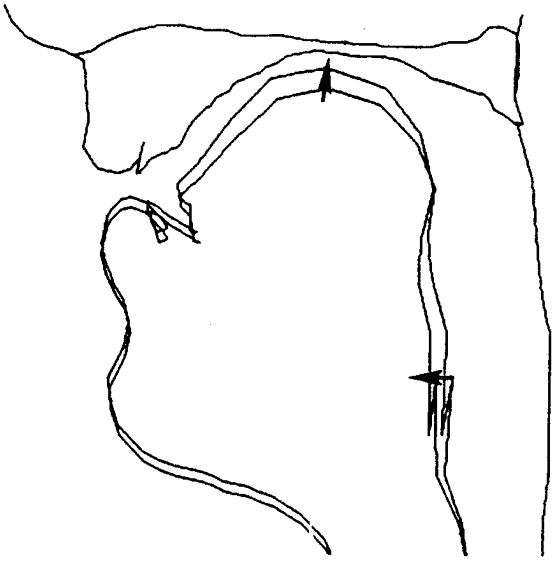
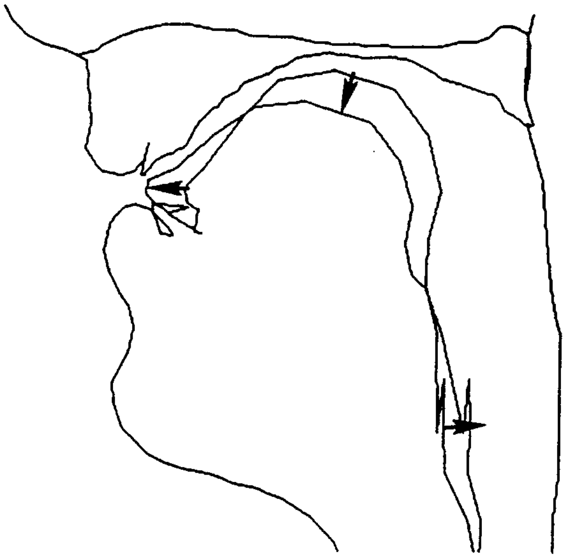
Title
WPP, No. 70

Publication Date
1988-07-01

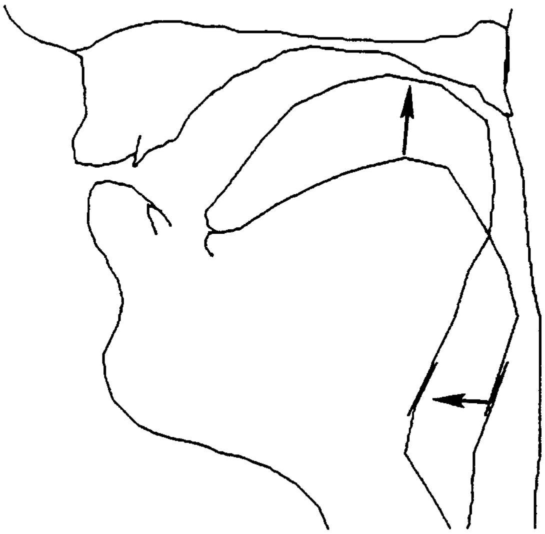


Number 70

July 1988



UCLA WORKING PAPERS IN PHONETICS



UCLA Working Papers in Phonetics 70

July 1988

Hierarchical features of the International Phonetic Alphabet

Paper presented at the Fourteenth Annual Meeting
of the Berkeley Linguistic Association

Peter Ladefoged
University of California, Los Angeles

Most American linguists think of the International Phonetic Association (IPA) as a conservative group of traditional phoneticians, whose concerns are very different from those of contemporary phonologists. They may be right in their assessment of the traditionalist nature of many IPA members (although it should be noted that the Council of the IPA now contains several younger linguists); but they are certainly wrong in viewing the concerns of the IPA as being different from their own. Throughout its hundred year existence the main endeavour of the IPA has been to provide accurate phonetic descriptions of languages, based on phonological principles as then understood. The Association's alphabet was devised as a tool to aid in the formulation of such descriptions. The alphabet has been continually revised, and there is today a strong movement to bring the present alphabet more in line with contemporary thought. A convention for this purpose is being held in Kiel, Germany, in August 1989, and all who are interested (whether members of the IPA or not) are welcome to attend. (Further particulars are available from: IPA Convention, Linguistics Department, UCLA, Los Angeles, CA 90024-1543.)

The linguistic foundations of the alphabet are evident throughout its history. In 1900 the IPA published an *Exposé des principes* containing a table showing the recommended alphabet. This table was set up so that it included "les *sons distinctifs* de toutes les langues étudiées jusqu'ici". (My italics.) The principles mentioned in the title of this and other early publications of the IPA (*Aims and principles*, 1904; *Exposé des principes*, 1905) were all concerned with language teaching. It is not until the 1908 *Exposé des principes* that, in addition to a section "Principes pédagogiques," there is also a section "Principes de transcription pratique." This section notes that: "Pour chaque langue, on représente les sons distinctifs, et ceux-là seuls." Similarly, the 1912 English version, in a section headed "principles of transcription for languages hitherto not transcribed," notes, long before the phoneme became a popular notion: "It is necessary to ascertain what are the *distinctive* sounds in the language, i.e. those which if confused might conceivably alter the meanings of words." (Italics in the original.) The corresponding section in the 1922 *L'Écriture phonétique internationale* uses the then new term 'phoneme' saying: "Pour chaque langue, on représente les *phonèmes* ou sons distinctifs, et ceux-là seuls." (Italics in the original.). The latest (1949) edition of the *Principles* makes as its first point: "There should be a separate letter for each distinctive sound; that is, for each sound which, being used instead of another, in the same language, can change the meaning of a word. "

The second principle in the current edition is also relevant to contemporary phonological concerns, in that it presupposes the existence of a set of universal phonetic categories, making it meaningful to equate sounds in different languages. It says: "When any sound is found in several languages, the same sign should be used in all. This applies also to very similar shades of sound." This principle is especially important when taken into account with another IPA practice which has never been formally stated as a principle, perhaps because it is regarded as too obvious to mention. This is the principle that the symbols of the alphabet should be defined in terms of general phonetic categories very much of the kind that we now regard as features. Phonetic theory in the early days of the IPA was greatly influenced by the work of Sweet and Bell, both of whom had developed systems for classifying all the sounds that were known to be able to distinguish meanings in the world's languages. Bell's *Visible Speech* (1867) and Sweet's *Handbook of Phonetics* (1877) provided iconic symbols for showing the combinations of articulatory elements present in a sound. These same elements (or at least a subset of them) were used to define the symbols of the alphabet. Throughout its history the alphabet has consisted of symbols defined in terms of intersections of phonetic categories (features). Most of the symbols are defined by the terms naming the rows and columns of the charts, and by the convention that when there are two items in a single cell the first one designates a voiceless sound (if there is a single item in a cell it is always voiced). In addition a few symbols and several diacritics are defined by supplementary notes. The whole work -- principles, charts, symbols and notes -- constitutes the IPA's theory of phonetic description.

Given this background we may now compare an IPA description with a feature specification of the kind that is nowadays more common. The location of [m] in the chart explicitly indicates:

+ voiced
+ bilabial
+ nasal

and, by means of the labels along the lefthand edge of the chart, the fact that this is a consonant made with the pulmonic airstream mechanism. In much the same way, Chomsky and Halle (1968:5) note that they will use symbols as "informal abbreviations for certain feature complexes." For them this symbol would be a shorthand way of designating the feature values:

+ voiced
+ nasal
+ anterior
\- coronal
+ sonorant
etc.

In both cases several other feature specifications are implied. IPA [m] implies [- dental, - alveolar, etc.; - implosive, -click, etc.]; and in SPE (Chomsky and Halle 1968), it is made clear that there are also a number of other features such as [Glottalic] the values of which, like some of those noted in (2), can be determined by marking conventions. Bearing these two approaches in mind, we may consider the extent to which the similarities between them could be increased in any future revision of the IPA *Principles*, symbols, charts, and accompanying notes.

The nature of feature systems

The first point to emphasize is that the two approaches are very different in some of their basic premises. It is true that they both describe segments in terms of features, and in some cases, such as Nasal, they both use the same terms. But, as has been shown by Halle and Ladefoged (1988), the hierarchical organization of the IPA feature set is very different from that of SPE or contemporary phonologies. In particular, the IPA has separate charts for vowels and consonants, whereas it is a major point of SPE and other phonologies in the same tradition that both vowels and consonants should be described in terms of the one set of features. As will be made clear below, I think both positions are correct. Another major difference between the two theories is that the IPA has little internal organization to the set of place categories other than (in some charts) the grouping together of some immediately adjacent places. Contemporary phonologies (e.g. Clements 1985, Sagey 1986, Halle 1988) recognize that there is far more structure imposed by the articulatory system. Considerations such as these led Halle and Ladefoged (1988) to propose that the major features that should be characterized by the symbols of the alphabet should be as shown in figure 1.

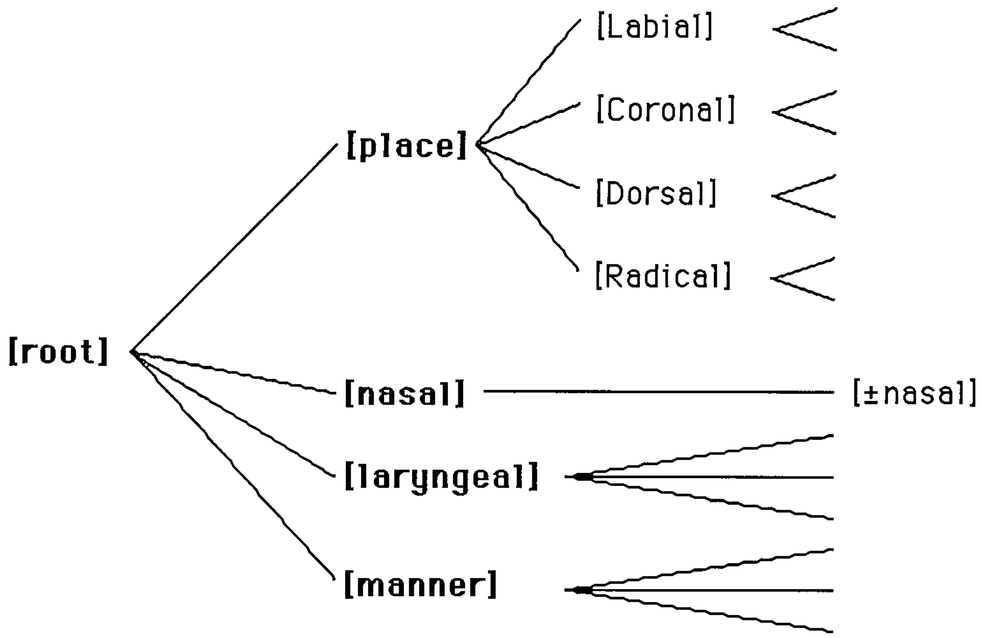


Figure 1. The hierarchical organization of the major features.

These are clearly the major features of segmental phonology; but, equally obviously, they are not sufficient for describing the sound patterns of languages. In the remainder of this paper I will sketch some of the additional structure that I consider necessary for phonological specifications. I will then consider briefly how this information should be regarded in a revised version of the International Phonetic Alphabet.

A hierarchical feature structure

An overview of the proposed structure is given in figure 2. This tree structure represents a conjecture about the phonological resources, the features and their relations, that are available to the languages of the world at the level of the segment. It should be emphasized that this figure gives only a tentative, incomplete view of the relations among features. Nevertheless, it forms part of a statement defining the phonological possibilities that can occur. The arrangement of features into a tree structure has also been used by phonologists for other purposes, notably the grouping of properties that co-occur in spreading rules. In this paper, however, the aim is simply to provide a way of representing the major constraints on phonological segments. This aim is very much in the spirit of the IPA tradition. As we have noted, the International Phonetic Alphabet has always been an attempt to represent all and only the distinctive sounds in the world's languages. We would now like to go a step further and list not just the sounds but also the features that characterize them, together with the constraints on feature interaction. In this sense figure 2 is a first step towards defining the possible phonological segments in the world's languages.

In order to serve this purpose a great deal has to be added to the tree structure in figure 2. In the first place we need to state the conventions governing the possible

Type	Hyper feature	Major node	Feature	Traditional term	Brief description
			Voice	*Voiced*	periodic low frequency energy
				Voiceless	- absence of such energy
			Grave		aperiodic low frequency energy
					- absence of such energy
Auditory			Sibilant	*Sibilant*	aperiodic high frequency energy
					- absence of such energy
				High vowel	low F1
			Height	*Mid-high vowel*	- mid low F1
				Mid-low vowel	- mid high F1
				Low vowel	- high F1
			Brightness	*Front unrounded*	high (F2' - F1)
				Back rounded	- low (F2' - F1)
			Sonorant		periodic well-defined formants
					- no periodic formant structure
			Rounding	*Rounded*	decreased lip opening
		Labial		*Unrounded*	- spread lips
					protruded
			Protrusion	*Bilabial*	- neutral
				Labiodental	- retracted lower lip
				Linguolabial	(tongue between lips)
				Laminal	laminal
	Place		Apicality	*Apical*	- apical
		Coronal -		*Retroflex*	- sublaminal
				Dental	advanced
			Anterior	*Alveolar*	- neutral
				Postalveolar	- retracted tongue tip/blade
		Dorsal	Front	*Front* *Palatal*	front
				Back	- back tongue body
				Velar / high vowel	high
			High	*Uvular / mid V*	- mid
			(Low)	*Pharyngeal / low V*	- low tongue body
		Radical		*Epiglottal*	retracted tongue root
					- advanced tongue root
				Stop	complete closure
			Interrupted	*Fricative*	- nearly complete closure
Physiological				*Approximant*	- approximation of articulators
			Lateral	*Lateral*	predominantly lateral airflow
	Manner			*Central*	- no lateral airflow
					held gesture
			Dynamic	*Trill*	- vibrating
				Tap	- ballistic
				Flap	- ?
	Oro-nasal		Nasal	*Nasal*	lowered soft palate
					- raised soft palate
				Creaky	stiff
			Stiffness	*Modal*	- neutral
			(Slack)	*Breathy*	- slack vocal cords
	Laryngeal			*Aspirated*	open
			Glottal aperture	*Unaspirated*	- narrowed
			(Closure)	*Glottal stop*	- closed vocal cords
		Pulmonic	Fortis	*Fortis*	segmental lung power
				Lenis	- no increased lung power
	Airstream			*Implosive*	downward
		Glottalic	Glottal		- no glottal movement
			movement	*Ejective*	- upward glottal movement
		Velaric	Click	*Click*	oral suction
					- no oral suction

Figure 2. A hierarchical arrangement of feaures forming part of a definition of the set of possible phonological segments in the languages of the world.

paths through the tree; but at the moment we do not know what these are. The general convention for reading the tree structure is that the maximum possible set of phonological segments is achieved by taking each path through every node except the terminal nodes (the features) where one of a set of choices has to be made. But this convention does not apply to the nodes in the third column. The Place node dominates a set of features such that for most sounds only one path has to be selected, but for some sounds more than one may be selected, and, arguably, for some, such as a glottal stop, none of the options is selected. Similar remarks apply to the choice of airstream mechanism, as will be elaborated later.

There are several other cases, such as the properties dominated by the Manner node, in which the inter-relations are too complex to be given in the form of a tree structure. There are also cases for which the level of our ignorance is such that we cannot even indicate the immediately superior nodes for a given sound (e.g. in the description of linguolabial consonants; Maddieson 1987). It is readily apparent that figure 2 is far from complete, and numerous additional statements are needed to define the limits of possible feature combinations. However, this should not cause us to overlook the fact that the figure does list many of the possibilities, and many of the required hierarchical properties *are* formalized by the lines indicating the necessary dominance relations.

As a further guide to the interpretation of this figure, very rough descriptions of the features are given in the extreme right column, the properties connected by dashes indicating sets within which choices must be made. Note that the choices are often binary, but on a number of occasions (e.g. for Protrusion and Apicality) there are three possible terms. To the left of these, in the penultimate column, there is a set of italicized terms indicating the traditional (usually IPA) terms. These terms do not have any formal standing within the theory of phonetic description being advanced here. They are simply useful (and familiar) terms summarizing certain feature combinations. Thus *Postalveolar* is equivalent to [- anterior, +coronal], and *Velar* to [+high, +dorsal].

The first and most important difference between the trees in figures 1 and 2 is that the latter tree contains another whole branch specifying auditory properties. All the features in figure 1 are ultimately defined in terms of actions of the vocal organs. But there are many important natural classes of sounds that arise because sounds have certain auditory properties in common. It is somewhat ironic that this great insight of the Prague school, much touted by Jakobson, Fant and Halle (1951), should now be overlooked by the phonologists who are their successors. The present situation arises partly because of the view of phonology first seen in SPE (Chomsky and Halle 1968), in which features are considered to be mental entities. From this point of view it is just a matter of exposition as to whether features are defined in articulatory or acoustic terms. But this is simply not true. Segments get grouped together into natural classes not because of some general mental property, but because of specific properties relating to the way sounds are heard, or to the way they are produced. Of course all features have both articulatory and acoustic properties in the sense that features are linguistic units that characterize the lexical items of a language. These lexical items have to be capable of being both spoken and heard. But it does not follow from this that we should consider the linguistic function of a feature as being required in both domains.

I do not want to overstate my case in this matter. Chomsky and Halle are correct in considering that for many aspects of sounds the correlation between the auditory properties and the physiological properties is so great that it really does not matter whether we define the feature in auditory or articulatory terms. Thus the feature Voice can be defined equally well in either way as is done by Jakobson and Halle (1956:30):

"acoustically -- presence *vs.* absence of periodic low frequency excitation; genetically -- periodic vibrations of the vocal cords *vs.* lack of such vibrations." In the list in figure 2 I have, somewhat arbitrarily, chosen to put Voice among the auditory features.

No such arbitrary choice is possible for some features, such as Nasal. The articulatory correlate is clear (lowering of the velum); but despite enormous pressure from speech pathologists, who need a simple way of measuring the degree of nasality of a sound, nobody has been able to suggest an acoustic attribute common to all nasalized sounds. Again, I am not saying there are no acoustic correlates of nasality; obviously there are, else we would not be able to hear whether a given vowel is oral or nasal. But from the point of view of how segments can be grouped into natural classes, it is not the diverse acoustic properties that are the basis of the grouping, but the fact that all nasal sounds are *produced* with something in common.

Auditory feature definitions

A large proportion of the features required for phonological purposes are defined, as Nasal and the other features shown in figure 1, in terms of articulatory properties. But, just as there is no definition of the acoustic correlates of Nasal that is useful for phonological purposes, so there are other features that have no phonologically useful articulatory correlates. We will begin our more detailed examination of the features listed in figure 2 by considering these auditory features. One of the most well known is the feature Grave, which groups some Labial and Dorsal sounds in accordance with their spectral characteristics. Sounds such as [p,k,f,x] are produced in very different ways, but they sound similar because they have a comparatively large amount of aperiodic acoustic energy in the lower part of the spectrum. This similarity is reflected in morphological alternations such as those in Bantu languages (e.g. Rutooro; Ladefoged, Glick and Criper 1972) and historical changes such as English [x] to [f] in words such as 'rough, tough,' a change that is completely inexplicable in articulatory terms.

Chomsky and Halle discarded the feature Grave because they found it did not provide a satisfactory basis for characterizing differences in place of articulation. This is undoubtedly true; from an articulatory point of view the feature Grave does not distinguish the appropriate natural classes. But this does not mean that it fails to characterize a natural class of sounds from an auditory point of view. Throwing out Grave just because it does not have a useful articulatory correlate is as bad as it would be to throw out Nasal just because it does not have acoustic correlates that themselves form a basis for a natural class.

Note that the feature Grave as proposed in this paper is not exactly the same as the feature proposed by Jakobson, Fant and Halle (1951). Their definition was "the predominance of one side of the significant part of the spectrum over the other." It was intended to include both consonants and vowels. My feature Grave is in practice restricted to obstruents (and, perhaps, voiceless approximants) because it stipulates that the auditory characteristic of a Grave sound is that there is salient *aperiodic* energy in the lower part of the spectrum. In speech, this type of energy occurs only in stop bursts and fricatives (and, perhaps, a voiceless labial-velar approximant). There is no auditory property of this sort that links particular vowels with particular consonants. (But there are, of course, links between particular vowels and consonants specified by the articulatory features High, Low and Back.)

Note also that this definition of Grave implies that [- grave] sounds are not necessarily Acute in the old Jakobsonian sense. All sounds that do not have a significant amount of aperiodic energy in the lower part of the spectrum are [- grave], irrespective of whether they have a significant amount of aperiodic energy in the upper part of the spectrum or whether they do not have any aperiodic energy at all.

Another auditory feature that is of importance in grouping consonants I have here called Sibilant, following the traditional phonetic usage. It is not exactly equivalent to the Jakobsonian feature Strident in that the feature Strident has also been used to distinguish [f,v] from [ɸ,β], thus resulting in the rather unnatural class of strident sounds [f,v,s,z,ʃ,ʒ]. So as to make make the difference in definition plain, I have retained the traditional term Sibilant, which has long been used (e.g. by Holder 1669, and many phoneticians after him) to identify the class of sounds [s,z,ʃ,(ʒ)].

It is interesting to consider whether it might be possible to give an articulatory definition of this feature, in that Sibilant sounds are always pronounced with the jaw raised so that there is a very narrow gap between the upper and lower front teeth. The high frequency aperiodic acoustic energy that gives rise to the auditory characteristics of this feature is due to the jet of air striking this narrow gap (Catford 1977, Shadle 1985). However, the fact that sibilant sounds have an articulatory attribute in common is an unlikely cause for their acting together in historical changes and morphological alternations. There is no evidence showing that jaw position is a salient characteristic of sounds causing them to be grouped together, whereas the auditory grouping of these sounds is evident in the perceptual confusion data of Miller and Nicley (1955) and its reanalysis by Shepard (1972), and in the perceptual similarity judgments reported by Ingram (1975).

It is appropriate at this point to consider what is at issue in claiming that a certain feature (e.g. Sibilant) should be defined in auditory rather than acoustic terms. It is not a matter of there is or is not a feature of this kind. There is little doubt that sibilants form a natural class of sounds that act together in phonological rules. Nor is it a matter of formal evaluation of rules. Given that there is a feature sibilant the system for evaluating its use within a phonology will be the same irrespective of its phonetic attributes. What is at stake is whether the auditory definition provides a better explanation for the grouping than a definition in terms of the articulatory attributes. Until there is some evidence for the shared articulatory properties being the reason for this grouping, it seems preferable to continue to maintain that the well attested salient auditory characteristics are the basis for the natural class.

The most outstanding features of the auditory type are properties of vowels. A problem that arises in discussing these features is that it has not been generally recognized that vowels have both articulatory *and* auditory properties. Hence the same name has been used for something that should be regarded as two distinct features. I will use the term (Auditory) Height to refer to an attribute that has as its acoustic correlate the frequency of the first formant. The other auditory feature of vowels is here called Brightness, a term ("Helligkeit") used by Trubetzkoy (1929, 1939), and more recently by Fischer-Jørgensen (1985). The acoustic correlates of Brightness may be taken to be the difference in frequency between the first formant and F2', a form of the second formant modified so as to account for the influence of the third formant. Algorithms for determining F2' have been given by Bladon and Fant (1978). From a physiological point of view, Brightness is a combination of all three articulatory vowel features, Front, Back, and Round. High front unrounded vowels have the highest value of Brightness, low back neutral vowels have a mid value and high back rounded vowels have the lowest value.

The explanatory power of the two auditory features for vowels is best exemplified by the dominance of the five vowel system [i e a o u]. Languages as diverse as Swahili, Spanish, and Hawaiian have five vowels, with qualities something like [i e a o u]. These and only these vowels are used by approximately 20% of the world's languages (Maddieson 1984). From an articulatory point of view, there is no reason why front unrounded and back rounded vowels should be more common than the reverse combinations. Phonologists who regard all features as having only articulatory definitions have no explanation for the remarkable facts of vowel distribution. There should be no doubt that in order to form the correct phonological classes of vowels these sounds have to be characterized in both physiological and auditory terms. The action of the body of the tongue in the production of a vowel is specifiable in terms of physiological features that are also applicable to consonants (and thus show the relations between vowels and consonants). But this does not preclude there being additional auditory features that are applicable only to vowels.

The remaining auditory feature listed in figure 2 is Sonorant. This is another very necessary feature that it is hard to define in articulatory terms. The notion 'spontaneous voicing' (Chomsky and Halle 1968) does not get at the essence of what it is that causes vowels, nasals, laterals and some approximants to be grouped together. Better articulatory statements can be made in terms of the function of the articulatory system as a whole: sonorant sounds are those in which the vocal cords are vibrating and there is no significant build up of oral pressure. But there is no evidence that vocal cord vibrations plus lack of pressure form a salient characteristic. Sonorant sounds are clearly related by having a periodic, well-defined, formant structure. Their grouping is not because they are made alike, but because they sound alike.

There are almost certainly other auditory features that will have to be included in future lists such as that in figure 2. One of these is the feature Rhotacized, which is associated with a lowering of the frequencies of the third and fourth formants. As has been shown by Lindau (1985) many forms of *r* share this characteristic. The fact that (as she also shows) some forms of *r* do not does not preclude rhotacization being an auditory feature that links some sounds in a natural class. Another possible auditory feature is Liquid, grouping together some kinds of rhotic and lateral sounds.

The organization of articulatory features

As most of the proposed articulatory features are well known, we need not consider explicit definitions of all of them. There is, however, much to be said about their hierarchical organization. The basic division into five hyper-features reflects the standard practice of articulatory phonetic description as seen in many textbooks. Abercrombie (1957) for example, notes that sounds can be described in terms of the place of articulation, the manner of articulation, the oro-nasal process, the state of the glottis, and the airstream mechanism. The same organization is apparent in Pike (1943), and has been taken over by Ladefoged (1971, 1982). The division of the Place node into four major nodes has received less formal recognition but it also has a respectable ancestry in, for example, Firth (1957). As has been noted elsewhere (Halle 1988, Halle and Ladefoged 1988), the fact that there are four major nodes attached to the hyper-feature place arises because these are the four independent articulatory possibilities. The further division of these major nodes is less clear, and full of complications. For example, the actions of the lips are extremely complex in that, in addition to being closed vertically as in a normal bilabial stop, they can also be protruded and rounded. Not all combinations of rounding and protrusion are possible. The feature system needs to be able to express the fact that bilabials and labiodentals can be rounded, and bilabials (but not labiodentals) can also be protruded. One way of doing this is by regarding Rounding and Protrusion as two separate possibilities, with Protrusion being a three valued feature accounting for the distinction between bilabial and labiodental sounds, as well as for the difference between Swedish high rounded vowels.

Below the Coronal node there are two features, Anterior and Apicality. The feature Anterior allows us to differentiate among places along the roof of the mouth, and thus distinguishes dental, alveolar, and postalveolar articulations. The three way division offers an appropriate way of showing within a single feature the low level allophonic variations that occur in such words as 'eighth, eight, tray' which in many pronunciations have dental, alveolar, and postalveolar allophones of /t/. Apicality distinguishes between articulations made with the blade of the tongue, the tip of the tongue , and the underside of the blade (cacuminal retroflexes).

The Dorsal node dominates the features necessary for specifying consonants made with the body of the tongue. These features also characterize some aspects of vowels. I have retained the terms High (Low) and Front (Back) for these physiological features as shown in figure 2, although it is not at all clear that the classes of vowels defined by tongue body positions are the same as those defined by the traditional terms which correspond more to the auditory features. We should also note that the features High and Front are multivalued features, each describing an ordered set of possibilities, although they could also be regarded as complexes of binary features, if binary terminal nodes are required. With this in mind the feature Low has been listed in parentheses. The Radical node also has implications for both vowels and consonants but it is as yet unclear how these should be formalized.

As has been noted by Sagey (1986), combinations of the major place nodes within a single segment are not uncommon. Labial plus Dorsal articulations as in [kp, gb, ŋm] are the best known; Lingual plus Dorsal articulations occur in clicks; and Radical plus Dorsal articulations occur in some Caucasian fricatives (Catford 1977). Following a suggestion made by Keating (1988) I have shown the traditional term palatal as representing a complex segment with both post-alveolar coronal and front dorsal attributes. What are traditionally known as secondary articulations (labialization, palatalization, velarization, pharyngealization) can be regarded as combinations of two different places involving different manners of articulation. There are 15 possible single and multiple combinations of the four major nodes within the hyper feature Place; we do not know how many of these can or do occur.

This leads us to a brief consideration of an interesting formal problem. Recall that figure 2 is intended to be part of a descriptive statement determining the possible phonological contrasts in the languages of the world. Every sound has to be able to have some value of each of the terminal nodes (the features). Thus each sound is either voiced or voiceless, it is either grave or it is not, etc. Note that this use of the tree structure cannot be maintained unless we allow some features to have three (or even more) values, so that, for example, the choices below Anterior consist of the set of mutually exclusive possibilities dental, alveolar and postalveolar, and those below High include the mutually exclusive possibilities high, mid, and low, for vowels and velar, uvular, and pharyngeal for consonants.

More work is obviously needed in the characterization of the set of possible manners of articulation. The hierarchical structure of the features dominated by this node is extremely hard to formalize. The first division I have suggested in figure 2 provides us with the three possibilities stop, fricative and approximant. As these items form a set of mutually exclusive possibilities, each of them can be considered as a distinct value of a single feature, here called Interrupted (a name I am not very happy with). As I noted (and then rejected) earlier (Ladefoged 1971:55): "These values form a linearly ordered set, by means of which we [can] give an explanatory account of lenition phenomena, in which stops weaken to fricatives, and a further weakening gives rise to approximants." This arrangement was rejected earlier because it did not permit fricative to be regarded as a value of a separate feature that could be added to stops for the characterization of affricates. Now, however, it seems best to regard affricates as sequences of feature specifications which can, if appropriate, occur within a single timing slot.

The next division among manner features provides the distinction between central and lateral sounds. Different values of the feature Lateral can occur with each of the values of Interrupted. Distinctions between central and laterally released stops are common (e.g. in Mayan languages); clicks are also forms of stops which utilize the central-lateral opposition. Central and lateral fricatives such as [s] and [ɬ] occur in Zulu and Welsh. Central and lateral approximants such as [ɹ] and [l] contrast in many languages, including most forms of English.

In addition to the more usual manner features I have suggested a new feature Dynamic (again, a name I am not very happy with) to account for distinctions and groupings among stops, trills, taps, and flaps. It seems likely that there is a natural class of this kind, but its internal organization is not completely clear. There is allophonic variation among stops and taps in many languages, including English. Similar variation among trills and taps occurs in languages such as Hausa; and diaphonic variation among forms of /r/ occurs in, for example, forms of Scottish English -- Ayrshire Scottish will have a trill where other forms of Lowland Scottish English have a tap or a flap. I am not certain whether the distinction between a tap and a flap is worth pursuing. I noted earlier (Ladefoged 1971) that "A flap is ... distinguished from a tap by having one articulator strike another in passing while on its way back to its rest position, as opposed to striking immediately after leaving its rest position [in a tap]." But this may be only an incidental difference between taps and flaps, as flaps (if defined as in the quoted sentence) always have a more retracted articulation than taps. It may therefore be appropriate to consider a flap as a tap with a different place of articulation. (Again, I am still uncomfortable with this, as the dynamics of the two gestures are so very different.)

In this paper little will be said about the Oro-Nasal and the Laryngeal hyper-features. The first of these is straightforward and needs no elaboration. The second is too complex, and too specialized, to be discussed here. I suggest that both Stiffness and Glottal Aperture are multivalued features (and if binarity is considered necessary, then it can be done by the addition of extra features, as indicated in figure 2 by the terms in parenthesis). The proposed features are similar but not identical to those proposed by Halle and Stevens (1971); they reflect more nearly the parameters proposed by Stevens (1988).

The two features beneath the laryngeal node are not in themselves sufficient for characterizing all the phonologically significant states of the glottis. Just as the articulatory features High (Low), Back and Round do not of themselves explain why vowel systems are as they are, so too the features Stiffness and Glottal Aperture do not provide a direct way of explaining why most sounds are either voiced or voiceless. There has to be a separate feature accounting for these two very natural classes of sounds. As we noted above, this feature could be given either an auditory or a physiological definition. At the moment it seems that both sets of properties distinguish the same classes of sounds, although further phonological evidence may later be forthcoming to show that one or other of these definitions provides slightly better groupings. Irrespective of whether it is considered to be an auditory or a physiological feature, there is no doubt that the feature Voice is a very necessary determiner of phonological classes.

All sounds should also be considered as having some particular airstream mechanism. It might seem as if there is no need to specify the presence of the pulmonic airstream mechanism, as it is present in all sounds; even clicks and ejectives still have a positive subglottal pressure. It is, however, necessary to note that some sounds have an increase in lung power associated with them. For example, Dart (1987) has shown that Korean so-called fortis stops have a significant increase in pulmonic pressure. Both the non-pulmonic airstream mechanisms occur in conjunction with the pulmonic mechanism (and sometimes, as in !Xóõ, in conjunction with each other as well). The glottalic airsteam mechanism has three mutually exclusive possibilities: ejective [tʼ] as in Amharic, simultaneous glottal and alveolar stop [ʔt] as in my final allophones of /t/, and glottalic ingressive [tʼ<] as in Owerri Igbo.

There are many constraints on feature combinations that are not made explicit by the paths through figure 2. Some of these are absolute constraints. For example pharyngeal nasals (to use a shorthand label for [+nasal, stop interrupted, low front dorsal] are an impossibility, as are labial and radical laterals. Some other combinations of feature values are best regarded as phonological impossibilities. For example ejective nasals (to use a shorthand label for [+stop, +nasal, ejective glottalic]) can be made, but they certainly do not appear. Yet other combinations indicate another form of overspecification in figure 2. There are combinations of values of features that can be used as ways of distinguishing the sounds of one language from those of another, which have not been observed to be used contrastively within a single language. For example there is no known contrast between a voiceless alveolar lateral fricative [ɬ] and a voiceless alveolar lateral approximant [l]; but Maddieson and Emmorey (1984) have shown that some languages consistently use one of these possibilities and others the other. Distinctions such as these should be given some special status (or perhaps omitted altogether) in a theory providing an account of all possible phonologically contrastive segments. There are also the problems concerned with defining possible paths through the tree that we noted in connection with places of articulation. We can now see that there are similar problems with the airstream node, through which one may take one or more possible paths. Bearing all these points in mind, we must obviously regard figure 2 as only a limited part of a theory specifying phonological segments. It is however a first step.

