the values for other synthesizer parameters, to the synthesizer.

3.1 Token Structure

Each token in a stream is a collection of user-defined fields and values. Each token has at least a name field. It is the value of the name field that is displayed for each token in the deltas shown above. Tokens can be given other fields as well, as shown below for a phoneme token named m:

(6) name: m
place: labial
manner: sonorant
class: cons
nasality: nasal

where cons = “consonantal”. All tokens in a given stream have the same fields, but of course the values for the fields can differ. Also, the value for a field can be undefined, when the value is not relevant for the token. Fields can be of different types, as discussed below. In most of the sample deltas, only the value of the name field is displayed, but it should be kept in mind that the tokens may have other fields as well.

A field together with a particular value is called an attribute. Thus the phoneme token illustrated above has the attributes <name: m>, <place: labial>, <manner: sonorant>, and <nasality: nasal>. The non-name attributes of a

2. A future version of Delta is planned that will allow tokens in the same stream to have different fields. This next version will also allow tokens to be represented as trees of features, as in the model of autosegmental phonology proposed by Clements (1985). See Section 6, “Final Remarks”, for more details.
token are also called features, so that we can speak of the token named m as having the features `<place: labial>, <manner: sonorant>`, etc. In general, token features are distinguished from token names in this paper by being enclosed in angle brackets. When the value of a field is unambiguous (i.e., when it is a possible value for only one field in the stream in question), the field name can be omitted in Delta programs, so we can also consider the m to have the features `<labial>, <sonorant>`, and so on. This abbreviated form for features will be used throughout the paper.

3.2 Delta Definitions

The first thing that a Delta rule-writer must do is give a delta definition. A delta definition consists of a set of stream definitions that define the streams to be built and manipulated by the program (rules). Figure 1 shows fragments of possible phoneme and F0 stream definitions for Bambara. All text following a double colon `:::` to the end of a line is a comment that is not part of the actual stream definition.

****************************

Insert Figure 1 here

****************************

The phoneme stream definition defines the tokens in the phoneme stream as having name, place, manner, class, nasality, voicing, height, and backness fields. (The stream names in the program fragments in this paper are all preceded by a percent sign.) The name field is a name-valued field; the place, manner, height, and backness fields are multi-valued fields; and the class, nasality and voicing fields are binary fields. A name-valued field is a field that contains the token names of some stream as possible values.3 A multi-valued field is a field that has more than two possible values and is not name-valued and not numeric (see the next paragraph). A binary field is a non-name-valued field that has exactly two possible values, such as `<nasal>` or `<~nasal>`, where "~" is Delta notation for "not".

3. It is not only the name field that can be name-valued. For example, Hertz et al. (1985) gives an example of a text stream definition that defines a name-valued field called default_pronunc with the names of the tokens in the phoneme stream as possible values, and shows how this field might be used in a Delta program for English text-to-phoneme conversion to synchronize default pronunciations (phoneme tokens) with text characters.
A binary field is always defined by specifying only one of the two possible values. The opposite value is assumed.

The phoneme tokens in this example do not have any numeric fields. A numeric field would be defined by following the field name with a keyword specifying the kind of number that can be a value of the field, as shown in the F0 stream definition, where the name field is defined as having integers as possible values. Thus a field can be both name-valued and numeric; all other field types are mutually exclusive. The keywords for numeric values are the same as the numeric type specifiers in C.

Below the field definitions for the phoneme stream are a set of initial feature definitions. These definitions assign values to tokens with particular names. When a token with a particular name is inserted into a delta stream, the token’s initial field values are automatically set, as discussed below.

If a token is not given a value for some field by an initial feature definition, it is automatically given an initial default value. For binary fields, the initial default value is the binary value not specified for the field in the stream definition. For example, in this case, all tokens not given the value `<cons>` for the class field will automatically be given the value `<~cons>`, and all tokens not given the value `<~voiced>` for the voicing field will automatically be given the value `<voiced>`. For multi-valued fields, the default value is `<undefined>`, a built-in value automatically defined for any multi-valued field. For numeric fields (other than numeric name fields), the default value is `<0>`. For name-valued fields, the default value is GAP, a built-in value automatically defined for any name-valued field. Gaps (i.e., tokens named GAP) are generally used as special “filler tokens” that separate tokens that would otherwise be considered adjacent, as discussed below.

3.3 Sample Program—Synchronizing Tokens

A Delta program consists of a delta definition followed by a set of procedures that operate on the delta. Figure 2 shows a short sample program that reads a sequence of phoneme tokens representing a Bambara word from the terminal into the phoneme stream, and synchronizes a C token in the CV stream with an initial consonantal token in the phoneme stream. (Later it will be shown how to apply the rule across the entire delta, synchronizing a C token with each consonantal phoneme.)
This program consists of a single procedure called `main`. Every program must have at least a procedure called `main`, where execution of the program begins. When the program begins execution, the delta has the following form (assuming that the streams shown are those defined by the delta definition):\(^4\)

\[
\begin{align*}
(7) & \quad \text{phrase:} & \text{| |} \\
& \quad \text{word:} & \text{| |} \\
& \quad \text{morph:} & \text{| |} \\
& \quad \text{phoneme:} & \text{| |} \\
& \quad \text{CV:} & \text{| |} \\
& \quad \text{syllable:} & \text{| |} \\
& \quad \text{tone:} & \text{| |}
\end{align*}
\]

The first program line,

\[
(8) \quad \text{read } @\text{phoneme;}
\]

reads a sequence of phoneme token names from the terminal and places the tokens in the phoneme stream. The fields of each phoneme are set as specified in the phoneme stream definition. For example, given the phoneme stream definition in Figure 1, if the sequence `muso` is entered, the delta would have the following form after the read statement has been executed:

\[
\begin{align*}
(9) & \quad \text{phrase:} & \text{| |} & \text{| |} & \text{| |} \\
& \quad \text{word:} & \text{| |} & \text{| |} & \text{| |} \\
& \quad \text{morph:} & \text{| |} & \text{| |} & \text{| |} \\
& \quad \text{phoneme:} & \text{| |} & \text{| |} & \text{| |} \\
& \quad \text{name:} & \text{m} & \text{u} & \text{| |} \\
& \quad \text{place:} & \text{labial} & \text{----} & \text{| |} \\
& \quad \text{manner:} & \text{sonorant} & \text{----} & \text{...} \\
& \quad \text{class:} & \text{cons} & \text{~cons} & \text{| |} \\
& \quad \text{nasality:} & \text{nasal} & \text{~nasal} & \text{| |} \\
& \quad \text{voicing:} & \text{voiced} & \text{voiced} & \text{| |} \\
& \quad \text{height:} & \text{----} & \text{high} & \text{| |} \\
& \quad \text{backness:} & \text{----} & \text{back} & \text{| |} \\
& \quad \text{syllable:} & \text{| |} & \text{| |} & \text{| |} \\
& \quad \text{tone:} & \text{| |} & \text{| |} & \text{| |}
\end{align*}
\]

---

4. Earlier papers about Delta showed initial deltas with a gap in each stream. The current version of Delta lets rule-writers specify in the stream definition whether a stream should be initialized to contain a gap or nothing. "Nothing" is assumed by default.
The dashes for the height and backness fields in the m and the place and manner fields in the u represent the value <undefined>.

The next statement,

\[(10) \ [\%\text{phoneme } \_\text{left } <\text{cons}> !\text{^ac}] \rightarrow \text{insert } [\%\text{CV } C] \text{^left...^ac};\]

is a rule. A rule consists of a test and an action. The action in this rule is separated from the test by a right arrow (\rightarrow).

The test portion of the rule,

\[(11) \ [\%\text{phoneme } \_\text{left } <\text{cons}> !\text{^ac}]\]

is a delta test, which tests the delta for a particular sequence of tokens and sync marks. Sync marks are referred to in Delta programs by means of pointer variables (also called pointers), such as \text{^left} and \text{^ac} in our sample rule. Pointer variable names in this paper always begin with a carat (^). The variable \text{^left} is a built-in pointer that always points at the leftmost sync mark in the delta, while \text{^ac} (“after consonant”) is a user-defined pointer whose use is explained below.

The sync mark where testing starts is the anchor of the delta test, and is marked by an underscore before the appropriate pointer name. In the above test, \text{^left} is the anchor, so testing starts at the leftmost sync mark in the delta.

The test looks in the phoneme stream immediately to the right of \text{^left} for a token that has the feature <cons>. If such a token is found, the expression !\text{^ac} sets \text{^ac} to point at the sync mark immediately to its right. Since our sample phoneme stream does start with a consonantal token, the delta test succeeds and \text{^ac} is set to the sync mark following the token:

\[(12) \ \text{phoneme: } | \ \text{m } | \ \text{u } | \ \text{s } | \ \text{o } | \]
\[\text{CV: } | \]
\[\text{^left } | \text{^ac}\]

The action of the rule,

\[(13) \ \text{insert } [\%\text{CV } C] \text{^left...^ac};\]

inserts a token named C into the CV stream between \text{^left} and \text{^ac}.
(14) phoneme: | m | u | s | o |
  CV: | C |
  ^left ^ac

The expression ^left...^ac, which specifies where the insertion is to take place, is called the insertion range.

Note that the sync mark pointed to by ^ac, originally defined only in the phoneme stream, is now defined in the CV stream as well. In general, when an insertion is made between two sync marks in the delta, each sync mark is put into the insertion stream if it does not already exist in that stream. A sync mark that is put (defined) in a new stream is said to be projected into that stream.

The final statement in our sample program,

(15) print delta;

displays the delta, showing only the token names in each stream. (Other print statements can be used to display features.) The sample program ends after this statement.

3.4 Adjacent Sync Marks

Adjacent sync marks in a stream act like a single sync mark for purposes of testing the stream. Consider, for example, the morph and tone streams of delta (1):

(16) morph: | root | | root |
  tone: | L | H | L | H |

The following delta test would succeed, despite the intervention of two sync marks between the roots in the morph stream:

(17) [*morph _^left root root]

Thus the floating L tone is in effect “invisible” in the morph stream. It could be made visible by placing a gap between the adjacent sync marks. If a gap were present, the above test would fail.
3.5 One-Point vs. Two-Point Insertions

The insert statement presented in Figure 2 specified a two-point insertion, which places a token sequence between two sync marks already existing in the delta. An insert statement can also specify a one-point insertion, which places a token sequence to the right or left of a single sync mark. For example, given the delta

(18) phoneme:  m u s o  
syllable:  syl syl  
tone:  L L  
      1 2 3 4 5 6

the one-point insert statement

(19) insert [%tone H] ...^5;

would insert a H token into the tone stream just before sync mark 5, automatically creating a new sync mark before the inserted H token:5

(20) phoneme:  m u s o  
syllable:  syl syl  
tone:  L H L  
      1 2 3 4 5 6
(new sync mark)

Note that the new sync mark is unordered with respect to sync marks 2, 3, and 4.

3.6 Sync Mark Ordering

Within each stream, all the sync marks have an obvious left to right ordering. Across streams, however, two sync marks may or may not have a relative left to right ordering. Consider, for example, the following delta:

(21) morph:  root  
phoneme:  m u s o  j a a b i  
syllable:  syl syl  syl syl  
tone:  L H L H  
      1 2 3 4 5 6 7 8 9 10 11 12

---

5. In the examples of one-point insertions in earlier papers about Delta, the new sync mark bounding the inserted token was automatically projected into all streams in which the sync mark designated by the range pointer was defined. The system has been changed so that by default the new sync mark is only defined in the insertion stream. The project option to the insert statement can be used to cause one-point insert statements to work the way they did in earlier versions of Delta, as illustrated in example (55).
In this delta, sync mark 4 is to the left of sync marks 6 through 12, because sync mark 6 is in the tone stream after sync mark 4, and is also in the phoneme stream before sync marks 7 through 12. Similarly, sync mark 4 is to the left of sync mark 8, because sync marks 6 and 7 are to the right of sync mark 4 in the tone stream, and to the left of sync mark 8 in the phoneme stream. However, sync mark 4 is not ordered with respect to sync mark 5, because sync mark 5 does not exist in the tone stream, and there is no sync mark between sync mark 4 and 5 that is to the right of 4 or the left of 5 or vice versa. (Sync mark 4 could just as well have been displayed after sync mark 5.) By the same logic, sync mark 4 is also not ordered with respect to sync mark 2 or 3.

Delta (21) might be posited as an early form for the word muso during the derivation of its surface representation, as explained below. It gives the tone pattern L H to the first root, without synchronizing either tone with a particular syllable.

Delta's merge statement can be used to merge two unordered sync marks into a single sync mark, creating the appropriate synchronizations. For example, if ^3 points at sync mark 3 in delta (21) and ^4 at sync mark 4, the statement

(22) merge ^3 ^4;

would produce the following delta:

(23) morph: | root
phoneme:   | m u s o   |
  syllable: | syl syl   |
tone:     | L H L |
          |  1 2 3 5 6 7 |

(4)

It will be assumed in the remainder of this paper that in the examples that contain deltas with numbered sync marks, ^1 has been set to point at sync mark 1, ^2 at sync mark 2, etc.

3.7 Contexts

The ordering of sync marks is important for determining sync mark contexts, which are in turn important for determining where sync marks can be legally projected and for determining the relationship between tokens across streams, as explained below. The context of a sync mark in stream x is the portion of stream x where the sync mark
could be put without causing it to cross the closest sync mark to its left or right. For example, in the delta

\[
\begin{align*}
\text{morph:} & \quad \text{root} \\
\text{phoneme:} & \quad m \ u \ s \ o \\
\text{syllable:} & \quad syl \ syl \\
\text{tone:} & \quad 1 \ L \ 2 \ H \ 3 \ 4 \ 5 \ 6 \ 7
\end{align*}
\]

the context in the morph stream of sync mark 2 is the root token; the context in the phoneme stream of sync mark 4 is the sequence of phonemes and sync marks between sync marks 1 and 6 (recall that sync mark 4 is unordered with respect to sync marks 2, 3, and 5); the context in the syllable stream of sync mark 5 is the second syl token, and so on;

The left context in stream \(x\) of a sync mark \(n\) is the sync mark that is the left boundary of the context in stream \(x\) of sync mark \(n\). Thus in delta (24), the left context of sync mark 2 in the morph stream is sync mark 1, and the left context of sync mark 5 in the syllable stream is sync mark 3. Similarly, the right context in stream \(x\) of a sync mark \(n\) is the sync mark in stream \(x\) that is the right boundary of the context in stream \(x\) of sync mark \(n\). If sync mark \(n\) is defined in stream \(x\), its context, left context, and right context in stream \(x\) is itself.

The Delta language has the operators `\` and `/` for taking the left and right context of a sync mark. For example, the statement

\[
(25) \ ^x = \ \_ \ / \ \_ \ %\text{syllable} \ ^5;
\]

sets \(^x\) to point at the sync mark that is the left context of \(^5\) in the syllable stream—at sync mark 3 in the case of delta (24).

In the following test, the context operators are used to test whether a H tone is the only token between the left context in the tone stream of sync mark 5 and the right context in the tone stream of sync mark 6. This test can be applied to delta (24) to determine whether the vowel o (surrounded by sync marks 5 and 6) is “contained in” a H tone.

\[
(26) \ [\%\text{tone} \ _\ (\ \_ \ \%\text{tone} \ ^5) \ H \ (\ / \ \%\text{tone} \ ^6)]
\]
(The multiple references to the tone stream could be eliminated by specifying in a previous statement a stream to be assumed if no other is specified. In general, Delta lets users set default streams for different purposes.) This test would fail in the case of delta (24), since the o is “contained in” a sequence of two tones, L H, but would succeed in delta (23), where the merging of sync marks 3 and 4 has provided an ordering of sync mark 5 with respect to the sync mark between the two tones.

3.8 Time Streams

While sync marks can be used to specify gross temporal relationships, such as whether one token is before, after, partway through, or concurrent with another, time streams are needed to make precise temporal specifications—for example, that a token in one stream begins 75% of the way through the duration of a token in another stream, or 50 milliseconds after its end. The tokens in a time stream are different from other (non-gap) tokens in that sync marks can be projected into the middle of them. When a sync mark is projected into a time token, the token is automatically divided into the appropriate pieces. For example, if a sync mark is placed a quarter of the way through a token with a duration (name) of 100, the token would get divided into a token with duration 25 and a second token with duration 75.

A time stream is defined very much like any other stream, as illustrated by the following definition of a time stream called duration:

```
(27) time stream $duration;
    name: int;
    end $duration;
```

The keyword time in the first line indicates that the stream is a time stream.

Any number of time streams can be defined and used in a single delta. For example, a rule-writer might use one time stream for slow speech and another for fast speech, or a rule-writer might use one time stream for actual milliseconds, and another for abstract phonological “time”. Consider, for example, a hypothetical language that is like Bambara in all respects except that intervocalic consonants are ambisyllabic. In the following delta for muso in this language, an abstract time stream called time is used to represent the s as ambisyllabic:
(28) syllable: | syl | syl |
phoneme:   | m  | u  | s  | o  |
time:      | 1  | 1  | .5 | .5 | 1 |
duration:  | 70 | 120 | 200 | 150 |

Note that given such a representation of ambisyllabic, the context operators can be used to determine the phonemic composition of the syllables. The phonemes that comprise a syllable are those phonemes that are between the left context in the phoneme stream of the sync mark before the syllable and the right context in the phoneme stream of the sync mark following the syllable.