The Symbols of the International Phonetic Alphabet

To conclude, we must return to the question of what should be symbolized within the International Phonetic Alphabet. The basic answer is that we should regard the traditional terms as part of shorthand labels for feature combinations. With this in mind the symbols may be taken as depicting intersections of terms which are themselves defined in terms of features. I would like to see symbols arranged in terms of several distinct charts. For example one chart might show the cardinal vowels (and perhaps some additional symbols) in terms of the two dimensions of Height and Brightness. Another chart would show how these same (and perhaps some additional) symbols relate to the features High (Low) and Front (Back). Much of this display would be fairly similar to our present charts. The major difference would be that the symbols would be explicitly defined as being equivalent to combinations of features.

Acknowledgements

My thanks are due to my colleagues in the UCLA phonetics group, notably Bruce Hayes, Susan Hess, Marie Huffman, Michel Jackson, Pat Keating, Mona Lindau, and Ian Maddieson.

References

Abercrombie, David. 1967. Elements of general phonetics. Chicago: Aldine.
Bell, A. Melville. 1867. Visible speech. London: Simpkin, Marshal.

Bladon, R. Anthony, and C. Gunnar Fant. 1978. A two-formant model and the cardinal vowels. Quarterly Progress and Status Report, Royal Institute of Technology, Stockholm 1978/1: 1-8.

Catford, John C., 1977. Fundamental problems in phonetics. Bloomington: Indiana University Press.

Chomsky, Noam and Morris Halle. 1968. The sound pattern of English. New York: Harper and Row.

Clements, G. Nickerson. 1985. The geometry of phonological features. Phonology Yearbook 2:225-52.

Dart, Sarah. 1987. An aerodynamic study of Korean stop consonants: Measurements and modeling. Journal of the Acoustical Society of America 81.1:138-147.

Firth, J.R. Papers in Linguistics: 1934-1951. London: Oxford University Press.

Fischer-Jørgensen, Eli. 1985. Some basic vowel features, their articulatory correlates, and their explanatory power in phonology. Phonetic lnguistics: Essays in honor of Peter Ladefoged. ed. by Victoria Fromkin. Orlando: Academic Press.

Halle, Morris. 1988. The immanent form of phonemes. Giving birth to cognitive science: a festschrift for George A. Miller. ed. by William Hirsp. Cambidge, U.K.: Cambridge University Press.

Halle, Morris and Peter Ladefoged. 1988. Some major features of the International Phonetic Alphabet. UCLA Working Papers in Phonetics. Forthcoming.

Halle, Morris and Kenneth N. Stevens. 1971. A note on laryngeal features. Quarterly Progress Report of the Research Laboratory of Electronics, M.I.T. 101:198-213.

Holder, William. 1669. The elements of speech. London.

Ingram, John, C. 1975. Perceptual dimensions of phonemic recognition. Ph.D. dissertation, University of Alberta, Edmonton, Alberta.

Jakobson, Roman, C. Gunnar Fant, and Morris Halle. 1951. Preliminaries to speech analysis. Cambridge: M.I.T. Press.

Jakobson, Roman, and Morris Halle. 1956. Fundamentals of language. The Hague: Mouton.

Keating, Patricia. 1988. Palatals as complex segments: x-ray evidence. UCLA Working Papers in Phonetics. 69: 77-91.

Ladefoged, Peter. 1982. A Course in Phonetics. New York: Harcourt Brace and Jovanovich.

Ladefoged, Peter. 1971. Preliminaries to linguistic phonetics. Chicago: Chicago University Press.

Ladefoged, Peter, Ruth Glick and Clive Criper. 1972. Language in Uganda. London: Oxford University Press.

Lindau, Mona. 1985. The story of /r/. Phonetic Linguistics: Essays in honor of Peter Ladefoged. ed. by Victoria Fromkin. Orlando: Academic Press.

Maddieson, Ian. 1984. Patterns of sounds. Cambridge: Cambridge University Press.

Maddieson, Ian. 1987. Revision of the IPA: Linguo-labials as a test case. Journal of the International Phonetic Association. 17.1: 26-30.

Maddieson, Ian and Karen Emmorey. 1984. Is there a valid distinction between voiceless lateral approximants and fricatives? Phonetics 41: 181-190.

Miller, George and Patricia Nicely. 1955. An analysis of perceptual confusions among English consonants. Journal of the Acoustical Society of America. 27: 338-352.

Pike, Kenneth. 1943. Phonetics: A critical analysis of phonetic theory and a technic for the practical description of sounds. Ann Arbor: University of Michigan Press.

Sagey, Elizabeth. 1986. The representation of features and relations in non-linear phonology. Ph.D. dissertation, M.I.T., Cambridge, Mass.

Shadle, Christine. 1985. The acoustics of fricative consonants. Ph.D. dissertation, M.I.T.

Shepard, R.N. 1972. Psychological Representation of speech sounds. Human communication: A unified View. Ed. by Edward E. David and Peter B. Denes. New York: McGraw-Hill.

Stevens, Kenneth N. 1988. Modes of vocal-fold vibration based on a two-section model. Vocal physiology: Voice Production, Mechanisms and Functions, ed. by Osamu Fujimura. New York: Raven Press.

Sweet, Henry. 1877. A handbook of phonetics, including a popular exposition of the principles of spelling reform. Oxford: Clarendon Press.

Trubetzkoy, N.S. 1929. Zur allgemeinen theorie der phonologischen vokalsysteme. Travaux du Cercle Linguistique de Prague 1:39-67.

Trubetzkoy, N.S. 1939. Grundzüge der Phonologie. Travaux du Cercle Linguistique de Prague 7.

The many interfaces between phonetics and phonology

Paper presented at the Sixth International Phonologyy Conference, Krems, Austria, July 1988

Peter Ladefoged

The nature of the interface between phonology and phonetics depends on the kind of phonological description that is being made. A phonology might be: (1) A description of a single language requiring only features that are distinctive. (2) A description that permitted comparison of rules and patterns of sounds observed in different languages, the description being in terms of a set of universal phonetic features. (3) A description of not only rules and patterns, but also of the particular sounds of each language. This kind of description will make use of additional phonetic features that are never contrastive within a single language. (4) Part of an account of how the underlying forms of words and sentences become realized as movements of the vocal organs. (5) Part of an account of how the auditory sensations (or the physical descriptions) of sounds become associated with mental representations of the underlying forms of words and sentences. This paper will develop the notion that the interface between phonetics and phonologies of types (2) and (3) is very different from that between phonetics and phonologies of types (4) and (5).

Even a glance at the abstracts of this meeting shows that there are several different views of the interface between phonetics and phonology. This is as it should be as there are several different kinds of phonological description, each being valid for a different purpose. A phonologist might wish to describe the patterns of sounds that occur in a single language, or to compare the sound patterns that occur in different languages, or to describe not only the patterns but also the individual sounds on a dialect by dialect basis.

We will consider first the case in which the sound pattern of a single language is described. In order to do this appropriately sounds have to be grouped together into classes that occur in phonological rules. Generally these classes are defined in terms of phonological features that have a physiological basis. Thus this interface between phonology and phonetics is primarily defined by the physical definitions of the features. But this is not always the case; sometimes certain sounds are grouped together in ways that cannot be justified by reference to a single, or even a group, of physical properties. It might just be the result of historical circumstances that some sounds, which no longer share any particular phonetic defining characteristics, nevertheless still pattern together. An example of this kind of phonological description occurs whenever the vowels of English are described in terms of the feature Tense. This is a notoriously difficult feature to define; but it nevertheless specifies a very real mental grouping that has to be considered part of the sound pattern of English.

Another kind of phonology might prohibit the use of a feature that had no physical basis on the grounds that a phonology of this kind is too unconstrained, and too difficult to learn. Given the possibility of ad hoc features of the sort described above it would be possible to devise abstract phonologies that described all sorts of pretty patterns that might have no physical correlate of any kind. It is arguable that such phonologies might be hard or even impossible to learn. If there are no physical cues for groupings of sounds, how could children, who have no access to the history of the language, internalize patterns that have only a historical basis? The answer is that from the age of 6 or 7 on some English-speaking children at least do have access to the history of the language. It is represented in the orthography. There is considerable evidence (Moskowitz 1973, Ohala 1974, Jaeger 1986) that people can use orthographic knowledge as the basis for forming phonological classes.

What, then, can we say about the interface between the phonetics and the phonology of a single language, in which some features may have no physical basis? We can substantiate the psychological reality of these features by showing that they are needed in productive rules, or by phonological experiments as described by contributors to a recent voulme (Ohala and Jaeger, 1986). But does their mental reality provide us with a problem in discussions of the interface between phonetics and phonology? I would think not, precisely because they do not have any role in that interface. They are mental objects defined by the phonology itself, and are not part of phonetics. They are like the mathematical symbol *i*, the square root of -1, useful concepts that have no physical reality but which have great value in explaining the way things work.

A second kind of phonological description is one that allows one language to be compared with another by using a universal set of phonological features. A non-physical arbitrary grouping of segments that simply reflects historical events in a particular language or group of languages cannot be used in phonologies of this sort. If we are to be concerned with the general nature of human language, then we must consider only phonological descriptions made in terms of features that have a physical basis that could apply to all languages. Comparative phonologies that are restricted in this way are, of necessity, different from those of a single language in which *ad hoc* features might be permitted.

There are two different kinds of universals that comparative phonologies might help elucidate: those that are concerned with the segmental inventories of languages, and those that are concerned with the types of rules that can occur. Many contemporary phonologists (e.g. Clements 1985, Halle 1988) are more interested in the types of rules that can occur than in phonological inventories. But it seems to me that one of the tasks of phonological theory is to help explain why languages have the sounds that they do. The best account of the phonological properties that have to be described is that of Maddieson (1984).

As an example of the facts that demand explanation, consider the types of vowel systems that occur in the world's languages. Maddieson (1984) has shown that over 20% of all languages have vowels somewhat like [i,e,a,o,u]. I do not know how many different vowel qualities might have been used contrastively within languages. Even if we keep to fairly well distinguished vowel regions, comparable in size to those used by Maddieson in defining the five vowels [i,e,a,o,u], it is not difficult for me as a phonetician to make a reduced set of primary cardinal vowels, the corresponding secondary cardinal vowels with the opposite degree of lip rounding, and perhaps a further three rounded and another three unrounded central vowels. Many of these vowels have qualities that are sufficiently far apart to permit an additional distinct vowel between them. Furthermore each of these vowels is potentially modifiable with a variety of secondary characteristics such as nasalization, rhotacization or sulcalization. As a minimum estimate we can say that there are about 50 broadly distinct vowels that languages might have chosen as the vocalic elements of their segmental inventories. The likelihood of the same five being chosen so frequently is therefore comparable with the likelihood of playing poker and finding that one hand in five always had the Ace, King, Queen, Jack, and Ten of Spades. It is therefore an absolutely astounding fact that so many language have the vowels /i,e,a,o,u,/, and any theory of phonology that does not offer an explanation for this fact must be considered to be seriously lacking.

Descriptions of vowels in terms of the articulatory features High, Low, Back and Round throw little light on this problem. There is no explanation in terms of these features of why back rounded vowels are more likely than back unrounded vowels. The solution to this problem that I have suggested elsewhere (Ladefoged 1988) is to retain the articulatory features, High, Back, and Rounded, and also to add features for vowels that would specify their auditory characteristics. Comparative phonologies should also include the auditory features Height and Brightness. The acoustic correlate of Height is the frequency of the first formant (F1). The other auditory feature of vowels, Brightness ("Helligkeit"), was originally proposed by Trubetzkoy (1929, 1939), and has been discussed more recently by Fischer-Jørgensen (1985). The acoustic correlates of Brightness may be taken to be the difference in frequency between the first formant and F2', a form of the second formant modified so as to account for the influence of the third formant. Algorithms for determining F2' have been given by Bladon and Fant (1983). From a physiological point of view, Brightness is a combination of all three articulatory vowel features, Front, Back, and Round. High front unrounded vowels have the highest value of Brightness, low back neutral vowels have a mid value and high back rounded vowels have the lowest value. The reason for the predominance of the vowels [i,e,a,o,u] is that these vowels are maximally opposed to one another in the most important perceptual attributes of vowels, Height and Brightness.

Implicit in this proposal is the claim that an interface between phonetics and phonology must permit some features to use auditory properties of sounds and others to use physiological properties for grouping sounds into natural classes. I have discussed this point at greater length elsewhere (Ladefoged 1988), and will simply note here that there are many important natural classes of sounds that arise because sounds have certain auditory properties in common. This is one of the great insights of the Prague school, discussed by Trubetzkoy (1939) as well as by Jakobson, Fant and Halle (1951). It has been suggested (Chomsky and Halle 1968) that it is just a matter of exposition as to whether features are defined in articulatory or acoustic terms. But this is not always correct. Although many natural classes can be defined in terms of either the auditory or the physiological properties of the segments contained within them, other natural classes depend on specific properties relating to the way sounds are heard, or to the way they are produced.

This distinction is shown in the hierarchical arrangement of possible phonological features presented in Figure 1. [This is a revised version of the figure in the paper on "Hierarchical features of the International Phonetic Alphabet" in this issue.] In the upper part of the figure are some of the features that require auditory definitions. They include, in addition to the vowel features discussed above, Grave, which groups some Labial and Dorsal sounds in accordance with their spectral characteristics, Strident (or Sibilant, to use the more traditional term that goes back to Holder (1669), and Sonorant (which, if a non-acoustic definition is required, can be defined only as a cover term summarizing a number of articulatory possibilities).

Returning to the discussion of the vowel features, another point which is implicit in the proposal for the features Height and Brightness must be discussed here. Both these features are non-binary; each of them permits a range of values along a scale. Thus the feature Height has (at least) three possibilities: [high], [mid] and [low]. The process of going from [low] to [mid] is the same as the process of going from [mid] to [high]. Only with this notion of scalar features can we maintain the notion of a two-dimensional (as far as these features are concerned) vowel space. Without this notion there is little likelihood of a true explanation not only of why vowel systems are as they are, but also of many phonological processes such as the English vowel shift and the changes in vowel quality that occur between dialects. It is, of course, possible to describe all such changes in vowel quality in terms of binary features as illustrated by Chomsky and Halle (1968). But any rule using binary features that has to account for [+high] becoming [-high] in the same circumstances as [-low] becomes [+low] inevitably misses a linguistically significant generalization. Simply by virtue of having to use two separate features it cannot show the unity of the process that is expressed

Type	Hyper feature	Major node	Feature	Traditional term	Brief description
				High vowel	low F1
			Height	*Mid-high vowel*	- mid low F1
				Mid-low vowel	- mid high F1
				Low vowel	- high F1
			Brightness	*Front unrounded*	high (F2' - F1)
Auditory				*Back rounded*	- low (F2' - F1)
			Grave		aperiodic low frequency energy
					- absence of such energy
			Sibilant	*Sibilant*	aperiodic high frequency energy
					- absence of such energy
			Sonorant		periodic well-defined formants
					- no periodic formant structure
			Rounding	*Rounded*	decreased lip opening
		Labial		*Unrounded*	- spread lips
					protruded
			Protrusion	*Bilabial*	- neutral
				Labiodental	- retracted lower lip
				? *Linguolabial*	(tongue between lips)
				Dental	advanced
			Anterior	*Alveolar*	- neutral
		Coronal		*Postalveolar*	- retracted tongue tip/blade
				Laminal	laminal
	Gesture		Apicality	*Apical*	- apical
				Retroflex	- sublaminal
		Dorsal	Front	*Front* *Palatal*	front
				Back	- back tongue body
				Velar / high vowel	high
			High	*Uvular / mid V*	- mid
			(Low)	*Pharyngeal / low V*	- low tongue body
		Radical	Tongue Root	*Epiglottal*	retracted tongue root
					- advanced tongue root
				Stop	complete closure
			Interrupted	*Fricative*	- nearly complete closure
Physiological				*Approximant*	- approximation of articulators
			Lateral	*Lateral*	predominantly lateral airflow
	Stricture			*Central*	- no lateral airflow
					held gesture
			Dynamic	*Trill*	- vibrating
				Tap	- ballistic
				Flap	- ?
	Oro-nasal		Nasal	*Nasal*	velic opening
					- velic closure
			Voice	*Voiced*	vibrating vocal cords
				Voiceless	- non-vibrating vocal cords
				Creaky	stiff
			Stiffness	*Modal*	- neutral
	Laryngeal		(Slack)	*Breathy*	- slack vocal cords
				Aspirated	open
			Glottal aperture	*Unaspirated*	- narrowed
			(Closure)	*Glottal stop*	- closed vocal cords
		Pulmonic	Fortis	*Fortis*	segmental lung power
				Lenis	- no increased lung power
	airstream			*Implosive*	downward
		Glottalic	Glottal		- no glottal movement
			movement	*Ejective*	- upward glottal movement
		Velaric	Click	*Click*	oral suction
					- no oral suction

Figure [illegible]. A hierarchical arrangement of phonological features.

by a rule of the form: [n high] → [n+1 high]. Why should the counterpart of : [+high] → [-high] be [-low] → [+low] rather than say, [+low] → [-low], or [-back] → [+back]? Similarly, if High and Low were truly independent binary features it would be difficult to give an explanatory account of vowel reduction in which both [+high] and [+low] vowels become simultaneously [-high, -low]. Current feature theories achieve the correct result by defining High and Low as deviations in opposite directions from the same starting point, a mid vowel. They also have to have a marking convention that prohibits a vowel from being simultaneously [+high] and [+low]. All this is equivalent to saying, in a rather cumbersome notation, that [high], [mid] and [low] form an ordered set of values on a single scale.

Good examples of vowel raising can be found in a number of languages spoken in Southern Africa. The phonetic facts are hard to state in terms of the conventional SPE features, which permit the specification of only three vowel heights. In the Nguni languages there are five vowels, which, in traditional IPA terms, have the qualities [i,ɛ,a,ɔ,u]. Each of these vowels is fairly similar to the corresponding cardinal vowel, except [a], which is retracted so that it is in between the cardinal [a] and [ɑ]. Preliminary data indicate that the unmodified qualities of these vowels are as shown by the solid points in figure 2. The two mid vowels [ɛ,ɔ,] have raised variants that occur whenever the following vowel in the word is [i] or [u]; their phonetic qualities are as indicated by the arrow heads in the figure. (The phonological conditions can be further elaborated, but this is sufficient for our purpose here.) We therefore have to have a rule of the form:

[mid-low] → [mid-high] / _ (C) [high]

This rule could be formulated using the SPE features High, Low and Tense, but this would be a purely arbitrary use of Tense to have exactly the same phonetic exponents as the features High and Low; the vowels [e,o] differ from [ɛ,ɔ,] in exactly the same way, physiologically and acoustically, as [ɛ,ɔ] differ from [a].

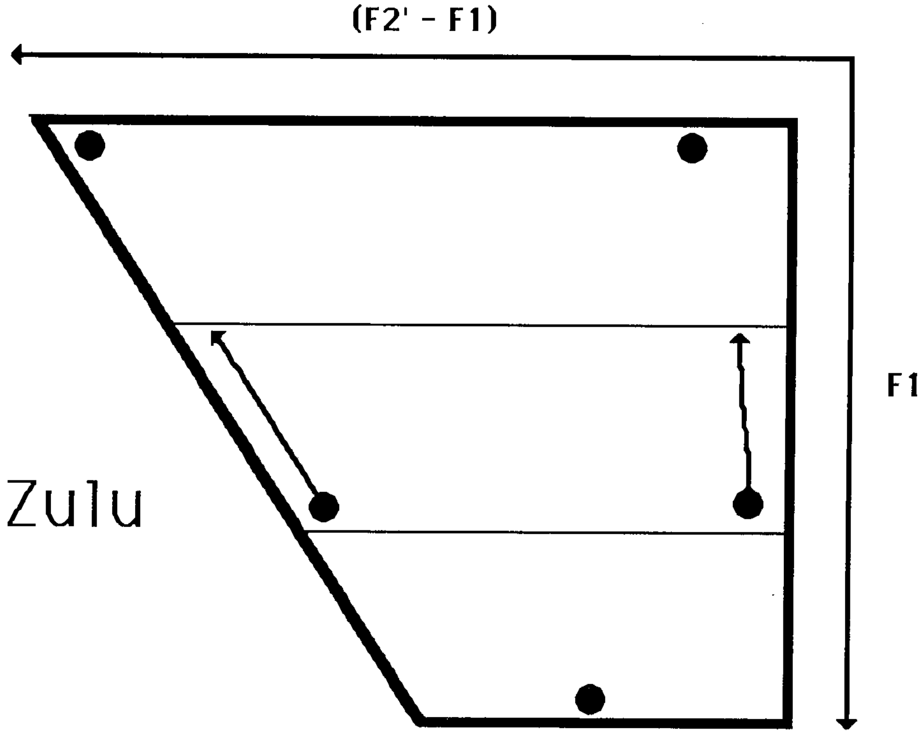


Figure 2. Vowel raising in Zulu.

The situation is complicated still further in languages of the Sotho group, which have seven underlying vowels [i,e,ɛ,a,ɔ,o,u]. In these languages all four of the mid vowels [e, ɛ,ɔ,o] have raised variants in similar circumstances to the raised variants in the Nguni languages. Preliminary data indicate that the phonetic facts are as summarized in figure 3. It seems obvious that we need a multi-valued feature vowel Height to make the correct generalizations.

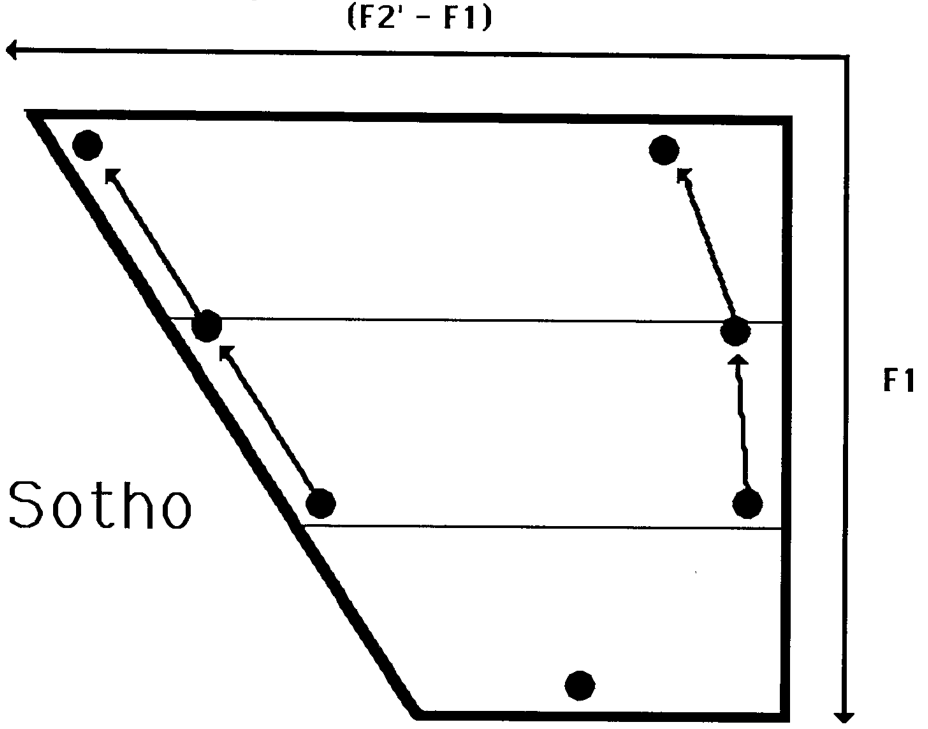


Figure 3. Vowel raising in Sotho.

The whole question of vowel specification is further complicated by the fact that what I have characterized as the difference between a binary and a multi-valued approach, others (e.g. Hulst 1988) regard as being binary versus single-valued. For me, multi-valued means that a single phonological feature has the possibility of one, and only one, value out of an ordered set of multiple values, ranging from a minimum amount of the property in question to a maximum. For those who use the notion of single-valued vowel features, each of these features is regarded not as a variable that can have different values, but as a component whose presence indicates a directional pull. It then becomes possible to denote the strength of the pull in a given direction by adding further components. The notation [i-a] (or some equivalent way of showing dominance) would indicate a vowel midway between [i] and [a]; but [ii-a] would indicate a vowel nearer to [i], and [i-aa] one nearer to [a]. This certainly makes evident the two dimensional nature of the basic vowel space; but it is not clear that it can capture the correct phonetic generalizations in the vowel raising cases discussed above.

We should also note that several of the physiological features shown in the lower part of Figure 1 have to be considered as having an ordered set of values in the sense that they can have only one of three (or more) mutually exclusive values, with the middle values being between the other values. A good example is Interrupted, the term I am suggesting for the feature that has the ordered set of values [stop], [fricative], and [approximant], and thus permits an explanatory account of changes from stop through fricative to approximant as exemplified in the pronunciation of the Danish word “Ladefoged” (barn steward). The original voiced stops in this word have weakened, first to fricatives [ð] and [ɣ], and then to approximants (frictionless continuants) for which there are no distinct IPA symbols, eventually becoming (at least in the case of the velar consonant) omitted altogether. Progressive changes from stop through fricative to approximant are not easy to explain in terms of binary features.

Further examples of multi-valued features in which there are at least three mutually exclusive values, arranged in an ordered series, include: Apicality, which specifies whether the blade, or the tip, or the underside of the tongue is involved; Anterior, with at least three values, dental, alveolar, and post-alveolar, all of which use the coronal part of the tongue; High, the tongue body (Dorsal) feature distinguishing velar, uvular and pharyngeal possibilities; and Stiffness (of the vocal cords), which orders the mutually exclusive phonation types, creaky, modal and breathy.

It seems to me that the evidence that some phonological features are sometimes multivalued is simply overwhelming. Many are binary; but in the case of some features we need scalar values both in phonological rules and in accounting for universal patterns of sounds among languages. There is no basis for the argument that all features must be binary because phonological rules have been shown to operate in terms of such features. Phonological rules have never been entirely binary. The fact that much of the patterns of sounds within languages can be described in binary terms does not militate against the fact that it is equally possible to use another formalism (as in SPE, in which about one third of the rules — those involving Stress — use a multi-valued notation) that permits scalar quantities.

I hope that phonologists will soon recognize that we are due for our own little Copernican revolution. It is possible to describe many of the observations of astronomy, and to predict future eclipses while still maintaining that the earth does not move and the sun goes round it. But as Galileo (1633) whispered after being forced to retract his Copernican heresies “Epur si muove [still it does move].” Like Galileo, I will not go to the stake for my belief. Still five vowel systems are most favored. Mid vowels are between high and low vowels. An interface between phonetics and phonology must allow some phonological features to have non-binary values.

I will not discuss other aspects of the features in Figure 1, as they have been described in more detail elsewhere (Ladefoged 1988). There are, however, a few general points that are particularly relevant to the theme of this paper. Firstly they are all phonological features, set up solely for the purpose of grouping sounds into natural classes by reference to phonetic phenomena. They are not meant to be parameters that could be used for specifying all the phonetic characteristics of each sound. Secondly, they are arranged in a hierarchical structure. This structure determines that all sounds must be specified in terms of each of the hyper-features, then one or more of the major nodes must be selected, and finally some of the features have to be given values, the subset of features involved being dependent on the previous choices. Even so, the terminal features are not orthogonal; the specification of the feature Voice, for example, depends on the values of the other laryngeal features. There is no reason why a phonological feature specification should not be redundant. The necessity for both physiological and auditory features makes this very apparent. As the sole object of having phonological features is to provide for the specification of natural classes, redundancy should not be a concern.

The nature of the interface between phonetics and phonology will also depend on the degree of phonetic detail which might be said to be part of the phonological description of a language. So far we have discussed relatively abstract phonologies of a single language, and more concrete phonologies that enable us to compare rules in different languages, and to account for tendencies towards patterns of sounds among languages. A third type of phonological description of a language is one that “should be able to characterize both the oppositions within a langauge ... and the contrasts between languages (all and only the features that mark the sounds of the language as being different from the sounds of other languages)” (Ladefoged

1972:275). This necessitates a different interface between phonetics and phonology because it is by no means apparent that we can describe the ways in which the sounds of one language are distinguished from another simply in terms of the features that are required for distinguishing lexical items, or for accounting for phonological universals, or for grouping sounds into the natural classes that occur in rules. There are many instances of small but reliable phonetic differences between languages that have not been found to be used for contrasting words within a single language. For example in a recent survey, I found that 27 out of 30 speakers of Californian English used an inter-dental [θ] in which the tip of the tongue was protruded between the teeth in words such as "think, thin," whereas 27 out of 30 speakers of different forms of Southern British English used a dental [θ] without tongue protrusion in these words. This is a highly significant difference from a statistical point of view. Speakers of American English must learn the one form, and speakers of British English must learn to use the other. But the difference between a dental and inter-dental [θ] is unlikely to be phonemically contrastive in any language, or referred to in any phonological rule, although it is a reliable marker of a difference in regional accent. There are many other examples of differences between languages that are not easily describable in terms of phonological features. If it is part of the phonology of a language to be able to specify differences between that language and other languages, then additional "features" will be required. I put the word "features" in quotes, because it is by no means clear to me that this is a proper goal for phonology, and that is is phonological features that are required. Such differences between languages are part of culture, and not of the languages themselves.

From my point of view we are also moving outside of phonology when we consider what people do when talking. Many of the phonologies that fall into one of the three types that we have considered so far are ostensibly accounts of articulatory actions. But, from Chomsky and Halle (1968) onwards, phonologists have generally maintained that they are really describing mental concepts and they are not concerned with a speaker's performance. This seems to me to be wholely proper. The only way in which one might even consider the possibility of a phonology that was concerned with articulatory actions is by means of a theory such as that propounded by Browman and Goldstein (1986); and their theory of phonology makes no pretense of being able to describe phonological processes such as velar softening or the vowel shift in English.

For at least the last 40 years it has been apparent that skilled, over-learned, actions are "triggered off as a whole" (Craik 1947). Producing vocal gestures is like producing any skilled action. For example, consider the actions of hitting a golf ball. The elements of a golf swing may be broken down into parts, and a new part substituted for a faulty old one. The new whole action may be practiced and practiced, at first moving slowly and eventually more rapidly. Finally the new action may become a truly integrated whole and used at will. Similarly in speaking it is possible to learn to substitute a uvular contact for the phonological unit /k/. If you train people to do this by getting them to imitate your pronunciation of "cat" as [qæt], "keep" as [qip], and so on, after a practice they will be able to generalize their actions beyond the particular learned words, and be able to talk about "Captain Cook" [ˈqæptən ˈqɷq] But it takes a very considerable learning period before they can talk really fluently in this way and deal with a sentence like "He quickly packed his bag while singing a song." As all of us who have taught a foreign language know, getting people to pronounce a sound in isolation is fairly easy, but getting them to integrate it into their regular conversation takes a lot of hard work.

If words really were just made up of isolated sounds, this should not be so. The difficulty comes because words have to be composed of large chunks before they can be pronounced. The situation is not like substituting a different letter for the letter attached to a particular key on a typewriter. When this has been done, every time you press that key you get the new letter. With sounds, the speaker does not have access to what corresponds to the individual keys on a typewriter.

One common way of dealing with this problem is to speak of "the speech code" (Liberman 1957), noting that speech sounds are encoded when produced. But why should we imagine that, in the normal course of speaking, this has to be done? The evidence we have just been considering provides us with no reason to believe that, for the purposes of speech production, words are stored as strings of sounds, each sound itself being stored as if it were a separate item attached to the mental equivalent of a typewriter key. As Aitchison (1987) has pointed out, words are stored in the mind in many different ways. For the purpose of talking they probably consist of larger than segment sized pieces, which can be taken out of storage as preprogrammed chunks that can be triggered off as a whole.

Even stronger objections can be made to any description of speech production that uses phonological features. Consider a comparatively simple task: reading a list of words aloud. Let us assume that part of this task involves finding the word in the mental lexicon, and obtaining its phonological representation. Further assume that this representation is (in part) in terms of features and observe what complications this leads us into. There are many features that cannot be interpreted as instructions to the articulators without knowing what the values of the other features are. For example what does [+consonantal] by itself tell us about where to move the articulators? And without knowing what other feature values are involved, [+continuant] is uninterpretable. Even the Place features do not give us enough information concerning the vocal tract shape. The tongue will have to be moved in one way if the dorsal feature [+high] is accompanied by [-continuant] (in which case there will be a stop closure with the back of the tongue against the soft palate), and in another way if it is accompanied by [+continuant] (in which case there will be several possibilities depending on whether other feature values specify a fricative, semivowel or vowel).

It would seem rather pointless to specify the instructions for articulating a word in terms of features, if the feature values had to be combined before anything could be done about forming the articulation. Why not store the articulatory instructions in terms of segments to begin with? And, as we have just noted, why not group the segments together in pre-stored packages for the efficient production of words?

I do not know the size of the chunks that people use for storing the articulatory components of words. The evidence from slips of the tongue, which nearly always involve the interchange of similar parts of the syllable, indicates that similar syllables may be stored together so that they can be confused in speech production; but slips of the tongue offer little support for anything that considers words to be stored simply as an ordered set of instructions for the equivalent of keys on a typewriter. For that matter, typing itself often does not consist of giving individual instructions for the operation of each key. MacNeilage (1985) has that shown, typing errors have certain similarities to speech errors; and these errors reflect the fact that typing, like speech, is a skilled motor action in which the act of typing a word often consists of giving a single command that initiates a whole set of learned motor movements of the fingers. As far as speech is concerned, my best bet at the moment is that the articulatory instructions for words are stored in syllable sized units. Smaller components of words, such as segments and features, are available for use by speakers, and can surely be part of a speaker's competence as expressed in phonological rules, but they are probably irrelevant to the act of speaking.

Features and segments are probably also irrelevant to the act of listening and the normal comprehension of speech. They may be involved in special tasks, such as reaction time experiments in which listeners are asked to respond as quickly as possible on hearing a given sound. But it is quite easy to show that listeners have extreme dificulty in locating the precise order of individual speech sounds without first taking in a larger chunk of speech, and interpreting it so that they can consider it as a whole word (or at least as a whole syllable). If they are asked questions about segments, they can answer them by mentally referring to something like the way they would spell the word. If listeners are put in a situation where this kind of first interpretation is not possible, they will often be unable to report the order of segments, as first demonstrated many years ago by Ladefoged and Broadbent (1960). These experimenters required subjects to listen to sentences which had an extra speech sound superimposed on them. Despite the fact that this extra sound was a perfectly normal speech sound (it was, in fact, an [s] cut out of another sentence in the same recording) listeners were often unable to report accurately on which word or between which word this superimposed sound occurred.