In the Delta language, a **time expression** is used to refer to a particular point in a time stream. A time expression consists of a time stream name, a pointer name, a plus or minus sign, and an expression representing a quantity of time. For example, given the delta

(29) phoneme:   | s  | o  | j  | a  |
CV:          | C  | V  | C  | V  |
nucleus:     ... | nuc | C  | nuc | ...
tone:       | H  |   | H  |   
duration:   | 200 | 150 | 140 | 200 |
| 1  2  3  4  5  6  7 |

the following time expressions would all refer to the instant midway between ^2 and ^3:

(30) \%duration ^2 + 75
(31) \%duration ^3 - 75
(32) \%duration ^4 - 75
(33) \%duration ^5 - 215
(34) \%duration ^2 + (.5 * dur(^2...^3))

where **dur(^2...^3)** in the last time expression uses the built-in function **dur**, which returns the duration between the specified points in the delta, in this case ^2 and ^3. In general, the expression after the plus or minus sign in a time expression can be an arbitrarily complex numeric expression.

Delta lets users define a particular time stream as the default stream, so that the stream name can be omitted in time expressions that refer to that stream. Users can also specify in a time stream definition whether time should be measured from left to right or right to left. When time is measured from left to right, a time expression with a positive stream offset, such as expression (30) above, refers to a point that is \( n \) time units to the right of the sync mark specified by the pointer variable, while one with a negative time offset refers to a point \( n \) time units to the left. When time is measured from right to left
(as it might be, for example, for a Semitic language, which is written from right to left), a
positive time offset specifies a time to the left, and a negative one a time to the right. It
will be assumed in the remainder of this paper that the deltas have a single time stream
called duration, that this stream is the default time stream, and that time is measured
from left to right.

Time expressions can be used anywhere ordinary pointer variables can be
used—for example, in the range of an insert statement. Thus, given delta (29) above, the
statements

\[(35) \quad n = (.4 \times \text{dur}(^2..^3));\]

\[
[\%\text{tone _ \}^2 \text{H} //^3] \rightarrow \text{insert \ [\%F0 160] \ (^2 + n)...(^3 - n);}\]

would produce the following result:

\[(36) \quad \text{phoneme:} \quad | s | o |
\text{CV:} \quad | C | V |
\text{tone:} \quad | \ldots | H |
\text{F0:} \quad | \ldots |
\text{duration:} \quad | 200 | 60 | 160 | 60 |
\quad ^1 ^2 ^3
\]

Notice that sync marks have automatically been placed at the appropriate timepoints in
the delta. In general, when a time expression is used where a sync mark is required, the
sync mark is automatically created in the time stream.

Delta has the special operator at for placing a token at a single point in time.
Thus if instead of the rule in (35), the rule

\[(37) \quad [\%\text{tone _ \}^2 \text{H} //^3]
\rightarrow \text{insert \ [\%F0 160] at \ (^2 + (.5 \times \text{dur}(^2..^3))});\]

had been executed, the result would be the following:

\[(38) \quad \text{phoneme:} \quad | s | o |
\text{CV:} \quad | C | V |
\text{tone:} \quad | \ldots | H |
\text{F0:} \quad | \ldots |
\text{duration:} \quad | 200 | 75 | 160 | 75 |
\quad ^1 ^2 ^3
\]

The at operator is special in that it will create two sync marks if no sync marks exist at
the specified point in time, it will add one sync mark if a single sync mark exists, and it
will use the outermost sync marks if several exist. Thus the at operator makes it easy to
place tokens in different streams at the same point in time, and synchronize the tokens with each other.

3.9 Forall Loops

The previous sections have included several examples of rules. Each of these rules was restricted to operating at one particular point in the delta. For example, rule (10),

(10) [%phoneme _^left <cons> !^ac] -> insert [%CV C] ^left...^ac;

is only tested against the first token in the phoneme stream. Usually, however, a rule is meant to apply at all appropriate points across the entire delta. Delta has several control structures for applying rules across the entire delta or selected stretches of the delta. One of the most useful of these is the forall loop, which performs the body of the loop each time the test at the top of the loop succeeds.

Consider, for example, the following forall loop, which applies rule (10) across the entire delta.

(39) loop forall [%phoneme _^bc <cons> !^ac];
    insert [%CV C] ^bc...^ac;
    pool;

The first time the loop is executed, the forall test at the top of the loop is tested with ^bc ("before consonant"), the anchor of the test, automatically set to ^left. If a consonantal phoneme follows ^bc, the loop body is executed, synchronizing a C token in the CV stream with the phoneme. After execution of the loop body, and also whenever the forall test fails to match, the advance pointer ^bc is automatically advanced to the next sync mark, and the test is repeated. This advancing is continued until the advance pointer hits the rightmost sync mark in the delta, in which case the loop terminates.

Assume that loop (39) is being applied to the following delta:

(40) phoneme: | m | u | s | o |
    CV:      | 1 2 3 4 5

First, ^bc would be set at sync mark 1. The forall test would succeed, setting ^ac at sync mark 2, and the body of the loop would insert a C token between ^bc and ^ac:
Then \(^{bc}\) would be advanced to sync mark 2. The forall test would fail, since a vocalic, rather than consonantal, phoneme follows. Pointer \(^{bc}\) would then be advanced to sync mark 3, the forall test would succeed (setting \(^{ac}\) at sync mark 4), and the body of the loop would insert a C token:

\[
\begin{array}{cccccc}
\text{phoneme:} & m & u & s & o \\
\text{CV:} & C & 1 & 2 & 3 & 4 & 5 \\
^b &  &  & c & ^ & a & c
\end{array}
\]

The loop would continue in this fashion until \(^{bc}\) reaches sync mark 5.

Users can override the system’s default assumptions about forall loop application by specifying the initial setting of the advance pointer, where the advance pointer should start from on subsequent iterations of the loop, a particular pointer to advance, the stream through which to advance, in which direction to advance (left to right or right to left), and so on. In general, the forall loop is a powerful construct with which users can specify precisely how their rules should apply.

### 3.10 Token Variables

The program fragments in previous sections have included several examples of pointer variables and one example of a numeric variable (example (35)). In addition, Delta provides token variables, which hold entire tokens. The following forall loop uses a token variable to replace with a single phoneme all token pairs consisting of two identical vocalic phonemes. For example, it replaces two adjacent a’s with a single a:

\[
\begin{array}{l}
\text{(43) loop forall [ ^phoneme _bv <~cons> !_vowel $vowel $vowel !^av ];}
\text{insert [ ^phoneme $vowel] ^bv...^av;}
\text{pool;}
\end{array}
\]

Assume that this loop is being applied to the following delta:

\[
\begin{array}{l}
\text{(44) phoneme:} & j & a & a & b & i \\
\end{array}
\]

(The two a tokens might be used in the initial underlying representation for jaabi to represent the long vowel [a:], as discussed below.) The forall test first looks for a vocalic
(<~cons>) phoneme following ^bv ("before vowel"). It will succeed when ^bv precedes the first a. The expression !$vowel puts a copy of the first a token in the token variable $vowel. (Token variable names in this paper always start with a dollar sign.) The next expression, $vowel, tests the next token to see whether it is the same as the token in $vowel—that is, whether it is also a. If so, ^av is set after this vowel, and the body of the loop inserts a copy of the token in $vowel between ^bv and ^av, thereby replacing the two vowels and their intervening sync mark with a single a:

(45) phoneme: | j | a | b | i |

### 3.11 Fences

It is often necessary to prevent a delta test from crossing a linguistic boundary, such as a morph boundary. Delta has a special construct, called a fence, for limiting the scope of a delta test. A fence is a pair of built-in pointer variables, ^lfence and ^rfence. ^lfence delimits the left side of the fence, and ^rfence delimits the right.

Fences are often used in conjunction with forall loops. For example, the forall loop below, whose body is itself a loop, adds a fence to loop (43) to restrict it to operating only on identical vowels that are within the same morph (see below for a simpler, unnested loop that accomplishes the same thing):

(46) loop forall [%morph _^lfence <> ^rfence];
    loop forall [%phoneme _^bv <~cons> !$vowel !$vowel !^av]
        advance from ^lfence;
    insert [%phoneme $vowel] ^bv...^av;
    pool;
pool;

The empty angle brackets in the outer forall loop match any token in the morph stream. Thus the outer loop surrounds the morph with ^lfence and ^rfence. For each morph, the inner forall loop reduces two identical vowels to a single one. The expression advance from ^lfence initializes the advance pointer, ^bv, to ^lfence, the beginning of the morph matched by the outer loop. The forall test in the inner loop will fail if it has to cross one of the fence variables in order to succeed. Thus it will not match vowels across a morph boundary.
In loop (46), the sole purpose of the outer forall loop is to set a fence. An alternative way to set the fence that does away with the outer loop altogether is to include the option `fence %morph` after the delta test in the inner loop:

(47) loop forall [%phoneme _^bv <-cons> !$vowel $vowel !^av] 
    insert [%phoneme $vowel] ^bv...^av;  fence %morph; 
    pool;

The expression `fence %morph` prevents the forall test from crossing a morph boundary, just as the outer loop did in the previous example.

3.12 Conclusion

This section has presented the delta data structure, and has shown various ways in which this data structure can be tested and manipulated. The concepts and constructs presented—streams, tokens, sync marks, contexts, delta tests, one-point and two-point insertions, time streams, forall loops, fences, and others—are by no means a complete inventory of Delta language features; rather, they have been strategically selected for presentation in order to provide enough information to follow the sample programs in the next two main sections, which show applications of the Delta language to Bambara and English.

4 Modeling Tone and Fundamental Frequency Patterns in Bambara

This section presents some sample programs that illustrate how Delta can be used to formalize and test a model of Bambara tone realization. This model is actually an integration of two separate models, a model of phonological tone assignment based on the work of Rialland and Sangare (1985), and a model of the phonetic realization of the phonological tones based on the work of Mountford (1983). Only the facts concerning monosyllabic, bisyllabic, and trisyllabic non-compound words in Bambara are considered. Words with four syllables exist, but are rare.

4.1 A Model of Bambara Tone Assignment

Each Bambara word has an inherent (unpredictable) tone pattern. Monosyllabic and bisyllabic words either have a low or a high tone on all the syllables. For example,
the word *muso* is inherently low-toned ([mûsò]), and the word *jaabi* is inherently high-toned ([jâ:bí]). The reason that *muso* was shown throughout this paper with a high tone on the final syllable will become clear in the ensuing discussion.

A trisyllabic morph can have one of several patterns: 1) It can be inherently high-toned or low-toned in the same way as the monosyllabic and bisyllabic words are. For example, the morph [gàlàmà] ‘ladle’ is inherently low-toned, while [sûngûrûn] ‘girl’ is inherently high-toned. (In the transcriptions, a word-final [n] or an [n] before a consonant marks the preceding vowel as nasalized.) 2) It can have the pattern low-high-low, as in [sàkènè] ‘lizard’. 3) It can have the pattern low-high or high-low. In this case, it is optional whether the first tone is associated with the first syllable and the second tone with the last two syllables, or the first tone with the first two syllables and the last tone with the last syllable. For example, the high-low morph *mangoro* ‘mango’ can be realized as [mângòrò] or [mângórò]. Only the second realization will be considered for purposes of the program shown below.

In addition to the inherent tones of morphs, Rialland and Sangare posit two other kinds of tones in underlying representations of Bambara utterances: 1) a floating low tone that follows definite noun phrases, as illustrated in delta (1) above, and suggested by the work of Bird (1966), and 2) a floating high tone suffix that follows all content morphs (as opposed to function morphs).

A floating low tone definite marker is not associated with any particular syllable in surface representations; it serves to trigger tonal downstep at the phonetic level, the lowering in fundamental frequency of following high tones.

The high tone suffix for content morphs is realized on different syllables, depending on whether the content morph is part of a definite noun phrase—that is, on whether a floating low tone immediately follows the content morph. If there is no floating low tone, the high tone is realized on the first syllable of the morph following the content morph, if there is one. Otherwise, it is realized on the last syllable of the content morph itself.

For example, consider the phrase *muso don* ‘It’s a woman’, which has an indefinite noun phrase, and hence, no floating low tone. In this phrase, the high tone associated with *muso* is added to the inherently low-toned *don*, producing the pattern
[mùsò ðòn], where the circumflex accent ^ represents the tone pattern high-low. (Here, the content tone has been added to the tone of the one-syllable morph. In a polysyllabic morph, the high tone simply replaces the inherent tone associated with the first syllable.) On the other hand, the phrase muso don ‘It’s the woman’, which has a definite noun phrase, is realized as [músó ðòn]. See Rialland and Sangare (1985) and Bird, Hutchison, and Kanté (1977) for a fuller description and analysis of Bambara tone patterns.

4.2 Formulation of the Tone Model in Delta

According to the model just presented, Bambara tone patterns consist of sequences of high and low tones. Each tone is either an inherent part of a morph, a floating definite marker, or a floating content marker, as illustrated in Figure 3, which shows several underlying representations of Bambara utterances in delta form. In the transcription in example 5 in the figure, the wedge ^ represents the tone pattern low-high.

***************

Insert Figure 3 here

***************

Given underlying forms like these, the correct tone pattern can be assigned in two main steps:

1. Assign each floating H tone to the appropriate morph. Attach it to the end of a preceding morph if a floating L tone follows. Otherwise, attach it to the beginning of the following morph.

2. Merge unordered pairs of syllable and tone sync marks from right to left until there are no unordered pairs left, thereby creating the appropriate synchronizations between tones and syllables.

Step 1 would create the forms shown in Figure 4 for the sample deltas in Figure 3:
Step 2 would operate on the deltas in Figure 4 to produce those shown in Figure 5.

Note that this step does not change delta 1, since this delta contains no sync marks in the syllable and tone streams that are unordered with respect to each other. In delta 5, the right to left merging of sync marks correctly leaves the first syllable synchronized with two tones, L and H. Left to right merging of the sync marks would produce the wrong result.

A final step might be to combine adjacent \( H \) tones and \( L \) tones into single \( H \) and \( L \) tones, but it does not make any difference for purposes of the program below that generates fundamental frequency patterns on the basis of the tone patterns whether a single tone is associated with several syllables, or each of the syllables has its own tone.

Step 1 can be accomplished in Delta by the forall loop shown in Figure 6.

The forall test

\[(48) \text{ ([$\text{tone } ^\text{bh} H \! ^\text{ah}] \& [\text{morph } ^\text{bh} ^\text{ah}])}\]

contains two parts. The first matches any \( H \) token after \( ^\text{bh} \), setting \( ^\text{ah} \) after it. The second part tests whether \( ^\text{bh} \) and \( ^\text{ah} \) point at adjacent sync marks in the morph stream—i.e., whether the \( H \) token matched by the first part is a floating tone.
The second step, which merges the appropriate sync marks in the syllable stream with those in the tone stream can be expressed as the forall loop shown in Figure 7.

Insert Figure 7 here

The outer forall loop invokes the inner simple loop for each token in the morph stream. A simple loop continues to repeat until a statement in the loop body, in this case an exit statement, causes the loop to terminate. The loop moves pointers \textsuperscript{\texttt{bt}} and \textsuperscript{\texttt{bs}} right to left through the morph, merging the appropriate sync marks, and terminating whenever \textsuperscript{\texttt{bt}} or \textsuperscript{\texttt{bs}} equals \textsuperscript{\texttt{bm}}—that is, when one of these pointers reaches the beginning of the morph.

4.3 Building Underlying Representations Using Dictionaries

The program fragments shown above assume an underlying representation in which each morph is associated with a tone pattern, and floating high and low tones are present where appropriate. A Delta program designed to test the above program fragments would have to build the appropriate underlying representations in the first place. To prevent users from having to enter underlying representations for the utterances they wish to test (a tedious task), the program could be designed such that the user only has to specify phoneme names, morph boundaries, and the unpredictable low floating tones. For example, our sample utterance \textit{muso jaabi} might be entered as follows:

(49) + muso ` + jaabi +

where + designates a morph boundary and ` a floating low tone. Two a tokens in succession represent the long vowel [a:], as discussed earlier. On the basis of this information, the program could easily build the following structure:

(50) morph: | root | root
phoneme: | m | u | s | o | j | a | a | b | i
CV: | C | V | C | V | C | V | V | C | V
tone: | | L |
The next step would be to insert the inherent morph tones and the floating high content tones. Since neither the inherent tones nor the content tones are predictable from this delta, a dictionary must be used. Delta has two kinds of dictionaries, action dictionaries and set dictionaries, or simply sets, both of which are useful for creating underlying representations from forms like the above.

An action dictionary contains token name strings (henceforth “search strings”) and associated actions. An action for an entry, which can consist of any legal Delta statements, is automatically performed whenever a search string is matched—that is, an identical token name sequence is looked up in the dictionary. For our purposes, an action dictionary named morphs might be defined to contain all morphs (represented in terms of phoneme names) and associated tone patterns. More specifically, the action for each morph might be an insert statement that inserts the appropriate tones for the morph into the delta, as shown in Figure 8.

******************************************************************************

Insert Figure 8 here

******************************************************************************

An action dictionary always has four pointer variables associated with it, in this case ^b, ^e, ^1, and ^2. The first two pointer variables take on the values of the pointers delimiting the token name sequence being looked up. The last two pointer variables are special pointers that can be used inside the dictionary search strings to isolate parts of the dictionary entries to which the associated actions need to refer. These special pointers are not used in this example, and can be ignored. See Hertz et al. (1985: 1597-1598) for an example of their use.

6. In earlier papers about Delta, the sample action dictionaries were unnamed, reflecting the then-current version of the system in which it was only possible to define a single, unnamed action dictionary. Delta has been enhanced to allow rule-writers to use any number of named action dictionaries.

7. Bambara contains sets of morphs that differ phonetically only in their tone pattern. The dictionary action for such morphs might prompt the user for the intended tone pattern. Alternatively, the user might annotate the input string with the intended tone pattern, or, in some cases, the system might be able to perform some syntactic analysis to determine the correct pattern automatically.
The following forall loop operates on a delta of the form shown in delta (50) above, looking up in dictionary morphs, the phoneme names associated with each morph in the delta:

(51) loop forall [%morph _^bm <> !_^am];
    find %phoneme _^bm..._ ^am in morphs;
    pool;

If the phoneme sequence is found, the dictionary automatically synchronizes the appropriate tones with the morph, creating, for example, the following deltas for muso jaabi and mangoro don:

(52) morph:       | root       | | root       |
    phoneme:   | m u s o   | | j a a b i  |
    tone:         | L         | | H         |

(53) morph:       | root       |
    phoneme:   | m a n g o r o   |
    tone:         | H         | | L         |

The next step would be to insert the floating high content tones after the content morphs. This step could be accomplished by adding the appropriate insert statement to the action of all content morphs in dictionary morphs. However, since there are so many content morphs, a simpler strategy would be to “mark” the non-content morphs as function morphs, and insert the content tone for all morphs that are not function morphs. The simplest way to mark the appropriate morphs as function morphs is to place them in a set.

A set contains search strings, but no actions. For example, one might define a set called function_morphs as follows:

(54) set function_morphs contains %phoneme: a, d on, ...;

(Since all of the morphs in set function_morphs are also in the action dictionary morphs, an alternative to the above set statement could be used in which each function morph in the action dictionary is followed by an expression that places the morph in set function_morphs.) Given set function_morphs, a rule could be added to forall loop (51) above to insert a H tone after any morph not found in the set:

(55) loop forall [%morph _^bm <> !_^am];
    ...  
    ~find %phoneme _^bm..._ ^am in function_morphs
      -> insert [%tone H] ^am...  project;
    pool;

The option project at the end of the insert statement specifies that the new sync mark after the inserted H tone should be projected into all streams in which ^am, the range pointer, is defined. This rule would operate on deltas (52) and (53) to produce the following underlying representations for muso jaabi and mangoro don:

(56) morph: root | root
phoneme: n u s o | j a a b i
CV: C V C V | C V V C V
tone: L | H L | H | H

(57) morph: root
phoneme: n an g o r o | d on
CV: C V C V C V C V | C V
tone: H | L | H L L

A final step in creating the underlying representation would be to reduce sequences of identical vocalic phonemes that are not separated by a morph boundary, such as the a a of jaabi, into a single phoneme, as shown earlier in example (47).

4.4 A Model of Bambara Fundamental Frequency Patterns

On the basis of the CV and tone streams, the appropriate F₀ pattern for the tones can be determined. This section considers one possible strategy for F₀ determination, based on the work of Mountford (1983). According to Mountford, high and low tones descend along relatively straight and independent lines, as shown schematically in the diagram of a sentence with the tone pattern H L L H L H L in Figure 9.

***************

Insert Figure 9 here

***************

The starting and ending frequencies of the baseline are relatively independent of the sentence duration. Thus the slope of the baseline varies with sentence duration. The sample program below uses a starting frequency of 170 hertz and an ending frequency of 130 hertz for the baseline. The starting and ending frequencies of the topline are a function of sentence duration. For the sake of simplicity, this detail is ignored in the
sample program, which assumes a starting frequency of 230 hertz and an ending frequency of 150 hertz.\textsuperscript{8}

My own very preliminary laboratory studies of Bambara tone patterns suggest that the F\textsubscript{0} tone targets are generally realized halfway through each syllable nucleus—that is, long vowels behave the same way as short vowels for the purpose of F\textsubscript{0} target placement. Furthermore, the F\textsubscript{0} targets tend to have no durations of their own, serving only to shape the overall F\textsubscript{0} contour.

4.5 Formulation of the Fundamental Frequency Model in Delta

The model of F\textsubscript{0} assignment just outlined can easily be implemented in Delta, as shown by the program fragment in Figure 10, which generates a descending topline and a descending baseline, and determines the appropriate F\textsubscript{0} values along those ramps. The program assumes that each phoneme was given a duration by earlier statements.

***************

Insert Figure 10 here

***************

The forall loop in this program would generate the following values for our sample delta for *muso jaabi* (assuming the segment durations shown and a sentence duration of 1080 milliseconds):

\begin{verbatim}
| morph: | root | s | o | j | a |
| phoneme: | m | u | C | V | C | V |
| CV: | C | V | C | V | C | V |
| nucleus: | nuc | nuc | nuc | nuc | nuc | ...
| syllable: | syl | syl | syl | syl | ...
| tone: | L | H | L | H |
| F0: | 165 | 195 | 172 |
| duration: | 170 | 60 | 60 | 200 | 75 | 0 | 75 | 140 | 195 | 0 | 195 |
\end{verbatim}

The program has been oversimplified for purposes of illustration. For example, it does not handle single syllables that have two tones, such as *don* in utterance 1 in Figure 5 above, nor does it compute the starting and ending frequencies of the high tone line as

---

\textsuperscript{8} Current research on downtrend in other languages suggests that this model may be oversimplified. Since Delta has fully general numeric capabilities, it should be equally well-suited for expressing other algorithms for computing F\textsubscript{0} values.
a function of sentence duration. Furthermore, it does not handle the tone raising and lowering phenomena that occur in Bambara for tones in particular contexts. Rialland and Sangare (1985) posit a rule that downsteps (lowers) a high tone and all subsequent high tones after a floating low tone. See also Mountford (1983) for some posited raising and lowering rules. Our sample program could easily be expanded to handle all of these phenomena.

In addition to F₀ values, a program designed for synthesis would have to compute or extract from a dictionary the values for other synthesizer parameters. Delta’s `generate statement` can be used to create a file of parameter values and durations for the synthesizer on the basis of the parameter streams and an associated time stream:

(59) generate (%duration, %F₀, %F₁, %F₂, %F₃, ...);

Also, users can tailor their program to their particular synthesizer setup by writing their own C programs to generate values for the synthesizer, and their own synthesizer driver. The ability to interface C programs at will also lets Delta be used in conjunction with demisyllable and diphone libraries.

5 Modeling English Formant Patterns

The section just completed illustrated how Delta can be used to formalize and integrate a phonological model of tone assignment and a phonetic model of tone realization. This section further demonstrates Delta’s flexibility by drawing from my own work on modeling English formant patterns to show how different hypotheses I have come up with over the years can be formulated in Delta. The section focuses on my most recent model, in which formant targets and intervening transitions are represented as independent durational units that are related to higher-level phonological constituents in well-defined ways. This model relies critically on the concept of synchronized units at both the phonological and phonetic levels.

5.1 Model 1: Linear Utterance Representations, Implicit Transitions

The first work I did on modeling English formant patterns was in the context of my students’ and my synthesis rule development. Our SRS synthesis rules for English, developed between 1978 and 1982, are based on the hypothesis that every phoneme (i.e.,
phoneme-sized unit) has an intrinsic duration that is modified according to such factors as segmental context and stress. Formant targets (usually two of them) are set in relation to the segment’s duration—for example, 20% and 80% of the way through the segment. For the most part, all durational adjustments, such as stretching before voiced segments, are made between the formant targets within given segments, so that the durations of the formant transitions between targets in adjacent segments remain constant.

5.2 Model 2: Multi-level Utterance Representations, Implicit Transitions

In the course of implementing a set of SRS rules based on the model just outlined, we realized that our formant target and duration rules might be simplified if we treated certain sonorant sequences that act as a single syllable nucleus, such as [ei], [ai], and [ar], as two units for the purpose of assigning formant targets, and as single units for other purposes, such as assigning amplitude patterns. (A syllable nucleus in this model consists of a vowel + a tautosyllabic sonorant, if such a sonorant exists.) Such a structure was impossible to represent straightforwardly with SRS, which relies on linear utterance representations, but is quite simple to represent with Delta, as shown below for the word ice:

(60) syllable: | syl
nucleus: | nuc | |
phoneme: | a | i | s |

By representing the [i] of syllable nuclei like [ai] as an independent phoneme token, the rules can assign a single target value for each formant—say, an F2 value of 2000 hertz—to all i’s, regardless of their context. Syllable nuclei can still be treated as single units where appropriate.

While this model simplifies the prediction of formant values, however, it leads to complicated rules for positioning the formant values with respect to the edges of segments, since the formant target positions for a token in one context (e.g., i as the sole component of the nucleus) are not necessarily the same as those in another (e.g., i at the end of a diphthong).

9. In all our SRS rules, not just those for English, some segments only have a single target for a particular formant, and others, like [h] in English and [r] in Japanese, have no targets for some formants. Non-initial English [h] is modeled by linear transitions connecting the last formant targets in the preceding segment and the first in the following segment.
5.3 Model 3: Multi-level Representations, Explicit Transitions

My recent durational studies support the view that formant transitions are relatively stable in duration. To avoid the complicated target placement rules alluded to above, however, I am exploring a new model, in which the transitions are represented as independent durational units, as shown below for the word *ice*:

\[
\begin{array}{c|c|c|c}
\text{syllable:} & \text{syl} & \\
\hline
\text{nucleus:} & \text{nuc} & \\
\text{phoneme:} & a & i & s \\
\text{F2:} & 1400 & 2000 & 1800 \\
\text{duration:} & 45 & 90 & 30 & 40 & 90 \\
\end{array}
\]

Note that the formant transitions are represented by adjacent sync marks in the phoneme and F2 streams, with intervening durations in the duration stream.

In this model, each formant target is synchronized with an entire phoneme token, and the duration-modification rules modify entire phoneme durations. In diphthongs, such as [ai], the duration rules modify only the first portion of the diphthong (e.g., the [a] of [ai]), thereby keeping the duration of the transition portion of the diphthong (e.g., from [a] to [i]) constant.

Below are some forall loops that might be used to insert formant values and formant transitions, and to modify phoneme durations in the appropriate contexts, building a delta like (61) above. These examples assume that earlier statements have inserted the appropriate initial phoneme durations so that the delta for *ice* looks as follows before the loops apply:

\[
\begin{array}{c|c|c|c}
\text{syllable:} & \text{syl} & \\
\hline
\text{nucleus:} & \text{nuc} & \\
\text{phoneme:} & a & i & s \\
\text{F2:} & \\
\text{duration:} & 75 & 30 & 90 \\
\end{array}
\]

The following forall loop matches each phoneme, synchronizing a value for each formant with the phoneme:
(63) loop forall [%phoneme _^1 <> !^2];

:: F2 values:

   if
   [ %phoneme _^1 high]   -> insert [%F2 2000] ^1...^2;
   [ %phoneme _^1 low]    -> insert [%F2 1400] ^1...^2;
   [ %phoneme _^1 alveolar] -> insert [%F2 1800] ^1...^2;
   fi;

...;

pool;

This loop operates on delta (62) to produce the following:

(64) syllable: | syl
nucleus:    | nuc
phoneme:    | a | i | s
F2:         | 1400 | 2000 | 1800
duration:   | 75 | 30 | 90

The next loop modifies the durations of vowels in certain contexts. In the case of delta (62), it shortens the duration of a to 60% of its previous duration, since it is in a nucleus that precedes a tautosyllabic voiceless segment. The forall option fence %syllable prevents the delta tests in the loop body from crossing a syllable boundary.

(65) loop forall [%phoneme _^bv <-cons> !^av] fence %syllable;

   [ %nucleus _^bv <nuc> [%phoneme <-voic>]]
   -> dur(^bv...^av) *= .6;

...;

pool;

where the expression

dur(^bv...^av) *= .6

is a shorthand for

dur(^bv...^av) = .6 * dur(^bv...^av)

This loop shortens the a in delta (64) from 75 milliseconds to 45:

(66) syllable: | syl
nucleus:    | nuc
phoneme:    | a | i | s
F2:         | 1400 | 2000 | 1800
duration:   | 45 | 30 | 90

A more complete program would modify the duration of consonants as well, depending, for example, on their position in the syllable and whether the syllable is stressed.
Finally, the following forall loop inserts transitions between adjacent phonemes:

(67)   loop for all [%phoneme _^1 <> !^2];
    if
        [_^1 <sonorant> <obstruent>]
            -> insert [%duration 40] ^2... project;
        [_^1 <sonorant> <sonorant>]
            -> insert [%duration 90] ^2... project;
    fi;
pool;

This loop operates on delta (66) to produce the following:

(68) syllable: | syl
    nucleus: | nuc
phoneme:   | a | i | s
F2:        | 1400 | 2000 | 1800
duration:  | 45 | 90 | 30 | 40 | 90

The transition rules presented here are hypothetical; at the time of writing (December 1987), I have not yet investigated in any systematic way what factors determine the transition durations. The transition durations probably depend as much on the place of articulation of the two phonemes as on their manner, or possibly even directly on the formant values they connect, or on other as yet to be discovered factors.

Independent developments in autosegmental CV theory (e.g., Clements and Keyser, 1983) have suggested some minor revisions to our current model, as illustrated by the following representation for ice, in which the diphthong-final i is differentiated from the vowel i by virtue of being synchronized with a C, rather than with a V:

(69) syllable: | syl
    nucleus: | nuc
CV:        | V | C | C
phoneme:   | a | i | s
F2:        | 1400 | 2000 | 1800
duration:  | 45 | 90 | 30 | 40 | 90

The remainder of this discussion will assume a representation like the one in delta (69).

Given a CV stream, one might hypothesize that every C and V is associated with a single value for each formant, reflecting the segment’s place of articulation. However, this assumption immediately gets us in trouble with [h], which Nick Clements and I studied recently over a period of two semesters. [h] exhibits no targets of its own; rather, it acquires its formant structure on the basis of the preceding and following segments, its formants looking very much like aspirated transitions from the last formant target in the
preceding segment to the first formant target in the following segment. Consider, for example, the schematic representation of the second formant pattern of the sequence \([g \ h \ a \ i \ s]\), as in *big heist*, in Figure 11.

The traditional way to segment this utterance would be into the four chunks shown in the figure. A problem caused by this segmentation, however, becomes apparent when we consider the second formant pattern of the sequence \([g \ a \ i \ s]\), the same sequence without the \([h]\), shown in Figure 12.

The second formant patterns of the two sequences are for all intents and purposes identical, except that the one with \([h]\) has aspiration, rather than voicing, during the transition from the \([g]\) to the target for the \([a]\). In order to write rules assigning a duration to the \([a]\) in each of these examples, one would have to posit a rule that shortens the vowel after \([h]\). Such a rule would be very unusual in English, since modifications to English vowel durations are not generally triggered by preceding segments. Furthermore, in order to generate the F2 pattern in the shortened vowel, one would have to move the target by precisely the amount that the vowel has been shortened.

Given the transition model posited above, we might hypothesize instead that \([h]\) adds no duration of its own, but rather, is realized by aspiration superimposed on the transition between the preceding and following segment, as illustrated in the following delta for *big heist*:

\[
\begin{align*}
\text{syl} & : & \text{nuc} \\
\text{nucleus} & : & \text{CV} : & \text{CV} : & \text{CV} : & \text{CV} : \\
& & & \text{CV} : & \text{CV} : & \text{CV} : \\
& & & \text{CV} : & \text{CV} : & \text{CV} : \\
& & & \text{CV} : & \text{CV} : & \text{CV} : \\
\text{phone} & : & \text{asp} & \text{asp} & \text{asp} & \text{asp} \\
\text{aspiration} : & & \text{V} & \text{V} & \text{V} & \text{V} \\
\text{F2} : & & \text{2000} & \text{1400} & \text{2000} & \text{2000} \\
\text{asp_amp} : & & \text{15} & \text{15} & \text{15} & \text{15} \\
\text{duration} : & & \text{50} & \text{60} & \text{45} & \text{90} & \text{30} \\
\end{align*}
\]
where \texttt{asp\_amp} is a stream containing amplitude values for aspiration. Here \texttt{[h]} is represented as a \texttt{C} in the \texttt{CV} stream, with an associated \texttt{asp} token in an aspiration stream, but no associated phoneme token. Since the tokens in the \texttt{phoneme} stream have values for place and manner of articulation, among others, \texttt{h} in this representation is phonologically unspecified for these attributes. The absence of \texttt{h} in the \texttt{phoneme} stream simplifies the transition rules, which can refer directly to the sequence \texttt{g\_a} in determining the transition duration.\textsuperscript{10} (The \texttt{C} is necessary in the \texttt{CV} stream, since /h/ functions phonologically as a consonant, preventing flapping of a preceding /t/ or /d/, for example.)

This model makes the prediction that the duration of the aspiration associated with \texttt{h} will be no greater than the duration of the transition between segments. For many speakers, however, \texttt{[h]} seems to consistently add a small amount of duration, on the order of 25 milliseconds, to the transition. For such speakers, an additional rule would be required that adds a fixed duration increment for \texttt{h} to the transition.