These and other experiments are the basis for my claim (Ladefoged 1981) that the act of listening does not involve the units that we have found so useful for phonological descriptions of languages. The usual answer to this mismatch between units is to say that both features and individual speech sounds are perceived as encoded in the stream of speech. But this notion is of no more help to us in discussions of speech perception than it is in discussions of speech production. It is about as true as saying that Arabic numerals consist of encoded binary numbers. If I ask you what is five times six you could go through a process equivalent to:

$$\begin{array}{lcrcr}
5 & = & 101 & & \\
 & + & 101 & & \\
 & = & 1010 & & \\
 & + & 101 & & \\
 & = & 1111 & & \\
 & + & 101 & & \\
 & = & 10100 & & \\
 & + & 101 & & \\
 & = & 11001 & & \\
 & + & 101 & & \\
 & = & 11110 & = & 30
\end{array}$$

In fact, if you were a computer, this would be a good representation of what happens when the input is 5 * 6 and the output is 30. But it is a very poor representation of how people do mental arithmetic. If we had been taught to do arithmetic in binary terms, then our brains might well have been doing the equivalent of these operations. But we were not, any more than we learnt to say the word “cat” by considering it as three phonological units. We talk before we can read. We do not need phonological units when describing speech perception and speech production. There is no direct interface between phonology and descriptions of this sort.

References

Aitchison, Jean. 1987. Words in the mind. London: Blackwell.

Bladon, R. Anthony, and C. Gunnar Fant. 1978. A two-formant model and the cardinal vowels. Quarterly Progress and Status Report, Speech Transmission Laboratory, Royal Institute of Technology, Stockholm 1978/1:1-8.

Browman, Cathe and Louis M. Goldstein. 1986. Towards an articulatory phonology. Phonology Yearbook 3:219-252.

Craik, K. 1947. Theory of the human operator in control systems. British Journal of Psychology, 38:56-61.

Chomsky, Noam and Morris Halle. 1968. The sound pattern of English. New York: Harper and Row.

Clements, G. Nickerson. 1985. The geometry of phonological features. Phonology Yearbook 2:225-52.

Fischer-Jørgensen, Eli. 1985. Some basic vowel features, their articulatory correlates, and their explanatory power in phonology. Phonetic Linguistics, ed. by Victoria Fromkin. Orlando: Academic Press.

Halle, Morris. 1988. The immanent form of phonemes. Giving birth to cognitive science: a festschrift for George A. Miller, ed. by William Hirsp. Cambridge: Cambridge University Press.

Holder, William. 1669. The elements of speech. London.

Jakobson, Roman, C. Gunnar Fant, and Morris Halle. 1951. Preliminaries to speech analysis. Cambridge, Mass.: M.I.T. Press.

Jaeger, Jeri J. 1986. Concept formulation as a tool for linguistic research. Experimental phonology, ed. by John J. Ohala and Jeri Jaeger, 211-235. Orlando: Academic Press.

Hulst, Harry van der. 1988. The phonetic and phonological basis of the simplex feature hypothesis. Sixth International Phonology meeting, Krems, Austria.

Ladefoged, Peter, 1972. Phonetic prerequisites for a distinctive feature theory. Papers in Linguistics and phonetics: To the memory of Pierre Delattre, ed. by Albert Valdman. The Hague: Mouton.

Ladefoged, Peter. 1988. The International Phonetic Alphabet as a hierachical feature system. Berkeley Linguistics Society (forthcoming).

Ladefoged, Peter and Donald Broadbent. 1960. Perception of sequence in auditory events. Quarterly Journal of Experimental Psychology. 12:162-170.

Liberman, Alvin M. 1957. Some results of research on speech perception. Journal of the Acoustical Society of America, 29:117-123.

MacNeilage, Peter F. 1985 Serial-ordering errors in speech and typing. Phonetic Linguistics, ed. by Victoria A. Fromkin. Orlando: Academic Press.

Maddieson, Ian. 1984. Patterns of sounds. Cambridge: Cambridge University Press.

Moskowitz , Breyne A. 1973. On the status of vowel shift in English. In T. E. Moore, ed., Cognitive development and the acquisition of language, 223-260. New York: Academic Press.

Ohala, John J. 1974. Experimental historical phonology. Historical linguistics II. Theory and description in phonology, ed. by J.M. Anderson and C. Jones., 353-389. Amsterdam: North Holland Publishing Co.

Ohala, John J. and Jeri Jaeger. 1986. Experimental phonology. Orlando: Academic Press.

Trubetzkoy, N.S. 1929. Zur Allgemeinen Theorie der phonologischen Vokalsysteme. Travaux du Cercle Linguistique de Prague 1:39-67.

Trubetzkoy, N.S. 1939. Grundzüge der Phonologie. Travaux du Cercle Linguistique de Prague 7.

Some major features of the International Phonetic Alphabet

Peter Ladefoged
University of California Los Angeles
and
Morris Halle
Massachusetts Institute of Technology

Abstract. Underlying the International Phonetic Alphabet there is a theory determining the categories (features) that are symbolized. We outline a revision of both the features and their organization in a hierarchy. This revision reflects the advances of the last century of phonetic research. Although there is disagreement about the lower levels of this hierarchy, the arrangement of the higher nodes is now reasonably clear.

The International Phonetic Alphabet is nearly 100 years old, and there is no doubt that it would be appropriate to revise it so as to bring it more in line with contemporary linguistic thought.[1] But unfortunately this is not a simple task as there are two conflicting requirements that need to be considered. On the one hand it seems advisable to retain the practical nature of the IPA symbols. Much of the utility of these symbols rests on the fact that they are based on the familiar Roman alphabet. Any radical revision that undermined this would make the symbols less useful as mnemonic devices. On the other hand, a fairly large scale revision is called for if it is going to take into account the advances that have been made in the study of phonetics during the last hundred or so years.

It seems to us that the most fundamental insight gained during the last century has been the realization that it is the features rather than the sounds that are the basic building blocks of spoken language. The idea that sounds are composed of features has been understood -- more or less -- at least since the seventeenth century (Fromkin and Ladefoged 1981), and is certainly implicit in the organic alphabets devised by Bell (1867) and by Sweet (1881), who noted that the set of iconic symbols represent 'the elementary actions by which all sounds are formed.' However, this insight was not made explicit until the work of the Prague school in the present century; and it is to this day not fully considered in all phonetic work. The failure to take this basic truth into account is in our view one of the major inadequacies of the IPA alphabet.

We are thus faced with a dilemma: an alphabet, particularly one based on the familiar symbols of the Roman alphabet, almost inevitably emphasizes the segmental nature of speech; but we want a set of symbols that reflect the features that are necessary for understanding linguistic processes. This paper will attempt to resolve this dilemma by emphasizing the symbolic nature of the elements of the alphabet. A symbol is something that stands for something -- in this case an intersection of feature values. The aspect of the revision of the IPA alphabet that we are most concerned with here is not the symbols themselves, but the organization of the set of categories -- the features -- which the symbols represent.

The IPA alphabet is defined in terms of a series of charts with their column and row names, together with accompanying notes and conventions; the whole system forms a theory of phonetic description. This theory is organized in such a way that each symbol on a chart reflects what may be regarded as a hierarchical structure of features. The first division is into consonants and vowels, with consonants normally being in the upper part of the chart, and vowels being in the lower part or on a separate chart altogether. Next, for consonants, there is a division into place and manner of articulation in accordance with the columns and rows; and, at the same time, there is a further sub-division whereby the order of the items within each cell indicates the state of the glottis; by convention, if there are two items, the one on the left is voiceless, and the one on the right is voiced. For vowels there is a different set of features, which may be referred to as [height], [frontness] and [rounding], each of which is then further sub-categorized. The main part of the IPA theory as contained in its charts may therefore be represented as a hierarchical structure as shown in figure 1.

This tree structure, together with the conventions for its interpretation, constitute a theory of phonetics specifying the nature of the sounds of speech. Before we offer our comments on a number of weaknesses of this theory, we must make explicit the conventions necessary for interpreting the hierarchical arrangement shown in Figure 1, which specifies sounds neither in the way that features were used for this purpose by Jakobson, Fant and Halle (1951), nor in the way that is common among contemporary phonologies that we will mention later. In the older distinctive feature

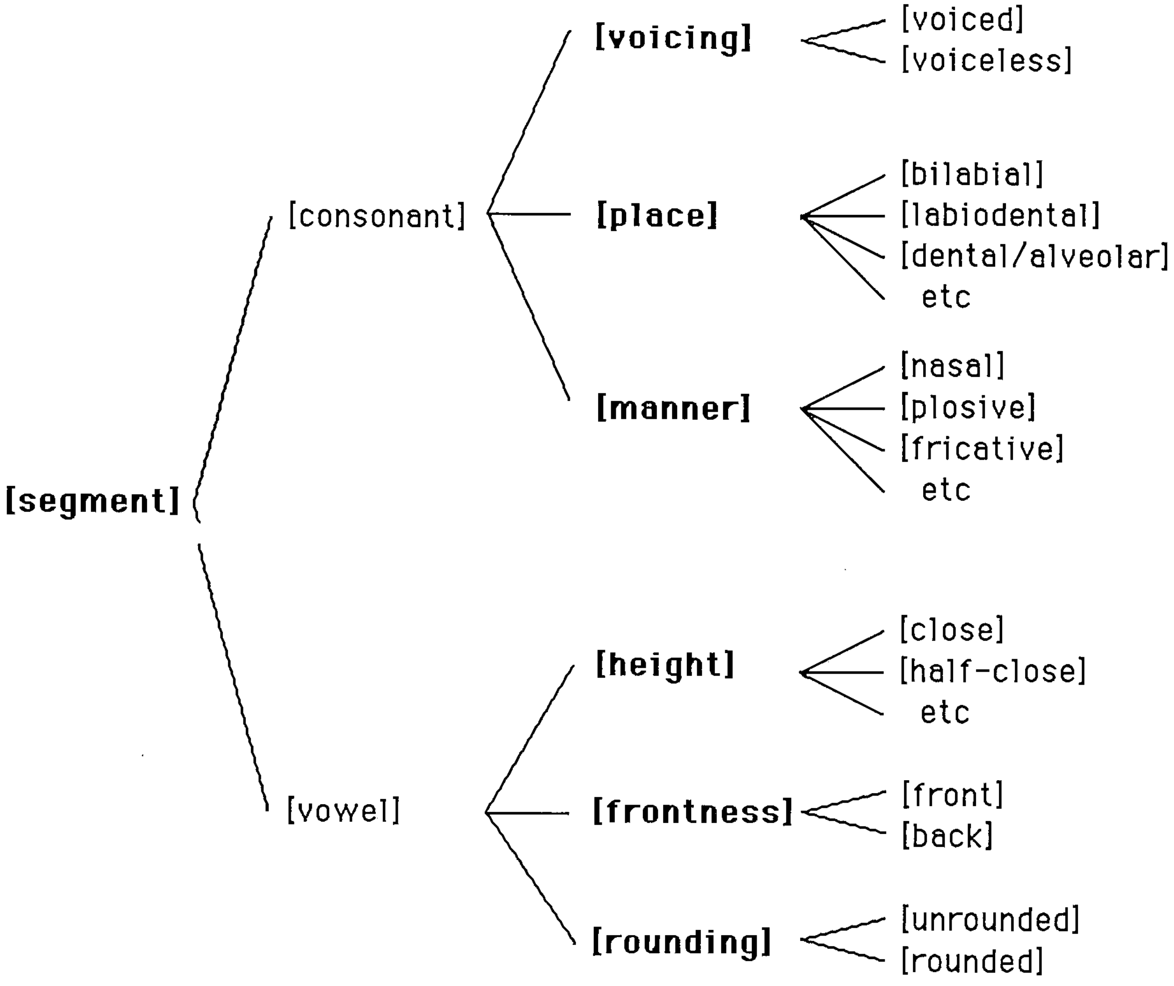


Figure 1. The IPA categories arranged as a feature hierarchy

work, each segment in a given language could be specified by taking one of the possible paths through a tree, each branch in which specified one or other of the two possible values of a feature. The tree shown in Figure 1 has to be interpreted in a different way. It is not a tree showing the set of segments that occurs in a single language; it is a structure delimiting human linguistic phonetic capabilities. In order to use the structure in this way one has to define the set of possible paths through the tree. There has to be a convention whereby one chooses among features in plain type, whereas one takes all the paths through the features represented in bold face. Thus the process of going from the first to the second node is one of choice; each sound is either a vowel or a consonant. But the next part of the specification of sounds involves going down all possible paths; consonants always have some specifiable degree of voicing, place of articulation, and manner of articulation, and vowels always have height, frontness and rounding. The next level again requires choice among a set of possibilities; consonants are either voiced or voiceless, etc, and either bilabial, or labiodental, or dental/alveolar, etc; and vowels have one and only one degree of height, etc. The IPA theory has these tacit conventions whereby in some circumstances one has to choose among alternatives (consonant or vowel, and when coming to the terminal nodes of the tree), and in other circumstances one has to go along all possible paths (intermediate nodes).

There are a number of major problems with the IPA theory. In the first place it begins by dividing segments into vowels and consonants, making it appear as if separate features were needed for these two classes of sounds. From our point of view it is obvious that vowels and consonants have a great effect on one another precisely because they are produced within the same mouth. We need a phonetic theory that allows us to explain why velar consonants are more likely to have an advanced articulation before front vowels than before back vowels, as they might well if there were no connection between the articulation of vowels and that of consonants. The same theory should make it evident that high vowels are more likely to be lowered after uvulars (as occurs in Serer, French, and many other languages) than for the reverse to occur so that mid vowels become raised after uvulars (which, to the best of our knowledge, never happens).

A second inadequacy of the IPA hierarchy is that is makes it appear as if voicing were irrelevant for vowels. It is true that voicing never distinguishes one vowel from another within the lexicon of a language. Nevertheless, voicing is a very necessary feature of the behavior of vowels in phonological rules, accounting for such well known facts as the intervocalic voicing of consonants.

Thirdly, in its charts the IPA restricts the application of the feature [nasal] to consonants (and, implicitly, to stop consonants). It does, of course, have a diacritic listed below the charts, enabling any segment to be marked as nasal. Again it seems preferable to us to reorganize the system so as to make it clear that nasalization is an additive component with a status more equivalent to that of voicing.

Finally, the IPA simply lists an ordered set of places of articulation without any internal organization. There are two considerations in which such an organization is evident within the articulations that occur in languages. Firstly, some adjacent places of articulation are more closely related than others. Thus, among the first three items in the list, bilabial, labiodental and dental/aveolar, there are only a few languages that distinguish between bilabial and labiodental articulations; but there are several that distinguish between labiodental and dental articulations. Secondly, in the description of sounds that have more than one place of articulation, some places can co-occur whereas others can not. For example, a bilabial articulation co-occurs with a velar articulation in the pronunciation of the Idoma word [àgbà] 'jaw'. But it does not make sense to speak of a bilabial articulation co-occuring with a labiodental articulation (although, of course, bilabial stops may be followed by labiodental fricatives in affricates).

Many of the above criticisms of the IPA theory arise because this theory does not explicitly take into account the fact that all speech sounds are produced by actions of a limited number of anatomical structures. These are the lips, the tongue tip and blade, the tongue body, the tongue root, the soft palate and the larynx. Following along lines suggested elsewhere (Clements 1985, Sagey 1986, Halle 1988), our first suggestion for the revision of the IPA theory is that this major functional organiztion should be made evident. One possible way of doing this is through a tree structure of the kind shown in part in figure 2. We have left nearly all of the lower part of this tree unspecified, as we do not entirely agree on how this should be done; as an exemplication of the notion of terminal features of some kind we have simply indicated the possibility of sounds being [±nasal]. We are, however, sufficiently in accord concerning the nature of the higher nodes to make it useful to discuss them.

The first node in the hierarchy is labeled 'root' (rather than 'segment' as in Figure 1) in accordance with the terminology introduced by Clements (1985). Beneath the root node there are four terms. The first is the place node dominating nodes for the four major active articulators. The terminal features which are dominated by these place hyper-features are articulator-bound in the sense that each of them can be implemented only by a given articulator. We have not named these terminal features, largely because we are not in complete agreement on how this should be done. (Some indications of how we would proceed individually are given in Halle 1988, and Ladefoged 1988.) We would both, however, specify vowels by means of the features of the dorsal and, to a lesser extent, the radical and labial articulators.

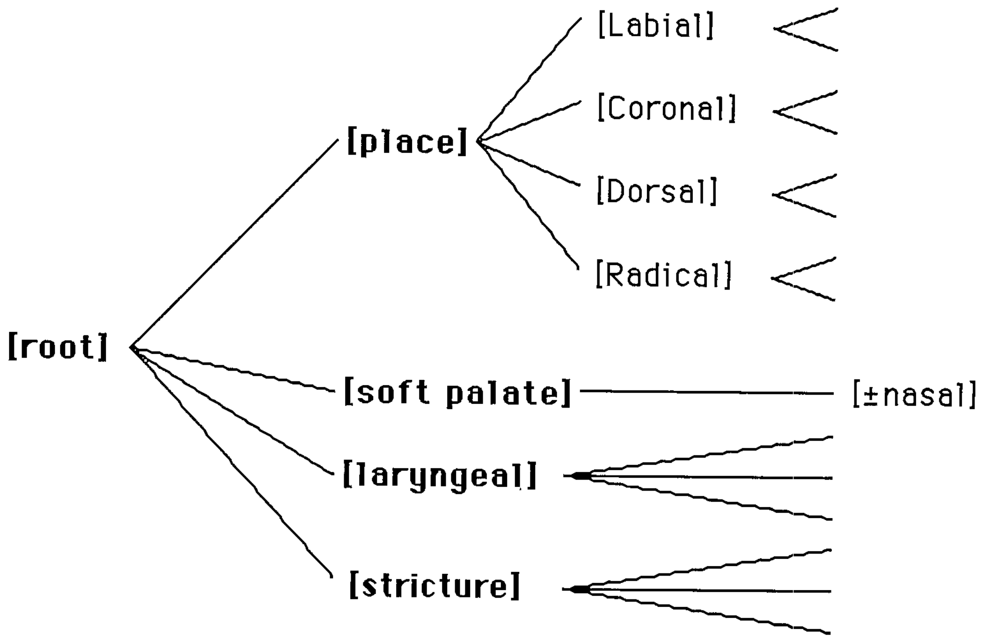


Figure 2. Part of the proposed hierarchy of features.

The notion of segments being articulator-bound is a very important aspect of this feature system. It is interesting to recall that earlier feature theories (Jakobson, Fant and Halle, 1951) suggested that features such as Grave and Compact could be realized articulatorily by a variety of maneuvers. It was shown in SPE (Chomsky and Halle 1968) and elsewhere that these features led to various phonological problems, and they were therefore replaced by Back, Low, and High, all of which are articulator-bound. It was surely no accident that it was precisely those features that specified the equivalent of place of articulation but were nevertheless not bound to particular articulators that turned out to be flawed.

We have shown the Soft Palate node as being dominated directly by the Root, although it would also be possible to introduce an extra node [Supra-Laryngeal], which dominated both [soft palate] and [place], distinguishing them both from the Laryngeal node. The Laryngeal node dominates a set of features that differentiate among voiced, creaky voiced, breathy voiced, and other sounds that differ in phonation type, as well as among sounds (primarily vowels) that contrast in tone.

The Stricture features specify the degree to which the vocal tract is constricted in producing a given speech sound. These features differ from all the others in that they are articulator free rather than articulator bound. In vowels the

Stricture features are implemented by the dorsal articulator; in [h] and [ʔ] by the laryngeal articulators; whereas in consonants the stricture features are implemeted by one of the four place articulators. To specify a speech sound it is therefore necessary to indicate the articulator charged with implementing its stricture feature. We have not had an opportunity to discuss the precise manner in which this is to be done. One possibility is by means of some version of the special pointer proposed in Sagey (1986), which would be located at the Stricture node, and would be set to point to the articulator that implements the stricture for the sound being specified.

We would like to emphasize that this hierarchical structure is part of a theory of phonological description. It is the first part of a statement of the complete set of phonological posibilities that speakers of different languages use. We have not yet worked out the conventions needed for using this tree stucture to determine all possible phonological segments. The conventions are somewhat similar to those for Figure 1, in that each sound is specified by a number of possible paths connecting the root node to the different terminal features, supplemented by a setting of the stricture pointer to a specific articulator.

The fact that we can agree on the general notions expressed in this paper has come about because of gradual developments in both our positions. Halle has been willing to abandon the enforcement of strict binarity on all nodes, and allow the Place node to dominate four other nodes (a fact that seems comparatively minor to him, as it is almost evident in his earlier work on redundancy in feature specifications). Ladefoged has abandoned trying to define "place" as a set of terms involving both an active and a passive articulator in acknowledgement of the paramount importance of the active articulator, which should be recognized before more detailed specification of the articulatory region. We still disagree on many particulars; and we are a long way from being able to present an agreed feature set. But who knows what the future may bring?

Footnote

1. A convention to consider possible revisions of the International Phonetic Alphabet will be held in Kiel, Germany, in August 1989. All who are interested (whether members of the IPA or not) are welcome to attend. (Further particulars are available from: IPA Convention, Linguistics Department, UCLA, Los Angeles, CA 90024-1543.)

Rerefences

Bell, A. Melville. 1867. Visible speech. London: Simpkin, Marshal.

Chomsky, Noam and Morris Halle. 1968. The sound pattern of English. New York: Harper and Row.

Clements, G. Nickerson. 1985. The geometry of phonological features. Phonology Yearbook 2:225-52.

Fromkin, Victoria, and Peter Ladefoged. 1981. Early views of distinctive features. Towards a history of phonetics. ed by Ronald J. Asher and Eugenie Henderson. Edinburgh: Edinburgh University Press.

Jakobson, Roman, C. Gunnar Fant, and Morris Halle. 1951. Preliminaries to speech analysis. Cambridge: M.I.T. Press.

Halle, Morris. 1988. The immanent form of phonemes. Giving birth to cognitive science: a festschrift for George A. Miller. ed. by William Hirsp. Cambidge, U.K.: Cambridge University Press.

Ladefoged, Peter. 1988. The International Phonetic Alphabet as a hierarchical feature system. Proceedings of the fourteeenth annual meeting of the Berkeley Linguistics Society. In press.

Sagey, Elizabeth. 1986. The representation of features and relations in non-linear phonology. Ph.D. dissertation, M.I.T., Cambridge, Mass.

Sweet, Henry. 1881. Sound notation. Transactions of the Philological Society 177-235.

Modeling articulatory-acoustic relations

A comment on Stevens's 'On the quantal nature of speech"

Peter Ladefoged and Mona Lindau
Phonetics Laboratory, Linguistics Department, U.C.L.A, Los Angeles, CA 90024-1543

Abstract.

This paper attempts to validate Stevens's quantal theory of vowel articulation by studying the acoustic effects of a wider and considerably more realistic and natural range of articulations than the simple tube model used by Stevens. Our model is implemented on a Macintosh computer and uses a vocal tract shape that is fairly similar to that of a human speaker. We studied the acoustic effects of a large number of articulatory vowel gestures. The results do not show much support for a quantal theory of vowel production. When articulatory gestures are varied in a simple way, the acoustic effects tend to vary monotonically with the articulatory variation, thus not producing stable plateau areas in the formant patterns.

1.Introduction

Stevens's Quantal Theory offers a number of intriguing insights into the nature of human language. With his unsurpassed knowledge of the acoustics of speech production, he has been able to point out several ways in which the non-linear relations between articulatory action and acoustic consequences might be exploited by speakers, so that they could allow small variations in articulation without suffering from major acoustic changes in the resulting speech sounds. He suggests that "this tendency for quantal relations between articulatory and acoustic parameters or between acoustic and auditory parameters is a principal factor shaping the inventory of articulatory states or gestures ..." But before we can accept this theory there are a number of things that must be done. Firstly we must examine the acoustic effects of a wider and more natural range of articulations. Stevens discusses articulatory-acoustic relations as demonstrated by a very simplified model of the vocal apparatus. He has shown the acoustic effect of varying the lengths, and to some extent the areas, of tubes representing the cavities of the vocal tract. We need to use a more realistic model of the shape of the vocal organs and, more importantly, to observe the acoustic results of varying differing shapes in ways that they might be varied by a real speaker. Only after having done this will we be able to say that there are certain sounds that remain acoustically stable despite potential variation in the articulations that produce them.

Secondly, we must check that any apparent acoustic stability is due to a simple articulatory movement. Whenever there is any change in the vocal tract shape there will virtually always be some kind of acoustic change. The only exceptions to this generalization occur when a speaker makes two simultaneous changes in vocal tract shape, the one compensating for the other. As has been shown by Atal, Chang, Mathews, & Tukey (1978) there can be simultaneous changes such as an increase in the degree of lip rounding and an increase in the height of the larynx which can cancel one another out. Riordan (1977) has shown that correlated changes of lip and larynx movement can be used by speakers to maintain acoustic constancy. But compensatory articulations of this kind are in fact rare. What is more to the point, a complex maneuver in which acoustic stability is the result of two opposing articulatory changes is not the kind of articulatory movement that Stevens is advocating as likely to be favored by languages.

A third area of concern is somewhat more complex. Stevens shows that part of the acoustic spectrum will remain constant when there are certain articulatory changes; but other parts of the spectrum will still be varying as a result of these same articulatory movements. Stevens concentrates on certain aspects of the acoustic structure. But we do not know whether those aspects of the sound that are acoustically stable are in fact those that are auditorily salient. Stevens shows that velar articulations will have a close approximation of F2 and F3 even when they are made over a wide range of articulatory closure locations. But is this approximation of the formants the significant cue for the perception of velars? Why should the auditory system pay more attention to the fact that F2 and F3 come together rather than to the frequency location of the locus, which, as Stevens notes, will change considerably as the location of the articulatory closure varies?

Another point to determine before we can accept Stevens's theory is whether the languages of the world actually use the favored articulatory gestures. This an enormous task that will not be attempted here, largely because we consider it premature. In this paper we will be concerned simply with the first step, checking whether a more realistic model of the articulatory system will produce the required quantal relationships.

2. Modeling vocal tract movements

Much of our recent work has been concerned with the relations that Stevens discusses in the first part of his paper. We have constructed a computer model of the vocal organs with which we can assess the acoustic results of different articulatory gestures. Stevens is well aware that his articulatory models are very schematic. We have tried to be more realistic in our description of the shape of the vocal tract and its possible movements. Our aim is to use a vocal tract shape that is fairly similar to that of a human speaker, and to control it in a way that it might be controlled in the production of speech. This latter consideration is a major stumbling block, in that we do not know what it is that the brain tells the vocal organs to do. But we can consider plausible alternatives; and we can make the variations in speech gestures in a way that are conceivable for a human to make.

In the model we have been using, the vocal tract is considered to be equivalent to 18 closely coupled tubes, each of variable length and cross-sectional area. The formant frequencies are calculated using a slightly modified version of the algorithm described by Liljenkrants and Fant (1975). The advantage of this algorithm is that it permits the specification of a variable length for each of the tubes. It is, however, not as sophisticated in its treatment of vocal tract losses as the later algorithm discussed by Wakita and Fant (1978).

The calculation of formant frequencies from a set of tubes has not been our main focus in the study of articulatory-acoustic relations. We have, however, taken pains to make sure that the lengths and areas of the tubes are properly specified. In our model the midline through each of the 18 sections of the tract is calculated for each sagittal view. Within each section, the length of this midline is taken to be the length of the equivalent cylindrical tube. The mean distance between the lower articulator and the upper surface of the vocal tract is then calculated for each section, using cross-sectional measurements at less than 0.5 mm intervals. The mean distances are converted into areas by reference to a look up table based on the measurements of Ladefoged, Anthony & Riley (1971), modified to bring them more into line with the mean measurements reported by Wood (1982). Figure 1 shows an example of the output of the program as implemented on a Macintosh computer. Note that the sections of the vocal tract are closer together near the tip of the tongue, where there may be more variation in tract shape; and that the length of the equivalent set of tubes reflects the high back vowel position.

We have sought to constrain the model's vocal tract shapes to those that could be produced by a human speaker; and we have also tried to make the movements from one shape to another in terms of controls that a person might plausibly use. These two goals have been met in part by specifying vocal tract shapes in terms of the the control parameters shown in Table I, combined with the options shown in Table II.

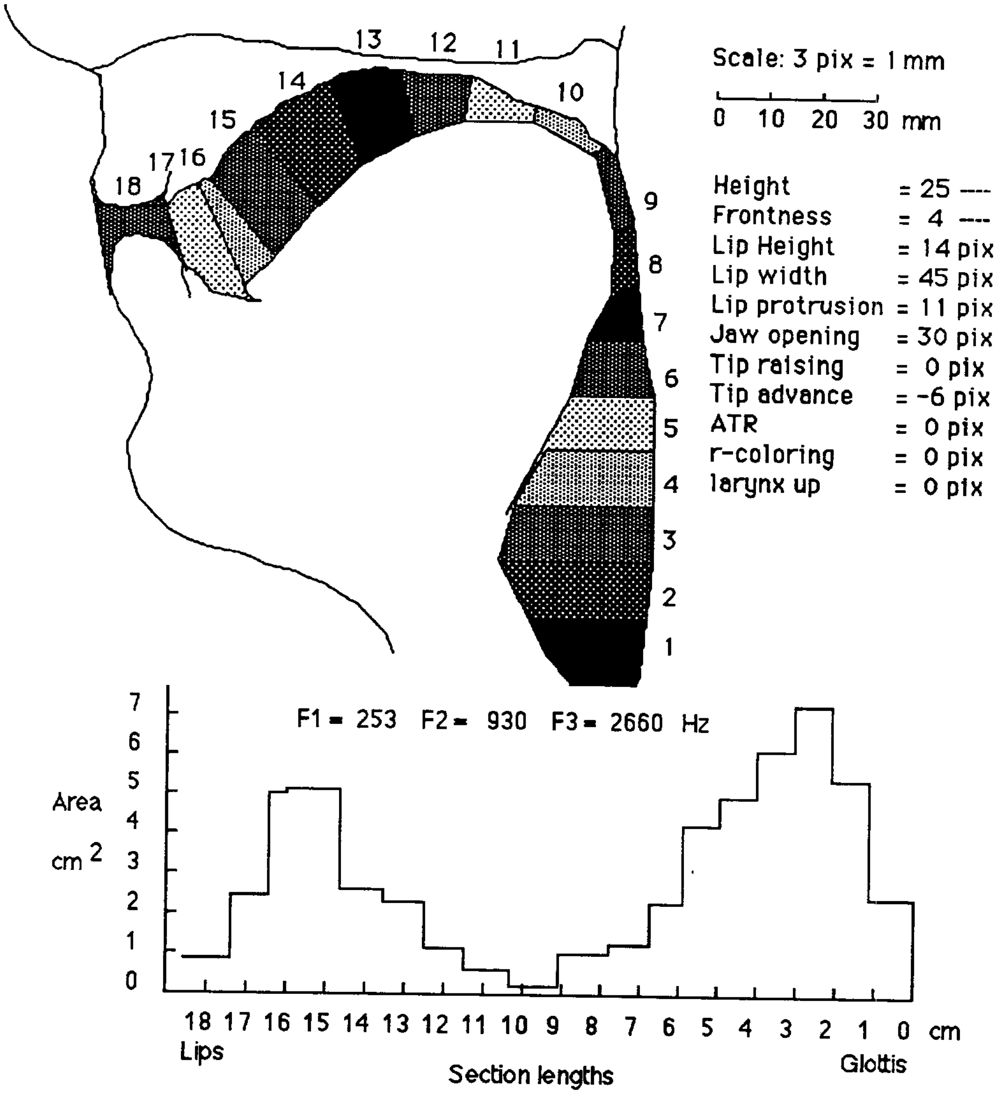


Figure 1. Sagittal diagram of the sections of the vocal tract and the corresponding area function for an [u] type vowel as produced by the UCLA vocal tract model. The calculated formant frequencies are also shown.

Table I. The parameters used for specifying vocal tract shapes.

1.	High-Low	**OR**	Front raising
2.	Front-Back	**OR**	Back raising
3.	Lip-Height		
4.	Lip width		
5.	Protrusion (rounding)		
6.	Jaw opening		
7.	Tip of tongue raising		
8.	Tip of tongue advancing		
9.	Advanced Tongue Root		
10.	Rhotacization (R-coloring)		
11.	Larynx raising		

Table II. Options in selecting or combining parameters of vocal tract shape.

1. Position of the body of the tongue determined by traditional tongue Height and Backness features.
2. Position of the body of the tongue determined by physiologically based factors.
3. Lower lip, jaw, and tongue body positions all independently specified.
5. Lower lip and jaw positions determined by vowel Height.
5. Lower lip and tongue body positions determined by jaw height.
6. Lip height, lip width, and protrusion all independently specified.
7. Lip width determined by protrusion.

Among the options, the first choice is between the use of a traditional feature based system for specifying the position of the body of the tongue, or the choice of more physiologically based parameters, such as the factors describing tongue shapes given by Harshman, Ladefoged & Goldstein (1977). If a traditional feature system is selected, then a further choice is possible between a separate specification of the jaw, lip, and tongue positions, and the specification of all these parameters together as a coordinative structure controlled by the traditional vowel Height feature. The latter possibility seems most plausible in the specification of English vowel sounds. Another option that is also possible is that this coordinated activity might be regarded as as determined by the parameter specifying jaw position, the tongue height and the lower lip position simply following along. In the production of sounds such as [s] that require a specific position of the lower jaw, this might be the more appropriate control structure.

Similar options exist for the specification of the lip opening. There are potentially three independent parameters: lip height (the distance in the mid-sagittal plane between the upper and lower lip); lip width (the distance between the corners of the lips); and protrusion (the lengthening of the lip tube). The latter two parameters can optionally be regarded as a coordinative structure, with an increase in lip protrusion also determining a decrease in lip width.

There is a problem in using traditional tongue Height and Backness features to specify vocal tract shapes. How do we interpret a two number specification of this kind as a specification of the position of 16 points on the tongue surface? No one has ever shown how these traditional parameters should be used to specify the shape of the tongue as a whole. However, it is well known (Ladefoged 1982) that they can be related to acoustic parameters in a direct way. Accordingly, in our model specifications of the Height feature are re-interpreted so as to give an appropriate value for F1 (using a look up table incorporating the inverse of the auditory scaling required), and specifications of Backness are similarly re-interpreted as values of (F2-F1). From these values of F1 and F2 a value of F3 is calculated, using the equation in Ladefoged and Harshman (1979). These three formant frequencies are then used to determine the degrees of Front-Raising and Back-Raising of the tongue, as described by Ladefoged, Harshman, Goldstein, & Rice (1978). The factors Front-Raising and Back-Raising are then used as described by Harshman et al. (1977) to generate the positions of the 16 points on the tongue.

This is a complicated way to get a specification of tongue shapes from values of the traditional terms; but, as far as we know, there is no better way to use values of the features Height and Backness to determine the locations of *all 16* points on the tongue. It is also interesting in that is a direct representation of the view proposed elsewhere (Ladefoged 1988) that the phonological feature Height should be regarded as having auditory correlates that determine the natural classes of sounds.

An alternative way of specifying tongue shapes is to consider the tongue to be controlled in terms of physiologically based factors, such as those described by Harshman et al (1977). The gesture that they call Front Raising (of the body of the tongue) has been shown by Jackson (1988) to be a component of descriptions of vowels in a number of languages. It is likely to be what he calls an articulatory prime, a basic component of speech gestures. Anatomically it is principally related to the pull of the genioglossus muscle. The second parameter described by Harshman et al. (1977), Back Raising (of the body of the tongue), is less well replicated in Jackson's study of a wider range of languages. It is perhaps not an articulatory prime, but a coordinative structure that could be decomposed into other primes. The principal anatomical correlate is the pull exerted by the styloglossus muscles. Our model allows the specification of vocal tract gestures in terms of what we consider to be anatomically plausible movements using these two factors.

3. Results of modeling articulatory-acoustic relations.

We have used a large number of the possible options within our model to attempt to validate Stevens's claims for the quantal nature of certain vowels. Unfortunately we have not been able to find much support for his position. In one set of experiments we tried various kinds of movements in the neighborhood of [i]. Figure 2 shows the results of separately varying the Height and the Backness features towards an [i] type vowel. Height is regarded as a coordinative structure that includes the movement of the tongue, jaw, and lower lip. This figure also shows the results of Front Raising and Back Raising movements of the tongue by itself. We tried similar additional experiments with all the options discussed above, all with very similar results. There is no way of producing the plateau in F2 that Stevens shows without making a much more complex articulatory maneuver.

The way in which one can produce a turning point in F2 is by moving the constriction further forward, as in Stevens's experiments with his simple tube model. But what is simple in terms of tubes is impossible for a human speaker to do without moving more than the body of the tongue. To get a constriction far enough forward it is necessary to add the action of additional parameters, such as raising and advancing the tip of the tongue, as well as adjusting the Height and Backness features so as to produce fronting and a slight lowering of the body of the tongue. Figure 3 shows articulatory movements of this kind. The left hand part of the figure shows a change in Height as the tongue moves towards an [i] position. If a movement of this kind is continued the tongue will make contact with the roof of the mouth. The only way of preserving a vowel like articulation is to use a more complex maneuver in which the blade of the tongue is advanced and raised, and the front of the tongue is lowered; this kind of change is shown in the right hand half of Figure 3. The acoustic results of these articulations is shown in Figure 4. The first seven points represents the changes in Height towards an [i] position, and the remaining points show a rising F1 and falling F2 as the blade of the tongue is advanced and raised. A complex maneuver of this kind occurs in the comparatively rare alveolarized front vowels that occur in Swedish and Chinese. Surely Stevens would not want to call these atypical vowels examples of quantal articulatory-acoustic relations?

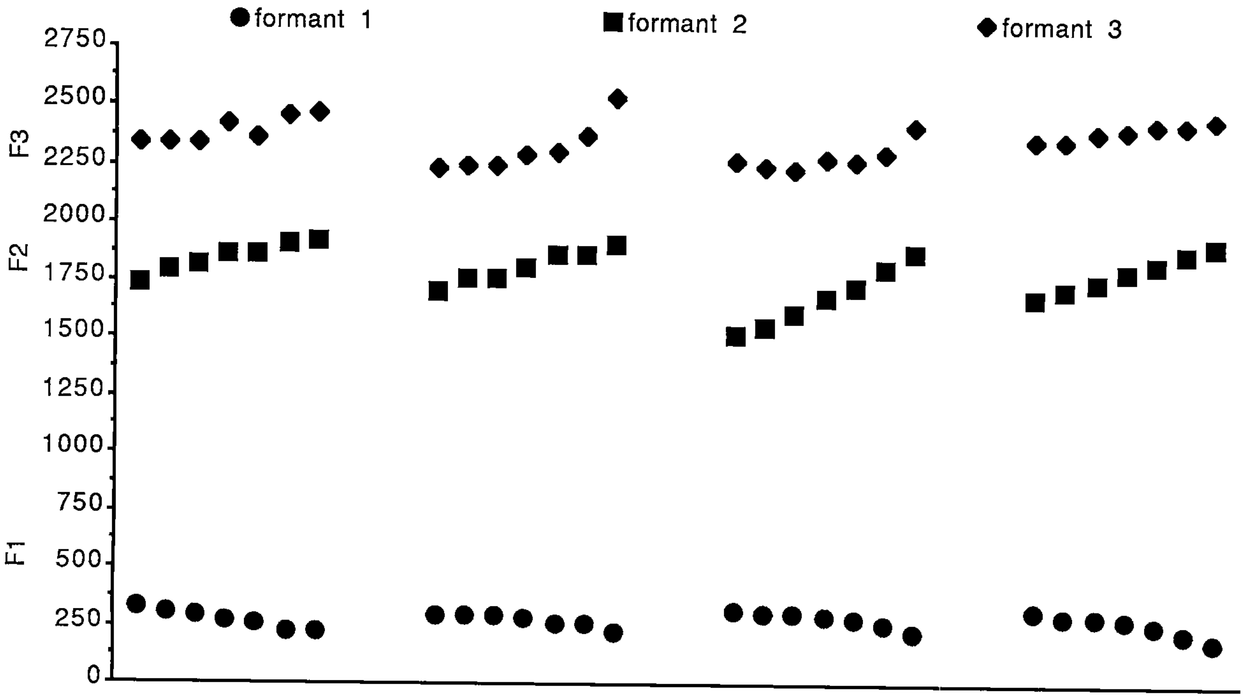


Figure 2. Acoustic results of separately varying the features of Height and Backness, and the parameters of Front Raising and Back Raising, moving towards an [i] vowel.

When separately varying the features Height and Back, and the parameters Front Raising and Back Raising in the neighborhood of [u] the results were similar to those for [i]. The formant frequencies varied monotonically with the articulatory variations, so that no plateaus of the kind required by Stevens's theory were found. We did, however, find an interesting acoustic stability in a totally unexpected region. Figure 5 shows the surprisingly stable formant frequencies that are produced when the tongue is varied from that of a relatively low back vowel to that of a relatively high back vowel while the lips remain rounded and protruded as indicated in Figure 6. A low back vowel with strong lip rounding similar to this is found in Assamese (an Indo-Iranian language), but it is, to say the least, not very common.

It may be seen that Stevens's theory provides a number of useful insights into the nature of articulatory acoustic relations but there is much that remains to be done before we can compare it with other accounts of why languages have the sounds that they do. In addition to the specific points concerning articulatory-acoustic relations we have discussed above, we note that Stevens's gives no role to the fact that some sounds are easier to pronounce than others. It seems to us that the sounds that actually occur in languages are the results of a number of factors. One of these may be the quantal relations documented by Stevens. Other involve the necessity for languages to balance the conflicting notions of economy of articulatory effort and the need for acoustic distinctiveness (Martinet 1955). The more recent development of this work by Lindblom (1984) is not incompatable with quantal theory. Perhaps when both have been fully developed we will know why we speak and hear the way that we do.

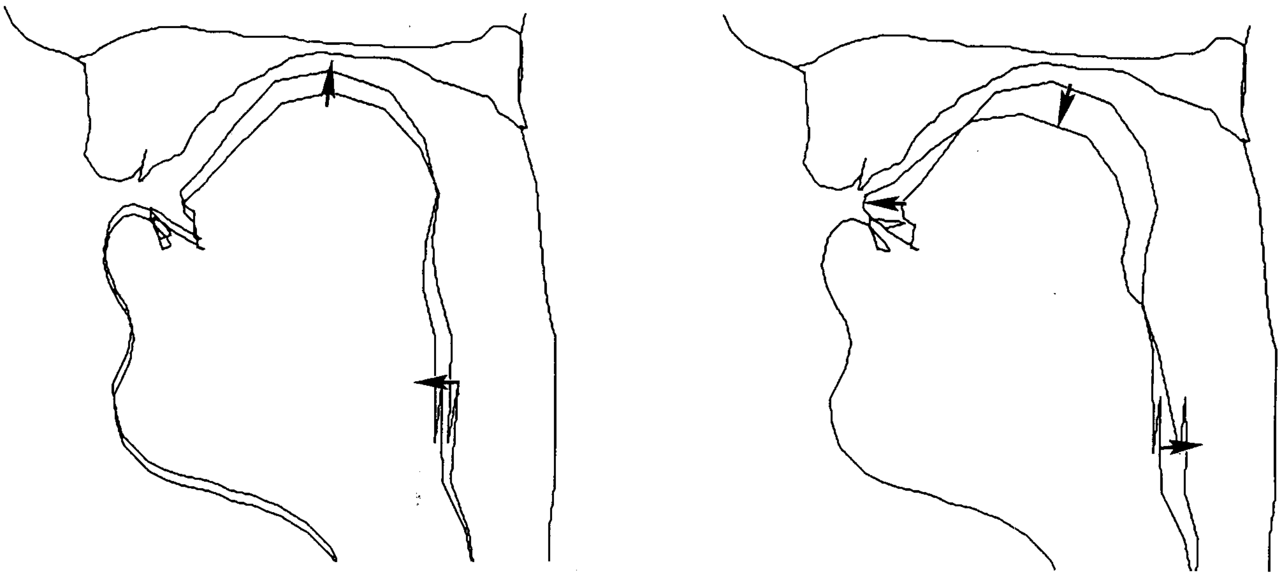


Figure 3. The lefthand part of the figure shows variation of the Height feature by moving the tongue, jaw, and lips towards an [i] position. The right hand part shows a more complex vowel articulation, moving away from the [i] position by raising and advancing the blade of the tongue towards an alveolarized high front vowel.

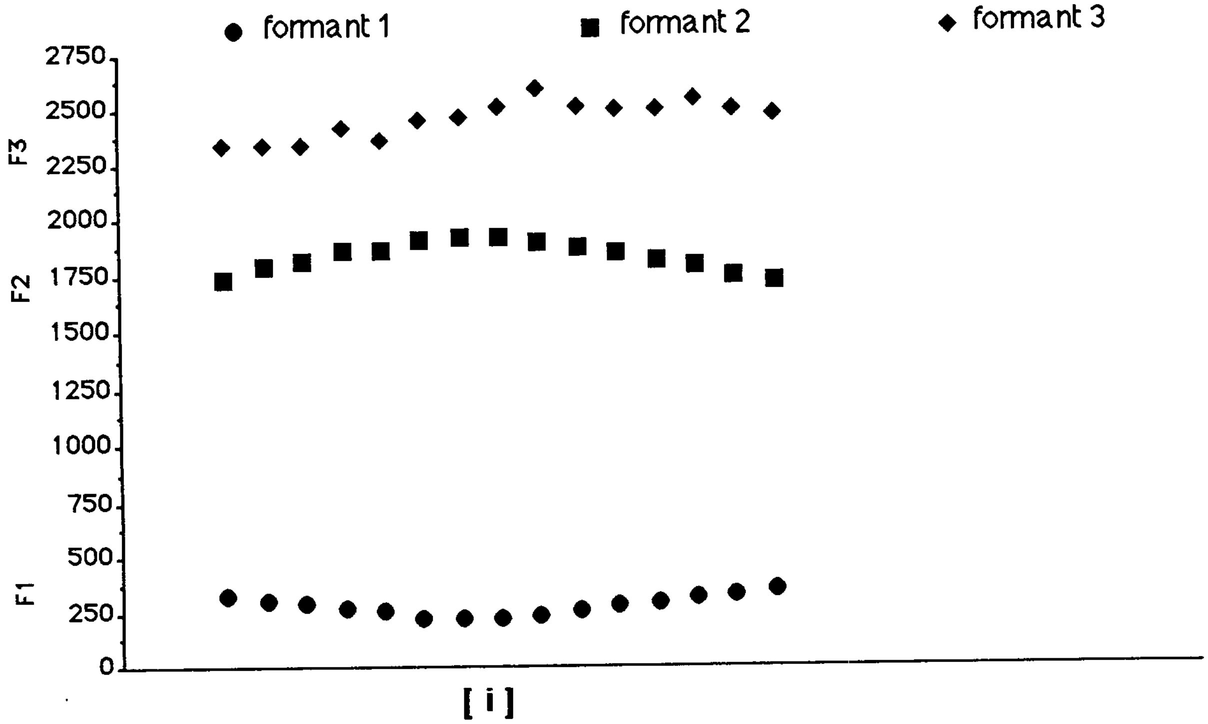


Figure 4. Acoustic results of the articulatory gestures in figure 3. The first seven points represent variation in Height towards [i], and the remaining points show a rising F1 and a falling F2 forming a plateau, as the blade of the tongue is raised and advanced.

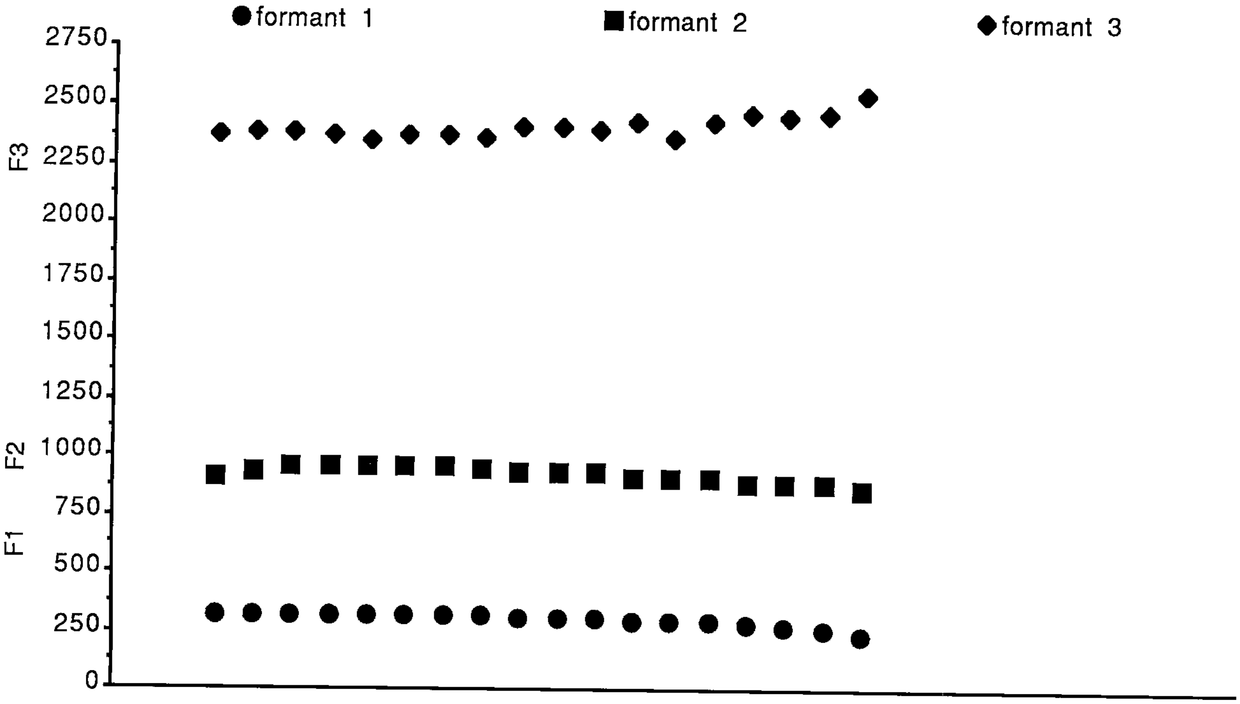


Figure 5. Acoustic results of varying the Back Raising parameter only from a low back tongue position to a high back tongue position. Note the very small effect on the formants from this relativelyl large articulatory movement.

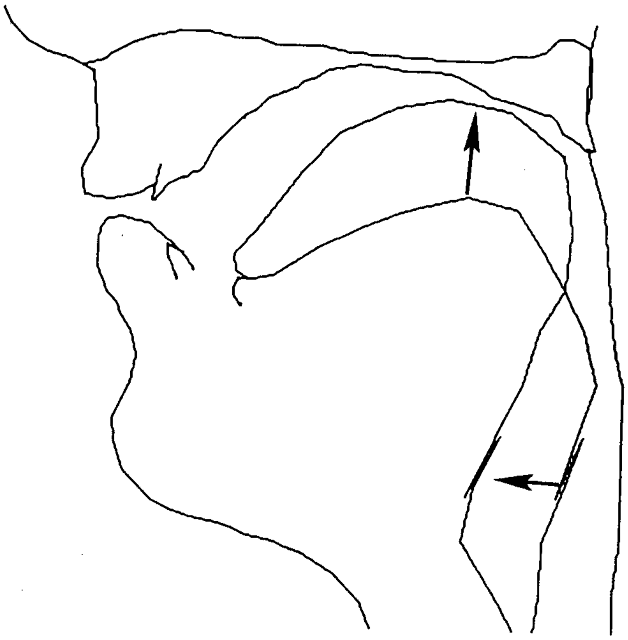


Figure 6. Varying the Back Raising parameter only from a low back tongue position to a high back tongue position with protruded lips.

References

Atal, B.S., Chang, J.J., Mathews, M.V.& Tukey, J.W. (1978) "Inversion of articulatory-to-acoustic transformation in the vocal tract by a computer-sorting technique." Journal of the Acoustical Society of America. **63** (5),1535-1555.

Harshman, Richard, Ladefoged, Peter, & Goldstein, Louis (1977) "Factor analyses of tongue shapes." Journal of the Acoustical Society of America. **62** (3), 693-707.

Jackson, Michel (1988) Phonetic theory and cross-linguistic variation in vowel articulation. UCLA: Ph.D. diss.

Ladefoged, Peter (1980) "What are linguistic sounds made of?" Language. **56** (3), 485-502.

Ladefoged, Peter (1982) A Course in Phonetics. Harcourt Brace & Jovanovich: New York.

Ladefoged, Peter (1985) "Macintosh usage for linguists" UCLA Working Papers in Phonetics. **60**, 84-86.

Ladefoged, Peter (1988) "The International Phonetic Alphabet as a hierarchical feature system." Berkeley Linguistics Society (forthcoming).

Ladefoged, Peter., Anthony, J.F.K. & Riley, C.(1971) "Direct measurement of the vocal tract." UCLA Working Papers in Phonetics **19**, 4 -13. Los Angeles: Phonetics Laboratory.

Ladefoged, Peter & Harshman, Richard (1979) "Formant frequencies and movements of the tongue." In Frontiers of speech communication research (B. Lindblom & S. Öhman, editors) pp. 25-34. London: Academic Press.

Ladefoged, Peter, Harshman, Richard, Goldstein, Louis, & Rice, Lloyd. (1978) "Generating vocal tract shapes from formant frequencies." Journal of the Acoustical Society of America. **64** (4), 1027-1035.

Liljenkrants, Johan & Fant, Gunnar (1975) "Computer program for VT-resonance frequency calculations." Speech Transmission Laboratory, Quarterly Progress and Status Report. **4**, 15-20.

Lindblom, Björn (1984) "Can the models of evolutionary biology be applied to phonetic problems?" In Proceedings of the Tenth International Congress of Phonetic Sciences. (M.P.R. van den Broeke & A.Cohen, editors), pp. 67-81. Dordrecht-Cinnaminson: Foris Publications.

Martinet, André (1955) Économie des changements phonétiques (2nd ed.) Berne: Franke.

Riordan, Carol (1977) "Control of vocal-tract length in speech." Journal of the Acoustical Society of America. **62**, 998-1002.

Wakita, Hisashi, & Fant, Gunnar (1978) "Toward a better vocal tract model." Speech Transmission Laboratory, Quarterly Progress and Status Report. **1**,9-29.

Wood, Sidney. 1982. X-ray and model studies of vowel articulation. Working Papers. **23**. Lund University: Department of Linguistics.

A view of phonetics

Peter Ladefoged

[Comments at the opening of the XIth International Congress of Phonetic Sciences, Tallinn, USSR]

At the opening of the Congress of Phonetic Sciences it is appropriate to begin by thinking about the nature of our field. What is a phonetician? I am often asked that question, and always find it difficult to answer. I have said many different things at different times in my life. When I was young, returning from a war, I thought of myself as a poet, concerned with the sounds of words. I went to university to learn about literature, but soon found that studying literature meant learning about the lives and times of the great poets and authors. Scholars of literature do not pay much attention to the sounds of the words that poets use.

Fortunately, I stumbled into phonetics where the sounds of words were the principle object of study; and I became a phonetician as well as a poet. But when I was asked what I was I soon found that not many people had heard of phonetics. It did no good to say that I was a phonetician, and, as it was not quite respectable to say that I was a poet, I would fall back on the fact that I had built a few amplifiers and other trivial circuits, and say that I was a speech engineer. Of course anybody who really knew about circuit theory would see through me at once, so when talking to them I would put it slightly differently and say that I was more of a physicist, concerned with the acoustics of speech. When a real physicist came along I would move again, and, relying on the fact that I had done some perceptual experiments, I would claim to be a kind of psychologist. But there are lots of real psychologists around, and when I ran into any one of them I would have to talk about another aspect of my work, the study of the anatomy of the vocal organs. Naturally, a real anatomist would look at my crude dissections and drawings, and realize instantly that I was not in that field. So then I would have to explain that I was more concerned with the functions of the muscles as studied by electromyography. But then the physiologists would catch me out in this field, and I would have to move in another direction, claiming to be more of a linguist, or language teacher. However, I know virtually nothing about syntax and semantics and all the other things that proper linguists talk about, so it is obvious I am not one of them either. Nowadays when asked what I do I sometimes just shrug and confess to being a closet hacker — a computer addict concerned with synthesizing and analyzing speech.

All these things are part of my life as a phonetician: communication engineering, physical acoustics, psychology, anatomy, physiology, linguistics, applied linguistics, computer science, and poetry. I am sure that this is true for many people at the Congress. So I can now sum up my original question: what is a phonetician? The answer is that we are phoneticians, we, the people who come to phonetics congresses, and know something about some of these diverse disciplines. None of us can know enough about all of them, which is why being a complete phonetician is an impossible task. But every four years we can get together and pool our knowledge. This is phonetics.

Revising the International Phonetic Association's Alphabet*

PETER LADEFOGED

PETER ROACH
(Dept. of Linguistics and Phonetics, University of Leeds, Leeds LS2 9JT, U.K.)

The present set of the Association's symbols had its origins nearly 100 years ago, shortly after the International Phonetic Association (IPA) was founded in 1887. It was revised many times in its early years, but in the last 40 years there have been few changes. As a result, it is now time for the Association to turn to this matter again. But before we do so, we would like to make it clear that in our view the Association should be concerned with far more than the management of a set of symbols. Just as no university course in phonetics should limit itself to teaching students how to make phonetic transcriptions, so equally the IPA is concerned with the whole science of phonetics. Phoneticians are not just people who can hear and produce a great variety of speech sounds. They are scholars who have studied the entire process of speech production and perception. They know how speech sounds form the medium of spoken language, and they continually relate their work to other fields such as general linguistics. They also know something about the practical applications of their work, ranging from pronunciation teaching to automatic speech recognition. Phonetics is an academic discipline, and the IPA is the organization of the group of scholars who are enaged in that discipline.

Having recalled our own background, let us also note that hundreds of scholars in many different disciplines rely on phonetic symbols to convey their meaning, without themselves knowing much about general phonetics. Some form of phonetic alphabet is essential for work in linguistics, speech pathology, computer speech processing, language teaching, anthropology, studies of ancient manuscripts, descriptions of social class, singing, criminal voice identification - the list of topics that require the use of phonetic symbols is obviously very lengthy. Some of these topics need special symbols that are suited only for them; but all of them require a common core for their basic needs. By far the most widely used common core is the International Phonetic Association's Alphabet (IPAA); this is the set of symbols that most scholars take as their starting point, and then, if necessary, augment with special symbols for their own needs.

When we come to revise the IPAA we must begin by recognizing that by far the majority of those who use the IPAA are not phoneticians, as is obvious from the list of users given above. Some language teachers, linguists, speech pathologists, communication engineers and language teachers may be interested in phonetics; but most regard the IPAA simply as a tool. Phoneticians who are members of the IPA must therefore regard themselves, at least in part, as manufacturers providing a product. If we want our product to be accepted, we must ensure that consumers are satisfied with what we provide. This will require a certain amount of customer research, noting both where the IPAA has been successful, and where it has failed. The successes are, perhaps, more extensive than is realized. There are a number of generally accepted standard usages, even among those who do not recognize that they are often following traditional IPA usage. For example, most English speaking linguists and phoneticians use basically IPAA vowel symbols, rather than the symbols commonly found in dictionaries of English; and all phoneticians use [k] rather than **c** for a voiceless velar stop. These are simply IPA usages.

The failures of the IPAA seem to be in small and disparate ways, with a variety of different causes, some of which will no doubt persist. For example, the influence of different national orthographies on the usage of particular symbols such as [j] and [y] are likely to be difficult to overcome. Europeans, many of whom need to distinguish front rounded vowels, and who are also accustomed to using the letter **j** for the first sound in words such as German *ja*, will no doubt continue to use these two symbols in the official IPAA way; but Americans, who are not known for their internationalism, will probably stick to their own orthographic **y** for the first sound in *yes*. We should in all such cases make the Association's position clear.

The very first step in revising the IPAA is, therefore, for the IPA to seek the cooperation of other organizations, such as national and international associations concerned with linguistics, speech pathology, language teaching and communication engineering. We have notified a number of these organizations of our intent to revise the IPAA, and asked for their comments on what they now find acceptable and unacceptable. We hope that all members will take it on themselves to see to the further publication of our plans.

The second major point that must be emphasized before we begin is that revision of the IPAA involves revising a theory of phonetics. It is not just a matter of getting agreement on what symbols to use. This is part of the problem; but there is also the much more difficult task of getting agreement on what to describe. Symbols stand for something; they are shorthand descriptions of sets of phonetic categories. Choosing the symbols required for an international phonetic alphabet is a simple task in comparison with choosing the categories that need to be represented by these symbols. The present IPA chart is a theory of phonetics specifying how sounds should be described in terms of particular articulatory categories. For example, [θ] stands for 'voiceless dental fricative' and [ɮ] is just a short hand way of writing 'voiced dental or alveolar lateral fricative'. We now have to consider whether these articulatory terms are still sufficient for our needs. There are many current phonetic theories ranging from acoustically based theories through more traditional IPA categories to elaborate articulatory notation systems. What do we want our symbols to symbolize? There are different requirements in different disciplines. Speech pathologists, for example, may want to think of sounds in terms of articulations, communication engineers in terms of acoustic categories, and linguists in terms of distinctive features. There is, however, a common core of needs which we will have to address when reviewing the underlying principles on which the IPAA is based. The theory represented by the IPAA should be high powered enough to take care of the complexities of phonetics, while at the same time allowing others to grasp the essence of our subject.

For any revision to be fully accepted, it must be not only satisfactory for general scholars and the multifarious special users of the IPAA; it must also have been fully discussed and approved by a large number of phoneticians who can be regarded as authorities. The elected Council of the IPA is the official group for this purpose. It should of course be supplemented by other well known phoneticians or linguists who, for one reason or another, are not members of the IPA. In addition, because many of the members of the Council are senior figures who are not the future leaders of the field, there must be ample provision for the younger members of the IPA to participate in the decisions.

The Editor has informed us that he is currently encouraging discussion of revisions in the *Journal.* We anticipate that in the next two years a number of papers and short notes will be published. Then, in the summer of 1989, we propose that there will be a working convention. This convention will not be like a Phonetic Congress, or a conference. Instead it will consist of a number of working groups, each of whom will prepare a report on one of the areas to be discussed below. Some of these reports will have been drafted before the meeting; others will be written at the time. All of them will be extensively discussed within the working groups.

The areas to be covered by the reports will include the nine topics enumerated below.

1. The principles on which an International Phonetic Association's Alphabet should be based. The historic, and, as we have not voted to alter them, the still current principles of the IPA are:

1. There should be a separate letter for each distinctive sound; that is, for each sound which, being used instead of another, in the same language, can change the meaning of a word.
2. When any sound is found in several languages, the same sign should be used in all. This applies also to very similar shades of sound.
3. The alphabet should consist as much as possible of the ordinaryletters of the Roman alphabet, as few new letters as possible being used.
4. In assigning values to the Roman letters, international usage should decide.
5. The new letters should be suggestive of the sounds they represent, by their resemblance to the old ones.
6. Diacritics should be avoided, being trying for the eyes and troublesome to read.

It is interesting that the first of these principles is an early formulation of the phonemic principle; and the second, which presupposes that the same sound can be found in different languages, is equivalent to a statement that there is a universal set of phonetic categories.

What is not explicitly stated in the principles themselves is what the alphabet is for. This is a fundamental concern that needs to be be made clear before attempting any revision of the IPAA. The group working in this area should perhaps consider whether we should oficially adopt something like the first paragraph of the 1949 edition of *The Principles:*

> The alphabet of the *Association Phonétique Internationale* is an alphabet on romanic basis designed primarily to meet practical linguistic needs, such as putting on record the phonetic or phonemic structure of languages, furnishing learners of foreign languages with phonetic transcriptions to assist them in acquiring the pronunciation, and working out orthographies for languages written in other systems or for languages hitherto unwritten. Numerous symbols and marks are also provided, by means of which many minute shades of sounds may be represented, and which thus render the alphabet well suited for use in scientific investigations, e.g. in dialectology, in the historical study of languages, and in comparative philology.

We would suggest that we should at least add investigations of man-machine interaction to the list of possible uses; but what else should be added - or left out?

2. The phonetic theory that the symbols represent. As we have discussed above, in any phonetic alphabet, the symbols are simply shorthand ways of representing sets of categories. There are now many different phonetic theories, ranging from acoustically based feature theories, through more traditional categories, to elaborate articulatory notation systems. It will be a challenge to come up with a set of categories that it would be useful to represent in terms of a limited set of symbols that can receive general recognition. This is undoubtedly the major problem in trying to revise the IPAA. It involves trying to agree on the form of phonetic descriptions. We will probably be able to do this only at a basic level; and we will probably have to be fairly conservative and avoid too many innovations. It may be appropriate to split this area into a number of of groups, one concerned with the general phonetic theory required for describing consonants, another with that required for vowels, and a third for suprasegmental aspects of speech. A fourth theoretical area is the question of symbolizing the time course of phonetic events. The IPAA has always been a set of symbols for segments without any temporal structure. There has been no authorized way of taking into account differences such as that between a fully nasalized vowel and a vowel that becomes nasalized part way through. It may not be possible, but we should at least consider whether we can agree on a phonetic theory that will permit the symbolization of the time course of phonetic events.

3. The common core of the IPAA. Historically, the IPAA has been used as a notation system by field linguists, teachers of phonetics, speech pathologists, communication engineers, and others, all of whom have slightly different needs. The working group considering this topic will be responsible for defining what is the most useful common core for the Association to promulgate.

4. The form of presentation of the IPAA. This problem is closely related to that in (2), as well as that in (3). The set of symbols has usually been presented in one or two charts (depending on whether vowels and consonants are treated as distinct), together with a collection of miscellaneous information on other symbols and diacritics. In this form the basic information concerning the alphabet has been made available on a single page, which has been reproduced in countless textbooks, encyclopedias, etc. Modern theories would seem to require a larger number of smaller charts, which still might be arrangeable on a single page. But, as noted in (2) above, we need to come to some agreement on how we should label the rows and columns on these charts;

and, as noted in (3), we need to agree on a common core that should be symbolized. The task of group (4) will be to consider the pedagogical and other requirements involved in presenting the IPAA.

5. Individual symbols and diacritics. There is a need for agreement on the shapes of individual symbols, and on additional symbols to represent sounds not previously considered (e.g. bilabial trills). We also need to standardize the shapes of some symbols that appear in more than one form (e.g. [g , g]) in the current alphabet. When considering new symbols we will have to discuss the basis for the relation between symbols and diacritics (a topic that overlaps with some of the other areas).

Discussions on symbol shapes extend beyond phonetic considerations. As the 1949 edition of *The Principles* states, we must pay attention "not only to the appropriateness of each letter from a phonetic point of view, but also to the suitability of letters from the psychological and pedagogical angles and as regards typographical harmony, the needs of the printer and written forms."

6. Past successes and failures of the IPAA. Some aspects of the IPAA have been more widely accepted than others. A survey of current books on phonetics, and perhaps more importantly of textbooks in related fields such as linguistics and speech pathology, will show that some aspects of the IPAA are in common use, whereas others have been disregarded. It is important for us to know where we have been successful, and where (and why) we have failed. Without knowing our strengths and weaknesses we will be unlikely to produce a viable new alphabet.

7. Computerization of the IPAA. There are several topics in this area. Firstly we should have an agreed ASCII computer coding for at least 128 IPAA symbols. Secondly we should consider offering a recommended way of selecting from these symbols when providing a simple phonemic transcription of well known languages in computer readable form. Thirdly we could provide symbols for a narrow, comparative, phonetic transcription of the kind required for annotating data bases used in speech computer projects. These symbols might include ways of representing different kinds of stop bursts, transitions, and many other items not normally considered within traditional phonetics. Fourthly (and, probably less importantly now that laser printers are becoming more common) we should try to standardize dot matrix forms of IPAA symbols. The IPAA was developed before the widespread use of typewriters, and the Association suffered because it never recognized the advisability of offering an alphabet that could be easily managed on a typewriter. We must not make the same kind of mistake by overlooking the need for computer compatibility.

8. Extensions of the IPAA. There is a need in some studies that are concerned with describing the speech of individuals (for example, speech pathology, speaker identification, singing) for symbols for sounds that are not part of any language, and are therefore not within the domain of the alphabet as originally conceived. We should consider how this need should be met.

9. Recorded illustrations of the IPAA. Thanks to John Wells and Susan Ramsaran we already have a cassette recording illustrating the symbols in the current chart. A similar recording should clearly be part of a revised IPA.A In addition it would be useful to provide examples of as many symbols as possible (at least all those used in well known languages) on a recording using real words spoken by native speakers of different languages. Both these types of recordings would, of course, have to be accompanied (as is the present Wells and Ramsaran recording) by a clear statement that each IPAA symbol can be used to represent many slightly different sounds, and that the recordings merely illustrate some possible examples of these sounds.

There are no doubt several other topics that will have to be considered at the proposed meeting. During the two years before this takes place there will be many discussions and views advanced that may lead to major revisions in the agenda. Our aim is to have a very well prepared convention that has been preceded by considerable ferment and thorough discussion. We hope that as many people as possible will attend and participate; a preliminary form for those wishing to participate is enclosed with this issue of the *Journal* (additional copies are obtainable from the Secretary). The IPAA is one of the most important and widely used tools in our field; we want to make sure that it is maximally useful.

* To appear in the *Journal of the International Phonetic Association.*

Prenasalized stops and speech timing.