An obvious question that arises is whether aspiration in stops might be represented in a way similar to that for \texttt{h}, as in the following delta for \textit{tie}:

\begin{verbatim}
(71) syllable: | syl | | nuc
CV: | C | | V | | C
phoneme: | t | | a | | i
aspiration: | asp |
F2: | 1800 | | 1400 | | 2000 |
duration: | 70 | | 60 | | 75 | | 90 | | 30 |
\end{verbatim}

Note that the token \texttt{asp} in this case is not associated with a \texttt{C} or \texttt{V} in the \texttt{CV} stream, reflecting its different phonological status from that of \texttt{h}. Note also that the \texttt{asp} token is not associated with any phoneme, which completely eliminates the much-debated problem of whether aspiration of stops in English should be considered for segmentation purposes as part of the preceding stop or part of the following vowel.\textsuperscript{11}

\textsuperscript{10} The phonetic properties of \texttt{[h]} provide excellent evidence that /h/ must be treated as phonologically underspecified, a point made by Keating (1985).

\textsuperscript{11} It is interesting to note that Peterson and Lehiste (1960) found that when aspiration is considered to be part of the vowel, the vowel is lengthened by about 25 milliseconds, the same duration increment we found for \texttt{[h]}.
This model is undoubtedly oversimplified in several ways. For example, aspiration does not always coincide with the entire transition; sometimes it stops before the end of the transition and sometimes after. Also, transitions for different formants do not always reach their targets at the same time. The model could be extended to handle these cases by adding additional duration tokens at the appropriate points. For example, a transition that is only partially aspirated could be modeled as two duration tokens, the first synchronized with an asp token in the aspiration stream, and the second with a voicing token in the voicing stream (not shown above).

Certainly the model has been oversimplified in other ways as well. The discussion has not been meant to give a full account of this model or even to be correct in all the details presented; the model is still very much under development at the time of writing. Rather, the discussion is intended to show how a model of this type, which relies heavily on synchronized phonological and phonetic units on different levels, can easily be accommodated in Delta, allowing linguists to test their ideas and revise them in the ways they see fit.

6 Final Remarks

This paper has presented selected features of the Delta programming language and two linguistic models formulated in that language, one for Bambara tone and fundamental frequency patterns and one for English formant patterns. The paper has focused on Delta’s central data structure, which represents utterances as multiple streams of synchronized units. This data structure gives rule-writers considerable flexibility in expressing the relationship between phonological and phonetic units.

The examples of programming statements in the paper have been meant to give the flavor of the Delta language, rather than to describe all possible Delta constructs. No complete programs were presented, other than the oversimplified program in Figure 2. Unlike the program in that figure, Delta programs generally consist of several procedures. Variables in Delta can either be local to a particular procedure or global to all procedures.

Delta has several kinds of statements not described in the paper. For example, in addition to the insert, delete, and merge statements for modifying deltas, Delta has a mark statement for changing the attributes of tokens and a project statement for projecting an existing sync mark into a new stream; and in addition to the forall loop and simple loop,
Delta has a *while loop*, which performs the body of the loop so long as the test at the top of the loop succeeds. Delta also has far more flexible input and output facilities than the simple read and print statements illustrated, and it has fully general numeric capabilities.

Delta tests can also include many kinds of expressions not illustrated, including expressions that test for optional occurrences of a pattern in the delta, for one of a set of alternative patterns, and for a pattern that must *not* be present for the test to succeed. Delta tests can even be temporarily suspended in order to execute arbitrary statements, a capability that adds enormous power to the language.

Delta also has many kinds of tests not illustrated at all, including procedure calls, which succeed or fail according to the form of the return statement in the called procedure. A particularly nice feature of the Delta language is its ability to group tests of different types into a single test—such as a delta test, a dictionary lookup, and a procedure call, making a single action dependent on the success of several tests.

The interactive source-level debugger has not been illustrated at all, but it is an essential part of the system that greatly enhances the system’s utility as a rule-development tool. With the debugger, users can issue commands to temporarily stop execution each time the delta changes, each time a particular variable changes, each time a particular line in the program is executed, each time a particular procedure is called, and so on. When the program stops, users can examine the contents of the delta, the contents of program variables, the contents of the run-time stack, and so on; they can also modify the delta, making any of the delta manipulations allowed by the Delta language. The debugger also gives users very flexible facilities for specifying the files that programs should read from and print to, and it lets users create logs of their terminal session and their program input and output.

A particularly interesting use of the debugger is as a well-formedness checker. Commands could be issued to the debugger that cause it to execute a particular procedure each time the delta or a particular stream in the delta changes. This procedure could test the delta to make sure that whatever manipulations have been made to it do not violate any well-formedness conditions of the language in question. Furthermore, the debugger could be instructed to print out any statement that causes such a condition to be violated, making it easy for rule-writers to find out exactly what parts of their program cause such constraints to be violated.
The debugger can be invoked interactively or via an `execmd statement` in a Delta program. A program might use an execmd statement, for example, to invoke the well-formedness commands discussed above, so that the program automatically tests for any violations of well-formedness conditions whenever it is executed.

At the time of writing (December, 1987) most, but not all of the features of Delta described in this paper have been implemented. At the time of publication, we expect to have implemented all of these features and to have begun implementing those planned for the next version of the system. This version will let users construct and test arbitrary data structures, such as trees, arrays, and graphs. A user might, for example, choose to represent the tokens in a particular stream as trees of features, as in the multi-dimensional model of autosegmental phonology proposed by Clements (1985).\textsuperscript{12} We also plan to add a macro processor to the system, which will let users tailor the syntax of their program statements to their particular needs.

In addition to enhancing the existing system, we are considering implementing a new rule interpreter that would let linguists interactively enter, test, and modify rules without recompiling the rules, much like SRS (Hertz, 1982) does. Like SRS, the system would be oriented to the novice rule-writer with no prior programming experience, and would have several built-in assumptions about how the rules should apply—for example, top to bottom through the rules, left to right across the delta. Unlike SRS, however, the rules accepted by the system would be in Delta form, and would thus be able to build and manipulate multi-level structures (deltas). This rule interpreter would be particularly useful as an instructional tool for beginning rule-writers, and would be a good stepping stone to the full-fledged system. At a later time, we may implement an interpreter for the entire Delta language.

Because the Delta System is still under development at the time of writing, little has been said about its performance or ease of use. We are encouraged, however, by our experience with an early version of the system. Students in speech synthesis classes felt comfortable writing programs with this version in anywhere from a few days to a few weeks. While large programs written with this version do not generally run in real-time,

\textsuperscript{12} In the current system, rule-writers can use C to construct data structures of their choice. For example, they could construct tree-structured feature representations for tokens by using a numeric field in a token as a pointer to a tree structure defined in C.
we have designed the current version of the system with real-time rule execution as a primary goal.

An immediate plan of mine is to use the Delta System to further explore the model of English formant transitions hypothesized above. (It is possible that at the time of publication of this paper, the model will have been altered in significant ways.) Users elsewhere plan to use Delta to explore many other kinds of models, including articulatory ones. Because of its ability to accommodate different models, the Delta System should help us increase our understanding of phonology and phonetics, and learn more about the interface between the two.

Acknowledgements

The author is the President of Eloquent Technology. The Delta System is an Eloquent Technology product, developed in cooperation with the Department of Modern Languages and Linguistics and the Computer Science Department at Cornell University. Several people have contributed to the system. Kevin Karplus and Jim Kadin continue to be instrumental in its design, and Jim Kadin also in its implementation.

I thank Mary Beckman, Nick Clements, John Drury, Greg Guy, Jim Kadin, Peggy Milliken, and Annie Riallend for their helpful suggestions on earlier drafts of this paper, and John McCarthy for his insightful comments on the version of this paper presented at the First Conference on Laboratory Phonology at the Ohio State University in June 1987. These comments motivated several of the revisions in this paper.

References


:: Phoneme Stream Definition:

stream %phoneme;

:: Fields and Values:

  name:    m, n, ng, b, d, j, g, p, t, c, k, ...,
          i, in, l, e, en, E, a, an, ...u, un;
  place:   labial, alveolar, palatal, velar, ...;
  manner:  sonorant, obstruent, ...
  class:   cons;
  nasality: nasal;
  voicing: ~voiced;
  height:  high, mid, low;
  backness: front, central, back;

:: Initial Features:

  m  has labial, sonorant, nasal, cons;
  n  like m except alveolar;
  ...
  j  has voiced, palatal, stop, cons;
  g  like j except velar;
  ...
  u  has back, high;
  un like u except nasal;

end %phoneme;

:: F0 Stream Definition:

stream %F0;
  name: int;  :: defines all integers as possible names
end %F0;

Figure 1. Sample Delta Definition.
:: Delta Definition:
<delta definition goes here>

:: Main Body of Program:

proc main();

:: Read phonemes from terminal into phoneme stream:
read $phoneme;

:: Synchronize a C token with an initial consonantal token:
[%phoneme _^left <cons> !_^ac] ^left...^ac;

:: Print the resulting delta:
print delta;

end main;

Figure 2. Sample Delta Program.
1) ‘It’s a woman’ (surface form: mūsò dòn):

morph: | root |
phoneme: | m | u | s | o |
syllable: | syl |
tone: | L |

2) ‘It’s the woman’ (surface form: mūsò dôn):

morph: | root |
phoneme: | m | u | s | o |
syllable: | syl |
tone: | L |

3) ‘answering the woman’ (surface form: mūsó jábî):

morph: | root |
phoneme: | m | u | s | o |
syllable: | syl |
tone: | L |

4) ‘It’s a mango’ (surface form: mángórò dòn):

morph: | root |
phoneme: | m | an | g | o | r | o |
syllable: | syl |
tone: | H |

5) ‘It’s the lizard’ (surface form: sākēnē dòn):

morph: | root |
phoneme: | s | a | k | e | n | e |
syllable: | syl |
tone: | L |

Figure 3. Underlying Deltas for Bambara Utterances.
1) ‘It’s a woman’ (surface form: mùsò dòn):

| morph: | root | root |
| phoneme: | m | u | s | o | d | on |
| syllable: | syl | syl | syl |
| tone: | L | H | L |

2) ‘It’s the woman’ (surface form: mùsó dòn):

| morph: | root | root |
| phoneme: | m | u | s | o | d | on |
| syllable: | syl | syl | syl |
| tone: | L | H | L | L |

3) ‘answering the woman’ (surface form: mùsó jábí):

| morph: | root | root |
| phoneme: | m | u | s | o | j | a | b | i |
| syllable: | syl | syl | syl | syl |
| tone: | L | H | L | H | H |

4) ‘It’s a mango’ (surface form: mágóró dòn):

| morph: | root | root |
| phoneme: | m | an | g | o | r | o | d | on |
| syllable: | syl | syl | syl | syl |
| tone: | H | L | H | L |

5) ‘It’s the lizard’ (surface form: sákêné dòn):

| morph: | root | root |
| phoneme: | s | a | k | e | n | e | d | on |
| syllable: | syl | syl | syl | syl |
| tone: | L | H | L | H | L | L |

Figure 4. Underlying Deltas after Floating High Tone Assignment.
1) 'It's a woman':

morph:      |  root    |  root    
phoneme:    |  m | u | s | o | d | on 
syllable:   |  syl | syl | syl 
tone:       |  L   |  H   |  L   

2) 'It's the woman':

morph:      |  root    |  root    
phoneme:    |  m | u | s | o | d | on 
syllable:   |  syl | syl | syl 
tone:       |  L   |  H   |  L   |  L   

3) 'answering the woman':

morph:      |  root    |  root    
phoneme:    |  m | u | s | o | j | a | b | i 
syllable:   |  syl | syl | syl | syl 
tone:       |  L   |  H   |  L   |  H   |  H   

4) 'It's a mango':

morph:      |  root    |  root    
phoneme:    |  m | a | n | g | o | r | o | d | on 
syllable:   |  syl | syl | syl | syl 
tone:       |  H   |  L   |  H   |  L   |  H   

5) 'It's the lizard':)

morph:      |  root    |  root    
phoneme:    |  s | a | k | e | n | e | d | on 
syllable:   |  syl | syl | syl | syl 
tone:       |  L   |  H   |  L   |  H   |  L   |  L   

Figure 5. Surface Forms Produced by Sync Mark Merging.
:: Forall floating H tones (^bh = "before H", ^ah = "after H")...

loop forall ([%tone ^bh H ^ah] & [%morph ^bh ^ah]);

if

:: If the floating H occurs before a floating L, move the
:: H tone into the end of the preceding morph. Otherwise,
:: insert the H tone at the beginning of the following
:: morph. Moving the H tone is accomplished by inserting
:: a new H tone and deleting the floating one.

([%tone ^ah L ^al] & [%morph ^ah ^al])
  -> Insert [%tone H ...^bh; ^ah...];

else -> insert [%tone H] ^ah...;

fi;

:: Delete the original floating H and following sync mark:

delete %tone ^bh...^ah;
delete ^ah;

pool;

**Figure 6.** Forall Loop for Floating High Tone Assignment.
:: For each morph (\texttt{^bm = "begin morph"}, \texttt{^am = "after morph"}) ...

loop forall [\texttt{%morph ^bm <> !^am}];

\texttt{^bs = ^am;} :: Set \texttt{^bs} (begin syllable) to \texttt{^am} (after morph)
\texttt{^bt = ^am;} :: Set \texttt{^bt} (begin tone) to \texttt{^am}

loop

:: Set \texttt{^bt} before the next tone token to the left. If
:: there are no more tone tokens in the morph (i.e., \texttt{^bt}
:: has reached \texttt{^bm}), exit the inner loop.

[\texttt{%tone !^bt <> _^bt}];
(\texttt{^bt == ^bm}) \rightarrow \texttt{exit};

:: Set \texttt{^bs} before the next syllable token to the left. If
:: there are no more syllable tokens in the morph, exit the
:: inner loop.

[\texttt{%syllable !_^bs <> _^bs}];
(\texttt{^bs == ^bm}) \rightarrow \texttt{exit};

:: Merge the sync mark before the tone and the sync mark
:: before the syllable:

merge \texttt{^bt} \texttt{^bs};

pool;
pool;

\textbf{Figure 7. Forall Loop for Sync Mark Merging.}
dict morphs(^b, ^e, ^1, ^2);

%phoneme:
:: morphs with low tones:
a, :: 'he'
b a l a, :: 'porcupine'
d o n, :: 'this is'
g a l a m a, :: 'ladle'
m u s o, :: 'woman'
... -> insert [%tone L] ^b...^e;

:: morphs with high tones:
s u n g u r u n, :: 'girl'
k u r u n, :: 'canoe'
j a a b i, :: 'answer'
j i, :: 'water'
... -> insert [%tone H] ^b...^e;

:: morphs with high-low patterns:
m a n g o r o, :: 'mud'
... -> insert [%tone H L] ^b...^e;

:: morphs with low-high-low patterns:
s a k e n e, :: type of lizard
... -> insert [%tone L H L] ^b...^e;

... %
end morphs;

Figure 8. Sample Morph Dictionary.
Figure 9. Baseline and Topline.
H_start_freq = 230;  :: Starting frequency of high tone line
H_end_freq = 150;   :: Ending frequency of high tone line
L_start_freq = 170;  :: Starting frequency of low tone line
L_end_freq = 130;   :: Ending frequency of low tone line

sent_dur = dur("left..."right);   :: Sentence Duration

H_slope = (H_end_freq - H_start_freq) / sent_dur;
         :: Slope of high tone line

L_slope = (L_end_freq - L_start_freq) / sent_dur;
         :: Slope of low tone line

:: Insert an F0 value for each nucleus:

loop forall [nucleus _^bn nuc !^an];

:: Compute the duration from the beginning of the sentence to
:: the point halfway through the V (the point where we wish
:: to compute and insert the F0 value):

half_nucDur = 0.5 * dur("bn..."an);
elapsed_time = dur("left..."bn + half_nuc_dur));

:: Compute the F0 value depending on whether the nucleus
:: is low- or high-toned:

if
    [%tone _\^bn H !^an]
    -> f0_val = H_slope * elapsed_time + H_start_freq;
    else -> f0_val = L_slope * elapsed_time + L_start_freq;
fi;

:: Insert the computed F0 value halfway through the nucleus:

insert [%F0 f0_val] at (^bn + half_nuc_dur);
pool;

Figure 10. Delta Program for F0 Target Assignment.
Figure 11. Second Formant Pattern for English [g h ai s].
Figure 12. Second Formant Pattern for English \([g \, ai \, s]\).
Articulatory Binding

John Kingston

1 The mismatch between phonological and phonetic representations

The division\(^1\) of the stream of speech into a string of discrete segments represents the number of places where contrast between the current speech event and others is possible. Discrete segments may not only be employed to represent phonological contrasts in speech events, however, but also in the plans for actually uttering them. Beyond isomorphism with phonological representations, the appropriateness of discrete segments at some level in the plans speakers employ to produce utterances is demonstrated by the fact that the vast majority of speech errors reorder entire segments rather than parts of segments or features (Shattuck-Hufnagel & Klatt 1979). This speech error evidence does not, however, reveal whether a segment-sized unit is employed at all levels in the plan for producing an utterance, nor what principles govern the coordination of articulators.