Paper presented at the 115th Meeting of the Acoustical Society of America
Seattle, Washington, May 16-20, 1988

Ian Maddieson

Prenasalized stops, that is, homorganic nasal+stop elements that behave as single phonological segments, raise a number of interesting questions concerning the relationship between phonological units and timing in speech. Do complex phonetic elements of this kind occupy the same duration as simpler elements, such as plain stops or nasals? Do prenasalized stops have the same timing pattern as a phonological sequence of nasal plus stop? How do prenasalized stops act with respect to rules which adjust the duration of neighbouring segments; for example, would a vowel before a prenasalized stop be shortened by the widespread rule which shortens a vowel in a closed syllable (Maddieson 1985)? It has been argued that the status of prenasalized stops as single segments is directly related to their duration. They have been defined as nasal+stop sequences with the duration typical of other single segments (Herbert 1986). He and Sagey (1986) in her dissertation on complex segments both indicate that they would expect phonological consonant sequences to have longer durations than single segments regardless of whether the single segments are phonetically simple or complex. On the other hand, Ladefoged and Maddieson (1986) suggest that there is no demonstrated phonetic difference in timing between nasal+stop sequences and prenasalized stops. Purported *language-internal* contrasts between these elements actually involve a difference between geminate and single nasals before stops, or between syllabic and non-syllabic nasals before stops. They suggest that deciding if a nasal+stop element is a prenasalized stop is not a phonetic issue but one which concerns solely the phonology of the language in question. Languages differ in the ways in which phonological structure interacts with timing, so timing may or may not be involved in this issue.

The Ladefoged and Maddieson view coincides with a principal point put forward by Browman and Goldstein (1986). At first glance, their data might appear to have answered many of the questions raised above. In the course of presenting a larger argument concerning phonological representations, they argue that prenasalized stops have similar temporal characteristics to nasal+stop sequences as well as to plain stops and that timing is not related to phonological status at a segmental level. Browman and Goldstein's demonstration of their point proceeds as follows. They monitored lower lip displacement in the set of real and nonsense words shown at the top of figure 1, containing either a single medial labial consonant or a labial nasal+stop sequence, then superimposed lip movement traces from a selected token of each word from a single speaker of English. The coincidence of the lip traces in figure 1 shows that "the overall envelope of the gestures is similar" and the consonantal portion has the same timing regardless of whether it is a single consonant or a consonant sequence. For data on prenasalized stops they used the Tanzanian Bantu language KiChaka (Chaga). Lip movement tracings from selected tokens of words beginning with simple labial segments were superimposed, together with one from a word

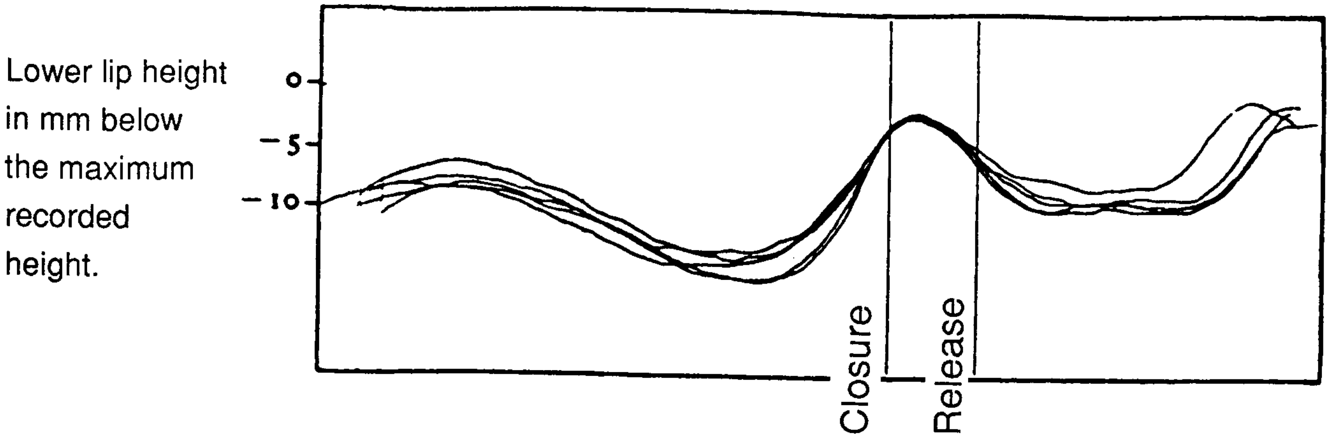


Figure 1. Superimposed lower lip movement trajectories of the English utterances *capper, cabber, cammer, camper* and *camber* (after Browman & Goldstein 1986). Trajectories are lined up at the peak of the labial articulatory gesture for each word. The value 0 is assigned to the highest lip position recorded in the session. Closure and release times indicated are for the utterance *capper.*

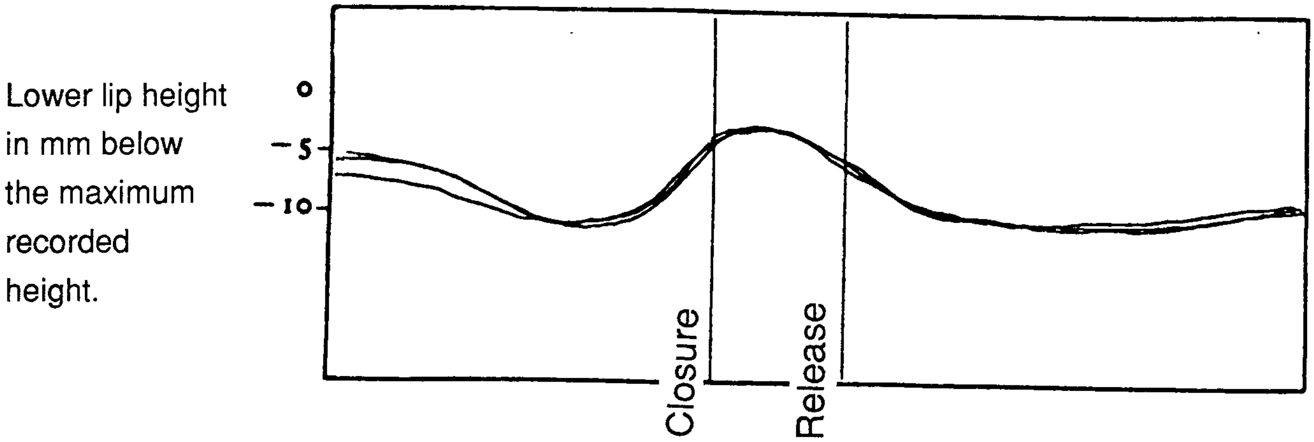


Figure 2. Superimposed lower lip movement trajectories of the KiChaka utterances /paka/ 'cat', /maka/ 'year' and /mbaka/ 'evil-doing' (after Browman & Goldstein 1986). Trajectories are lined up at the peak of the labial articulatory gesture for each word as in figure 1. Closure and release times indicated are for the utterance /maka/.

beginning with /mb/, as shown in figure 2. The data are from a speaker of the KiVunjo dialect. Again they point to the similarity between the various labial gestures and argue that plain stops and prenasalized stops in KiChaka are alike. We now have essentially established a syllogism. English nasal+stop clusters are like (English) plain labial stops and nasals; KiChaka prenasalized stops are like (KiChaka) plain labial stops and nasals; therefore English stop+nasal clusters are like prenasalized stops.

Browman & Goldstein also suggest that there are additional temporal regularities in English which can be explained on the basis of a common gesture underlying both plain stops and nasals and the nasal + stop sequences. They point out that, using the same set of English words, Vatikiotis-Bateson (1984) found stressed vowel duration was the same before the single consonants /p/ and /b/ as before the matched nasal+stop cluster (i.e. /p/ and /mp/; /b/ and /mb/). Similarly, vowels were shorter before both /p/ and /mp/ than before /b/ and /mb/. Thus, vowel length was determined by stop voicing but not by status as single consonant or cluster. An expected shortening effect of the cluster (Lindblom & Rapp 1973, Klatt, 1979) was not observed. As for KiChaka, they suggest that the phonological regularity of importance is that possible syllable onsets are those which consist of single oral articulatory gestures, whether or not nasality is involved.

Closer examination, however, suggests that there are some problems with Browman and Goldstein's approach to both English and KiChaka and these observations do not satisfactorily document a similarity between single stops and prenasalized ones.

Let us discuss the English case first. Pairs of words like "camper" and "capper" seem unsuited for testing effects of segmental organization, since there are good reasons for regarding the intervocalic consonant in words like "capper" as ambisyllabic (for example, /æ/ is a vowel that is normally limited to closed syllables). The vowel before /p/ is therefore in a syllable which contains a coda, just as the vowel in "camper" would be if the nasal + stop sequence has the articulatory organization of a cluster. The question of whether /mp/ is acting like a single consonant or a sequence cannot be answered by comparison with intervocalic /p/ since /p/ may be acting rather like a sequence itself, particularly with respect to effect on a preceding vowel.

The evidence from voicing-dependent vowel length adjustment rule is also unclear. If the similarity of application of these rules before plain stops and nasal + stop sequences is due to the fact that they share the same articulatory organization, that is, both are made with a single labial gesture with the same temporal characteristics, then we would expect this rule to be blocked from applying to the first vowel in pairs like the proper names 'Albert', 'Alpert', where a lingual gesture precedes the labial gesture. In fact, as is well known, in such cases there is not only a difference in the vowel duration, but also one in the duration of the lateral segments (Klatt 1979, Fourakis 1980). Such durational adjustments apply generally to liquids and nasals in English as well as to vowels (although Browman and Goldstein did not report this effect in the English materials they studied). Homorganicity is not a requirement for the rule to apply. Instead, only the voicing category of the last obstruent in the cluster is relevant. That the rule applies before nasal + stop sequences does not indicate that they have single-segment-like temporal organization.

On the other side of the equation, KiChaka is not entirely convincing as a representative of a language which has a phonological category of prenasalized stops. The word-initial nasals in nasal + stop onsets are generally separate morphemes -- prefixes specific to particular noun classes which may alternate with prefixes of CV or V structure. For example, the stem /-ku/ 'firewood' may be preceded by a singular prefix /u-/, or a 'mass' or collective prefix which takes the form of a nasal homorganic to the stem-initial consonant. These nasal prefixes are neither syllabic nor tone-bearing, but are simply incorporated in the syllable onset.[1] When preceded by a nasal, KiChaka stops become voiced (Nurse 1981: 142), hence the plural of 'firewood' is /ŋgu/. Otherwise - in absolute initial position or between syllabic segments - stops are always voiceless. The nasal prefix causes voicing assimilation of a following stop, by a rule which also governs morpheme-internal nasal + stop sequences, as in the noun /numba/ 'house'. There is nothing here to demonstrate the phonological unity of the nasal + stop elements. Their similarity in timing to initial plain stops is interesting, but doesn't clearly relate to the timing of phonologically unitary prenasalized stops, and is not known to have any relevance for the timing of adjoining segments, either at the phonetic level or in the operation of such phonological rules as compensatory lengthening.

Therefore the questions concerning timing of prenasalized stops remain somewhat open. More extensive and less ambiguous phonetic data is required to address them. Data from Fijian has been collected with a view to assisting in this aim. Fijian is a language in which prenasalized stops are persuasively analyzed as single segments from the phonological point of view (Geraghty 1983). It has stops of two types, voiceless oral and voiced prenasalized. These two types of stops have the same distributional patterns. Unlike KiChaka, no boundaries fall between nasal and stop and a post-nasal voicing assimilation rule cannot be motivated by alternations. If we set aside the prenasalized stops for the moment, all syllables are open, and no consonant clusters occur. If they were regarded as a segment sequence, prenasalized stops in initial and medial positions would constitute the only clusters, and in medial position, their nasal portion would constitute the only syllable-closing consonants in the language. In short, there is nothing to suggest that prenasalized stops behave as other than single phonological segments.

Utterance-initial and word-medial prenasalized stops at bilabial, alveolar, post-alveolar and velar places of articulation were examined in the set of words shown in table 1. The release of the postalveolar stops is sometimes trilled, so the Fijian orthographic convention of distinguishing them from the alveolar ones with an /r/ is adopted.[2]

The wordlist was constructed to provide the opportunity to measure the acoustic duration of intervocalic prenasalized stops and to compare that with the duration of certain plain stops and liquids as well as to allow effects on preceding vowel duration to be measured. Plain nasals were not included because it was feared that nasalization of adjoining vowels would create problems in demarcating acoustic segments. In the event, the duration of the trill /r/ proved to be impossible to measure reliably, and nasals would have been a lesser problem. Only the vowel /a/ was used in the wordlist in order to eliminate any intrinsic vowel duration variation.

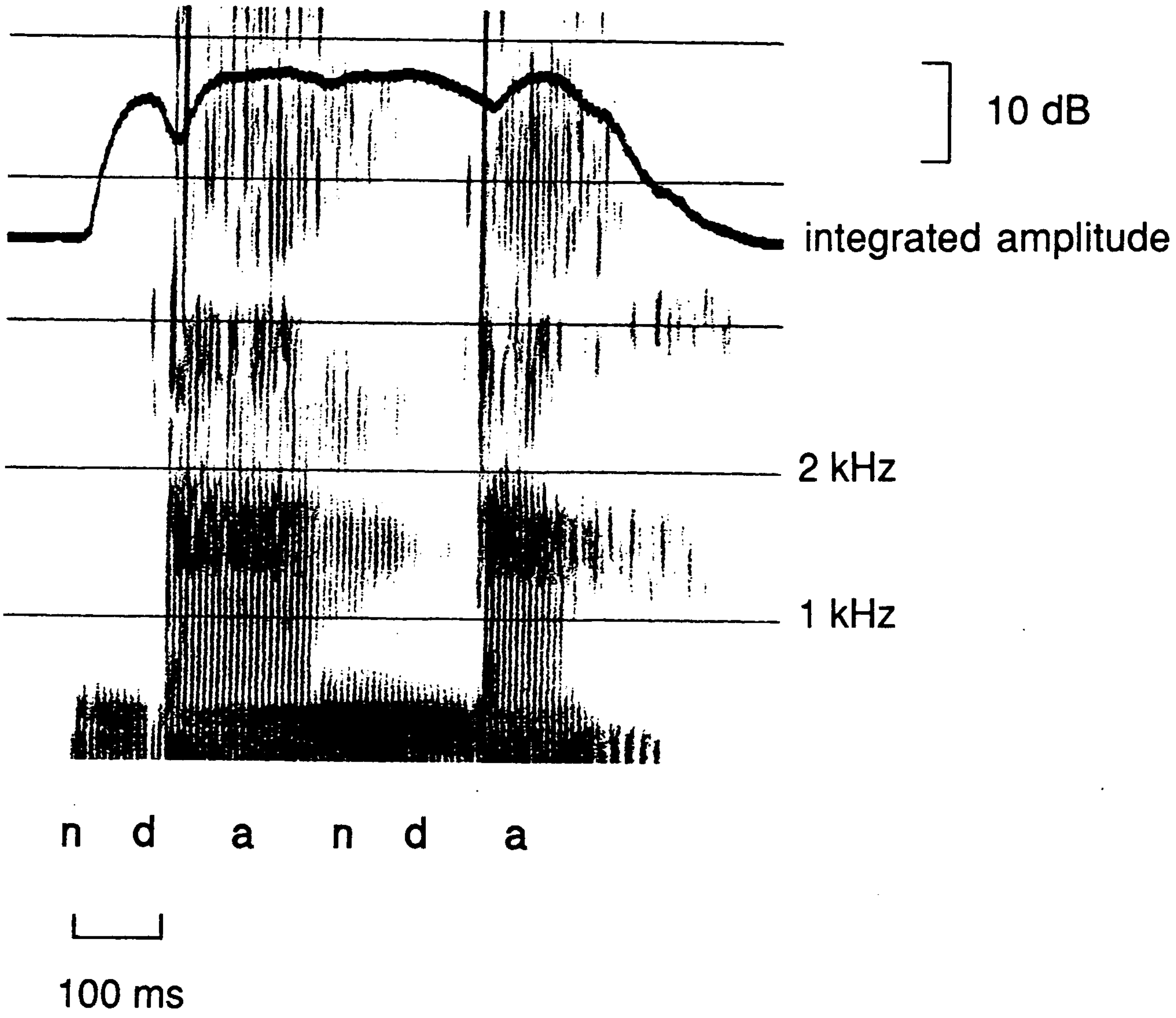


Figure 3. Broad-band spectrogram of the Fijian word /ndanda/ 'soft' spoken by a speaker of Standard Fijian. Duration and amplitude of initial and medial prenasalized stops can be compared on this figure.

Table 1. Fijian wordlist exemplifying prenasalized stops

	Initial	Prefixed	Medial	Initial & Medial
/mb/	mbaka	nambaka	kamba	mbamba
	"banyan tree"	(+ def. art)	"climb"	"slope"
/nd/	ndala	nandala	kanda	ndanda
	"dal"	(+ def. art)	"run"	"soft"
/ndr/	ndrata	mondrata	tandra	ndrandra
	"sit"	(+ 2 sg. pn.)	"dream"	*nonsense word*
/ŋg/	ŋgara	naŋgara	raŋga	ŋgaŋga
	"hole"	(+ def. art)	"show off"	"tough (of cooked fish)"

Prenasalized stops in both initial and medial positions of words were included to address another issue arising from the fact that the presence of a nasal portion in initial prenasalized stops is often not percieved by those who aren't native speakers of Fijian or another similar language. This effect was hypothesized to be due to a difference in acoustic duration and/or amplitude between prenasalized stops in these two positions, rather than to a perceptual process alone. To refine this hypothesis, the wordlist also included medial prenasalized stops derived by adding a prefix to words with initial ones so as to examine if any residual effect of the initial/medial contrast remained, as it might if stress placement was the relevant controlling factor. Note that stress falls on the first syllable of the lexical entry and does not shift forward to the prefix. Hence the prefixed medial prenasalized stops are pre-stress, like the initial ones, whereas the underlyingly medial prenasalized stops in unprefixed forms are post-stress.

The wordlist was compiled and recorded with the generous assistance of Dr. Paul Geraghty. Recordings were made in a sound-attenuated room belonging to the Fiji Broadcasting Service in the Government buildings in Suva, Fiji. Eleven speakers representing three main varieties of Fijian were recorded, with the majority (six) being speakers of what is considered the standard dialect, the others from Lau and Kandavu dialects. Each read the list twice, once across row by row and once down column by column.

Duration measurements were made on wide-band spectrograms of all the words spoken by the eleven subjects. A sample spectrogram is shown in figure 3. Acoustic duration of the prenasalized stops was measured from onset of nasal murmur to stop release (the stop component is usually very brief and not separately measurable). Durations of the other medial consonants whose acoustic boundaries could be determined were also measured, as were the durations of the initial nasals in the prefixed forms. In addition, the duration of most of the stressed vowels was measured from preceding consonant release to following consonant closure.

The resulting measurements were analyzed using an analysis of variance procedure. Means were compared using the Scheffé F-test. No significant differences were found between the dialect groups in the prenasalized stops, but dialects differ in the realization of some of the other consonants (for example, velar stops are replaced by velar fricatives in Kandavu). Hence some results are given pooled across all speakers, while others include only the Standard Fijian speakers. Durations of the medial prenasalized stops are given in table 2, together with the durations of the three other measureable intervocalic consonants, voiceless stops /t/ and /k/ and the lateral /l/. No significant difference was found between underlyingly medial prenasalized stops and those which are medial through prefixation, so measurements on these two categories have been pooled. Bilabial and alveolar prenasalized stops are significantly longer than postalveolar and velar ones, but the durations of these two groups of prenasalized stops are very comparable to the durations of the plain stops in the data, which also reflect the difference between alveolar and velar place of articulation. The alveolar lateral continuant falls within the same range of durations covered by the prenasalized stops. Hence we may conclude that the acoustic duration patterns of the prenasalized stops are very comparable to those of simple segments in Fijian.

Table 2. Mean duration of medial consonants

	mean	s.d.	n
All speakers			
/mb /	132	22.2	62
/nd /	131	23.8	63
/ndr /	114	25.5	65
/ŋg /	114	32.6	65
Standard Fijian speakers only			
/t /	125	19.5	24
/k /	116	30.1	24
/l /	117	12.3	21

Our measurements also indicate that vowel length before prenasalized stops is comparable to that before other consonants. Results are given in table 3. Overall, vowel duration remains consistently close to 140 ms in the various contexts measured. We may examine these data for effects of consonant voicing and prenasalization, place of articulation, and, to a limited degree, manner. Voicing and prenasalization are of course confounded, but vowel duration seems unaffected by whether a voiceless stop or voiced prenasalized stop follows. Mean duration of the /a/ vowel before the voiceless stop /t/ for the Standard Fijian speakers is 142 ms, while before the prenasalized stop with the same place of articulation it is 139 ms. Needless to say, these durations are not significantly different. Nor is there evidence to support a tradeoff between place-dependent consonant duration (table 2) and preceding vowel duration. Before /k/ the vowel duration of 130 ms is actually a little shorter than before /t/, even though /k/ is itself shorter than /t/. Although the vowel before /ŋg/ in /ŋgaŋga/ is a little longer than the overall mean at 148 ms, there is no general evidence that the voicing in the prenasalized stops conditions a longer preceding vowel, as the alveolar examples show. Further, before both /mb/ in /kamba/ and /ndr/ in /tandra/ mean vowel duration is 141 ms. The 18 ms difference before /k/ vs /ŋg/ therefore seems aberrant. It is possible that the shorter vowel before /k/ results accidentally from the fact that the tokens containing medial /k/ have a bilabial stop *before* the vowel. Shorter vowels occur after bilabials than alveolars in Danish (Fischer-Jørgensen 1980: 227) and perhaps in English (Halle & Stevens 1967) and other languages where bilabial closures are longer. Fijian did not show any effect of following consonant manner in the one case tested by our data: mean vowel duration before intervocalic /l/ was measured at 142 ms.

Table 3. Mean Stressed Vowel Durations before Consonants

Standard Fijian speakers only

environment		mean	s.d	n
/__ t/	(ndrata, mondrata)	142	28.0	24
/__nd/	(kanda, ndanda)	139	12.2	22
/__k/	(mbaka, nambaka)	130	22.2	24
/__ŋg/	(ŋgaŋga)	148	17.4	12
/__mb/	(kamba)	141	17.1	11
/__ndr/	(tandra)	141	15.9	12
/__l/	(ndala)	142	12.4	12

In these results we see evidence that Fijian vowels can be considered as isolated from the effects of following consonants, a view which is consistent with the idea that all the consonants we have examined, including the prenasalized stops, are single segments that belong exclusively to the syllable of which they form the onset. However, another possibility must be considered. If vowels were subject to a shortening effect when a cluster follows but also subject to a lengthening effect when a voiced segment follows, these two effects might cancel each other to produce comparable length before voiced nasal + stop clusters and voiceless single stops. These effects are usually regarded as (at least in part) compensatory adjustments, associated with the longer duration of clusters (producing shortening) and the shorter duration of voiced consonant closures (producing lengthening) (Fischer-Jørgensen 1964, Lehiste 1970, Lindblom & Rapp 1973). We note the absence in Fijian of the other type of widely reported compensatory adjustment, dependent on place of articulation of the following consonant. In view of this, we prefer to adopt the position that vowel duration is unaffected by following consonants, rather than to assume that they are affected by two separate but separately unobservable compensatory duration adjustments.[3]

As for the question of an initial/medial difference in the prenasalized stops, the initial ones do have a significantly shorter acoustic duration than the medial ones, the difference being 39 ms. The means are 95 ms for bilabial, 85 ms for alveolar, 83 ms for postalveolar and 72 ms for velar. The velar duration is significantly shorter than the bilabial ($p < .01$) and alveolar ($p < .05$) durations. Initial prenaslaized stops also have noticeably lower amplitude, as exemplified in the amplitude envelope on the spectrogram shown in Figure 3. As noted above, when underlying initial prenasalized stops are preceded by a prefix, they become indistinguishable in duration from underlyingly medial ones. Therefore it is clearly the difference in utterance position (initial vs medial) which controls the duration difference, not the position of the prenasalized stop in relation to stress (pre-stress vs post-stress). Parenthetically we might note that there is an indication that consonants initiating an unstressed syllable are shortened in Fijian: the initial nasals in unstressed prefix syllables are significantly shorter than initial prenasalized stops (/n/ = 45 ms, /m/ = 51 ms). The short acoustic duration of utterance-initial prenasalized stops is probably simply due to a delay in the onset of vocal cord vibration from a pre-speech setting, and may not reflect a briefer articulatory gesture. However, this effect surely contributes to the lesser salience of the nasal portion of prenasalized stops in initial position.

In conclusion, these results indicate that the role of prenasalized stops in the timing pattern of Fijian is generally similar to that of other single segments. They neither have longer duration themselves nor do they shorten a preceding vowel, as might be expected if they had a timing pattern similar to geminate consonants or consonant clusters. Prenasalized stops, at least in Fijian, therefore possess the timing characteristics stated by some linguists to be part of their definition. We have also shown evidence that Fijian temporal patterns appear to be strictly based on the CV syllable. Comparison with appropriate data from languages in which the phonological structure of nasal+stop elements is that of a sequence is required before conclusions can be drawn about whether, in general, timing patterns distinguish between prenasalized stops and nasal+stop clusters or demonstrate that they are alike at the level of production.

Acknowledgments

The research reported here was supported by the National Science Foundation through grants BNS 8704361 and BNS 8720098. I also gratefully acknowledge the hospitality of Dr. William Aalbersberg of the University of the South Pacific, and the assistance of Dr. Paul Geraghty of the Fijian Dictionary Project and the Fijian speakers who gave generously of their time.

Footnotes

1. On the other hand, syllabic nasals arise in KiChaka from contraction of an NV-syllable, and appear in a few loanwords, such as /mpaka/ 'boundary' (from Swahili).

2. Fijian orthography differs from our transcription in that the nasal portion of prenasalized stops is not represented; symbols that represent simple voiced stops are used except that /ŋg/ is written 'q'.

3. Unfortunately, a combination of the limitations of our data set and the restrictions on Fijian segments impede further investigation of these questions. If medial nasals, non-prenasalized voiced stops, and clusters of different types could have been examined, any effects of voicing and cluster structure could have been separated. Nasals might have been included in the wordlist, but the others are impermissible in Fijian.

References

Browman, Cathe & Louis M. Goldstein. 1986. Towards an articulatory phonology. *Phonology Yearbook* 3: 219-252.

Fischer-Jorgensen, Eli. 1964. Sound duration and place of articulation. *Zeitschrift für Sprachwissenschaft und Kommunikationsforschung* 17: 175-207.

Fischer-Jorgensen, Eli. 1980. Temporal relations in Danish tautosyllabic CV sequences with stop consonants. *Annual Report, Institute of Phonetics, University of Copenhagen* 14: 207-261.

Fourakis, Margaritis S. 1980. A phonetic study of sonorant-fricative clusters in two dialects of English. *Research in Phonetics (Department of Linguistics, Indiana University, Bloomington)* 1: 167-200g.

Geraghty, Paul A. 1983. *A History of the Fijian Languages* (Oceanic Linguistics Special Publication 19). Honolulu: University of Hawaii Press.

Halle, Morris & Kenneth N. Stevens. 1967. On the mechanism of glottal vibration for vowels and consonants. *Quarterly Progress Report, Research Laboratory of Electronics, MIT* 85: 267-270.

Herbert, Robert K. 1986. *Language Universals, Markedness Theory and Natural Phonetic Processes.* Berlin: Mouton de Gruyter.

Klatt, Dennis H. 1979. Synthesis by rule of segmental durations in English sentences. In *Frontiers of Speech Communicaation Research,* ed. B. Lindblom & S. Öhman. New York: Academic Press.

Ladefoged, Peter & Ian Maddieson. 1986. (Some of) The Sounds of the World's Languages. *UCLA Working Papers in Phonetics* 64.

Lehiste, Ilse. 1970. *Suprasegmentals.* Cambridge, Mass: MIT Press.

Lindblom, Björn & Karin Rapp. 1973. Some Temporal Regularities of Spoken Swedish. *Papers from the Institute of Linguistics, University of Stockholm* 21.

Maddieson, Ian. 1985. Phonetic cues to syllabification. In *Phonetic Linguistics,* ed. V.A. Fromkin. New York: Academic Press.

Nurse, Derek. 1981. Chaga/Taita. In *Studies in the Classification of Eastern Bantu Languages*, Thomas H. Hinnebusch, Derek Nurse & Martin Mould. Hamburg: Buske.

Sagey, Elizabeth C. 1986. *The representation of features and relations in nonlinear phonology.* Ph. D. dissertation. M.I.T., Cambridge, Mass.

Vatikiotis-Bateson, Eric. 1984. The temporal effects of homorganic medial nasal clusters. *Research in Phonetics* (Indiana University, Bloomington) 4:197-233.

Acoustic characteristics of tongue root position and vowel assimilation in Akan

Susan Hess

Paper presented at the 115th meeting of the Acoustical Society of America

The purpose of this research was twofold: 1) to find an acoustic measure that can be used to diagnose tongue root position in a language where a difference of tongue root position is systematic and significant, and 2) to apply this measure to cases of assimilation between vowels of conflicting tongue root positions in order to determine whether tongue root positions were affected.

Akan, a Kwa language spoken in Ghana, exhibits a form of vowel harmony in which the vowels in a particular word must be spoken with either an advanced or a retracted tongue root; hereafter referred to as either + or - ATR (for Advanced Tongue Root). In this paper we will look at data from Kwawu, an Akan dialect, which has the following vowels: [i,e, u, o] are [+ATR] vowels, while [ι, ε, ɷ, ɔ, a] are [-ATR] vowels. All [-ATR] vowels have variants when they occur in a syllable which precedes another word containing [+ATR] vowels. Two claims have been made about these variants. Dolphyne says that the unadvanced vowels preceding advanced vowels are replaced by the corresponding advanced vowel - thus /ε/ would be replaced by /e/. The essential change would thus be one of tongue root position. On the other hand, Clements argues that these [-ATR] vowels are only raised, /ε/ becoming more similar to /ι /. Assimilation thus would not involve the tongue root but merely raising of the tongue blade and front of the body.

Resolving this dispute required that we first define a measure which would provide a reliable diagnosis of tongue root position. Only acoustic measures were considered as phsiological data could not be obtained. Five kinds of measures were examined: formant frequency, formant bandwidth, duration, pitch, and relative amplitude.

Data were obtained from one male speaker of the Kwawu dialect of Akan, from Obomeng, Ghana. The word list was selected so as to obtain as nearly as possible a uniform set of words displaying a VCV syllabic shape with a common consonant and with the same tone.

1. Wordlist

Gloss	Word	Vowel	Tone
1. adversary	àsí	i	H
2. under, father	àsɩ́	ɩ	H
3. a yell	òsé	e	H
4. tiny shrimp	sɛ́sɛ	ɛ	H
5. your measure-of-gold	wɷ́tækú	æ	H
6. a war	ɔ̀sá	a	H
7. a fine person	òsó	o	H
8. a hole	àsɔ́	ɔ	H
9a. a species	èsú	u	H
9b. yam	òsú	u	H
10. the top	ɛ̀sɷ́	ɷ	H

Formants

Halle and Stevens (1969: 211) report that "the clearest and most consistent acoustic consequence of widening the vocal tract in the vicinity of the tongue root is a lowering of the first-formant frequency." Lindau (1975, 1979) found this to be true of vowel harmony pairs, that is vowels with roughly the same blade height which differ only in ATR value, e.g. i - ɩ, e - ɛ. However, she also noticed that the high [-ATR] vowels had very similar formant values to mid [+ATR] vowels. On the other hand, Jacobson (1978) found that the vowel harmony pairs in the Western Nilotic language DhoLuo did not differ much in their first formant value.

Turning to our Kwawu data, we find a situation similar to what Lindau found in another Akan dialect, Akyem. Formant means of four tokens of each vowel measured from wide-band spectrograms are given in (2).

2. Formant means (n=4)

vowel	F1	F2	F3	F4
i	260	2141	2678	3565
ι	309	1785	2438	3640
e	311	1875	2498	3605
ε	458	1564	2430	4000
æ	525	1570	2053	3523
a	630	1319	2137	3460
u	306	780	2225	----
ω	410	1198	2195	3170
o	423	1293	2455	3360
ɔ	524	1108	2065	3415

Formant plots of these four tokens of each vowel are shown in Figures 1 and 2. In Figure 1, a plot of F2 vs. F1, vowel harmony pairs are distinguished by their first formant frequency, e.g. i ι, e ε, but it is immediately noticeable that the formant frequencies for ι and for e overlap in space, as, to a lesser extent, do the frequencies for ω and o. Figure 2, which plots F2' (a weighted mean of F3 and F2) vs. F1, shows that the addition of information on F3 does not disambiguate these pairs of vowels. If raised ε and ɔ reach normal formant values for ι/e and ω/o, the formant values alone will not tell us whether partial assimilation of tongue blade height (ε > ι) or total assimilation, including tongue root position (ε > e), has taken place.

Vowel duration, consonant duration, and VOT

Next, we will consider various duration measures. Vowel duration measures for 10 tokens each of i, ι, e, ε are given in (3). Averages of vowel duration show that there is a tendency for [-ATR] vowels to be shorter than [+ATR] vowels. However, the duration ranges for the vowels overlap considerably, and it is clear that we would not be able to predict the [ATR] value of an unknown vowel from its duration. An analysis of variance of ι and e tokens shows that their duration difference does not reach the .01 significance level (p = .0392).

3.Vowel duration: i, ɩ, e, ɛ (n = 10)

gloss	word	V duration range		mean
<u>+ATR:</u>				
adversary	àsí	100-140		114
a yell	òsé	111-129		119
<u>-ATR:</u>				
father	àsɩ́	98-112	n=3	105
under		89-123	n=6	
shrimp	sɛ́sɛ	96-124	(second	106
		syllable)		

Consonant-related measures are given in (4). Measurements of the duration of s before ɩ and e are virtually the same. VOT measurements of p before ɩ and e are potentially more promising, but it was not possible to find a suitable set of words for all of the vowels after any stop consonant.

4. Consonant duration/VOT

gloss	word	duration (ms.) (s or VOT)	no. of tokens
father	àsɩ́	152.5	4
a yell	òsé	153	4
he spits	ɔ̀pɩ́	30	2
cat (nickname)	òpé	20	2

Pitch

Pitch shows little difference between +ATR and -ATR vowels. Means of pitch measurements of four tokens each of high-toned ɩ and e are shown in (5). From these measurements, it seemed unlikely that more tokens would show pitch to be a useful measure in identifying tongue root position, and thus it was not pursued.