Since more than one articulator is moving or otherwise active at any point in time in a speech event, the notion of a segment implies a specification for how these independent movements are coordinated. However, articulatory records of speech events cannot be divided into discrete intervals in which all movement of articulators for each segment begins and ends at the same time, nor can these records be obtained by a simple mapping from discrete underlying phonological segments. The continuous and coarticulatory properties of actual speech events instead make it very unlikely that the plans for producing them consist of nothing more than a string of discrete segments (Chman 1966, 1967; Fowler 1980; Browman & Goldstein 1985, 1986; but cf. Henke 1966; Keating 1985). Looking up from the

---

1. This paper will appear in J. Kingston & M. Beckman (eds.) Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech. Cambridge: Cambridge University Press.
phonetic to the phonological representation, the discrete phonological segment derived from the commutation of contrasting elements is at best covert.

Discreteness has, moreover, gradually disappeared from phonological representations, in two widely-separated steps. The first step broke the segment into a bundle of distinctive features, a move which can be traced back at least to Trubetzkoy (1939). Entire segments are still formally discrete, but they are no longer indivisible atoms. This step necessarily preceded breaking up the feature bundles themselves into linked tiers on which the domain of each feature is independently specified (Goldsmith 1976, Clements 1985, Sagey 1986). Discreteness survives in these models only in the timing units of the skeleta (Clements & Keyser 1983), and even the equation of the timing unit represented by the traditional segment with these timing units is presently in question (McCarthy & Prince 1986). Since the domain of feature specification in nonlinear phonological representations is both larger and smaller than the traditional segment, these representations remove any principled phonological reason to expect to find discrete segments in speech events. Discrete segments cannot be found in speech simply because they are not there in the phonological representation anymore than in the phonetic one. Giving up the idea of the segment as a discrete element in phonological as well as phonetic representations requires some other principles for coordinating articulations, however, since the evidence for such coordination is patent.

The principles of coordination which have been devised so far in nonlinear phonological models are largely formal in nature. They include one-to-one, left-to-right mapping (Goldsmith 1976, Clements & Ford 1979), the projection of P-bearing units (Clements & Sezer 1982), the no-crossing prohibition (Goldsmith 1976, Kenstowicz 1982, Sagey 1986), geminate integrity (Schein & Steriade 1986) or inalterability (Hayes 1986), the obligatory contour principle (Leben 1978, McCarthy 1986), and the shared feature convention (Steriade 1982). Parallel principles for coordinating articulations in the plans speakers employ in producing speech must also be discovered (see Browman & Goldstein 1985, 1986, to appear for other proposals). This paper presents such a phonetic principle of coordination, which constrains when glottal articulations in consonants occur relative to oral ones. I refer to these constraints as "binding".
2 The binding principle

The binding principle is intended to account for two fundamental asymmetries in the distribution of glottal articulations (see Maddieson 1984):

1. Stops are much more likely to contrast for glottal articulations than either fricatives or sonorants, and

2. Glottal articulations in stops are much more frequently realized as modifications of the release of the oral closure than its onset.

The principle attempts to account for these asymmetries in terms of two kinds of phonetic differences between stops and continuants.

First, both the state of the glottis and degree of constriction downstream may affect the pressure of the air inside the oral cavity, and thereby the acoustics of the sound produced. Changes in the size of the glottal aperture and the tension of the folds influence the resistance to air flow through the glottis, and therefore have the control of air flow through the glottis into the oral cavity as their proximate function. In obstruents, the distal function of the changes in glottal resistance which affect glottal air flow is to manipulate intraoral air pressure. The larger the glottal aperture and the lower the fold tension, the more glottal resistance will be reduced, the more air will flow through the glottis, and the more rapidly air pressure will rise in the oral cavity behind the obstruent articulation. This aerodynamic interaction between glottal and supraglottal articulations will be most dramatic in stops since the downstream obstruction of air flow is complete. The amount that intraoral air pressure is elevated behind the stop closure together with the size of the glottal aperture determine the acoustic character of the explosive burst of noise that occurs when the stop is released. Glottal aperture affects the acoustics of the burst both by the aerodynamic means just described and by determining what acoustic coupling there is between supra- and sub-glottal cavities. The acoustic character of the burst therefore at once depends on and cues the state of the glottis.

Second, the release of a stop is acoustically distinct from its onset. A burst of noise is produced at the stop release as a result of the sudden opening of the oral cavity. Because the acoustic character of the burst reflects the size of the
pressure buildup behind the obstruction (and acoustic coupling to the subglottal cavity), glottal articulations are expected to be coordinated with that part of the stop rather than its onset. No burst occurs at the release of a continuant's articulation because their obstruction of air flow is not complete. The release of a fricative or sonorant will therefore be the acoustic mirror image of its onset. With a less than complete obstruction, variations in glottal air flow will affect intraoral air pressure less, since the more open the articulation, the less air is trapped behind it. The goal of the glottal articulations which accompany continuant, especially approximant, articulations can therefore be only the proximate modification of the source. Fricatives are actually intermediate between stops and approximants, since their oral aperture is small enough to obstruct air flow, elevate intraoral air pressure, and accelerate flow through the oral constriction enough to create noisy turbulence. This elevation of intraoral air pressure depends upon a sufficiently large glottal aperture; voiceless fricatives exhibit the widest glottal aperture of any voiceless consonant, and even voiced fricatives are produced with a more open glottis than corresponding voiced stops or sonorants (Hirose & Gay 1972, Hirose, Lisker, & Abramson 1972, Collier, Lisker, Hirose, & Ushijima 1979). These larger glottal apertures would also alter the spectrum of the fricative noise by increasing the acoustic coupling between the supra- and sub-glottal cavities. How this coupling influences the acoustics and perception of fricatives cannot be gone into in this paper, beyond noting that all the acoustic effects of the wide aperture will be distributed across much of the fricative interval. Such distributed acoustic effects of glottal articulations probably characterize continuants in general, in contrast to stops, where they are confined to moment of the release and the short interval immediately following it. In general terms, the acoustic effects of glottal articulations in stops are disjoint from and follow the acoustic effects of the oral articulations, while in continuants they are much more nearly simultaneous.

The binding principle claims that a glottal articulation is more constrained -- more tightly "bound" -- in when it occurs the more the oral articulation obstructs the flow of air through the vocal tract: a glottal articulation

2. What is essential here is whether the supraglottal articulation causes pressure to build up above the glottis. The binding principle therefore does not distinguish between nasal and nonnasal sonorants, i.e. between segments in which air is allowed to escape through the nose from those where it escapes through the mouth.
is most tightly bound to a stop, since air flow is completely obstructed by the stop closure, while the lesser obstruction of a fricative or approximant allows the glottal articulation to shift with respect to the oral one. Since degree of obstruction is neither an articulatory nor acoustic continuum -- both the transition from approximation to constriction and from constriction to closure involve crossing of thresholds --, the binding principle partitions segments into distinct manners. Usually just two are distinguished: either stops vs. fricatives and sonorants, i.e. noncontinuants vs. continuants, or stops and fricatives vs. sonorants, i.e. obstruents vs. approximants. The dividing line appears to fall most often between stops and continuants, despite fricatives' dependence for noise production on an adequate glottal air flow.

The binding principle predicts that the timing of oral articulations determines when glottal articulations will occur and that timing depends on the continuancy of the oral articulation. The onsets and offsets of glottal articulations will be more or less synchronized with the onsets and offsets of oral articulations in continuants, but not in stops. In the latter, they will align with the release so as to modify it and the immediately following interval acoustically. Alignment of glottal articulations in stops with the release is apparent in voice onset time (VOT) contrasts, in which adduction of the glottis is timed with respect to the release of the stop. Glottal articulations are also aligned with the stop release in at least two other kinds of stops, breathy voiced stops and ejectives, which do not participate in VOT contrasts. Partial abduction begins with or just before the release of the stop in breathy voiced stops and tight glottal closure combines with reduction of oral volume to produce an intense burst at the release of an ejective. This alignment again has the effect of concentrating the acoustic effects of the glottal articulations in the immediate vicinity of the release.

The next section of this paper addresses the problems that pre-aspirated stops of the sort found in Icelandic present to the claim that glottal articulations bind to a stop's release rather than its closure. It is shown that both in the phonology and phonetics of Icelandic, pre-aspirates are not after all a problem for this aspect of the binding principle. The final part of the paper explores a number
of further problems with the binding principle and proposes tentative resolutions of them.

3 The problem of pre-aspirates

3.1 Introduction

3.1.1 The oral-glottal schedule in pre-aspirated stops

Icelandic has a class of stops which are traditionally called "pre-aspirated," since the glottis opens before the oral closure is made, producing a preceding interval of noise. Pre-aspirates are markedly rarer in the world's languages, at least as a contrasting type, than the common post-aspirates. Contrary to the prediction of the binding principle, this early abduction of the glottis may be bound to the onset of the oral closure, in contrast to the post-aspirated stops in which abduction is bound to the release of the closure. Since Icelandic also has post-aspirated stops, a contrast appears to exist in this language between stops whose glottal abduction binds to the stop closure and those where it binds to the release.

In Icelandic pre-aspirated stops (Petursson 1972, 1976, Thrainsson 1978, Löffqvist & Yoshioka 1981b, Chasaide p.c.), the glottis begins to open during the preceding vowel, partially devoicing it. Pre-aspiration generally has the timbre of that vowel. There is some uncertainty about when peak glottal opening is reached and when the folds are again adducted, however. Löffqvist and Yoshioka's data indicate a small glottal opening that peaks before the stop closure, with adduction coinciding with the beginning of the stop closure, while both Petursson's and Chasaide's data show a much larger peak opening that is aligned to the closure, with adduction not occurring until the stop is released or even somewhat after. Pre-aspirated stops in Löffqvist and Yoshioka's data invert the order of glottal and oral articulations found in breathy voiced stops such as occur in Hindi, in which the glottal opening (largely) follows the release of the oral closure (Kagaya & Hirose 1975, Benguerel & Bhatia 1980), while the pre-aspirates observed by Petursson and Chasaide are the mirror image — relative to the stop closure and
release -- of post-aspirates, in which the glottis begins to open at the onset of the oral closure, reaches peak opening at release, and is only adducted noticeably later. In either case, the timing of glottal abduction relative to the oral closure in pre-aspirates contrasts with that found in another stop type in which the opening of the glottis is likely to be bound to the release of the closure, either breathy voiced or post-aspirated stops.

3.1.2 The phonology of pre-aspiration in Icelandic

Pre-aspiration in Icelandic (Garnes 1976, Thráinsson 1978), and perhaps in the other languages in which it is found, arises only from underlying clusters and not from single segments. On the surface, pre-aspirated stops contrast with unaspirated geminate or single stops between vowels and with post-aspirated stops, but before nasals or laterals the contrast is only between pre-aspirates and single unaspirated stops. The syllable boundary falls after the stop in the latter sort of clusters. Even when they occur alone, Icelandic pre-aspirates all retain one quite overt property of surface heterosyllabic clusters: only short vowels may precede them, which also indicates that they close the preceding syllable. On the other hand, vowels are long before post-aspirates in this language, whether occurring alone or followed by a glide or rhotic. Post-aspirates and clusters beginning with them must therefore belong only to the following syllable.

In all dialects, pre-aspirates arise from underlying geminate aspirates, but the sources of pre-aspiration or a close analogue of it are by no means restricted to those. In Southern Icelandic dialects, underlying clusters of a voiced fricative, nasal, lateral, or rhotic preceding an underlying aspirated stop are typically realized with the first segment is voiceless and the second unaspirated. In Northern dialects, a more restricted set of consonants undergoes devoicing before underlying aspires. In both sets of dialects, when the first consonant devoices, the second is underlyingly aspirated, i.e. [+ spread], and when the first consonant remains voiced on the surface, the following stop is invariably post-aspirated. Since devoicing of the first segment co-occurs with an obligatory absence of post-aspiration on the stop, it represents the shift of the [+ spread] specification to the first segment from the stop.
Furthermore, in most dialects the first stop in heterorganic clusters of stops where the second is underlyingly aspirated is realized as the corresponding voiceless fricative, and again, if the stop spirantizes, the second may not be post-aspirated. Since voiceless fricatives demand a very wide glottal aperture to elevate air flow through the glottis sufficiently to produce turbulence downstream through the oral constriction, this realization can also be seen as a shift of [+ spread] to the first element of the cluster. Finally, as in many other languages, Icelandic stops do not contrast for aspiration after /s/, a segment which demands at least as wide a glottal aperture as the other voiceless fricatives.

In general terms, not only may just one segment bear a [+ spread] specification in surface forms, but only one segment in the cluster needs to (or can) be specified as [+ spread] in underlying forms. In all these cases, underlying specification of the second member of the cluster as [+ spread] is phonetically realized as a property of the first rather than the second element of the cluster by a quite general procedure in the language. In all dialects, clusters in which the final stop is underlyingly [+ spread] contrast with others in which the final stop is not [+ spread]. In the latter sort of cluster, the preceding consonant remains voiced uniformly.

Since it arises only in clusters whose second member is [+ spread], pre-aspiration cannot contrast directly with post-aspiration. The latter arises only when an underlying [+ spread] stop occurs in absolute initial position in word or directly after a vowel. Generalizing, pre-aspiration arises in syllable codas in heterosyllabic clusters whose second member is [+ spread] as a result of a shift of this specification to the preceding coda. On the other hand, post-aspiration is only found in singleton [+ spread] stops or those which are the first member of a tautosyllabic cluster, i.e. in onsets rather than codas. Formulaically, /aC.pʰa/ is realized as [aCh.pa], while /aa.pʰ(C)a/ undergoes no change. The failure of pre-aspirates to ever contrast directly with post-aspirates within morphemes in Icelandic eliminates them as a problem for the binding principle, at least insofar as the phonology of the language is concerned. (The analysis outlined above is a generalization of Thráinsson's (1978). I am indebted to Donca Steriade for making the significance of devoicing in the initial segments of clusters ending in [+ spread] stops clear to me.)
3.2 Pre-aspirates as a test of the binding principle in the phonetic component: The covarying durations of glottal abduction and flanking oral articulations

3.2.1 Introduction

An experiment was designed to determine whether the abduction of the glottis is still after all part of the same articulatory unit as the following stop closure in pre-aspirated stops, since there is no a priori reason why the units of articulation should be identical to underlying ones. If pre-aspiration is coordinated with the following stop rather than the preceding vowel, then glottal articulations must be allowed to bind phonetically to the closure as well as the release in stops, despite the absence of any event as salient as the release burst at the beginning of the closure.

This experiment assumes that if two articulations are coordinated with one another, their individual durations will covary across global changes in segment duration, due to changes in rate or prosodic context, i.e. coordinated articulations should exhibit relational invariance (Tuller & Kelso, 1984). Specifically, if abduction of the glottis is coordinated with an adjacent or overlapping oral articulation, then the duration of abduction should covary with that of the oral articulation. If the binding principle is correct, then the duration of an abduction which overlaps the release of a stop closure, as in post-aspirated stops, should correlate positively with the duration of the closure, but no necessary correlation should be observed when abduction does not overlap the release, as in pre-aspirated stops. Pre-aspiration may instead be coordinated with the preceding vowel and its duration might covary with the duration of that segment. Alternatively, glottal abduction in pre-aspirated stops may be coordinated with no oral articulation. The binding principle does not distinguish between the two possibilities.
3.2.2 Methods

A single female speaker of Icelandic was recorded producing words containing medial pre- and post-aspirated stops under conditions that would change segment durations. Since the binding principle predicts coordination of glottal abduction with the preceding stop, specifically its burst, in post-aspirated stops, covariation of the durations of oral and glottal articulations in them can be compared with that in pre-aspirated stops. Medial breathy-voiced and post-aspirated stops were also recorded from a single female speaker of Hindi under similar conditions. Since abduction overlaps or follows the stop release in both breathy voiced and post-aspirated stops, these Hindi data provide an additional test of the positive prediction of the binding principle that abduction will be coordinated with a release burst that it overlaps temporally.

All tokens in both languages were produced medially in short frame sentences, with the sentences read in a different, random order for each repetition. The Icelandic speaker produced a variety of real words containing pre- and post-aspirated stops between the first, stressed syllable and the second. (Lexical stress occurs obligatorily on the first syllable in Icelandic words.) The Hindi speaker produced nonsense words of the shape [a__a] or [a__apa] with the stops in the blanks. In alternate readings, the Icelandic speaker put focus on the test word itself or on the immediately preceding word. Segments are expected to be longer in words in focus than in those not in focus, particularly in the vicinity of the stressed syllable. Final lengthening would lead one to expect longer segments in the Hindi words where only one syllable follows the stop than when two do. Both of these manipulations, though they are quite different from one another, should therefore affect the duration of the stops, and their component articulations. In addition, for both languages, each way of reading the words was produced at self-selected moderate and fast rates. This orthogonal variation in rate will also affect segment duration, though more globally than focus or the number of following syllables. Multiple repetitions of each utterance type were collected from each speaker under each condition.