5. Pitch measurements (means of four tokens in Hz)

vowel	F0 onset	F0 offset
ι	177	156
e	177	160

Spectral balance

Two measures of spectral balance (the relative amplitudes of different components of the spectrum) were checked to see if there were any damping effects or voice quality effects associated with the different tongue root positions. In both cases, amplitude measurements were based on spectrographic power spectra of vowels taken at a point one-third of the way through the vowel. The difference in amplitude between the first and second formants (F1-F2) was measured to get some indication of overall spectral balance, but was found to be random and inconsistent (See 6: a second value indicates a frame taken 10 ms. after the first to check consistency). The difference in amplitude between the first and second harmonics (H2-H1) was measured to check for any signs of breathiness, as literature on other African ATR harmony systems suggests that breathiness may be associated with advanced tongue root vowels. With breathy voice the vocal cords are vibrating more loosely, and we would expect to find a markedly decreased fundamental in relation to the second harmonic than with modal voice, as described by Kirk et. al. 1984 for Jalapa Mazatec. Kwawu [+ATR] vowels did not appear perceptibly breathy, and in fact no correlation with tongue root emerges from the measurements shown in (7).

6. F1-F2: individual measures (in dB)

i	2.3	ι	3.8	e	5.5
	5.5		6.8, 2.5		4.0, 4.2
	8.4, 3.0		4.1, 2.5		1.0
	6.0, 0.2		4.8		3.0

7. H2-H1 means of four tokens of each vowel (in dB)

i	-2.54
ι	.8
e	4.1
ε	-1.17
æ	-1.0
a	-1.74

Formant bandwidth

The last measure to be examined, formant bandwidth, proved to be more useful. This was first examined using the same power spectra generated for the amplitude measures and measuring bandwidths of the first three formants at -6dB from the peak amplitude within each formant. The bandwidth of the first formant proved to be the easiest to measure and moreover showed consistent differences between high and mid vowels of the different harmony sets (i.e. ι e, ω o) for these tokens. The difference in bandwidths in the different harmony sets was correlated with the strength of the third harmonic: [+ATR] vowels displayed a weaker third harmonic and a narrower F1 bandwidth, while [-ATR] vowels displayed a strong third harmonic and a broader F1 bandwidth. Figure 3 shows power spectra from one token each of ι and e. Preliminary studies (Hess 1987) attributed bandwidth differences to the relative strength of the third harmonic. However, with the vowel assimilation cases, to be discussed in the next section, it was found that the relative strength of the third harmonic was not an adequate predictor of bandwidth.

The bandwidth differences found in spectra were pursued with a computer measure of bandwidth which allowed bandwidth to be measured at different points throughout the course of the vowel more easily. Here I will only report on results for 14 tokens of ι and 12 tokens of e, the vowels relevant to the assimilation issue. These tokens were sampled at 20 KHz using the CSpeech program (Milenkovic 1987) on an IBM XT and LPC analyzed. The bandwidths and frequencies were obtained by extracting the roots of the resulting z-polynomials. The LPC order of analysis used was

18, as this was the lowest number which gave good LPC spectra.

8. F1 frequency and bandwidth

Vowel	F1 frequency (mean)	F1 Bandwidth (mean)	
ɩ	411.5	74.7	n = 14
max	*433.7*	*99*	
min	*382.3*	*45*	
e	366.7	52.9	n = 12
max	*414.7*	*68.3*	
min	*333.7*	*32*	
	p = .0001	p = .0003	

The results are somewhat surprising. LPC analysis of more tokens of this one pair of vowels revealed a more consistent difference in the frequency of the first formant. A one-factor ANOVA shows that differences in both F1 and bandwidth are highly significant. However, frequency and bandwidth are only mildly correlated (r = .663), allowing us to conclude that differences in the location of the first formant are not causing the differences in bandwidth. It is important to notice that the ranges of both F1 frequency and bandwidth values for the pair of vowels overlap to a great extent, with the F1 overlap being somewhat greater than the bandwidth overlap. We would thus not want to depend solely on one measure. Figure 4 shows a scattergram of the data summarized in (8).

One hypothesis for this difference in bandwidth is the presence of a glottal zero in advanced tongue root vowels. X-rays (Lindau 1975, 1979) of other dialects of Akan show that advanced vowels have a much lower larynx position than unadvanced vowels. This lower position would allow laxer vocal cords (but not lax enough for breathy voice) and thus a greater possibility of sub-glottal coupling.

Above we have looked at five types of measures with an eye to distinguishing vowels of different harmony sets. We have found that pitch and relative amplitude measures either show no real difference between the targeted vowels or do not show a consistent difference that could be linked to articulatory differences. VOT and vowel duration measurements display tendencies that differentiate [+ATR] and [-ATR] vowels, but as ranges of the different vowels overlap to a large degree, these measurements are not suited for use as discriminatory criteria. Measurements of the first formant and the bandwidth of the first formant are most strongly correlated with vowel type, but neither is sufficient by itself.

Vowel assimilation

We will now use this finding to look at data that address the second question posed at the beginning, namely, whether assimilation affects vowel height or ATR value. I investigated the effects of assimilation on the mid -ATR vowel [ɛ]. If assimilation involves a change in tongue root position, we would expect to find similar formant values and bandwidths to the vowel [e], while if it only involves the blade, we would expect to find characteristics similar to [ι]. The sentences used are given in (9). 9a, c, e contain contexts for assimilation (high +ATR vowels) and b,d,f are controls ([a]).

9. Assimilation data

a. Yaw stands beans under the table	Yàw dὶ asɛ́ sí pʊ́n nʊ́ àsί
b. Yaw spreads beans under the table	Yàw dὶ asɛ́ sàm pʊ́n nʊ́ àsί
c. Yaw takes beans' insides to Kumase	Yàw dὶ asɛ́mu kɔ́ kùmásι
d. Yaw gives beans to Afua	Yàw dὶ asɛ́ ma afúá
e. Yaw takes certain beans to Kumase	Yàw dὶ asɛ́bi kɔ́ kùmásι
f. Yaw brings beans to Kumase	Yàw dὶ asɛ́ bá kùmásι

(10) shows formant frequencies and spectra for the ɛ vowel in these sentences. For comparison, formant averages of ι and e from the word list (1) are shown at the bottom.

10. Formant measures of ɛ assimilation examples (from wide-band spectrograms)

frame-token	syllable	F1	F2	F3	F4
a. asɛ si -1	sɛ	340	1590	2300	3670
asɛ si -2	sɛ	410	1710	2520	3750
b. asɛ sam -1	sɛ	410	1510	2300	3690
c. asɛmu -1	sɛ	350	1750	2450	3360
asɛmu -2	sɛ	375	1725	2390	3560
d. asɛ ma -2	sɛ	470	1550	2450	3500
e. asɛbi -1	sɛ	375	1880	2550	3540
asɛbi -2	sɛ	350	1890	2540	3525
f. asɛ ba -1	sɛ	480	1550	2420	3550
asɛ ba -2	sɛ	460	1620	2470	3560
mean before i, u		*367*	*1758*	*2458*	*3568*
mean before a		*455*	*1558*	*2410*	*3575*
	ι	309	1785	2438	3640
	e	311	1875	2498	3605

Mean F1 frequency for ɛ before i, u is lowered but does not get as low as the F1 averages for ι and e. The F2 average for ɛ before i, u is more similar to ι than to e. Formant measures of assimilated [ɛ] show that assimilaiton is not complete. The downward movement in F1 suggests that there is a partial assimilation of tongue blade/body height. In order to determine tongue-root position, the ɛ vowels in the sentences in (9) were subjected to computer analysis of bandwidth and frequency in the same manner as before. Results are shown in (11):

11. ε F1 frequency and bandwidth

frame	F1 freq.	s.d.	F1 band.	s.d.	no. of tokens
asɛ si	419	21.7	47.5	16.9	8
asɛ sam	499	17.3	90.8	23.5	10
asɛmu	393	10.4	43.9	12.3	7
asɛ ma	486	13.5	95.7	15.7	7
asɛbi	380	10.9	44.4	9.1	11
asɛ ba	470.3	21.2	84.5	15.4	7
mean before i, u					
	395.5	*22.1*	*45.2*	*12.3*	*26*
mean before a					
	488.4	*18.1*	*86.2*	*24.9*	*24*

LPC analysis has given us slightly higher averages for F1 frequency than those obtained through measurement of wide-band spectrograms in (10), but this is not uncommon. Interestingly, bandwidth measures reveal a consistent difference between the two sets of vowels: ε before i, u and ε before a. A scattergram of these results is shown in Figure 5. These two sets have bandwidth differences of about the same magnitude as was found for ι, e: ε before i, u has narrower bandwidth, comparable to the +ATR vowel e, and ε before a has a wider bandwidth comparable to the -ATR vowel ι. This suggests that the cases of assimilation of ε before i, u are produced much in the same way as [+ATR] vowels.

In this section we have presented evidence concerning the nature of assimilation in Akan which allows us to settle one question raised by Clements and Dolphyne. Formant measurements of ε before i,u shows us that the assimilation is a partial one in terms of vowel (tongue body) height, as F1 for these vowels does not get as low as F1 for ι or e. On the other hand, bandwidth measurements suggest that ε before i,u is produced in the same manner as [+ATR] vowels, which is in closer agreement with Dolphyne than Clements.

Acknowledgements:

I would like to thank Lloyd Rice for his help in developing a program for extracting bandwidths and both Lloyd and Ian Maddieson for invaluable discussion.

References

Clements, G. N. 1981. Akan vowel harmony: a nonlinear analysis. *Harvard Journal of Phonology* 2: 108-177.

Dolphyne, Florence. 1987. Akan phonology ms.

Halle, Morris and Kenneth Stevens. 1969. On the feature "Advanced Tongue Root". *Quarterly Progress Report* 94: 209-215. MIT Research Laboratory of Electronics.

Hess, Susan. 1987. Acoustic characteristics of the vowel harmony feature and vowel rasing in Akan. UCLA *Working Papers in Phonetics* 68: 58-72.

Jacobson, Leon. 1978. DhoLuo vowel harmony: a phonetic investigation. UCLA *Working Papers in Phonetics* 43.

Kirk, Paul, Peter Ladefoged, and Jenny Ladefoged. 1984. Using a spectrograph for measures of phonation type in a natural language. UCLA *Working Papers in Phonetics* 59: 102-113.

Lindau, Mona. 1975. Features for vowels. UCLA *Working Papers in Phonetics* 30.

Lindau, Mona. 1979. The feature expanded. *Journal of Phonetics* 7: 163-176.

Milenkovic, Paul. 1987. *CSpeech Users' Manual.* Department of Electrical and Computer Engineering, University of Wisconsin, Madison.

Universals of nasalization: Development of nasal finals in Wenling

Susan Hess

In Wenling, a Wu dialect[1] of Wenling County in southern Zhejiang Province of China, historical nasal finals[2] (that is, syllables containing a final nasal consonant, "NC") have undergone extensive change. All of the most likely outcomes are exhibited: retention of the original NC, retention but shift of place of articulation of the NC, vowel nasalization with loss of the NC, and vowel denasalization, resulting in an open oral syllable. In addition, there have been accompanying changes of vowel height and backness.

Since much interest has been focussed on phonetic universals of nasals and nasalization, some particularly stated with reference to diachronic change, it is interesting to see how a detailed study of one dialect with extensive changes in historic nasal consonants confirms or challenges posited universals. Four main areas will be examined below: relation of vowel height to nasalization, denasalization, post-vocalic NC place perception, and vowel height changes in original V-NC sequences. If the descriptions of the changes that occurred in this dialect are motivated by considerations of universality, it would dictate a particular chronology of changes. Internal and comparative evidence will be used to examine the plausibility of such a reconstruction.

Among the generalizations concerning nasals and nasalization, those proposed by Chen (1975) are specific in reference to Chinese. Among Chen's proposals, those which will be relevant to the Wenling data include: vowel nasalization has led to vowel height changes, nasal consonants tend to merge in the order m > n > ŋ, and nasalization proceeds from low vowels to high vowels and from finals with anterior nasals to those with posterior nasals. In this paper it will be argued that Chen's proposals do not receive support from the Wenling facts, although superficial inspection of the data may give the appearance of confirming them.

There is a relative scarcity of truly historical works on Chinese dialects, particularly of non-Peking dialects. A secondary goal of this paper is to provide a detailed study against which other Wu dialects can be compared.

Sources for reconstruction

There are essentially two possible starting points for reconstructing the phonological history of a Wu dialect. One is to use a reconstruction of Ancient Chinese[3]; the other to use a reconstuction of Proto-Wu (which postdates Ancient Chinese) as provided by Ballard (1969) or Chang (1971). At first glance it would appear preferable to use the later stage, i.e. Proto-Wu, as a basis. However, neither reconstruction of Proto-Wu is satisfactory, as both make assumptions about the merging of some Ancient Chinese finals and propose specific (but different) reconstructions that are not necessarily warranted.

Chang's paper presented a reconstruction of historical changes in one dialect point, Wenzhou. The starting point for his derivations was his reconstruction of Proto-Wu, which he said was formulated by selecting Ancient Chinese categories represented by contrasts in Wenzhou and in other Wu dialects. He does not discuss at all how the selection was made nor how he arrived at his phonemic (phonetic?) values for these categories.

Ballard, in his 1969 dissertation, was more interested in theoretical aspects of historical phonology and approached the task of reconstruction in a different way, namely through comparison of thirteen Wu dialect points. His reconstruction of Proto-Wu is much different from Chang's; but once again, there is insufficient discussion as to how the proposed values were determined. The bulk of both Chang's and Ballard's papers are concerned with the derivations from Proto-Wu to modern dialects, but neither link subsequent changes (post-Proto-Wu) to earlier ones which led up to Proto-Wu. Thus their reconstructions miss things which are general Wu processes.

Although there is an abundance of studies providing analyses of synchronic phonological problems, there is a paucity of published data that would aid one in doing further historical research via the comparative method in Wu dialects. There are published vocabularies for only a few dialect points, such as Leqing in Zhejiang and Shanghai and Suzhou in Jiangsu. Ballard does not indicate the sources of his data, nor does he publish data for the thirteen diapoints used in his study. Spoken languages of the Wu dialect group vary greatly from town to town and from county to county; an ideal historical study would canvass a large number of dialect points, as Chao (1928) did in his survey of Wu dialects. He does not provide a lexicon with which one could do further historical work, however.

The other possible starting point for approaching Wu historical studies is to use Ancient Chinese. One method, advocated by Paul Denlinger (ms.), is to use structural information from the Ancient Chinese system, rather than any particular reconstruction thereof, as an aid in doing historical research. This approach is useful in organizing data, but is not an alternative in any real sense. It provides categories, and some information about some of the categories, but does not provide enough interpretable information if one wishes to trace the development of a single modern dialect.

In the absence of a full body of available dialectal materials or of intermediate historical records, it has been decided to use a reconstruction of Ancient Chinese. There are many reconstructions of Ancient Chinese; most are refinements of Karlgren's reconstruction (1954). The most recent reconstruction of Ancient Chinese, Pulleyblank (1984), goes much further than merely making additional refinements. He shows the need for dividing Ancient Chinese into two smaller periods to reflect a shift in dialect base of the standard language by the end of the seventh century. As a result of this periodization, Pulleyblank re-evaluated the sources used for reconstruction and has provided a much more informative interpretation of the phonological systems of the two periods and the changes between them. It is the latter period, Late Middle Chinese (LMC, which Pulleyblank puts as roughly from the end of the seventh century to the beginning of the tenth) that is most relevant to reconstruction of modern dialects, as there is evidence that LMC phonological categories largely overlaid those of Early Middle Chinese (EMC), even in the South. In the reconstruction below, we will begin with Pulleyblank's reconstruction for LMC. This does not imply a commitment to the phonological structure in terms of the particular syllabic model that he proposes.

It will be useful to review some of the characteristics of LMC finals before proceeding. Vowels were divided into four grades of openness in rime books of the late Tang and early Song periods.[4] This information, together with Sino-xenic correspondences and modern dialectal information, has provided the basis for reconstructing vowel qualities in Chinese historical phonology. Pulleyblank's reconstruction of vowel qualities largely agrees with previous reconstructions. One innovation is his reconstruction of long vowels for Grade II rimes, which gives a contrast between long and short low vowels. A second innovation is his use of postvocalic glides [j, w, ă] to replace the second element in diphthongs. While use of [j, w] in this way causes no difficulties, it is not clear what [ă] is. [ă] (or [ʜ]) is proposed as an analogue of [j, w] at the glottis. It is used to back a preceding [a] ([aă] = [ɑ]), to indicate a final velar consonant as opposed to a palatal one, and as a means of obtaining symmetrical syllable structure, particularly by providing a closing consonant to open syllables. The existence of [ă] is perhaps debatable, but as its distribution is predictable, it actually does not create any problems in deriving the finals in Wenling.

LMC had a symmetrical series of nasals and voiceless oral syllable-final stops at four places of articulation: bilabial, alveolar, palatal, and velar. Pulleyblank reconstructs a glide before all velar finals; palatal finals are represented as sequences of the glide [j] and a velar [ŋ, k], and actual velar finals are represented as sequences of [ă] and [ŋ, k]. A complete set of correspondences of EMC/LMC forms and Wenling finals is given in the Appendix. Since we are not primarily concerned with the development of initial consonants, we will not discuss their reconstruction and correspondences here.

It should be noted that Wenling appears to be something of an isolate. It has certain features which mark it as a Wu dialect and thus it has obviously shared some of the innovations of this dialect area, but on the other hand, certain crucial developments in nasal finals in Wenling are not reflected in other dialects for which there is data. In the discussion below, data from four other Wu dialects will be used to point out which developments would appear to be true of Wu dialects in general and which appear to be specific innovations in Wenling.

Data on Wenling

Data were obtained between 1983 and 1985 from two speakers of the dialect who were living in Taipei. These speakers were chosen as being the most representative speakers available. They are:

1. SYM, male, 87 years old in 1985. Mr. Sun was born in Tangtun, Changshou xiang, in the northeastern part of Wenling County (長壽鄉 塘呑). He went to a neighboring county for high school (Linhai 臨海縣), to Nanjing for college and then fought in the Sino-Japanese War. He has been in Taiwan continuously since around 1950.
2. LC, female (wife of SYM), 74 years old in 1985. She was born in Roheng Zhen (箬橫镇) in the central part of Wenling County. She lived in Wenling continuously, except to go to Linhai county for high school, until her move to Taiwan in 1950.

Despite 35 years of residence in Taiwan, the couple continue to speak the Wenling dialect at home, and their children and grandchildren also speak it. This fact reassured me that they are still fluent in this dialect. In addition, they speak Mandarin poorly, that is, they make few adjustments in their Wenling speech in order to convert it into 'Mandarin'. This led me to believe that exposure to Mandarin had not affected their command of this dialect.

SYM and LC speak the same dialect, but there are some systematic differences in their pronunciation of words from certain rime classes. This information will be indicated in the Appendix by placing LC's reflexes in parentheses.

The data consist of approximately 1400 monosyllables recorded in isolation. Chao (1930) was followed in choosing words to represent historical categories. These monosyllables retain the same segmental form in disyllabic words.

four other Wu dialects will be used to point out which developments would appear to be true of Wu dialects in general and which appear to be specific innovations in Wenling.

Data on Wenling

Data were obtained between 1983 and 1985 from two speakers of the dialect who were living in Taipei. These speakers were chosen as being the most representative speakers available. They are:

1. SYM, male, 87 years old in 1985. Mr. Sun was born in Tangtun, Changshou xiang, in the northeastern part of Wenling County (長壽鄉 塘吞). He went to a neighboring county for high school (Linhai 臨海縣), to Nanjing for college and then fought in the Sino-Japanese War. He has been in Taiwan continuously since around 1950.
2. LC, female (wife of SYM), 74 years old in 1985. She was born in Roheng Zhen (箬橫镇) in the central part of Wenling County. She lived in Wenling continuously, except to go to Linhai county for high school, until her move to Taiwan in 1950.

Despite 35 years of residence in Taiwan, the couple continue to speak the Wenling dialect at home, and their children and grandchildren also speak it. This fact reassured me that they are still fluent in this dialect. In addition, they speak Mandarin poorly, that is, they make few adjustments in their Wenling speech in order to convert it into 'Mandarin'. This led me to believe that exposure to Mandarin had not affected their command of this dialect.

SYM and LC speak the same dialect, but there are some systematic differences in their pronunciation of words from certain rime classes. This information will be indicated in the Appendix by placing LC's reflexes in parentheses.

The data consist of approximately 1400 monosyllables recorded in isolation. Chao (1930) was followed in choosing words to represent historical categories. These monosyllables retain the same segmental form in disyllabic words.

The finals of this dialect are shown in (1) below. Major allophonic variants are placed in square brackets. [ɯ] occurs only after velars, [ə] elsewhere. The medial [y] occurs before round vowels, [i] before unrounded vowels.

1. Open finals:		i	y	e	ø	ɛ	ə ~ [ɯ]	a	ɔ	o	u
				ie	iø	iɛ	iə	ia	iɔ		iu
					=[yø]				=[yɔ]		
				ue	uø	uɛ		ua		uo	
Finals with nasal endings/ nasalization			yn		øn		ən	ã	ɔ̃	iŋ	oŋ
								iã	iɔ̃		ioŋ
									=[yɔ̃]		=[yoŋ]
							uən	uã	uɔ̃		uoŋ
	m̩	n̩	ŋ̍								
Finals ending with a glottal stop:		iʔ	yʔ		øʔ	ɛʔ	əʔ	aʔ	ɔʔ		
						iɛʔ		iaʔ	iɔʔ		
									=[yɔʔ]		
						uɛʔ		uaʔ	uɔʔ		

Finals in Wenling which derive from LMC nasal finals include all current reflexes with a final nasal consonant or nasalization, and certain open finals. Among the open finals above, those with [-ɛ] and [-ø] derive solely from original nasal finals, [-ie] derives both from original open finals and nasal finals. Finals with glottal stops derive from LMC finals ending in voiceless oral stops. As LMC nasal and checked finals formed two parallel series (parallel in vowel quality and place of final consonant), Wenling finals with glottal stops will be used to check whether processes were unique to nasal finals and related to nasality or not.

Development of nasal finals

Nasal finals (together with their paired checked finals) comprised nine LMC rime groups. In Wenling, the final voiceless oral stops have become glottal stops in all cases, but a final nasal consonant is retained in reflexes of only five of the nasal-final rime groups (although the place of articulation may not be the same). As for the other four rime groups, two have yielded oral open syllables in Wenling (i.e. those with [-ɛ, -ø] and some [-ie]) , and two have resulted in nasalized vowels ([-ã, -ɔ̃]). A number of questions present themselves: When did nasal consonants delete? Was there a single process of nasalization in these four rime groups which now lack NC's, followed by denasalization for two of them? Is there in fact any evidence for a transfer of nasalization to the vowel in the two rime groups which have open oral syllables today? When and why did the place of articulation of retained NC's shift? Did any of these changes interact with the vowel raising processes that will be discussed below? Are there articulatory or acoustic bases for any of these changes?

These questions will be addressed in quasi-chronological order. First, however, we will consider Chen's (1975) account of the general order of change of nasal finals in Chinese dialects. Chen's data consisted of a core of eighteen dialects^5, plus information on 1364 dialects from published surveys and personal sources. The published surveys are uneven in depth of treatment; many are concerned mainly with providing inventories of sounds and provide little in the way of a lexicon that can be used for checking correspondences. Chen began by looking at the distribution and the frequency of occurrence of nasalized vowels in synchronic systems, and concluded that "A basic pattern regarding the distribution of NV [nasalized vowels] emerges with unmistakable clarity: the decreasing order of frequency is low-mid-high and front-central-back" (p 88). A closer look at his figures suggests that the basic distinction is between low and non-low, as there is comparatively little difference in the occurrence of mid and high nasalized vowels. Chen gives the percentages of nasalized vowels occurring within the three heights low, mid, and high as 24%, 12% and 10% respectively. The latter two percentages are composed almost entirely of front vowels (Chen 1975: 89-93). Thus, in his tables, it is unrounded low vowels (front and back) which most frequently occur as nasalized vowels, followed by unrounded non-low front vowels. It is also striking that certain round vowels, y, ʉ, ø, ɔ (but not o) and the fricative vowels, , , rarely occur nasalized. It is not surprising that the fricative vowels are rarely nasalized, as they generally descend from LMC non-nasal finals, but this is not true of the round vowels above.

.

Diachronically, Chen assumes that in general the nine Middle Chinese[6] (MC) nasal finals have merged into five pre-modern (PM) finals (p 97), as shown in (2). It should be noted that the two PM finals containing /ə/ include cases such as in, yn, iŋ, etc, which Chen analyses as a medial + /ə/ + a nasal. In other analyses no ə is reconstructed and in phonetic fact these finals appear to have been pure high vowels.

2.

```
MC XIAN  SHAN   DANG  JIANG   GENG  ZENG   SHEN  ZHEN   TONG
    am    an     aŋ    ɔŋ      aɲ    əŋ     əm    ən     uŋ
     \    /      \     /       \     /      \     /      |
      \  /        \   /         \   /        \   /       |
PM    an            aŋ            əŋ           ən        uŋ
```

He then determines the order of nasal loss among the PM finals, that is the order of finals in terms of their likelihood of yielding nasalized vowels, to be: /an, $\{^{\text{aŋ}}/_{\text{ən}}\}$, əŋ, uŋ/ in decreasing order of frequency. Given this, he says that "the scope of nasalization is defined by two parameters: from low to high vowels, on the one hand, and from the anterior (-n) to the posterior (-ŋ) nasal endings on the other" (p 97-98). One question that must be addressed is whether his generalization is based on an accurate reconstruction of pre-modern finals. The data which are supposed to support this ordering do not do so conclusively, and suggest that either the reconstruction or the generalization requires refinement. His data (p 100) are given below in (3). The first three columns represent Chen's data, the fourth indicates the order in which these finals underwent nasal consonant loss as determined by Chen. Chen considers that there are alternate orderings between the second and third stages of nasalization ($\{^{\text{aŋ}}/_{\text{ən}}\}$), one dependant on vowel height and one dependent on nasal place. There are somewhat greater numbers of nasalized vowels for PM *aŋ, so Chen considers this to be the marginally preferable order. The last column shows another ordering suggested by the data (number of diapoints yielding NV's):

3.

PM	MC	# of diapoints yielding NV's	Chen's order	revised rank
an	am	412	1	1
	an	414	1	1
aŋ	aŋ	243	2; 3	2
	ɔŋ	189		3
ən	əm	190	3; 2	3
	ən	193		3
əŋ	aɲ	171	4	3
	əŋ	134		4
uŋ	uŋ	24	5	5

The different frequencies of nasalized vowels within PM finals aŋ and əŋ, traceable to earlier distinct finals which he believes to have merged (MC DANG/JIANG and GENG/ZENG respectively), suggest that there are numbers of diapoints for which this set of mergers is not accurate[7]. Given that four out of the eight dialect groups covered are subgroupings of Mandarin, there is also some question as to the representativeness or bias of the sample. Certain language areas may have shared a particular sound change, which in turn may influence the frequency of a certain nasalized vowel. The particular set of mergers proposed is certainly plausible for Mandarin dialects such as Pekingese, but is overly simplified with respect to dialects from other major dialect groups, which are under-represented in the sample. There are 718 Mandarin diapoints and only 376 diapoints from the remaining six major non-Mandarin dialect groups.

Above we have mentioned two of Chen's hypotheses: nasalization proceeds from low to mid to high vowels, and from the anterior /n/ to the posterior /ŋ/ nasal finals. In addition, he states that nasalization of vowels may cause them to lower or to raise, but with respect to the Wu dialects, he specifically mentions that nasalization has caused finals from the XIAN and SHAN groups to raise; and he claims that nasalized vowels have a certain lifespan - the earlier a particular nasalized vowel enters the system, the earlier it disappears; and that otherwise, denasalization proceeds from high vowels to low vowels. These claims will be discussed when we are considering the relevant data in Wenling.

Retention of nasal finals/shift of nasal place of articulation

We will first consider why nasal consonants should have been retained in some instances but not in others. As noted above, Wenling reflexes retain final nasal consonants from only five of the nine LMC rime groups containing nasal finals. In (4), a fuller set of correspondences is provided on the left for those which retained a nasal consonant, and a few example syllables of those which lost a nasal consonant appear on the right. (If there is more than one reflex, the most numerous reflex is given first):

4.

	Wenling	LMC	Rime Group
1	iŋ	im	SHEN
2	iŋ/søn	əm	"
3	kən/tøn	ən	ZHEN
4	øn/kuən	un	"
5	iŋ	in	"
6	yn	yn	"
7	ən/boŋ	əăŋ	ZENG
8	iŋ	iăŋ	"
9	iŋ	iajŋ	GENG
10	ioŋ	yajŋ	"
11	oŋ/kuoŋ	əwŋ	TONG
12	ioŋ	iwŋ	"
13	ioŋ	yawŋ	"

	Wenling	LMC	Rime Group
14	ɔ̃	aăŋ	DANG
15	ɔ̃	aawŋ	JIANG
16	ã	aajŋ	GENG

If we compare the forms on the left (1-13) with those on the right (14-16), we find a vowel height difference between the two in both the LMC forms and the Wenling reflexes. Nasal consonants appear to be retained after historical non-low vowels. The three apparent counter-examples, numbers 9, 10 and 13, are explained by rules assimilating short /a/ to a preceding high front vowel. (ă in 7 and 8 is not a low vowel but indicates that the following consonants are velars.) These forms are then consistent with Chen's hypothesis that nasalization proceeds from low to high vowels, which will be discussed in more detail in the following section.

However, the Wenling forms in (4) are not consistent with Chen's generalization that GENG and ZENG merge into PM əŋ and SHEN and ZHEN merge into PM ən, nor with his hypothesis that m merges with n, and n then merges with ŋ. In the process of reduction of these four rime groups, there are partial mergers of these finals (e.g. 1 and 5, 8 and 9), but not in the predicted directions, as there are either vowel quality or consonant differences from what Chen posits. The most serious counter-example is 7, where ŋ > n in the majority of the reflexes.

In other dialects, Chen derives the homogeneous vowel quality of PM ən and əŋ finals by analysing non-low vowels as sequences of glides + ə. For example, [in] would be analysed as /iən/. If we look back at the Wenling finals in (1), we find that place of articulation of a final nasal consonant is predictable from the main vowel. Decomposing iŋ into iəŋ and yn into yən yields forms with a common nuclear vowel (and one could further write rules deriving ŋ from n after iə), but this analysis would mean that place of the final nasal is somehow related to the glide, which Chen does not represent either, rather than to the main vowel. Chen's use of ə for the main vowel in these finals thus obscures rather than simplifies. In the instances where place of articulation of the final nasal has shifted, it is striking that we find ŋ after i and n after ə.

In Chao (1928), we find that 21 out of 33 Wu dialects also have -iŋ reflexes where Wenling does. (In four others, -in and -iŋ alternate, six have

-ɩn, one has -ɩ̃, and one shows no consistent pattern.) The shift of -im and -in to -iŋ could be a shared Wu innovation, but one also finds it in distant dialects, such as some Southwest Mandarin dialects in Yunnan. One potential explanation for this change comes from Zee (1979, 1981). He investigated the effect of vowel quality (i, e, a, o, u) on the perception of post-vocalic nasal consonants (m, n, ŋ) in noise. Two findings which are of particular relevance here are: (1) [m] frequently tends to be misidentified as [n] after [i] and [e]; and (2) [ŋ] tends to be misidentified as [n] after [i], but [n] also tends to be misidentified as [ŋ] after [i] and [e], the former tendency being somewhat stronger than the latter (1981: 39). That these errors are dependent on the vowel quality is shown by the additional finding that [m, n, ŋ] tend to be correctly identified after [a] in even the noisiest condition. In the cases where [m] was misidentified as [n], Zee suggests that perseverative coarticulation of the high front vowel with a bilabial nasal creates a constriction in the vocal tract somewhat similar to that of [n], influencing the quality of the nasal murmur (1981: 45). In the cases where [n] was misidentified as [ŋ], one could postulate that a tense [i] in particular could further reduce the size of the cavity in front of the velum, causing [n] to sound more velar-like. In fact almost all of the listeners' in > iŋ errors are from tokens by one of the two speakers used by Zee, and this speaker (LG) does have a decidedly higher F2 than the other speaker (GP) (1979: 143-4). Hence, Zee's findings suggest that there may be a perceptual basis for the shift of place of nasal consonant articulation in the development to Wenling. This would be one reason not to adopt Chen's reconstruction of PM finals with /ə/ as a nuclear vowel in these cases.

Nasalization and vowel height

Next, we will consider Wenling reflexes of historic nasal finals which yielded nasalized vowels or open oral syllables. We will first look at the relationship between nasalization and vowel height, an area that has been extensively discussed in the literature on phonetic universals. It has been found that nasalization can have a raising effect on low vowels and a lowering effect on high vowels (see for example Beddor 1983). A subset of finals in Wenling which either have nasalized vowels or may have had them at one time is shown in (5):

5.

Wenling	LMC	Rime Group
ie	am	XIAN
ɛ, iɛ	aam	"
ie	an	SHAN
ɛ, iɛ	aan	"
ɔ̃	aăŋ [ɑŋ]	DANG
ɔ̃	aawŋ	JIANG
ã	aajŋ	GENG

Although the first four Wenling forms above are not nasalized, some Wu dialects have nasalized reflexes and similar vowel qualities for these finals, prompting Chen to state: "The predominant direction in which the NV's have shifted is raising. This trend is particularly visible with low vowels followed by an anterior nasal ending among the Wu dialects. PM /an/ (from MC /am, an/), for instance, manifests itself as a low vowel in 12 out of the 33 diapoints surveyed by Chao (as /æ̃/ in 6, /æ/ in 4, /ã/, /A/ in one each), as a mid vowel in 20 (/ɛ̃/ in 1, /ɛ/ in 7, /E/ in 2, /E/ in 6, /ẽ/ in 1, /e/ in 13), and even as a high vowel /ĩ / in one case." (p 111) Chen's conclusion that nasalization causes vowel raising has been widely cited in the literature on phonetic universals of nasalization. However, we find that the observation that nasalization has caused raising is invalid with respect to Wu dialects when we consider reflexes of checked finals corresponding to these nasal finals, such as those in (6):

6.

Wenling	LMC	Rime Group
iʔ	ap	XIAN
ɛʔ, iɛʔ	aap	"
iʔ	at	SHAN
ɛʔ	aat	"
ɔʔ	aăk [ɑk]	DANG
ɔʔ	aawk	JIANG
aʔ	aajk	GENG

The vowel heights of the finals in (6) are the same as those in (5), indicating that nasalization is not responsible for the change in vowel height. These data do not invalidate the claims made with respect to the effects of nasalization on vowel height in general, but they do disqualify this dialect (and other Wu dialects) as instances of these effects.