The utterances were digitized at 10 kHz and measurements made from the waveform displays, rather than articulatory records, to the closest millisecond. Three intervals were measured in each case:
1. The duration of audible glottal abduction,

2. The duration of the flanking vowel plus the duration of audible glottal abduction, and

3. The duration of the flanking oral closure plus the duration of audible glottal abduction.

The interval of glottal abduction was taken to be the stretch of noise between the vowel and oral closure, except for pre-aspiration in Icelandic, where this interval included a breathy interval at the end of the preceding vowel as well as a following interval of noise before the stop closure. This interval is that portion of the actual abduction of the glottis which is audible; since in all cases the flanking stop closure overlaps some part of the abduction, the interval measured is only part of the total duration of the abductory gesture. The vowel was identified as that stretch where the waveform was intense, clearly periodic, and with the asymmetric periods characteristic of modal voice. The oral closure was identified as that interval of greatly reduced amplitude or silence between vowels.

The question immediately arises whether acoustic records can be used in this way to measure the duration of articulatory events. This question is especially pertinent here since that part of glottal abduction which overlaps with the oral closure cannot be observed directly in an acoustic record. However, I have assumed that changes in the duration of the observable part of the glottal abduction outside the interval of the oral closure will be in proportion to changes in its total duration. If this is true, measuring the observable part would reveal the same relative changes as measuring the entire interval of abduction.

3.2.3 Analysis

Correlations were calculated between the duration of audible glottal abduction and the acoustic durations of each of the flanking oral articulations combined with the duration of audible glottal abduction for each of the two stop types in the two languages. Correlations tend to be positive, often quite strongly so, when an interval is correlated with a larger interval which includes itself, but the portion of the obtained correlation which is due simply to this inclusion can be calculated and the significance of the difference between these obtained and
expected correlations determined (Cohen & Cohen 1983, Munhall 1985). If the obtained correlation is significantly larger than the expected value, then the durations of the two articulations can be assumed to be positively correlated, while if the obtained correlation is significantly smaller, then the two intervals are negatively correlated. Correlations that are more positive than expected are taken here to indicate coordination between the two articulations. These part-whole correlations avoid the negative bias which arises from errors locating the boundary when correlating the durations of adjacent intervals (Ohala & Lyberg 1976).

3.2.4 Results

Figures 1a-d illustrate the relationship between the duration of audible glottal abduction and the acoustic durations of the adjacent consonant and vowel articulations for each of the stop types, pooled across all conditions.

***

Figures 1a-d about here.

***

In each of these figures, the duration of glottal abduction, i.e. pre-aspiration, post-aspiration, or breathy-voice, is plotted on the vertical axis against the durations of both the flanking oral articulations, the vowel (in open squares) and stop closure (in filled squares), on the horizontal axis. The duration of each of the oral articulations on the horizontal axis is the combined acoustic duration of that oral articulation and glottal abduction since the correlations are part-whole. Regression lines for abduction by vowel and abduction by closure are plotted separately in each figure, and the formulae of each of these regression lines are given at the top of each figure, together with the observed correlations. As expected, all the correlations are positive, some quite strongly so. The obtained and expected correlations with indications of significant differences, if any, pooled across conditions (as plotted in the figures) are listed in table 1 below, for the two
different rates in table 2, and for the two different focus positions (Icelandic) or two different numbers of following syllables (Hindi) contexts in table 3:

<table>
<thead>
<tr>
<th></th>
<th>Icelandic</th>
<th>Hindi</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Preaspirated</td>
<td>Postaspirated</td>
</tr>
</tbody>
</table>

**Overall:**

<table>
<thead>
<tr>
<th>Condition</th>
<th>Icelandic</th>
<th>Hindi</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>187</td>
<td>78</td>
</tr>
<tr>
<td>AxC</td>
<td>.800 &gt; (.598)</td>
<td>.616 = (.619)</td>
</tr>
<tr>
<td>AxV</td>
<td>.917 &gt; (.848)</td>
<td>.728 &gt; (.381)</td>
</tr>
</tbody>
</table>

Table 1: Part-whole correlation between glottal abduction and adjacent closure or vowel duration: observed (expected), all data pooled. n is the number of tokens of each type; different n's are given for all the tokens of each stop type pooled for each condition. "AxC" indicates correlation between abduction and adjacent closure, "AxV" between abduction and adjacent vowel. ">" indicates the observed correlation is significantly (p < .05 or better) positive compared to the expected value, "<" that it is significantly negative, and "=" that there is no significant difference. The same conventions are used in all the other tables.

<table>
<thead>
<tr>
<th></th>
<th>Icelandic</th>
<th>Hindi</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Preaspirated</td>
<td>Postaspirated</td>
</tr>
</tbody>
</table>

**Rate:**

<table>
<thead>
<tr>
<th>Condition</th>
<th>Icelandic</th>
<th>Hindi</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>94</td>
<td>39</td>
</tr>
<tr>
<td>AxC</td>
<td>.674 = (.561)</td>
<td>.554 = (.663)</td>
</tr>
<tr>
<td>AxV</td>
<td>.891 &gt; (.806)</td>
<td>.762 &gt; (.394)</td>
</tr>
</tbody>
</table>

**Fast:**

<table>
<thead>
<tr>
<th>Condition</th>
<th>Icelandic</th>
<th>Hindi</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>93</td>
<td>39</td>
</tr>
<tr>
<td>AxC</td>
<td>.705 = (.684)</td>
<td>.678 = (.748)</td>
</tr>
<tr>
<td>AxV</td>
<td>.855 = (.812)</td>
<td>.632 = (.544)</td>
</tr>
</tbody>
</table>

Table 2: Part-whole correlation between glottal abduction and adjacent closure or vowel duration: observed (expected), condition: rate.
Icelandic | Hindi
---|---
Preaspirated | Postaspirated | Breathy voice | Postaspirated

Focus:

In focus:

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>Number of syllables:</th>
</tr>
</thead>
<tbody>
<tr>
<td>n=</td>
<td>93</td>
<td>40</td>
</tr>
<tr>
<td>AxC</td>
<td>.821 &gt; (.627)</td>
<td>.574 = (.597)</td>
</tr>
<tr>
<td>AxV</td>
<td>.942 = (.921)</td>
<td>.696 &gt; (.376)</td>
</tr>
</tbody>
</table>

Not in focus:

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>_a_apa:</th>
</tr>
</thead>
<tbody>
<tr>
<td>n=</td>
<td>94</td>
<td>38</td>
</tr>
<tr>
<td>AxC</td>
<td>.710 &gt; (.544)</td>
<td>.641 = (.692)</td>
</tr>
<tr>
<td>AxV</td>
<td>.871 = (.827)</td>
<td>.741 &gt; (.396)</td>
</tr>
</tbody>
</table>

Table 3: Part-whole correlation between glottal abduction and adjacent closure or vowel duration: observed (expected), condition: focus (Icelandic) and number of syllables (Hindi).

3.2.5 Discussion

The binding principle predicts no positive correlation between the durations of glottal abduction and the FOLLOWING stop closure in Icelandic pre-aspirates but just such a positive correlation between the durations of abduction and the PRECEDING closure for the other three stop types: Hindi breathy voiced and Hindi and Icelandic post-aspirated stops. Nearly complementary predictions are made regarding correlation between the durations of abduction and the adjacent vowel; no positive correlation is expected for stop types other than the Icelandic pre-aspirates, which may exhibit such a correlation as a result of coordination between the opening of the glottis and the articulation of the preceding vowel. Alternatively, abduction of the glottis in Icelandic pre-aspirates may not be coordinated with either of the two flanking oral articulations. The binding
principle does not require that glottal articulations be coordinated with oral ones, but instead predicts how they will be coordinated if they are.

As table 1 shows, for all the data pooled, correlations between pre-aspiration and BOTH flanking oral articulations were significantly more positive than expected. The observed correlation of .800 between abduction and the following closure is significantly more positive than the expected .598, and the observed correlation between abduction and the preceding vowel is also significantly more positive than expected: .917 vs. .848. Significantly positive correlations are also observed between abduction and the following vowel in Icelandic post-aspirated stops, but not between abduction and the preceding closure. In post-aspirated stops in Hindi, however, the correlation between abduction and the following closure is significantly more positive than expected. The Hindi breathy voiced stops exhibit no significant correlation between abduction and either the preceding closure or the following vowel. Even so, the difference between the observed and expected correlations for abduction with preceding closure in Hindi breathy voiced stops is in the same direction as for the Hindi post-aspirated stops. The correlations between abduction and the following vowel, on the other hand, are in the opposite, negative direction, though for neither type of stop are they significant.

Tables 2 and 3 show that, with one exception, the difference between observed and expected correlations observed in each condition separately do not differ in DIRECTION from those obtained when all the data were pooled. The exception is that significantly positive correlation between abduction and the following vowel in Hindi breathy voiced stops when followed by two syllables (table 3), which is otherwise consistently negative with respect to the expected correlation. The same pattern of SIGNIFICANT differences was not always found, however, in each of the conditions as when all the data were pooled. Though in the same direction as in other conditions, no significant differences between observed and expected correlations were obtained for any of the stops at the fast rate; elsewhere, however, at least some of the differences are significant. One consistent and troublesome result is that no significant difference in either direction was found between abduction and the preceding closure for Icelandic post-aspirated stops in any of the experimental conditions. This failure is especially worrisome since post-aspirated stops in Icelandic do show significantly positive correlations between abduction and the following VOWEL in three out of four conditions, and
in the fourth the difference between observed and expected correlations is in the same direction. Also, the near absence of significantly positive correlations for both of the Hindi stops in any of the conditions reveals generally weak covariation between the duration of abduction and that of either of the flanking oral articulations in this language. Finally, the significantly positive correlation between abduction and the preceding closure obtained for the Hindi post-aspirated stops when all the data were pooled is not found in any of the conditions taken separately.

If all correlations which differ from the expected value in one direction are considered, not just those which are significant, then the binding principle fails in two ways in Icelandic:

1. Pre-aspiration correlates positively with the following closure, but
2. Post-aspiration does not correlate positively with the preceding closure.

This pattern of results is exactly the opposite of what was predicted by the binding principle. For both pre- and post-aspirated stops, abduction correlates positively with the flanking vowel, which is neither predicted nor excluded by the binding principle. The Hindi data weakly support the binding principle, on the other hand; both breathy voice and post-aspiration correlate positively with the preceding closure but either do not correlate or correlate negatively with the following vowel.

What these data reveal is an essential difference between the two languages, rather than between stops which differ in the order of the glottal and oral articulations. By this measure in Icelandic, glottal abduction appears to be coordinated with the adjacent vowel and for pre-aspiration also with the following oral closure, while in Hindi, glottal abduction is coordinated with the preceding oral closure but not the following vowel. The relatively small number of significant correlations suggests one of two things: either this method of evaluating coordination is too coarse, perhaps because abduction is only measured partially, or that coordination between glottal and oral articulations, where it does exist, does not exhibit strong relational invariance.

The most troubling results are the double failure of the binding principle in Icelandic: the absence of the predicted positive correlation of abduction with
the preceding closure in post-aspirated stops and the presence of the excluded positive correlation of abduction with the following closure in the preaspirated stops. In the next section, the coordination between glottal and oral articulations that this latter positive correlation appears to indicate is shown instead to be an illusion. The failure of the binding principle with respect to the Icelandic post-aspirates remains genuine, however, unless it can be shown that relational invariance is not an appropriate measure of coordination.

3.3 Breathiness and noise: the two components of pre-aspiration in Icelandic

3.3.1 Introduction and methods

A closer look at pre-aspiration reveals that the coordination with the following oral closure is illusory. As illustrated by the waveform of the transition from a vowel to a pre-aspirated stop (in the word stökkur 'brittle, nom. masc.') in figure 2a, preaspiration has two components: a breathy interval followed by an interval of noise.

***

Figure 2 about here.

***

The breathy interval was identified as that where the shape of a glottal period changed from asymmetric to sinusoidal, while noise was simply that interval where the waveform no longer appeared periodic; both intervals could be identified consistently by eye. Breathy voicing is a product of partial abduction of the folds, which increases air flow through the glottis sufficiently to produce noise and also increases the negative slope of the spectrum of the glottal wave. Compared to modal voice, in which the higher harmonics are more intense than the first, in breathy voice, a shortening of the closed phase of the glottal cycle relative to the open phase produces a source spectrum in which the first harmonic is more intense than higher harmonics (Bickley 1982, Ladefoged 1982). The more intense first harmonic produces the characteristic nearly sinusoidal shape of the waveform. This initial partial abduction is not great enough to extinguish voicing, however. That
only happens somewhat later, when the vocal folds become too far apart to vibrate any longer and only noise is produced. Vocal fold vibration thus ceases substantially before the beginning of the following oral closure. Of the two components of pre-aspiration, Chasaide (1986) has argued that it is the breathy interval and not the noise which is the most salient cue to identifying a stop as pre-aspirated in Icelandic and Scots Gaelic.

The waveform in figure 2a is for a token in focus spoken at a moderate rate, i.e. under conditions when segment durations are expected to be longest. Figure 2b is the waveform for an out-of-focus token of the same word spoken quickly, i.e. one in which segment durations are expected to be shortest. This comparison shows that the duration of the noisy component is reduced much more drastically than the duration of the breathy component, a consistent feature of these data. The difference between these two waveforms is in fact characteristic of the effects of varying rate or focus location; the breathy component remains essentially inert while the noisy component changes in proportion to changes in other segment durations. Changes in the durations of the two components of pre-aspiration were examined quantitatively in the data collected from the Icelandic speaker and the changes in their individual durations compared to changes in the total duration of pre-aspiration, the duration of the preceding vowel, and the following stop closure, again via part-whole correlations.

3.3.2 Results

The breathy component varies in duration much less than does the noisy component, across the experimental conditions. The noisy component itself varies in the expected direction, being markedly shorter at fast than slow rates and when the word is not in focus than when it is. Figures 3a-c illustrate the relationship of the duration of the breathy and noisy components to the total duration of pre-aspiration, the duration of the preceding vowel plus pre-aspiration, and the duration of the following closure plus pre-aspiration for all the data pooled.
The results of correlating the durations of the breathy and noisy components to each of these intervals are compared with expected values in tables 4-6. These are again part-whole correlations.

<table>
<thead>
<tr>
<th></th>
<th>Breathiness</th>
<th>Noise</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall: n = 188</td>
<td>-0.202 &lt; (0.247)</td>
<td>0.929 &gt; (0.878)</td>
</tr>
<tr>
<td>Rate:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Slow: n = 94</td>
<td>-0.471 &lt; (0.298)</td>
<td>0.922 &gt; (0.827)</td>
</tr>
<tr>
<td>Fast: n = 94</td>
<td>-0.108 &lt; (0.318)</td>
<td>0.884 = (0.855)</td>
</tr>
<tr>
<td>Focus:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes: n = 93</td>
<td>-0.018 &lt; (0.217)</td>
<td>0.963 &gt; (0.900)</td>
</tr>
<tr>
<td>No: n = 95</td>
<td>0.339 = (0.365)</td>
<td>0.780 = (0.721)</td>
</tr>
</tbody>
</table>

Table 4: Part-whole correlations between breathy and noisy components of preaspiration and total duration of preaspiration: observed (expected).

<table>
<thead>
<tr>
<th></th>
<th>Breathiness</th>
<th>Noise</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall: n = 188</td>
<td>-0.004 &lt; (0.370)</td>
<td>0.915 = (0.929)</td>
</tr>
<tr>
<td>Rate:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Slow: n = 94</td>
<td>-0.230 &lt; (0.404)</td>
<td>0.898 = (0.915)</td>
</tr>
<tr>
<td>Fast: n = 95</td>
<td>0.082 &lt; (0.423)</td>
<td>0.885 = (0.906)</td>
</tr>
<tr>
<td>Focus:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes: n = 94</td>
<td>0.127 = (0.273)</td>
<td>0.958 = (0.962)</td>
</tr>
<tr>
<td>No: n = 94</td>
<td>0.446 = (0.487)</td>
<td>0.861 = (0.874)</td>
</tr>
</tbody>
</table>

Table 5: Part-whole correlations between breathy and noisy components of preaspiration and duration of preceding vowel: observed (expected).
<table>
<thead>
<tr>
<th></th>
<th>Breathiness</th>
<th>Noise</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall:</td>
<td>n = 187</td>
<td>.047 &lt; (.215)</td>
</tr>
<tr>
<td>Rate:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Slow:</td>
<td>n = 94</td>
<td>.077 = (.251)</td>
</tr>
<tr>
<td>Fast:</td>
<td>n = 93</td>
<td>.000 &lt; (.299)</td>
</tr>
<tr>
<td>Focus:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes:</td>
<td>n = 93</td>
<td>.087 = (.147)</td>
</tr>
<tr>
<td>No:</td>
<td>n = 94</td>
<td>.459 &gt; (.240)</td>
</tr>
</tbody>
</table>

Table 6: Part-whole correlations between breathy and noisy components of pre-aspiration and duration of following closure: observed (expected).