Processes of nasalization

In this section we will consider whether there was a single process of nasalization which affected all four LMC rime groups yielding nasalized or oral vowels in Wenling, whether it occurred in stages as claimed by Chen, or whether there is in fact any evidence for supposing that there was a transfer of nasalization in the two rime groups which have open oral syllables today. A simplified but representative set of correspondences is shown in (7); checked finals corresponding to the historic nasal finals are also included in the two rightmost columns. Parentheses indicate that those reflexes only occur with a subset of initials, an example of which is provided.

7.	Wenling	LMC	Rime group	Wenling	LMC
1	ɔ̃	aăŋ [ɑŋ]	DANG	ɔʔ	aăk
2	(kuɔ̃)	waăŋ [wɑŋ]	"	(kuɔʔ)	waăk
3	iã	iaăŋ [iɑŋ]	"	iaʔ	iaăk
4	(fɔ̃)	yaăŋ [yɑŋ]	"	(vɔʔ)	yaăk
5	ɔ̃, iɔ̃	aawŋ	JIANG	ɔʔ, iɔʔ	aawk
6	ã	aajŋ	GENG	aʔ	aajk
7	(huã)	waajŋ	"	(ɦuaʔ)	waajk
8	ie	am	XIAN	iʔ	ap
9	ɛ, iɛ	aam	"	ɛʔ, iɛʔ	aap
10	ie, ɛ	iam	"	iʔ	iap
11	ie	an	SHAN	iʔ	at
12	ɛ, iɛ	aan[8]	"	ɛʔ	aat[8]
13	ie	ian	"	iʔ	iat
14	ø, uø, uɛ	uan	"	əʔ, ɛʔ, uɛʔ	uat
15	uɛ	waan	"	(kuɛʔ)	waat
16	ø, iø	yan	"	yʔ	yat

Chen has a number of hypotheses that would be relevant to this data:

1) He assumes that Middle Chinese finals have undergone a particular set of mergers, as discussed above; 2) Given the results of these mergers, nasalization proceeds from low to mid to high vowels, and from the anterior /n/ to the posterior /ŋ/ nasal finals; 3) Nasalized vowels have a certain lifespan - the earlier a nasalized vowel enters the system, the earlier it disappears; 4) Otherwise, denasalization proceeds from high vowels to low vowels. Although the last two are concerned with denasalization, they are related to the way in which Chen formulates his nasalization hypothesis, and we will consider them below as well. For instance, as there is a greater predominance of oral reflexes from PM /an/ < MC /am, an/ (XIAN/SHAN) in Chen's data than from other PM finals, (3) is based on the premise in (2) that PM /an/ nasalized earlier.

With respect to the first of these, it was suggested earlier that there may be a problem in assuming that there was a particular set of mergers of LMC/MC finals relevant to all dialects (see (2)). This is illustrated in the Wenling reflexes in (7) above. Reflexes from the DANG rime group are divided into two distinct groups (1, 2, 4 vs. 3), only one of which has merged with JIANG. In addition, part of the GENG rime group (other GENG reflexes are shown in (4)) has the same nuclear vowel as a portion of the DANG rime group (3), rather than with ZENG as postulated. The course of these mergers leading to the Wenling reflexes is not arbitrary, but is linked to the vowel raising process to be discussed later. From inspection of the reflexes in (7), there would also appear to be no evidence in this dialect that *am merged with *an as a result of a change m > n. Reflexes of these two groups are currently the same - oral vowels - but this could have resulted from a rule which affected both simultaneously, such as the loss of both m/n after a. This argument is perhaps clearer if we consider the LMC syllables ending in p, t, c, k. All have become glottal stops in Wenling. There is no synchronic evidence to suggest that there were stages in this replacement (p > t, t > c, c > k, k > ʔ). It is just as plausible to posit a rule whereby all were simultaneously replaced by ʔ, as happens synchronically in a number of languages, e.g. Toba Batak.

The third and fourth hypotheses suggest two courses of denasalization -

one linked to length of time and the other to vowel height. The first seems somewhat implausible at face value, as there is no evidence to suggest that languages 'know' how old a form is. However, Chen's generalization can be explained in another way. Namely, nasalization of a subset of vowel-nasal finals introduces a distinction between contextual and non-contextual nasalized vowels. Subsequent nasalization of other vowel-nasal finals potentially creates a distinction between lightly nasalized (old) and heavily nasalized (new) vowels. The lightly nasalized vowels might then be more likely to denasalize.

Chen suggests that the two denasalization processes are alternatives to each other. If both were equally relevant, we would predict that in a dialect where extensive nasalization had taken place, nasalized mid vowels would be the most common: the low vowels nasalize first, and should thus denasalize early; the high vowels nasalize last but would be the first to undergo vowel-height-sensitive denasalization. Chen's data do not show any such cases, which is in itself curious and worthy of further investigation. Chen's data also do not show which is the preferred method of denasalization.

The hypothesis linking nasalization to vowel height receives independent support in the phonetic literature. Studies by Moll (1962), Ohala (1975), Clumeck (1975) and others show that mid and low vowels are more likely to be articulated with a somewhat lowered velum, though they differ on the explanation for this fact. Moll (1962: 34) has suggested that it is due to the muscle connections between the tongue and the soft palate. Ohala (1975:299) doubts Moll's suggestion since a number of electromyographic studies (e.g. Lubker 1968) show that the muscles which control the elevation of the velum actively control the variations in velic elevation for the various vowels. Alternatively, it has also been suggested that velic lowering during the articulation of low vowels occurs because nasalization causes less of an acoustic disruption on low vowels and thus that low vowels tolerate nasalization better. As Ohala points out (1975: 299), this by itself is not a reason for lowering the velum when low vowels are articulated. In vowel-nasal consonant sequences, there is of necessity some lowering of the velum during the vowel in anticipation of the nasal consonant. Ohala postulates that in some languages, such as American English, earlier velic lowering with low vowels may originally have been accidental but tolerated, and later became a necessary part of the contrast. Of course this hypothesis, like any other in historical linguistics, does not allow predictions about the direction of change in any language or dialect.

The second of Chen's claims concerning the nasalization process, namely that nasalization proceeds from finals with anterior nasal consonants (e.g. -n) to those with posterior nasal consonants (e.g. -ŋ) does not seem to have outside support in the literature. On the contrary, Ohala postulates that alternations between velar nasals and nasalized vowels should be more common than alternations of other nasal consonants with nasalized vowels. This is because velar nasals have negligible side cavities and thus negligible anti-resonances, which should make the velar nasal more like a nasalized vowel (1975: 296). There are languages which synchronically exhibit alternations of the type Ohala predicts, e.g. in Akan, words ending in -m are stable while those ending in a velar nasal vary with nasalized vowels. Chen's hypothesis is based on inferences from his survey while Ohala's hypothesis is based on the acoustic nature of different nasals. Unfortunately, there is no experimental data which would show which is more likely.[9] We will consider both of these with respect to the Wenling data below.

There are three plausible basic hypotheses (plus variations on the three) about the course of nasalization which could account for the Wenling data. The first of these is that there could have been a single process of nasalization which either affected all low vowels ([4 high] on a multi-valued height scale) or all low and low mid vowels ([4, 3 high]) depending on the way the relationship to vowel raising is viewed. This hypothesis asserts that place of the nasal is irrelevant to the likelihood of nasalization, and that a low vowel height is the main factor. The virtue of this approach is that it allows unitary expression of the nasalization process. (In order to get the fronted and backed raised versions, i.e., ie, ɛ and ɔ̃, we would have to assume that /a/ was contextually fronted ([æ]) and backed ([ɑ]) according to the place of the nasal consonant before it was lost.) The first variation (low [4 high] vowels nasalized) would create a rather odd vowel system which had frontness, length and nasalization contrasts on low vowels, but mainly frontness contrasts at other heights (and possibly a long and short [ĩ]). One difficulty with both variations would be motivating the denasalization process, as there would appear to be no reason why ɛ should denasalize and ɔ̃ should remain nasalized.

A second possibility is that -m and -n deleted after the low fronted vowel [æ] without transferring nasalization, and only vowels before a velar nasal became nasalized. This would resolve the denasalization question by rendering it unnecessary. It is also consistent with Ohala's hypothesis concerning the greater likelihood of velar nasal finals giving rise to nasalized vowels.

Finally, the third possibility is the one suggested by Chen; namely following mergers of MC finals into PM finals, *an (< *am, *an) nasalized first, followed by *aŋ (< *aŋ, *ɔŋ). Following raising, the nasalization on finals derived from *an (< *am, *an) deleted because it had nasalized earlier.

The key to this choice lies in the extensive raising processes affecting most finals. These processes will be examined next.

Raising of low vowels

In the course of development of Wenling from LMC, raising of low vowels was a widespread process - far more extensive than in Pekingese for example - yet there were some finals which did not undergo raising. In Wenling, most back *[ɑ]'s (aă), both long and short, in open syllables or before velar finals, have been raised to mid back (rounded) vowels ([o, ɔ]). Front *[a]'s before [w] also raise to [ɔ]. Before -j and bilabial and alveolar consonants, most other front *[a]'s have become mid or high front vowels. The relevant correspondences are given in (8). Parentheses enclosing items in the reflex column indicate that these reflexes only occur with a subset of initials, an example of which is provided.

8.	reflex	LMC	rime group
a. aǎ back raising			
1	o	aǎ	GUO
2	o	waǎ	"
3	o	(j)aaǎ	JIA
4	o	waaǎ	"
5	ɔ	(u)aw	XIAO
6	ɔ	(j)aaw	"
7	iɔ	(j)iaw	"
8	ɔ̃	aǎŋ	DANG
9	ɔʔ	aǎk	"
10	(kuɔ̃)	waǎŋ	"
11	(kuɔʔ, ɔʔ)	waǎk	"
12	(ɔ̃)	aaǎŋ	"
13	(fɔ̃)	yaǎŋ	"
14	(fɔʔ)	yaǎk	"
15	(i)ɔ̃	aawŋ	JIANG
16	(i)ɔʔ	aawk	"
17	(ɕioŋ)	yajŋ	GENG
18	ioŋ	yawŋ	TONG
19	iɔʔ	yawk	"
a retention			
20	ia	iaǎ	JIA
21	a	aaj	XIE
22	a	(j)aaj	"
23	(kua)	waaj	"
24	ã	aajŋ	GENG
25	aʔ	aajk	"
26	(ɦuã, uoŋ)	waajŋ	"
27	(ɦuaʔ)	waajk	"
28	iã	iaǎŋ	DANG
29	iaʔ	iaǎk	"

	reflex	LMC	rime group
a front raising			
30	e	aj	XIE
31	e	uaj	XIE
32	ɛ/iɛ	aam,aan	XIAN/SHAN
33	ɛʔ/iɛʔ	aap, aat	"
34	uɛ	waan	SHAN
35	uɛʔ	waat	"
36	øn/ie	am,an	XIAN/SHAN
37	øʔ/əʔ	ap, at	"
38	ie	iam, ian	"
39	iʔ	iap, iat	"
40	iø	yan	SHAN
41	yʔ	yat	"
42	ø	uan	"

Earlier it was suggested, but not demonstrated, that LMC /a/ became fronted or backed according to place of final consonant articulation. Reflexes in (8) show that /a/ was fronted before /m, n, j, p, t / and that elsewhere, i.e. before /ă, w, jŋ, ăŋ, jk, ăk /, it was backed. (If we chose not to adopt Pulleyblank's ă and his use of jŋ/jk for ɲ/c, /a/ would be backed in open syllables and before /w, ɲ, ŋ, c, k /.)

Next, we will try to find an explanation as to why some low vowels raised and others did not. Three possible hypotheses are: 1) short low vowels raised and long low vowels did not; 2) vowel quality differences caused some finals to raise and others to remain low; 3) differences in endings caused some finals to raise and others to remain low. We will show that each of these hypotheses is too general to account for the developments. Instead, reference must be made to all three.

The idea for the first hypothesis comes from Pulleyblank's (1984) explanation of height differences in Southern reflexes of two pairs of rimes in two different rime groups. Northern dialects such as Pekingese show no difference in these reflexes, as shown in (9):

9.

EMC	LMC	Peking
əj	aj	ai
aj	aj	ai
əm	am	an
am	am	an

Riming of ninth century poets confirms that [ə] in these finals had merged with the corresponding [a] finals (p 111). However, Southern dialects reflect the EMC distinction in these rimes after certain initials. Pulleyblank proposes that in these dialects, short [a] lengthened after front (i.e. [+coronal]) initials prior to the lowering of [ə] to [a], and for this reason did not undergo the same raising process. In Wenling, and in other Wu dialects such as Wenzhou, the reflexes of *aaj < *aj are the same as those of *aaj in Grade II, as the hypothesis predicts. Examples are given in (10):

10.

EMC	LMC	Wenling		Wenzhou[10]	
XIE rime group					
əj	aj	de 代	kie 該	dɦe 代	ke 概
aj	aj		kie 蓋		ke 蓋
	aaj	ta 帶		ta 帶	
ɛ^{r}j	aaj	tsa 齋	ka 界	dʑɦa 豺	ka 界
a^{r}j	aaj	za 柴	ka 解	dʑɦa 柴	ka 佳

Among the finals in (8), the reflexes of XIE (21-23) and GENG (24-27) rime groups support this hypothesis. These are reconstructed as having long vowels and they do remain low. The only other reflexes to retain a low vowel are those of the JIA (20) and DANG (28-9) rime groups, which have short vowels and would be expected to raise. The latter are a well-defined class: they are reconstructed as having a high front unrounded vowel and a back element. One could hypothesize that raising was neutralized in this environment.

On the other hand, there are many forms which are reconstructed as having long vowels but which undergo raising in Wenling. These include LMC finals: (j)aaǎ (3), waaǎ (4), jaaw (6), aaǎŋ (12), aawŋ (15), aawk (16), aam (32), aap (33), aan (32), aat (33), waan (34), waat (35). Reflexes of these forms in Pekingese retain low vowels. The length hypothesis is adequate for explaining reflexes in dialects such as Pekingese, but not for the development in Wenling. One is faced with three choices at this point: positing a shortening rule that affects only these forms, supplementing the length hypothesis with some other explanation for these forms, or discarding the length hypothesis.

A shortening rule would need to apply to the finals in the paragraph above, but not to the forms ending in *aaj (21-3) or *aajŋ (24-7). Shortening would then apply everywhere except where there was a prevocalic i or a postvocalic j [11]. A shortening rule of this sort suggests that not length but the presence of a high front unrounded element was a preservative factor. Such a factor would tie in better with the second hypothesis mentioned at the beginning of this section, namely, vowel quality.

In his discussion of the GUO rime group, Pulleyblank states, "The rounding of aǎ to ɔǎ, which definitely separated the two groups [GUO/JIA], must have taken place between the eleventh century and the thirteenth. . . . In spite of its late start in Northern Chinese, it has spread to all parts of the country and there are few traces of unrounded LMC -aǎ in any dialect. . ." (pp 106-7). It is possible that changes in vowel quality (which may also have included changes in height, as in the change from [ɑ] to [ɔ]) preceded vowel raising rules. Extending this to the development of Wenling finals, we would posit that all LMC finals with aǎ (1-4, 8-14), long or short, except those with a high front unrounded element, changed to (ɔ)ɔǎ. In addition, all low vowels which preceded w also became ɔ (5-7, 15-6, 18-9), but this need not have occurred at the same time. It is difficult to imagine what the difference between ɔǎ (or [ɔ]) and ɔw would be, but reflexes of the XIAO and GUO groups have not merged in Wenling (ɔǎ > o; ɔw > ɔ). The vowel quality hypothesis works well for those finals which have back reflexes in Wenling, but not for those which have front reflexes. We will now consider a subset of these together with comparative data from other modern Wu dialects given in (11). Leqing and Wenzhou are geographically close to Wenling (all three are in southern Zhejiang Province) while Suzhou and Shanghai are further away in Jiangsu Province. (Gr. in the table below stands for Grade.)

11. Gr.	EMC	LMC	Wenling	Leqing[10]	Wenzhou[10]	Suzhou[12]	Shanghai[13]
XIAN rime group							
I	əm	am	貪 tʰøn[14], dɛ 潭 tɕie 感	te, dɛ ke	tʰø, dɦø kø	-, dE kø	tʰø, - kɔ
I	əp	ap	雜 zøʔ[14], təʔ 答 tɕiʔ 蛤	-, de ke 鴿	zø, tø kø 鴿	-, taʔ kɤʔ	zaʔ, taʔ kəʔ
I	am	aam am	dɛ 談 tɕie 敢	dɦɛ ke	dɦɔ kø	dE kø	dɛ kø

I	ap	aap	tɛʔ 塔	ta	tɔ	taʔ	taʔ
		ap	ɦiʔ 盍				
II	ε^{r}m	aam	dzɛ 站	dʑɦɛ	dʑɦɔ	zE	zɛ
			kiɛ 減		kɔ	tɕɩ/kE	
II	ε^{r}p	aap	tsʰɛʔ 插		tɕʰɔ	tsʰa	tsʰaʔ
			kiɛʔ 恰	kɛ/ka 夾	ka, kɔ 夾	tɕiaʔ/kaʔ	
II	a^{r}m	aam	sɛ 衫		sɔ	sE	sɛ
			kiɛ 監	kɛ/tɕie	kɔ	tɕɩ/kE	kɛ
II	a^{r}p	aap	kiɛʔ 甲	ka	kɔ	tɕiaʔ/kaʔ	tʃiaʔ

SHAN rime group

I	an	aan[8]	tɛ 單	tɛ	tɔ	tE	tɛ
		an	tɕie 干	kø	kø	kø	kø
I	at	aat[8]	dɛʔ 達	dɦa	dɦɔ	daʔ	daʔ
		at	tɕiʔ 割	ke	kø	kɤʔ	kəʔ
II	ε^{r}n	aan	sɛ 山	sɛ	sɔ	sE	sɛ
			kiɛ 間	kɛ	kɔ	tɕɩ/kE	tʃi
II	ε^{r}p	aat	sɛʔ 刹	sa	sɔ		
II	a^{r}n	aan	kiɛ 諫	kɛ 姦	kɔ 姦		kɛ 姦
II	a^{r}p	aat	sɛʔ 殺	sa	sɔ	saʔ	saʔ

There are two observations to be made: the first is that most forms have undergone raising (the exceptions are three dialects where checked finals which had long vowels did not raise); the second is that there are relative height differences in the Wenling reflexes (and other Wu dialects) for LMC finals reconstructed with short and long vowels. As noted earlier, the consonantal endings -m, -p, -n, -t played a role in fronting the vowels in these finals, but they could not have had a role in causing height differences in reflexes of LMC at and aat, for example. The most likely candidate is the reconstructed length difference. All finals in 10 and 11 above with reconstructed long vowels in LMC have [ɛ], [a], or [ɔ] as a main vowel; those with short vowels have [e], [ø], [ə], or [ɤ] as a main vowel. Pulleyblank posited lengthening of *aj and *am after front ([+coronal]) initials. If length is responsible, lengthening would also have to be posited with the *an final and after labial initials in all three rimes (i.e., *aj, *am, *an), since these do not show the higher reflexes.

The height difference in Wenling and Wenzhou reflexes between a < *aaj on the one hand, and ε < *aam, *aan and ε? < *aap, *aat on the other, must still be explained. It is noteworthy in these two dialects that finals ending in a nasal consonant and those ending in a voiceless stop changed in the same manner. Either a shortening rule came into play in this environment (before m, n, p, t as opposed to j), or the raising rule was generalized to include long vowels in this environment. Either rule would still have to apply after the finals with short vowels had already raised, so we will need to posit two stages of raising. In the first stage, only those finals which had short vowels in LMC raised; in the second, those finals which had long vowels and which were followed by a true consonant were raised, and the first set, with original short vowels, were raised an additional step.

We have seen that no one hypothesis - length, vowel quality, or consonantal endings - is adequate for explaining the raising vs. non-raising of low vowels. All have played a role in determining the course of development. Above it was shown that length was crucial to the raising of front vowels, but not necessarily so to back vowels, as vowel quality or contextual environment could be playing a decisive role. However, shortening of all long vowels did take place at some point in the language and thus we do have to posit one or more shortening rules at some point. The front vowel cases suggest that we may want to posit a general shortening rule which would take place after raising of short vowels but before the next stage of raising. However, the back vowels suggest that we may need two shortening rules. If we shorten back open ɑ (Pulleyblank's aă) first, and have it raise in the first raising step, we eliminate the possibility of confusing the GUO and XIAO finals at the low mid height (ɔă vs. ɔw) by raising the former first.

In formalizing the raising rules, one could analyse the vowel system as having either three or four vowel heights. A four-vowel height analysis, using a multi-valued height feature, allows a unitary treatment of raising, while a three-vowel height analysis, which considers low mid vowels to be low, must treat raising of front and back vowels separately. Both analyses will be sketched out below.

If we assume three vowel heights and two degrees of backness, a LMC [+low, -round] vowel is [+back] before [ă] and [-back] elsewhere. The change of ɑ > ɔ would involve a change only in the feature [round]. The change from a > ɛ on the other hand, must invoke an additional feature, such as [tense], [lowered] or [mid] to create the necessary distinction. It is then difficult if not impossible to write one rule that would allow one to simultaneously raise low > low mid and low mid > high mid, although this would seem to be an intuitively plausible story.

Under a multi-valued four vowel height analysis (1 high = highest, 4 high = lowest), simultaneous raising of short [4 high] to [3 high] and short [3 high] to [2 high] can be expressed as: [high, +short] > [-1 high] where > 2 . This formalization does a more satisfactory job of capturing the process with a minimum of complexity.

In summary, raising would involve the following steps:

1. fronting of /a/ before -m, -n, -j;
2. backing of /a/ before w;
3. shortening of back [ɑː];
4. raising of short [4 high] vowels to [3 high];
5. shortening of front a before -m, -n ;
6. [αhigh, +short] > [α-1 high] where α > 2,

A constraint must be placed on the back vowels at either step 4 or 6 above.

At step 4 the constraint would be formulated as: /a/ did not raise before a [+ high, + back] segment (w, k, ŋ); at step 6 it would be formulated as: /ɔ/ did not raise before a [+ high, + back] segment. Data concerning nasalization will show that the latter formulation would have to be modified and that the former is preferable.

Nasal finals

We will now turn back to a consideration of the three hypotheses concerning the course of nasalization. One area in which the three hypotheses differed was in whether LMC *am (XIAN) and *an (SHAN) became nasalized as a result of nasal consonant loss or whether the nasal consonant simply deleted. The latter would seem preferable if we merely observe the reflexes in (11), as this would explain why the reflexes in (11) all have oral vowels. However, there are two pieces of evidence which suggest that these two finals did undergo nasalization. The first is that, of the 33 Wu dialects surveyed in Chao (1928), ten have nasalized reflexes for LMC *am and *an. The second is internal evidence from Wenling concerning the palatalization and affrication of velars (k > tɕ). Palatalization normally occurred in all Grade III and IV rimes, which had high front medials, but not in Grades I and II. (12) illustrates the normal pattern with reflexes of words in the XIAO rime group. Those above the dashed line do not palatalize while those below do:

12.	Group	Rime	Grade	Ex.	EMC	LMC	Wenling
	XIAO	豪	I	告	kaw	kaw	kɔ
		肴	II	敲	kʰaʳw	kʰaaw	kʰɔ
		- - -	- - -	- - -	- - -	- - -	- - -
		宵	III	驕	kjiaw	kjiaw	tɕiɔ
		蕭	IV	叫	kɛu	kjiaw	tɕiɔ

In the XIAN (LMC -am) and SHAN (LMC -an) rime groups, we unexpectedly find alveopalatal initials in Grade I as well, instead of the expected velar initials. More surprising is the fact that this development is not shared by the Grade I rimes in the XIE (-aj) rime group as rimes of the XIE group have the same nuclear vowel and syllabic shape and would be expected to behave in a parallel fashion with XIAN and SHAN. Examples are shown in (13):

13. Group	Rime	Grade	Ex.	EMC	LMC	Wenling
XIE	咍	I	該	kəj	kaj	kie
	泰	I	蓋	kaj	kaj	kie
	皆	II	界	kɛ^{r}j	kaaj[15]	ka
	佳	II	解	ka^{r}j	kaaj	ka
	齊	IV	計	kɛj	kjiaj	tɕi
XIAN	覃	I	感	kəm	kam	tɕie
	合	I	蛤	kəp	kap	tɕiʔ
	談	I	敢	kam	kam	tɕie
	咸	II	減	kɛ^{r}m	kaam[15]	kiɛ
	洽	II	恰	kʰɛ^{r}p	kʰaap	kʰiɛʔ
	銜	II	監	ka^{r}m	kaam	kiɛ
	狎	II	甲	ka^{r}p	kaap	kiɛʔ
	監	III	檢	kjiam	kjiam	tɕie
	業	III	刦	kɨap	kiap	tɕiʔ
	添	IV	謙	kʰɛm	kʰjiam	tɕʰie
SHAN	寒	I	干	kan	kan	tɕie
	曷	I	割	kat	kat	tɕiʔ
	山	II	間	kɛ^{r}n	kaan[15]	kiɛ
	刪	II	諫	ka^{r}n	kaan	kiɛ
	仙	III	件	gjian	gjian	dʑie
	薛	III	傑	gjiat	gjiat	dʑiʔ
	元	III	建	kɨan	kian	tɕie
	月	III	竭	gɨat	giat	dʑiʔ
	先	IV	見	kɛn	kjian	tɕie
	屑	IV	結	kɛt	kjiat	tɕiʔ

The high front medial which now occurs in Grade II rimes (e.g. kiɛ) must have arisen after the palatalization process which applied to Grade III and IV rimes had lapsed. In (11) we can see that none of the other four Wu dialects show any trace of a high front medial in Grade II rimes. These other dialects also do not show any trace of a medial in Grade I rimes, but Wenling reflexes show that its ancestor dialect must have had a high front element before the palatalization process lapsed. If reflexes of LMC *am, *an had become oral through nasal deletion before the raising process, we would have expected reflexes of these finals in Wenling to be the same as the reflexes of *aj, but they are not. If, on the other hand, the nasal consonant had been retained during the raising process, we would have expected reflexes of *am, *an to have merged with reflexes of LMC finals with non-low vowels (see 4) and to have retained a final nasal consonant today. Since they did lose the final nasal, but did not merge with *aj, there must have been a difference between the vowels of syllables derived from *am, *an, and from *aj. The obvious difference to suggest is that the vowels from the nasal finals were nasalized before raising to [2 high].

While we are discussing the *am, *an finals, one small matter to be resolved is whether the palatalization of their LMC velar initials was due to a high front medial glide (e.g. ĩẽ) or to a high front main vowel (ĩ). There is some evidence that it may have been due to the latter. Reflexes of the checked finals *ap, *at are the same in Wenling (iʔ) as reflexes of checked finals which derive from finals with high vowels in LMC and from finals with short low vowels which assimilated to a high glide and became high vowels. The nasal finals paired with these checked finals preserved their nasals and have the reflex iŋ. These forms are shown in (14):

14.

Group	Rime	Grade	Ex.	EMC	LMC	Wenling
SHEN	侵	III	今	kjim	kjim	tɕiŋ
	緝	III	急	kjip	kjip	tɕiʔ
ZHEN	真	III	巾	kjin	kjin	tɕiŋ
	質	III	吉	kjit	kjit	tɕiʔ
	欣	III	斤	kɨn	kin	tɕiŋ
	迄	III	乞	kʰɨt	kʰit	tɕʰiʔ
ZENG	蒸	III	凝	ŋiŋ	ŋiăŋ	ɲiŋ
	職	III	極	gjik	gjiăk	dʑiʔ
GENG	庚	III	京	kiajŋ	kiajŋ[16]	tɕiŋ
	陌	III	逆	ŋiajk	ŋiajk	ɲiʔ
	青	III	經	kɛjŋ	kjiajŋ	tɕiŋ
	锡	III	激	kɛjk	kjiajk	tɕiʔ

Given the merging of checked finals as -iʔ, we need to posit a raising rule for the reflexes of *ap, *at from [2 high] to [1 high]: eʔ > iʔ. In view of the generally parallel development of checked and nasal finals, we prefer to generalize the rule to cover the reflexes of *am, *an finals so that ē > ī. Subsequently, ī diphthongized and denasalized, yielding ie.

Having ruled out the second hypothesis concerning the course of nasalization (i.e., -m, -n > Ø), we will consider whether nasalization was a unitary process, affecting all low and possibly also low mid vowels, or whether nasalization occurred in stages as a function of vowel height and nasal place of articulation. (7) will be repeated below for ease of reference in the following discussion:

7.	Wenling	LMC	Rime group	Wenling	LMC
1	ɔ̃	aăŋ [ɑŋ]	DANG	ɔʔ	aăk
2	(kuɔ̃)	waăŋ [wɑŋ]	"	(kuɔʔ)	waăk
3	iã	iaăŋ [iɑŋ]	"	iaʔ	iaăk
4	(fɔ̃)	yaăŋ [yɑŋ]	"	(vɔʔ)	yaăk
5	ɔ̃, iɔ̃	aawŋ	JIANG	ɔʔ, iɔʔ	aawk
6	ã	aajŋ	GENG	aʔ	aajk
7	(huã)	waajŋ	"	(ɦuaʔ)	waajk
8	ie	am	XIAN	iʔ	ap
9	ɛ, iɛ	aam	"	ɛʔ, iɛʔ	aap
10	ie, ɛ	iam	"	iʔ	iap
11	ie	an	SHAN	iʔ	at
12	ɛ, iɛ	aan[9]	"	ɛʔ	aat[9]
13	ie	ian	"	iʔ	iat
14	ø, uø, uɛ	uan	"	əʔ, ɛʔ, uɛʔ	uat
15	uɛ	waan	"	(kuɛʔ)	waat
16	ø, iø	yan	"	yʔ	yat

If we posit multiple stages of nasalization according to Chen's place criterion, then the finals with an anterior nasal, i.e. 8-16 in (7) would nasalize first, followed later by the nasalization of those with non-anterior nasals, i.e. 1-7. Nasalization of the latter might be further divided into two stages depending on a vowel height difference (recall that these are assumed to be back vowels), according to Chen's other hypothesis concerning order of nasalization. This would be the case if some of these back vowels had raised before nasalization occurred. The ones which remained low would nasalize before the raised ones did. Note that the reflexes from the anterior nasal finals now show oral vowels. A potential benefit of positing multiple nasalization processes is that we could account for this difference between oral and nasalized reflexes of these two sets by a modified nasal lifespan hypothesis, as discussed earlier: the more recently nasalized vowels, all of which would have derived from -ŋ or -ɲ (Pulleyblank's /jŋ/), would be more strongly nasalized and the lightly nasalized vowels (derived from *am, *an) would then denasalize. This explanation is plausible only if the lightly nasalized vowels denasalized as the finals with -ŋ, etc. were becoming phonemically nasalized, as it is rare for a language to maintain a difference in degree of nasalization between two sets of non-contextually nasalized vowels (though it has been claimed for at least one Chinantec dialect (Palantla, Merrifield 1963). Given that we want to maintain this ordering of denasalization of front vowels with respect to nasalization of back vowels, nasalization of the back vowels would have to wait until the raising processes were complete, as we have shown that we need to posit retention of nasality in ĩ < *am, *an. After raising had occurred, we would have to posit two separate stages of nasalization for the back vowels as a function of vowel height: first ã < aŋ, aɲ, then ɔ̃ < ɔŋ. If this were the case, we might expect ã to denasalize as ɔ̃ became phonemically nasalized, through further application of the modified nasal lifespan hypothesis. This expectation is not borne out. This failure may just be an instance of the non-absolute nature of diachronic tendencies; or we may have incorrectly stated the course of back-vowel nasalization. Chen's hypotheses are thus not invalidated by this argument, but we have one other hypothesis to which it can be compared in finding the most plausible explanation for the development in this dialect.

The remaining hypothesis posited a unitary process of nasalization, either of low vowels ([4 high] vowels) as opposed to non-low vowels, or of [3,4 high] vowels simultaneously. Evidence from studies on velic lowering which show that there is in general a direct correlation between vowel height and velum height might incline us to favor the former; however, in support of the latter, Clumeck (1975: 135) finds that the velum is low for both mid and low vowels as opposed to high vowels in American English. We have already seen that the relevant condition for final nasal retention in Wenling is [2 high] or higher, this indirectly suggests that the relevant condition for nasalization is [3 high] or lower. There is also internal evidence to suggest that nasalization of the back vowels took place after the first raising step of short [4 high] vowels to [3 high] vowels: [ɑ] in open syllables (reconstructed by Pulleyblank as ending in ǎ) raises to mid high, while [ɑ] before LMC -w, -k, -ŋ only raises to low mid. It is somewhat more plausible to attribute the lack of raising to the presence of a member of a class of any velar consonant (velar stop, nasal or labial-velar glide) rather than to a member of a class consisting of nasalized vowels, a velar stop (-k) and a labial-velar glide. There is no need to posit nasalization in stages as Chen does to account for the Wenling data. Not doing so allows simpler rule statement, is consistent with phonetic evidence that low and mid vowels are more likely to be articulated with a lowered velum, and is preferable on internal grounds. Denasalization would then be posited to depend on the front/back parameter, as this is clearly the most likely condition in this dialect, rather than on the intrinsic life-span of nasalized vowels. It is difficult to say if the front/back parameter generalizes to other dialects, as Chen's data is not presented in a way that this could be examined. If *am, *an were fronted in other Chinese dialects, Chen's generalization that the finals derived from these are the first to denasalize in Chinese dialects would also be explainable by a frontness condition on denasalization. Chen's lifespan hypothesis is undesirable on logical grounds, and it appears to be concealing other factors.

Since nasalization is posited to apply to [3, 4 high] vowels simultaneously, it would be inserted after step 5 in the original set of steps involved in raising. The constraint on raising of back vowels would need to be expressed in Step 4.

1. fronting of /a/ before -m, -n, -j;
2. backing of /a/ before -w;
3. shortening of back [ɑː];
4. raising of short [4 high] vowels to [3 high], with the constraint that /a/ did not raise before a [+ high, + back] segment;
5. shortening of front a before -m, -n;

   INSERT: nasalization of [3, 4 high] vowels;
6. [αhigh, +short] > [α -1 high] where $\alpha > 2$.

Conclusion

In conclusion, we have given an internally motivated analysis of the development of nasal finals in Wenling in the context of other historical developments within this dialect. Our analysis does not support the claims made by Chen with respect to the development of nasal finals in Chinese dialects. Nasalization is argued to affect low and low mid vowels in one unitary process, rather than in stages as a function of vowel height and nasal place of articulation. This conclusion is supported by phonetic evidence which indicates that both mid and low vowels are more likely to be articulated with a lowered velum. Denasalization in this dialect can most plausibly be stated as depending on the front/back parameter rather than on the life-span of nasalized vowels. Second, we have shown that, contrary to claims made by Chen, raising of vowel height in Wenling and other Wu dialects is not a result of nasalization, but a function of general raising processes. Third, we provided evidence that the shift of nasal place of articulation (m > n > ŋ) after i in Wenling and other Wu dialects has a perceptual basis. Fourth, we have provided a general description of a raising process which supports a multi-valued height feature. We have also demonstrated that Chen's assumptions about a particular course of mergers of MC finals is inadequate and misleading.