The duration of the breathy component correlates negatively with the total duration of pre-aspiration (table 4) when all the data are pooled and when correlations are calculated for each condition separately. The observed correlation for the noisy component, on the other hand, is nowhere significantly different from the expected value. The correlation between the breathy component and the preceding vowel (table 5) is also significantly negative overall and in all conditions, except that where the word was not in focus. Even there, the observed correlation is less than the expected value. On the other hand, the correlation between the noisy component and the preceding vowel is significantly positive in the overall correlation and the two conditions where a longer vowel is expected: at the slow rate and when the word was in focus. (For both components, the nonsignificant correlations are in the same direction compared to the expected values as the significant ones.) Finally, the breathy component exhibits a significant negative correlation with the following stop closure (table 6) overall and in the two conditions when a shorter closure was expected (again, the nonsignificant correlations are in the same direction). The correlation of the noisy component is significantly positive overall and when the word was in focus; otherwise, the observed correlations are not significantly different from the expected values (here, no consistent direction is evident in the nonsignificant correlations).
3.3.3 Discussion

Since the part-whole correlations between the noisy component and pre-aspiration were nowhere significantly more positive than expected, the noisy component can be identified as the single significant component of the measured duration of pre-aspiration. On the other hand, the breathy component must be independent of pre-aspiration duration, since its duration correlates negatively. With respect to the flanking oral articulations, a similar pattern emerges; the breathy component tends to correlate negatively with each of them, while the noisy component either exhibits no correlation or a positive correlation. More generally, the duration of the breathy component stays the same or increases slightly as the other articulations get shorter, while the duration of the noisy component varies proportionally in the same direction as the other articulations. These facts allow a different interpretation of the apparent positive correlation between the duration of pre-aspiration and that of the following closure, which removes the problem these Icelandic data presented to the binding principle.

Figure 4a schematically represents the timing of the oral and glottal articulations which will produce the waveform in figure 2a (cf. Petursson 1976). As sketched in figure 4, at faster rates or when the word is not in focus (cf. figure 2b), the interval between the two successive oral articulations is shortened; the closure occurs earlier than at slower rates or in focus. This earlier closure overlaps more of the interval of wide glottal abduction which produces the noisy component of pre-aspiration and thereby shortens the interval during which it is audible. Since this noisy interval is the principal component of pre-aspiration, this rate- or focus-induced variation in the amount of overlap produces the observed positive correlation between pre-aspiration and the following closure. However, since the oral articulation is simply sliding with respect to the glottal one, that positive correlation does not indicate coordination between the two after all. Instead the timing of the glottal articulation remains more or less unchanged as the interval between the oral articulations varies, with the closure overlapping more or less of the interval of wide glottal aperture which produces the noisy component of pre-aspiration. However, slight lengthening of the breathy component at the same time that the noisy component is being shortened indicates that the schedule of the glottal articulation is not entirely invariant. At faster rates or out of focus, the small abduction which produces breathiness occurs earlier with respect to the end
of the vowel articulation and the transition to the wider abduction which produces noise alone takes longer. What is significant here is that the timing of the glottal articulation is being adjusted with respect to the preceding vowel, though it apparently is not adjusted with respect to the following consonant.

The complementary variation between the two components of pre-aspiration suggests that they may trade off perceptually in conveying the fact that the stop is pre-aspirated. With the longer breathy interval, pre-aspiration is anticipated to greater extent in shorter vowels, perhaps to ensure that some minimum interval of non-modal phonation occurs, while with longer vowels, the upcoming oral closure is also later, so a longer noisy component will occur. These data from production thus do not conflict entirely with Chasaide's claim that breathiness is more important than noise for identifying a pre-aspirated stop. They instead suggest that the relative importance of the two components, as measured by their relative durations, varies in a complementary way with the acoustic duration of flanking oral articulations.

This account of Icelandic pre-aspirates supports the suggestion of Browman and Goldstein (to appear) that apparent assimilations and deletions which occur in casual speech arise simply from an acoustically obscuring overlap of the movement of one articulator -- the apparently assimilated or omitted articulation -- by the sliding movement of another over it, rather than being a product of actually substituting or omitting articulatory gestures. In their view, the shift from a careful to a casual speech style changes the relative timing of articulatory gestures but not their shape or occurrence. It is not at all surprising that this suggestion should generalize to the effects of varying rate and focus location in these Icelandic data, since the differences between casual and careful speech undoubtedly parallel those between fast and moderate rates of speaking or between a word out of focus and one in focus.

3.4 Summary

The Icelandic data appeared at first to cast serious doubt on the binding principle in two ways. First, pre-aspiration correlated positively with the following stop closure, and, second, post-aspiration did not correlate with the preceding stop closure. In both kinds of stops, furthermore, the duration of audible abduction of
the glottis correlated positively with the adjacent vowel. However, as closer examination of the variation of duration in the two components of pre-aspiration has shown, there is no actual coordination between glottal abduction and the following oral closure in Icelandic pre-aspirates, at least as reflected in relational invariance. Instead, the relative timing of the two flanking oral articulations vary with respect to one another, occurring closer together at faster rates or out of focus. This variation changes the duration of the audible interval of noise in pre-aspiration without affecting the duration of the breathy component proportionally because the oral closure overlaps more or less of the interval of wide glottal abduction that produces the noisy component. Since the noisy component is the principal component of pre-aspiration, this variation in the amount of overlap of glottal abduction by the oral closure leads to an APPARENT covariation between the durations of pre-aspiration and the following oral closure, suggesting coordination of the two articulations. The more correct view is that the glottal articulation is not coordinated at all with the following oral closure. Its coordination with the preceding vowel, on the other hand, tends towards invariance, though there is evidence of a small deceleration of glottal abduction at fast rates and out of focus. The Icelandic pre-aspirates do not therefore pose a problem for the binding principle in the phonetics any more than in the phonology.

The Hindi data turned out much closer to the predictions of the binding principle in that in both breathy voiced and post-aspirated stops, glottal abduction only correlates positively with the preceding stop closure.

The post-aspirates of Icelandic remain problematic, since they failed to show even the slightest sign of covariation in duration between glottal abduction and the preceding oral closure of the sort predicted by the binding principle.

4 Binding sites: Predictions of the binding principle and problems

The single ballastic opening of the glottis which occurs in the production of single voiceless consonants could bind either to the beginning or the end of the oral constriction. In fricatives, the beginning and end of the oral constriction are acoustically symmetric -- and the glottal abduction is simply intended to produce a sufficient and continuous flow of air --, so the binding principle can choose one or the other as the favored binding site. In stops, on the other hand, these sites are
asymmetric: the dramatic reduction in amplitude at the stop closure is quite different from the burst of noise at the release, and the binding principle predicts that the opening of the glottis binds to the release, because it is the acoustic character of the burst which the glottal articulation is intended to modify.

Observations of the timing of glottal abduction in single consonants and in clusters appear to indicate, however, that glottal articulations are not always timed with respect to oral ones in ways predicted by the binding principle -- stops show coordination of glottal articulations with the closure as well as the release and fricatives exhibit more restrictive timing of glottal articulations than expected:

1. Glottal abduction is coordinated with the closure rather than the release of the stop in voiceless unaspirated stops, and with the beginning of the constriction in voiceless fricatives;

2. As a corollary, the binding principle predicts that glottal articulations will bind to the stop in a cluster of a stop and a fricative (regardless of their order) and to the last stop in a cluster of stops, because often only the last is audibly released. However, the manner of the first rather than the last obstruent in a cluster, i.e. whether it is a continuant, generally determines when the peak glottal opening will occur, despite the first segment's frequent lack of an audible release, whether because it's a fricative or because it's a non-final stop. In some languages, furthermore, peak opening velocity occurs at the same time relative to the onset of the cluster articulation for clusters beginning with stops as well as fricatives, implying that the closure rather than the release is the general binding point for glottal articulations in consonants of both manners in clusters; and

3. Closure durations in aspirated stops vary inversely across languages with the duration of aspiration, implying that the timing of glottal abduction is relatively invariant, instead of shifting to stay close to the oral release.

The second and third problems are elaborated below; see Louis Goldstein's commentary for much more extensive discussion of aspects of the first two problems.
Considering the speaker's ACoustIC goals unsurprisingly resolves the first of the problems for the binding principle. Binding the peak opening in aspirated stops to the oral closure instead does yield a reliable contrast in how their bursts sound compared with those of unaspirated stops. Since the glottis is adducted by the time the stop is released in voiceless unaspirated stops, abduction serves only to ensure an absence of vocal fold vibration during the closure and will not positively modify the acoustics of the burst and what follows. The weak burst of unaspirated stops may be acoustically neutral, i.e. phonetically underspecified and lacking in any unique acoustic signature, in contrast to the positively modified burst of aspirated and some other kinds of stops (such as ejectives) in which the glottal articulation is coordinated with the release. Unaspirated stops cannot, of course, be treated as underspecified in ARTICULATION, since their successful production is only possible if the abduction of the glottis is scheduled so as to be complete at or before the stop is released. If the burst is acoustically neutral in unaspirated stops, however, then the binding principle, or a restricted form of it would not constrain when the glottal articulation occurs. This restricted form of the binding principle applies not to all stops but only to those in which the timing of the glottal articulation is controlled so as to positively alter the acoustics of the burst.

There is some further evidence that the timing of glottal articulations is acoustically less crucial in unaspirated stops. Flege (1982) has shown that the timing of ADduction in the voiceless unaspirated realizations of utterance initial [+voice] stops in English does not predict the onset of vocal fold vibration. Some speakers who occasionally or always produce voiceless unaspirated stops in this position always adduct the folds substantially before the stop release, but, nonetheless, there is no voicing until the stop is released. Unlike these stops in English, voiceless unaspirated stops in most of the other languages discussed here are the principal rather than just an optional allophone of the one of the types of stops in the language. Furthermore, in none of these languages is there any variation in the timing of abduction and consequently adduction in their unaspirated stops. These differences in the timing of glottal abduction between the unaspirated stops of these languages and of languages such as English are acoustically irrelevant, however, at least insofar as the acoustics of the burst is concerned, since the glottis is adducted in all of them at the stop release.
The markedly different aerodynamic requirements of fricatives, especially sibilants, suggest that the relative timing of their glottal articulation is governed by quite different considerations than apply to stops. Both the early peak glottal opening and its large size in voiceless fricatives yields the high glottal air flow and hence high oral air flow needed to produce noisy turbulence. Noise also characterizes aspirated stops, but unlike in fricatives this noise occurs late in the articulation of the segment rather than early. These aerodynamic requirements downstream thus demand a different, earlier timing of glottal abduction relative to the oral constriction in fricatives than stops.

Clusters of voiceless obstruents present a picture quite like that of single segments. Lofqvist and Yoshioka (1981a) compared the size and velocity of the glottal opening in single obstruents and obstruent clusters in which just a single opening occurred for three languages, Swedish, Japanese, and Icelandic. In all three, peak opening was earlier in fricatives than stops, and in clusters beginning with fricatives than stops. English is similar (Yoshioka, Lofqvist, & Hirose 1981). In all four languages, the manner of the initial obstruent in the cluster, i.e. whether it was a continuant or not, determines the timing of the glottal gesture. These timing

3. Voiceless fricatives can be divided (coarsely) into a frication portion followed by an aspiration portion. It is, however, uncertain whether the transition from frication to aspiration is a product of relaxation of the oral constriction, reducing oral air flow directly, or of adduction of the glottis, reducing oral air flow through reduction of glottal air flow. If the latter is the source of this asymmetry, then voiceless fricatives are the articulatory complement of aspirated stops, at least in the relative timing of their oral and glottal articulations. If voiced fricatives are similarly asymmetric, then they are the articulatory complement of voiced stops, in which typically the first part is voiced, but the later portion may exhibit devoicing.

4. The obstruent clusters examined in Japanese included strings derived by devoicing of an intervening vowel. Vowel devoicing is the only means by which "clusters" whose members differ in continuancy may be obtained in this language. These appeared to exhibit the same temporal coordination between oral and glottal articulations as underlying clusters, which in Japanese consist only of geminates, i.e. consonants with the same specification for continuancy (and place).
patterns disconfirm the predictions of the binding principle in two ways. First, it is the first rather than the last member of the cluster which determines when peak glottal opening occurs and, second, stops occurring later in the cluster do not attract abduction to their releases, away from a preceding fricative.

These four languages also resemble one another in that the opening velocity of the glottis is consistently slower for sequences beginning with stops than fricatives. Furthermore, peak velocity is reached at the same time relative to the preceding end of voicing for all sequences beginning with the same manner of consonant. The slower velocity of abduction in stops undoubtedly accounts for their later peak opening compared to fricatives. However, peak velocity does not necessarily occur at the same time relative to voicing offset for stop as fricative initial sequences in all the languages.

The velocity peak is reached at the same time for both stop and fricative initial clusters in Japanese and Swedish, but Swedish differs from Japanese in having a broader velocity peak in its stops than in its fricatives, while both stops and fricatives have a similar narrow velocity peak in Japanese. In Icelandic, on the other hand, the peak opening velocity peak is uniformly later for stops than fricatives. Lower opening velocity alone accounts for the later peak opening in stops than fricatives in Japanese, but in Swedish stops the sustained velocity peak leads to a larger as well as later abduction, and in Icelandic, the later peak velocity in stops compared to fricatives augments the effect of stops’ lower velocity of opening in delaying when peak glottal opening occurs. Despite the lack of agreement in the timing of glottal articulations among these languages, these data are a problem for the binding principle since for all three languages the velocity peak is controlled for stops as well as fricatives relative to the beginning of the consonant’s articulation, not its release. These data suggest that the peak glottal opening in stops only appears to coordinated with the release because it is later than in fricatives.

Other kinds of clusters of voiceless obstruents may exhibit multiple openings of the glottis, with partial adduction between them, for each segment that requires a high glottal air flow to produce noise (Löfqvist & Yoshioka 1981a,b, Yoshioka, Löfqvist, & Hirose 1981, 1982). Multiple openings are controlled somewhat differently than single ones. Single openings in single segments and clusters are produced by the posterior cricoarytenoids and the interarytenoids contracting reciprocally in a classical antagonist pattern: contracting the posterior
cricoarytenoids opens the glottis and then contracting the interarytenoids closes it; activity in the interarytenoids is suppressed when the posterior cricoarytenoids are active and vice versa. However, in clusters where the glottis opens more than once, activity is almost entirely suppressed in the interarytenoids throughout the cluster. The glottis either opens and closes with a waxing and waning of activity in the posterior cricoarytenoids alone or the small adductions come from slight increases in interarytenoid activity. These data are also problematic for the binding principle since the principle predicts the consolidation of all glottal gestures in a cluster into a single one, bound to the last stop in the cluster, rather than multiple openings. Consolidation in such clusters is evident only in the near total suppression of adductory activity in the interarytenoids.

According to the binding principle, if the release of a consonant articulation produces a salient acoustic event, such as a burst, a glottal articulation will bind to the release because the state of the glottis determines the release’s acoustic character. Since in clusters of stops often only the final stop in stop clusters is audibly released, the binding principle predicts tight coordination with this last stop but not with nonfinal stops in such clusters. (By "final", I mean here the last stop in a cluster preceding a vowel or sonorant; in word final stop clusters, the last stop is often unreleased as well.) The latter are, in many languages, often not audibly released. The neutralization of contrasts in glottal articulations in nonfinal members of stop clusters which also occurs in many languages could be a product of the failure of a potential glottal articulation to bind to those members because they are not fortunate enough to be released audibly.

The binding principle predicts that in obstructent clusters of mixed continuancy the stop rather than the fricative will determine when the peak glottal opening will occur because only the stop is released with a burst. That an initial fricative controls the timing of the glottal opening in fricative plus stop clusters is a real problem for the binding principle, despite the fact that a temporal compromise occurs between the two segments, as indicated by the shift of the peak glottal opening to a point later than it would occur in the fricative alone. This compromise is not sufficient, however, to prevent the neutralization of the aspiration contrast in the following stop. In fact, glottal abduction in voiceless obstructent clusters with just a single glottal opening is generally earlier in clusters beginning with fricatives than in those beginning with stops. This is true regardless of the manner of the following obstructent. That the manner of the first rather than
the last obstruent in a cluster determines when the peak glottal opening will occur clearly conflicts with the predictions of the binding principle.

On the other hand, these differences between clusters beginning with fricatives and those beginning with stops may be an artifact. In all the data available on the timing of glottal articulations in obstruent clusters beginning with a fricative, that fricative is a sibilant. Sibilants are the noisiest of fricatives and they also have strong peaks in their spectra. In their high intensity and clear timbre, sibilants grossly resemble vowels, and these two properties make them detectable and identifiable at the syllable periphery. Producing such intense, spectrally distinct noise demands, in addition to a sufficiently small oral aperture to produce turbulence (Stevens 1971, Shadle 1985), a large enough glottal aperture to produce a high glottal air flow. Sibilants may therefore be exempt from the constraints imposed by the binding principle since their requirements for a high glottal air flow probably exceed those of stops.

Browman and Goldstein (1986) also suggest that sibilant plus stop clusters are a special case, though in their view this is because sequential oral articulations are incorporated into the domain of a single opening of the glottis, rather than because of the demand for high glottal air flow to produce high oral air flow by the sibilant. Though all voiceless fricatives may require a large glottal aperture and be expected to incorporate a following stop, non-sibilant fricatives do not generally occur external to stops in syllables. Since only sibilants and not all fricatives, much less all continuants, are incorporated with following stops within single glottal gestures, this coordinative structure, with the loss of an aspiration contrast in the stop it produces, cannot be generalized to other sequences of a continuant followed by a stop.

Turning now to the third problem, Hutters (1985) presents evidence from five languages that the duration of aspiration varies inversely with the duration of the stop closure, apparently because the stop is released before peak glottal opening is reached in the languages with shorter closures, Danish and Hindi, but not until after peak glottal opening in the languages with longer closures, Swedish, English, and Icelandic. These data indicate that the glottal gesture is ballistic and relatively invariant in its timing across these five languages. The timing of the oral closure instead varies cross-linguistically and certainly does not appear to dictate when the peak glottal opening will occur. If the timing of the glottal gesture is roughly invariant in all five languages, the length of the aspiration interval depends on
when the oral release occurs relative to the glottal gesture, rather than the glottal gesture being timed with respect to the oral one. This suggestion acquires further support from Hutters' Danish data, in which the interval from the onset of the oral closure to peak glottal opening is nearly constant for stops differing in place of articulation, even though closure and aspiration durations vary inversely across places of articulation. The explanation for these differences cannot be found in the phonologies of these languages, since Icelandic and Danish employ aspiration in quite parallel ways to distinguish their two classes of stops from one another, as do Swedish and English.

Closure durations also differ systematically between aspirated and unaspirated stops, being longer in the latter than the former. The duration of closure in an unaspirated stop is in fact approximately equal to the closure plus aspiration interval of an aspirated stop (Hutters 1985, also Weismer 1980). In English, the duration of constriction in voiceless fricatives is also approximately equal to the duration of closure in unaspirated stops (Weismer 1980), but both the timing of the gesture and its magnitude must be specified because there is an early and large peak in voiceless fricatives compared to the early but small peak in voiceless unaspirated stops and to the late and large peak in voiceless aspirated ones. The two kinds of stops still cannot be produced with the same abductory gesture since the peak opening is much larger in the aspirated than unaspirated stops. It is also not possible to specify velocity alone, since that would not predict the differences in the size of the opening between aspirated and unaspirated stops.

Finally, the fact that closure durations vary across languages in aspirated stops with complementary variation in the duration of aspiration does not mean that glottal articulations are not bound to oral ones. So long as the peak glottal opening occurs at some constant interval from the oral release, then it can be said to be bound to it. The binding principle does not require that this interval be the same for all languages. Significant counterevidence would be a demonstration that closure duration could vary in a single language without proportional variation in the timing of the peak glottal opening. In this connection, we might appeal to a difference between macro-binding, the pattern of covariation in the timing of one autonomous articulation with respect to another, and micro-binding, a specification that the bound articulation occur at a constant interval from the event to which it
bends, here, peak glottal opening with respect to the stop release (the distinction is Ohala's, p.c.).

5 Conclusion

The binding principle claims glottal articulations will bind more tightly to oral ones in stops than in continuants and that a glottal articulation would be coordinated with the release of the oral articulation because in that way the release would be shaped acoustically by the glottal articulation and thus convey the nature of that glottal articulation. Support for the first part of the principle has not been presented in this paper, and it will be taken up elsewhere.

The second part of the principle was tested first against data from Icelandic and Hindi. One problem, the possibility of binding of glottal abduction to the onset of a following stop closure in the pre-aspirates in Icelandic was shown to be only apparent problem. Since pre-aspiration is a generally a phonetic property of heterosyllabic clusters ending in an underlying [+ spread] stop -- through a shift of the glottal abduction to the first member of the cluster --, while post-aspiration arises only when such stops occur singly or in tautosyllabic clusters, pre- and post-aspiration may never contrast in the phonology of Icelandic. The apparent coordination of the glottal opening with the following oral closure was shown to be artifact of variation in the amount of overlap between the oral closure and the interval when the glottis was completely abducted; no relational invariance between glottal abduction and oral closure was actually found. The absence of evidence of coordination between glottal abduction and the preceding closure in Icelandic post-aspirates remains unresolved. The data from Hindi more closely followed the predictions of the binding principle, but this suggested a difference in coordination schedules between the two languages rather than the contrast between stops of different types predicted by the binding principle.

The unresolved problems may not expose weaknesses of the binding principle so much as they reflect difficulties in determining the tightness of binding. In particular, measuring covariation in duration may not accurately reveal how or what events are coordinated with one another. This would be especially likely if it were not the onset and offsets of articulations which are coordinated but rather something like their peak displacements. The evidence reviewed above suggests that it is peak opening of the glottis which exhibits invariance with
respect to some part of the oral articulation it accompanies. Unless abduction were strictly ballistic, coordination of peak opening with the stop release would probably not be clearly observable in measurements of covariation in duration.

Implicit in the traditional matrices of features employed until recently to represent the component articulations of segments is the claim that each articulatory gesture begins and ends simultaneously. The glottal and oral articulations of unaspirated stops and voiceless fricatives actually exhibit just this pattern of coordination, glottal abduction beginning at nearly the same time as the oral closure of constriction is complete and adduction being complete at the point when the closure or constriction is released. Furthermore, peak glottal opening is coordinated in unaspirated stops and fricatives with the onset of closure or constriction. Voiceless fricatives and aspirated stops contrast in the timing of glottal abduction, a difference which is perhaps best expressed in terms of contrasts in opening velocity, the fricatives being produced with much rapider opening than the stops. Abduction in aspirated stops is, furthermore, coordinated with the release of the stop; it varies in when it occurs with respect to the onset of closure, but not with respect to the release.

The explanatory power of the binding principle arises from the fact that the contrast between aspirated stops on the one hand and voiceless fricatives or unaspirated stops on the other depends on TIMING differences. These timing differences are what is controlled in conveying these contrasts; their results are a markedly different acoustic quality between the burst and the brief interval after it in aspirated and unaspirated stops or between the early noise of fricatives compared to late noise in aspirated stops. The binding principle assures that these acoustic differences exist.

Most troublesome in the long run for the binding principle is the fact that the timing of glottal abduction in voiceless obstruent clusters is determined by the manner of the initial rather than the final obstruent. The difficulty, of course, arises from the fact that that obstruent's release may not be acoustically signalled by a salient, transient event such as a burst. The more general problem for the binding principle posed by such data is that it suggests that glottal and oral articulations are scheduled with respect to the beginning of the articulatory unit rather than looking ahead to its end, despite the possibility that no audible release may occur in the middle of the cluster to convey the state of the glottis. The difficulty presented by these data is only reduced somewhat by the fact that in all
the clusters beginning with fricatives, the fricative is a sibilant, which demands such a large glottal aperture that it may override the binding principle. This raises the question of how this principle applies in a grammar.

Considering stops alone, the binding principle may be limited to single stops and not extend to clusters of stops. Rather than enforcing a consolidation of glottal gestures into a single gesture aligned with the audible release of a final stop in a cluster, the glottal gesture of each stop may bind to its release, whether or not it is audible. (Only when the cluster contains a word boundary are there separate openings; otherwise, it appears that only a single glottal abduction occurs.) If the possibility of an audible release, rather than its actual occurrence, is sufficient to determine when glottal abduction occurs, then the binding principle only determines the schedule of articulation within the domain of a single segment, even if that segment forms part of a larger articulatory unit, as it does in a cluster of voiceless stops in which the separate glottal opening gestures of each member appear to have been consolidated into a single opening. The schedule of glottal articulations remains what it would be if the segment occurred alone.\footnote{This result closely resembles Browman and Goldstein's (1986) demonstration that the movement of the lower lip in bilabial articulations is the same for [p], [b], [m] and [mb], [mp], i.e. whether the accompanying soft palate articulation is single or double. The acoustic result which the binding principle is intended to assure, a burst positively modified by the glottal articulation bound to it, is apparently dispensable in clusters, perhaps because the glottal articulation is predictable within most clusters in the languages investigated (I had originally hoped that the binding principle would provide an account of this predictability). Restricting the binding principle to single consonants shows that neither it nor other constraints are solely---

\footnotetext{5}{Strictly speaking, this is not true, since as the evidence of sibilant plus stop clusters shows the peak in the glottal abduction occurs at the boundary between the sibilant and stop articulations, a shift to a later point compared to the fricative occurring alone.}
responsible for determining when glottal articulations occur relative to oral ones in consonants. Coordination arises instead out of a combination of constraints.

Acknowledgments

What follows has been much improved by the concerted efforts of readers of earlier versions, alphabetically: Mary Beckman, Ailbhe Ni Chasaide, Nick Clements, Louis Goldstein, Johanna Nichols, John Ohala, Donca Steriade, and Kenneth Stevens; you, the reader, should be grateful for their efforts. Kim Silverman's advice regarding the statistical treatment of these data was invaluable. I am also grateful to the speakers of Icelandic and Hindi whose speech is described here, Gudrun Thorhallsdottir and Veneeta Srivasta. Final responsibility for everything here remains, of course, with me. The work reported in the section on preaspirates was carried out in the Phonetics Laboratory in the Department of Modern Languages and Linguistics at Cornell University with support from the College of Arts and Sciences.
References


Figure Legends

Figure 1: Correlation plots for durations of glottal abduction with the adjacent closure and vowel for all data pooled: a) Icelandic pre-aspirated stops; b) Icelandic post-aspirated stops; c) Hindi breathy voiced stops; d) Hindi post-aspirated stops.

Figure 2: Waveforms of the end of a vowel before a preaspirated stop: a) in the word stökkur "brittle, nom. masc." spoken in focus at a moderate rate and b) of the same word, but out of focus at a fast rate. Note the marked reduction in the duration of noisy component, compared to the inertness of the breathy component.

Figure 3: Correlation of duration of breathy and noisy components with total duration of pre-aspiration in Icelandic; b) correlation of duration of breathy and noisy components with combined duration of preceding vowel and pre-aspiration; c) correlation of duration of breathy and noisy components with combined duration of preceding vowel and pre-aspiration.

Figure 4: a) A diagram of the relationship between the oral and glottal articulations in pre-aspirated stops (cf. Petursson 1976), b) variation in the interval between the onsets of the oral articulations, without varying the glottal articulation, leading to varying amounts of overlap of the interval of wide glottal abduction which produces the noisy component of pre-aspiration by the stop closure.
CLOSURE (solid) $y = -12.9565 + 0.531x$, $r = 0.83$
VOWEL (open): $y = 3.4519 + 0.2534x$, $r = 0.38$
CLOSURE (solid): $y = -8.6663 + 0.4273x \quad r = 0.79$

VOWEL (open): $y = -10.5888 + 0.4871x \quad r = 0.67$
BREATHY (solid): $y = 27.6877 - 0.0057x \quad r = 0.01$

NOISY (open): $y = -27.6877 + 1.0057x \quad r = 0.92$
BREATHTY (solid): \[ y = 35.1074 - 0.0632x \] \( r = 0.20 \)

NOISY (open): \[ y = -54.5172 + 0.7312x \] \( r = 0.93 \)
BREATHTHY (solid): $y = 26.2474 + 0.0077x \quad r = 0.03$
NOISY (open): $y = -20.871 + 0.399x \quad r = 0.71$
Research Facilities in the Cornell Phonetics Laboratory

G. N. Clements, S. Hertz, and J. Kingston

The Phonetics Laboratory of the Department of Modern Languages and Linguistics (DMLL), Cornell University, is a fully-equipped laboratory for the study of speech. Its facilities are aimed at providing an integrated environment for the study of the phonological and phonetic properties of human languages through a variety of experimental and computational approaches.

1. Overview of the Lab's Facilities

The Phonetics Laboratory has a variety of computers and specialized hardware and software for speech analysis and synthesis. Its computers include a Sun 3/160C Workstation running the Unix operating system, an IBM-AT running DOS, a PDP-11/73 running RSX-11, a PDP-11/40 (rapidly becoming obsolete) running RT-11, and a Macintosh II. We are also considering attaching a second Sun workstation to our Sun 3/160C. We have interfaced all our computers together so that they can easily pass data back and forth and share printers and other facilities. We currently use a Hewlett Packard Laserjet printer, and will soon acquire an Apple Laserwriter. A more detailed description of the computer configurations is given in section 2.

In addition to general-purpose programs, such as word processors and statistics packages, the computers have the following specialized software for the study of speech:

- ILS: the Interactive Laboratory System, a set of self-contained interactive digital signal processing programs developed by Signal Technology, Inc. for the acoustic analysis of speech, which displays results in both graphic and alphanumeric form. This software has been customized to some extent by writing interactive command files for frequently-used procedures. It currently runs on our PDP-11/73.

- SRS: the Speech Research System, an interactive speech synthesis rule development system developed by Susan Hertz at Cornell between 1974 and 1982, which is presently running on the PDP-11/40.
• SRS text-to-speech rules for English and Japanese. This consists of two synthesis rule sets developed with SRS: the English rules developed by Susan Hertz with help from students, and the Japanese rules developed by M. Beckman, O. Fujimura and S. Hertz. The English rules take ordinary text as input and produce values for an OVE IIIId formant synthesizer as output. The Japanese rules take a romanized input string annotated with intonation marks as input and produce synthesizer values as output.

• Delta: a rule development system designed for developing and testing morphological, phonological, phonetic, and speech synthesis rules that operate on multi-level utterance representations consisting of synchronized units of the user's choice, such as phrases, morphs, syllables, phonemes, articulatory parameters, durations, and so on. The Delta system was developed by Hertz and her associates at Eloquent Technology in cooperation with the DMLL and the Computer Science Department, Cornell. It currently runs on the IBM-AT, the Sun Workstation, and on a VAX-780 in the Computer Science Department, and will soon be ported to the Macintosh II.

• TFED: the Time Function Editor, a program developed by Nirvonics for editing and analyzing synchronized data streams, which currently runs on the IBM-AT.

• A custom-designed experiment runner that allows the experimenter to specify a variety of formats for presenting stimuli and acquiring data in perception experiments, which runs on the IBM-AT.

• A custom-designed pitch tracker using the AMDF method, optimized by center-clipping the data before applying the algorithm, which will run on the IBM-AT.

• A version of the Klatt software synthesizer developed by Dennis Klatt, which will run on the IBM-AT.

• Various on-line wordlists and pronouncing dictionaries, including the Brown-Lancaster/ Oslo/Bergen corpus of Modern English, the SRL pronunciation dictionary, and the American Heritage word frequency list.
Our specialized speech hardware includes:

- Kay’s DSP real time digital Sona-graph model 5500, a programmable workstation for recording, processing, displaying, editing and resynthesizing speech signals. The DSP sonograph produces real-time spectrographic, FFT, waveform, amplitude, LPC analysis/synthesis, \( F_0 \) and other analysis displays on a high resolution monitor. It can simultaneously analyze two signals in real time in a split screen mode, and stores 2 Mbytes of sampled speech (about 50 seconds) in memory. It prints gray scale sonagrams to a Mitsubishi video copy processor.

- Voice Identification’s Precision Methods (PM) real-time pitch analyzer, a microprocessor controlled device which displays \( F_0 \) and intensity traces as well as digital data on a CRT monitor, running PM 100, which displays the \( F_0 \) curve and/or the intensity curve for any two utterances in split screen mode, and PM 201, which displays the \( F_0 \) curve of an utterance on the upper half of the screen and the intensity curve on the lower half. It prints hard copy to a Mitsubishi printer.

- Kay’s 6061B analog Sona-graph, a standard spectrum analyzer, largely rendered obsolete by the DSP Sona-graph described above.

- an Ove IIId formant synthesizer, used in conjunction with the speech synthesis software described above.

We have also finished installing new speech analysis software for the Sun Workstation which provides an alternative to ILS.

The Phonetics Lab is located in the basement of Morrill Hall, where it shares space, personnel, equipment and many facilities with the Cornell Language Laboratory. The shared facilities include a soundproof recording studio, several reel-to-reel and cassette tape players/recorders, and a variety of audiovisual facilities, including VHS video playback/record units for foreign and domestic tapes and a JVC video camera. The Language Lab’s collection of taped foreign language instructional and study materials is one of the world’s largest and provides an additional research and teaching resource. The two labs share a fulltime Technical Coordinator and an Administrative Aide, and the Phonetics Lab has a Systems Programmer/Analyst.
2. Specifics of the Computer Configurations

2.1. The Sun Workstation

The Sun 3/160C runs the Unix operating system. It has C, FORTRAN and PASCAL compilers and an Ethernet connection to the IBM-AT, a Digital Signal Processor, and a high-resolution color graphics display, with a 141 Mbyte disk (we are thinking about getting a second one) and 8 Mbytes of RAM. It is connected to campus computers through a UUCP (modem) and has dial-in capability from offcampus terminals.

The Sun Workstation is a relatively new acquisition. We are still in the process of obtaining needed software and hardware, such as analog-to-digital and digital-to-analog converters.

2.2. The PDP 11/73

The PDP 11/73 runs the RSX-11M-Plus operating system, allowing several people to use it at the same time. The 11/73 has a Skymink array processor, uses Data Translation A/D & D/A boards for data acquisition, and is connected to a Tektronix 4105 color graphics terminal. Presently, the 11/73 is the principal site for digital analysis of speech acoustics, using the ILS software. The 11/73 is also connected to the IBM-AT, which may be used as a graphics terminal through a Tektronix 4014 emulator, thus allowing access to ILS and the other software on the 11/73.

2.3. The IBM-AT

The IBM Personal Computer AT is designated to run the TFED waveform editing software, the pitch extraction software, and the experiment runner described above. It also has Data Translation A/D & D/A boards, and can be used directly both to acquire and emit signals.

2.4. The Macintosh II

The Apple Macintosh II has 2 megabytes of RAM and an internal 40-megabyte hard disk. Apart from general purpose applications it is designated as a terminal to the Sun Workstation.