Acknowledgements

I would like to thank Ian Maddieson, Patricia Keating and Peter Ladefoged for their comments, and Liao Rong-rong for her calligraphy.

Endnotes

[1] The term dialect point or diapoint is perhaps more appropriate for refering to the spoken language of a relatively small area such as a town or a county. This term is especially appropriate here as the spoken languages of small areas vary widely throughout the Wu dialect area and they have not been grouped into an accepted set of larger sub-dialects. 'Wu dialects' will be used as a shorthand for 'Wu dialect points'.

[2] Throughout this paper, I will be adopting the terminology developed by Chinese scholars to refer to syllable structure. The primary division is between the *initial*, that is, the initial consonant, and the *final*, which comprises the remainder of the syllable. The final can be further subdivided into the *medial*, an i̯, y or u glide, the *main vowel*, and the *ending* , a nasal or stop segment.

[3] Ancient Chinese (also called Middle Chinese) is generally considered to refer to the standard language of the Sui and Tang dynasties (A.D. 581-618, 618-907 respectively). Pulleyblank divides this one lengthy period into Early Middle Chinese and Late Middle Chinese.

[4] These rime tables include: Yunjing (preface dated 1161), Qiyinlue (published in a book dated 1162), Sishengdengzi (997-1126), Qieyunzhizhangtu (1176-1203) and Qieyunzhinan (1336). They have generally been considered to be systematizations and summaries of the information contained in the rime dictionaries. In the rime tables, each table contains larger groupings of rimes divided into *deng* ("grades") and each rime is united with its four tones on one table. This information is given on the vertical axis of the table. On the horizontal axis, initials are labelled (by representative characters) across the top, and initials are also labelled as to voicing or phonation type. Example characters illustrating the different initial and final categories are filled in on the table itself. In addition, individual tables are also labelled *kaikou* ("open-mouthed") or *hekou* ("closed-mouth"), indicating the absence or presence of a labial medial, and *nei* ("inner") or *wai* ("outer"), which has proved more difficult to interpret.

Later rime tables, such as the Sishengdengzi group the tables into 16 *she* or rhyme groups, which indicate that the rimes comprising a *she* must be very similar.

[5] Chen's data are from DOC (see Streeter, M. 1972. DOC, 1971. A Chinese dictionary on computer. Computers and the humanities, 6: 250-270.), which in turn is from the Lexicon compiled by Beijing daxue.

[6] Middle Chinese is synonymous with Ancient Chinese; see footnote 3. As the forms which Chen posits for Middle Chinese finals are different from Pulleyblank's, Middle Chinese (MC) will be used to refer to Chen's posited forms.

[7] With respect to the difference in numbers for MC /aŋ/ and /ɔŋ/, Chen says that the difference is "due to the under-representation of /ɔŋ/ words in the comparative lexicons or a handful of words made available in the literature, so that the lower number for /ɔŋ/ reflects not so much the different behavior of /ɔŋ/ as the silence or lack of positive evidence regarding the final. MC /ɔŋ/ has always been a minority among the MC finals. . . " (p 101) Given this, one would also expect a disparity between the /əm/ and the /ən/ finals, as the /əm/ rime was also a particularly small one. Methodologically, it would have been better to exclude those diapoints for which information was not available for each MC final in question.

[8] This final is a generalization of the lengthening rule proposed by Pulleyblank for other rime groups in southern dialects.

[9] Possible experimental and perceptual tasks for testing these two hypotheses include: 1) investigating whether a is more nasalized before -m, -n than before -ŋ in different languages, or alternatively, if -m, -n are weaker, perhaps by testing perceptive salience in noise, and 2) clipping post-vocalic nasals and have subjects identify which are most like nasalized vowels.

[10] Nakajima, 1983.

[11] Pulleyblank 1984 reconstructs a prevocalic j after velar initials, which accounts for reflexes in Pekinese; Wenling forms do not show any trace of it and it perhaps never developed in its ancestral dialect.

[12] Beijing daxue, 1959.

[13] Hashimoto, 1971.

[14] The Wenling forms ending in øn and øʔ are a bit odd and atypical at least in comparison with reflexes from Wenzhou, Leqing, Shanghai, and Suzhou. These Wenling forms have merged with reflexes of LMC finals containing an [ə]: *ən and *əm. It seems likely that this group of words did not undergo the general LMC lowering of ə to a in these rimes.

[15] See note 11. Medial -j- has been omitted in all Grade II forms in this table.

[16] In the GENG reflexes containing a, there are two arguments for supposing that a underwent assimilation. First, if a is assumed to have assimilated between two high front elements, giving a long high front vowel, we can explain such contrasts as Wenling tɕi < *tsiaj and ts < *tsi : frication occurred with short i rather than with long iː. Second, *-ya- sequences regularly give Wenling /-iø-/([yø]), which suggests assimilation rather than deletion. Wenling /ioŋ/([yoŋ]) < *yajŋ, yawŋ can be explained as backing of the vowel before a velar after assimilation had taken place: yajŋ > yøjŋ > yøŋ > yoŋ. (Alternatively: yajŋ > yyjŋ > yyŋ > yoŋ. The latter would have the advantage of allowing a simple feature spreading rule in the first step, but makes expression of the last step, and derivation of yø reflexes, more difficult.) A rule which causes either partial or total assimilation of a after y also explains why *yawŋ does not give yɔ̃ as would otherwise be expected. One rime, *yaă (Wenling [y]), seems to be an exception to general patterns discussed so far, as the reflex is neither yø nor yɔ. a in this case does seem to have totally assimilated to the preceding y.

Appendix: Correspondences between EMC, LMC and Wenling finals

In the table below, Group = rime group (she攝); R = rime; G = grade (deng 等); EMC = Early Middle Chinese; LMC = Late Middle Chinese (the latter two from Pulleyblank 1984). In the coumns under "Wenling", Main = main reflex; Lab(ial initials), Vel(ar + laryngeal initials), Ret(roflex sibilant initials), Alv(eolar initials) columns will be used to list exceptions to the main reflex. (These categories refer to LMC initials.) If this applies to a specific type of initial within these broader categories, e.g. labiodental as opposed to bilabial, or velar as opposed to laryngeal, an example syllable from modern Wenling will be given to specify the type of initial to which the exception applies. A hyphen under any of these columns indicates that that type of initial did not occur with that final. In some instances, a rime may have so few words that there will only be data for relexes with one type of initial. In such cases, there will be a hyphen under Main and the reflex will be listed under the appropriate initial column. Parentheses under any of the Wenling columns indicate Lin Chih's pronunciation where different from that of Sun Yi-mou.

	Group	R	G	EMC	LMC	Wenling:				
						Main	Lab	Vel	Ret	Alv
果	GUO	歌	I	aă [ɑ]	aă	o (u)	-	ɯ,o	-	
		戈	I	waă	uaă	o (u)	u	u, ŋɯ	-	
		戈	III	uaă	yaă	-		hy		
假	JIA	麻	II	a^r	(j)aaă [aː]	o				
		麻	II	wa^r	waaă	-		o, ua (uo)		
		麻	III	iaă	iaă	ia	-		-	*tɕ: o, ia
遇	YU	模	I	ɔ	uă	u		ku, ŋɯ, ŋ		-
		虞	III	uă	yă	y	fu			
					uă	-			u	
		魚	III	ɨă, uă	iă, (yă)	y		y, ŋɯ		*tr: y,
					əă, (uă)				u, o	

Group	R	G	EMC	LMC	Main	Lab	Vel	Ret	Alv
蟹 XIE	咍	I	əj	aj	e		kie	-	na
	泰	I	aj	aj	a		kie, he	-	
	灰	I	wəj	uaj	e		ue	-	
	泰	I	waj	uaj	e		ue	-	
	皆	II	$\varepsilon^{r}j$	(j)aaj	a				
	佳	II	$a^{r}j$	(j)aaj	a		a, o		
	夬	II	$a^{r}j^{h}$	(j)aaj	-	a			
	皆	II	$w\varepsilon^{r}j$	waaj	-		ua		
	佳	II	$wa^{r}j$	waaj	-		ua, uo		
	夬	II	$wa^{r}j^{h}$	waaj	-		ua		
	祭	III	(j)iaj	(j)iaj	i			-	
	廢	III	uaj	yaj	-	fi			
	祭	III	(j)wiaj	(j)yaj	y	-		-	
	齊	IV	ɛj	(j)iaj	i			-	
	齊	IV	wɛj	(j)yaj	-		y		
止 ZHI	支	III	(j)iă	(j)i̢, r̩, z̩	i̢	i̢, e			
	脂	III	(j)i	(j)i̢, r̩, z̩	i̢		-		
	之	III	ɨ	i̢, r̩, z̩	i̢	-			
	微	III	ɨj	i	-		i		
	支	III	(j)wiă	yj̢, uj	y	e, i		i̢, iø	le
	脂	III	(j)wi	(j)yj̢, uj	y	e		e	le
	微	III	uj	yj	-	fi	y (ue)		
效 XIAO	豪	I	aw	(u)aw	ɔ			-	
	肴	II	$a^{r}w$	(j)aaw	ɔ				
	宵	III	(j)iaw	(j)iaw	iɔ			-	
	蕭	IV	ɛu	(j)iaw	iɔ	-		-	
流 LIU	侯	I	ow	əw	ə	iə	iə	-	
	尤	III	uw	iw, əw	iu	fu, iu, iə	ɲiə	ə (also tr+)	
	幽	III	jiw	jiw	-	iɔ, iu	iu		

Group	R	G	EMC	LMC	Main	Lab	Vel	Ret	Alv
咸 XIAN	談	I	am	am	-		tɕie		ɛ
	盍		ap	ap	-		iʔ, ɛʔ		ɛʔ
	覃	I	əm	am	-		ie, ɛ		ɛ, øn
	合		əp	ap	-		iʔ		ɛʔ, øʔ, əʔ
	銜	II	a^{r}m	(j)aam	ɛ	-	kiɛ		-
	狎		a^{r}p	(j)aap	-		kiɛʔ, ɦɛʔ		
	咸	II	ɛ^{r}m	(j)aam	ɛ	-	kiɛ		
	洽		ɛ^{r}p	(j)aap	ɛʔ	-	kiɛʔ		
	鹽	III	(j)iam	(j)iam	ie			-	
	葉		(j)iap	(j)iap	iʔ			-	
	嚴	III	ɨam	iam	-		ie		
	業		ɨap	iap	-		iʔ		
	凡	III	uam	iam	-	fɛ			
	乏		uap	iap	-	fɛʔ			
	添	IV	ɛm	(j)iam	ie			-	
	帖		ɛp	(j)iap	iʔ			-	
山 SHAN	寒	I	an	an	ɛ	-	ie	-	
	曷		at	at	ɛʔ	-	iʔ	-	
	桓	I	wan	uan	ø		uø, uɛ	-	
	末		wat	uat	-	ɛʔ	uɛʔ	-	əʔ
	刪	II	a^{r}n	(j)aan	ɛ		kiɛ		-
	鎋		a^{r}t	(j)aat	ɛʔ		Ø (no data)		-
	山	II	ɛ^{r}n	(j)aan	ɛ		kiɛ		-
	黠		ɛ^{r}t	(j)aat	ɛʔ		Ø		-
	刪	II	wa^{r}n	waan	-		uɛ		
	鎋		wa^{r}t	waat	-		uɛʔ		yʔ
	山	II	wɛ^{r}n	waan	-		uɛ	-	
	黠		wɛ^{r}t	waat	-		uɛʔ	-	
	仙	III	(j)ian	(j)ian	ie			-	
	薛		(j)iat	(j)iat	iʔ			-	
	元	III	ɨan	ian	-		ie		
	月		ɨat	iat	-		iʔ		

Group	R	G	EMC	LMC	Main	Lab	Vel	Ret	Alv
	仙	III	(j)wian	(j)yan	iø	-			lø
	薛		(j)wiat	(j)yat	yʔ	-	yʔ, ɦiaʔ	-	lɛʔ
	元	III	uan	yan	-	fɛ	yø		
	月		uat	yat	-	fɛʔ	gyʔ,ɲyɔʔ,ɦiɛʔ, ɦiaʔ		
	先	IV	ɛn	(j)ian	ie			-	
	屑		ɛt	(j)iat	iʔ			-	
	先	IV	wɛn	jyan	-		ø, iø		
	屑		wɛt	jyat	-		yʔ		
蓁 ZHEN	痕	I	ən	ən	-		ən		øn
	没		ət	ət					
	魂	I	wən	un	øn	ən	uən	-	
	没		wət	ut	-	ɛʔ	uɛʔ		əʔ,yʔ
	臻	II	in	ən	-			iŋ	
	櫛		it	ət	-			ɛʔ	
	真	III	(j)in	(j)in	iŋ			-	
	質		(j)it	(j)it	iʔ			-	
	欣	III	ɨn	in	-		iŋ		
	迄		ɨt	it	-		iʔ		
	諄	III	(j)win	(j)yn	yn	-		-	nøn
	術		(j)wit	(j)yt	yʔ	-		-	liʔ
	文	III	un	yn	-	fən	yn		
	物		ut	yt	-	fɛʔ	yʔ, yɔʔ		
深 SHEN	侵	III	(j)im	(j)im	iŋ				
				əm				øn, iŋ	
	緝		(j)ip	(j)ip	iʔ			-	
曾 ZENG	登	I	əŋ (əăŋ)	əăŋ	ən	oŋ			
	德		ək	əăk	ɛʔ	ɔʔ	əʔ		
	登	I	wəŋ	uăŋ	-		(ŋ)		
	德		wək	uăk	-		uɔʔ		
	蒸	III	iŋ (ɨăŋ)		iăŋ	iŋ			-
	職		(j)ik(ɨăk)		iăk	iʔ			-

Group	R	G	EMC	LMC	Main	Lab	Vel	Ret	Alv
				əǎk	-			ɛʔ	
	職	III	wik(wɨǎk)	yǎk	-		uɔʔ		
梗 GENG	庚	II	a^{r}jŋ	(j)aajŋ	ā				
	陌		a^{r}jk	(j)aajk	aʔ		ā, iŋ	ɛʔ (əʔ)	
	耕	II	ɛ^{r}jŋ	(j)aajŋ	ā				(*tr: i°)
	麥		ɛ^{r}jk	(j)aajk	aʔ				(": aʔ)
	庚	II	wa^{r}jŋ	waajŋ	-		ɦu ā		
	陌		wa^{r}jk	waajk	-				
	耕	II	wɛ^{r}jŋ	waajŋ	-		ɦuoŋ, ŋ		
	麥		wɛ^{r}jk	waajk	-		ɦuaʔ		
	庚	III	iajŋ	iajŋ	iŋ			ā	-
	陌		iajk	iajk	-		iʔ		
	清	III	(j)iajŋ	(j)iajŋ	iŋ			-	
	昔		(j)iajk	(j)iajk	iʔ				
	庚	III	wiajŋ	yajŋ	-		ioŋ		
	陌		wiajk	yajk	-				
	清	III	jwiajŋ	jyajŋ	-		iŋ~ioŋ		
	昔		jwiajk	jyajk	-		iɔʔ		
	青	IV	ɛjŋ	jiajŋ	iŋ			-	
	錫		ɛjk	(j)iajk	iʔ			-	
	青	IV	wɛjŋ	jyajŋ	-		ɦiŋ		
	錫		wɛjk	jyajk	-				
宕 DANG		I	aǎŋ[ɑŋ]	aǎŋ	ɔ̄			-	
	鐸		aǎk	aǎk	ɔʔ			-	
	唐	I	waǎŋ	waǎŋ	-		u ɔ̄		
	鐸		waǎk	waǎk	-		uɔʔ, ɔʔ		
	陽	III	ɨaǎŋ	iaǎŋ	i ā			-	
				aaǎŋ				ɔ̄	
	藥		ɨaǎk	iaǎk	iaʔ			-	
	陽	III	uaǎŋ	yaǎŋ	-	ɔ̄	k ɔ̄,hi ɔ̄,hu ɔ̄		
	藥		uaǎk	yaǎk	-	ɔʔ			

	Group	R	G	EMC	LMC	Main	Lab	Vel	Ret	Alv
江	JIANG	江	II	aʳwŋ	aawŋ	ɔ̄, i ɔ̄	ɔ̄	-	-	i ɔ̄
				(œʳwŋ)		jaawŋ	-		ɔ̄	
					waawŋ	-		i ɔ̄		
		覚		aʳwk	aawk		ɔʔ	-	-	iɔʔ
				(œʳwk)		jaawk	-		ɔʔ	
					waawk	-			iɔʔ	
通	TONG	東	I	owŋ	əwŋ	oŋ		uoŋ, ŋ̍	-	
		屋		owk	əwk	ɔʔ		uɔʔ (ɔʔ)	-	əʔ
		冬	I	awŋ	əwŋ	-				oŋ
		沃		awk	əwk	ɔʔ		uɔʔ	-	
		東	III	uwŋ	iwŋ, əwŋ	ioŋ	foŋ		oŋ	loŋ
		屋		uwk	iwk, əwk	iɔʔ	fɔʔ	iɔʔ, yʔ		lɔʔ
		鍾	III	uawŋ	yawŋ	ioŋ	foŋ		-	loŋ
		燭		uawk	yawk	iɔʔ	-		-	lɔʔ

Bibliography

Spellings of all titles and names of PRC authors are given in pinyin; spellings of names of authors in Taiwan follows their personal preference. Spelling cross-references are provided for names which might cause confusion.

Ballard, William L. 1969. *Phonological history of Wu.* Berkeley: U. of California, Ph.D. diss.

______. 1971. Scenarios of change in Chinese dialectology. *Anthropological Linguistics* 13: 125-157.

Beddor, Patrice Speeter. 1983. *Phonological and phonetic effects of nasalization on vowel height.* Ph.D. diss. Reproduced by Indiana University Linguistics Club.

Beijing daxue zhongguo yuyan-wenxuexi yuyanxue jiaoyanshi (eds.) (Linguistics group, Department of Chinese Linguistics and Literature, Beijing University). 1959. *Hanyu fangyin zihui* (Lexicon of Chinese dialect pronunciations). Beijing: Wenzi gaige chubanshe.

Chan, Marjorie Kit Man. 1980. *Zhongshan phonology: a synchronic and diachronic analysis of a Yue (Cantonese) dialect.* U. of British Columbia, M.A. thesis.

Chang, Kun. 1970. Wenchow historical phonology. *Bulletin of the Institute of Ethnology* (Academia Sinica Taipei) 32: 13-76.

______. 1971. Phonological aspects of Chinese dialectology. *Tsing Hua Journal of Chinese Studies* NS 9.1-2: 192-215.

______. 1974. Ancient Chinese phonology and the *Ch'ieh-yun*. *Tsing Hua Journal of Chinese Studies* NS 10.2: 61-82.

______. 1983. Hanyu fangyan zhong biyin yunwei de xiaoshi (Disappearance of nasal endings in Chinese dialects). *Bulletin of the Institute of History and Philology*, Academia Sinica 54.1: 3-74.

Chao, Yuan Ren. 1928. *Xiandai Wu yu de yanjiu* (Studies of the modern Wu dialects). Beijing: Qinghua Xuexiao Yanjiuyuan.

______. 1930. *Fangyin diaocha biaoge* (Tables for investigation of dialect pronunciations). Beijing: Institute of History and Philology, Academia Sinica.

______. 1941. Distinctions within Ancient Chinese. *Harvard Journal of Asiatic Studies* 5: 203-33.

______. 1967. Contrastive aspects of the Wu dialects. *Language* 43.1: 92-101.

______. 1968. *A Grammar of Spoken Chinese.* Berkeley.

Chen, Li. 1842. *Qieyun kao.* Reprinted in Yinyunxue congshu, Taipei: Guangwen shuju. 1966.

Chen, Matthew. 1973. Cross-dialectal comparison: a case study and some theoretical implications. *Journal of Chinese Linguistics* 1.1: 38-63.

______. 1975. An areal study of nasalization in Chinese. *Journal of Chinese Linguistics* 3.1: 16-59. Also in C. Ferguson et.al. (1975) 81-124.

Chen, Yinke. 1949. Cong shishi lun Qieyun. *Lingnan xuebao* 9: 1-18.

Chou, Fa-kao. 1954. Lun gudai hanyu de yinwei (On the phonemes of old Chinese). *Bulletin of History and Philology,* Academia Sinica 25:1-19.

______. 1983 *Zhongguo yinyunxue lunwenji* (Collected essays on Chinese [historical] phonology). HongKong: Chinese University of Hong Kong Press.

Clumeck, Harold. 1975. A cross-linguistic investigation of vowel nasalization: an instrumental study. In C. Ferguson et. al. (1975) 133-152.

Denlinger, Paul. Pre-modern Chinese: material and methods. ms.

Ding, Bangxin. See Ting Pang-hsin.

Dong, Tonghe. See Tung, T'ung-he.

Ferguson, Charles A., Hyman, Larry M., Ohala, John J., eds. 1975. *Nasálfest: Papers from a symposium on nasals and nasalization.* Stanford: Language Universals Project.

Foley, James. 1975. Nasalization as universal phonological process. In Ferguson et. al. (1975) 197-212.

Hangzhou Daxue Zhongwenxi Fangyan Diaocha Zu (Dialect investigation group, Department of Chinese Literature, Hangzhou University). 1959. Wenling fangyan (Wenling dialect). *Hangzhou Daxue Xuebao* 3: 151-205.

Hashimoto, Mantaro J. 1965. *Phonology of Ancient Chinese.* Ohio State University, Ph.D. diss. Reprinted as nos. 10 and 11 of Study of Langauges and Cultures of Asian and Africa Monograph Series, Tokyo, 1978 and 1979.

______. 1971. *A guide to the Shanghai dialect.* Center for Applied Linguistics, microfilm series no. 061851.

Huang, Cuibo. 1931. *Huilin Yiqie jing yinyi fanqie kao.* Beijing: Institute of History and Philology, Academia Sinica.

Karlgren, Bernhard. 1954. *Compendium of phonetics in Ancient and Archaic Chinese.* Reprinted from *Bulletin of the Museum of Far Eastern Antiquities* 26: 211-367.

Li, Fang Kuei. 1971. Shanggu yin yanjiu (Studies on Archaic Chinese phonology). *Tsinghua Journal of Chinese Studies* 9: 1-61. Translated by G.L. Mattos as Studies on Archaic Chinese and printed in *Monumenta* Serica 31 (1974-75): 219-97.

______. 1973. Languages and dialects of China. *Journal of Chinese Linguistics* 1.1: 1-13.

Li, Jung. See Li Rong.

Li, Rong. 1956. *Qieyun yinxi* (The phonological system of the Qieyun). Beijing: Kexue chubanshe.

______. 1966. Wenling fangyan yuyin fenxi (Phonological analysis of the Wenling dialect). *Zhongguo Yuwen* 1: 1-9.

Lindau, Mona. 1978. Vowel Features. *Language* 54.3: 541-563.

Long, Yuqun. See Lung Yu-ch'un.

Lubker, J. F. 1968. An electromyographic-cinefluorographic investigation of velar function during normal speech production. *Cleft Palate Journal* 5: 1-18.

______, B. Fritzell, and J. Lindqvist. 1971. Velopharyngeal function: an electromyographic study. *Speech Transmission Laboratory-Quarterly Progress and Status Report* 4/1970: 9-20.

Lung, Yu-ch'un. 1979. *Yunjing jiaozhu* (Corrected and annotated *Yunjing*). Taipei: Yiwen yinshuguan.

Luo, Changpei. 1931. Zhi che cheng niang yinzhi kao (Examination of the phonetic values of zhi che cheng niang). *Bulletin of the Institute of History and Philology*, Academia Sinica 3: 121-57. Reprinted in Luo 1963.

______. 1931b. Dunhuang xieben Shouwen yunxue canzhuan ba (On the preface to the Dunhuang fragments of Shouwen's Rime Study). *Bulletin of the Institute of History and* Philology, Academia Sinica 3: 252-62. Reprinted in Luo 1963.

______. 1933. Shi nei wai zhuan (Explaining inner and outer turning) *Bulletin of the Institute of History and Philology*, Academia Sinica 4: 209-26. Reprinted in Luo 1963.

______. 1963. *Luo Changpei yuyanxue lunwen xuanji* (collection of selected essays on linguistics by Luo Changpei). Beijing: Zhongguo shuju.

Merrifield, William R. 1963. Palantla Chinantec syllable types.

Anthropological Linguistics 5.5: 1-16.

Moll, Kenneth L. 1962. Velopharyngeal closure on vowels. *Journal of Speech and Hearing Research* 5: 30-7.

Nakajima, Motoki. 1983. *The Southern Chekiang Dialect: a linguistic study of the phonology and lexicon.* Asian and African Lexicon 14. Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa.

Nord, L. 1976. Perceptual experiments with nasals. *Speech Transmission Laboratory-Quarterly Progress and Status Report* (Royal Institute of Technology, Stockholm) 2-3/1976: 5-8.

Ohala, John J. 1974. Experimental historical phonology. *Proceedings of the First International Conference of Historical Linguistics*, Edinburgh. North Holland Press, vol. 2: 353-89.

______. 1975. Phonetic explanations for nasal sound patterns. In C. Ferguson et.al. (1975) 289-316.

Pulleyblank, E.G. 1970-71. Late Middle Chinese. *Asia Major* 15: 197-239 and 16: 121-68.

______. 1981. Some notes on Chinese historical phonology. *Bulletin de l'#Ecole Française d'Extrême-Orient* 69: 277-88.

______. 1983. Middle Chinese reflexes of Old Chinese final palatals, labial-velars and uvulars. *Journal of Asian and African Studies* 25: 45-60.

______. 1984. *Middle Chinese: a study in historical phonology.* Vancouver: U. of British Columbia Press.

______. 1986. CV phonology and diachronic change as illustrated in the history of Chinese. Paper presented at the 19th International Congress of Sino-Tibetan Languages and Literature.

______. 1986. Some theoretical issues in CV phonology with special reference to the history of Chinese. *Canadian Journal of Linguistics* 31:

225-265.

Ruhlen, Merritt. 1975. Patterning of nasal vowels. In C. Ferguson et.al. (1975). 333-352.

Ting, Pang-hsin. 1982. Phonological features for classification of the Chinese dialects. *Tsing Hua Journal of Chinese Studies* 14.1-2: 257-83.

______. 1985. Zhongguyin yu cunhua wenbai zitong de guanxi (On the relationship between the colluquial and literary systems of Cunhua [Danzhou] and Ancient Chinese phonology) ms.

Tung, T'ung-he. 1977. *Hanyu yinyunxue* (Chinese phonology). Taipei: Wenshizhe chubanshe.

Wang, Li. 1935. *Hanyu yinyunxue* (Chinese phonology). Reprinted 1956 Beijing: Zhonghua shuju.

Wang, William S.Y. 1969. Competing sound changes as a cause of residue. *Language* 45: 9-25.

Wright, James. 1975. Effects of vowel nasalization on the perception of vowel height. In C. Ferguson et. al. (1975) 373-388.

Yip, Moira Jean W. 1982. Reduplication and C-V skeleta in Chinese secret languages. *Linguistic Inquiry* 13: 637-661.

Zee, Eric P. 1979. Effect of vowel quality on perception of nasals in noise. *Working Papers in Phonetics* 45: 130-149.

_____. 1981. Effect of vowel quality on perception of post-vocalic nasal consonants in noise. *Journal of Phonetics* 9.1: 35-48.

Zhang, Shilu. 1938. *Zhongguo yinyunxue shi* (History of Chinese historical phonology). Reprinted 1978 Taipei: Taiwan shangwu yinshuguan.

Zhao Yintang. 1956. *Dengyun yuanliu* (Origins and changes of rime tables). Shanghai: Shangwu yinshuguan.

Zhou, Fagao. See Chou Fa-kao.

Zhou, Zumo. 1966. *Wen xue ji.* Beijing: Zhonghua shuju.

Anticipatory length

Diana Van Lancker, Jody Kreiman, & Dwight Bolinger*

1. Introduction

1.1 Background

A well-known effect on syllable length is the presence or absence of a syllable-final consonant: length increases with voicing, most markedly in pre-pausal monosyllables, e.g. rib-rid-rig versus rip-rit-rick. Is there a similar effect on the syllable in question given conditions in the following syllable?

In a series of impressionistic studies (1963, 1965, 1981, 1986) and in one limited experimental study (Dasher and Bolinger, 1982), Bolinger made the claim that (1) a syllable that contains a full vowel followed by a syllable likewise containing a full vowel will be longer than (2) a syllable containing a full vowel followed by a syllable containing a reduced vowel. To abbreviate, we mark (1) (S)trong and (2) (W)eak, and we refer to a syllable containing a full vowel as a full syllable and one containing a reduced vowel as a reduced syllable.

This lengthening effect is alleged to hold, in continuous speech, regardless of the location of stress or word boundary, and regardless of word length or morpheme division. It is thus predicted that in the sentences

(a) A workman like(S) Jack can lift a heavy weight.
(b) A workmanlike(S) jack can lift a heavy weight.
(c) A workmanlike(W) jacuzzi is a nice thing to have.

the syllable like will be longer in (a) and (b) than in (c); that in

She yelled, "What kind of sandwich do you want?," and I
yelled back,
(d) "Ham(S) on rye!"

(e) "Ham(W) and rye!"

ham will be longer in (d) than in (e); and that in

(f) Hey coroner, when you finish that au(?)topsy put your
au(W)tograph on it.

those who pronounce autopsy with medial shwa will not lengthen the au- of either autopsy or autograph, but those who use [a] in autopsy will make that au- longer, contrasting with both their own autograph and the shwa pronunciation of autopsy.

1.2. The vowel series

The vowels, full and reduced, are sufficiently defined, for our purpose, by enumerating those of the reduced series, which are fewer in number. they include the syllable sonorants of words like battle, batten, batter, and chasm and the three contrasting final vowels in Calley, calla, callow, of which the central member, the shwa of calla, is the most frequent. We hypothesize that all the reduced vowels function similarly and the full vowels (which include diphthongs) likewise, in respect of the point we are investigating.

1.3. Simple and complex contexts

In the one situation, namely that of successive monosyllables, the length difference we hypothesize is uncontroversial. Thus Van Draat (l910, 14) pointed out that in Money makes the mare go, mare is longer than in Money makes the mare to go. Our purpose is to test the hypothesis under more general syntactic and morphological conditions. If, as Lehiste (l972, 2021) observes, "the temporal structure of the utterances seems to depend most of all on their syllable structure", the effect should be relatively constant regardless of such other factors.

2. Methods

2.1. The test sentences

Two sets of sentences were devised, typed on cards, and randomized (see Appendices). Words and phrases were chosen to provide minimal pairs, one member of which contained a target syllable (S) and the other a target syllable (W).

The aim of the first set was to expose the target syllables to a wide variety of environments, particularly involving stress, for example, initial stress versus initial nonstress (e.g. the *ban*- of *bandit* versus the *ban* of *bandanna*), initial stress versus stressed monosyllable (*rotate* versus *wrote it*), initial stress versus initial stress (*mandragon* versus *mandolin*), and final nonstress versus final nonstress (*electron lineup* versus *electron alignment*). As this was exploratory, in a few cases there were variables other than the one to be tested that might unduly affect length. For example, in the pair

> It's(S) quite a grotesque idea.
> It's(W) a quite grotesque idea.

there is potential interference in the immediate phonetic environment: the complex syllable structure of *quite* as against the simple structure of *a* exaggerates the strong-weak contrast, producing a bias toward the hypothesis. On the other hand, in the pair

> The ban(S)danna is green.
> The ban(W)dit is mean.

the emotive content of the second member, which came across quite obviously in the speakers' productions, was conducive to the lengthening of the already stressed *ban*- of *bandit*, creating a bias against the hypothesis. Although the results suggest that these effects canceled each other out, we nevertheless designed the second test to correct them as far as possible within the limits of natural speech.

2.2 Procedure

Each stimulus set was recorded by two different speakers. All four speakers were naive as to the purpose of the experiment. The speakers read each sentence first silently, and then aloud into a dynamic microphone attached to an Ampex tape recorder in a sound-attenuated booth. They were instructed to say all test sentences as naturally as possible. Wide-band spectrograms were made of the tape-recorded utterances using a Kay digital spectrograph. Boundaries of the target syllables were marked independently by two phoneticians; for each syllable, a measurement was made to the nearest half-millimeter. Measurements were rechecked twice by two different phoneticians, and any discrepancies resolved by referring to the established measurement criteria.

Identification of speech segments from spectrograms requires an algorithm to be applied consistently throughout the measurement process. As virtually all measurements were made on minimal pairs, criteria could easily be applied equally to both test syllables. Several different approaches would serve equally well. The following criteria for establishing syllable boundaries in our data were adopted and applied consistently across all measurements.

Sentence-initially, for voiced segments (voiced consonants and vowels), beginnings of syllables were measured from the onset of voicing. (No voiceless target segments appeared sentence-initially in these stimuli.) Medially in the sentence, onsets of stop consonants were measured from the beginning of the closure, and nasals from the onset of the nasal formant. The points of terminus of test syllables were measured for stop consonants at the end of the closure/beginning of the burst (release); for all other segments, ends of syllables were found at the onset of the following segment (usually a stop).

3. Results

3.1 Collective effects

Mean durations for test syllables --(S) versus (W)--are given in Table 1 (in millimeters) for the two sets of sentences spoken by the two pairs of speakers.

Two-way ANOVAs were performed on data from the two speakers for each stimulus set to determine whether the durational differences were independent of speaker. For set one, no significant main effect of speaker was observed ($F(1,59)=0.32$), indicating that the patterns of absolute durations were comparable for both speakers. A significant main effect of following syllable was seen ($F(1,59)=8.83$, $p < 0.01$) but no significant speaker by (S) versus (W) interaction was observed ($F(1,59)=0.07$, $p < 0.01$). This indicates that syllables followed by (S) syllables were significantly longer than those followed by (W) syllables for both speakers.

	(S)	(W)
STIMULUS SET 1		
Speaker 1	**30.00** **(9.9)**	**27.50** **(8.9)**
Speaker 2	**29.2** **(8.9)**	**26.5** **(8.2)**
STIMULUS SET 2		
Speaker 3	**35.39** **(7.09)**	**31.31** **(7.39)**
Speaker 4	**39.06** **(8.08)**	**37.18** **(11.79)**

Table 1. Means and standard deviations of durations, comparing (S) and (W) test syllables (in millimeters).

3.2 Partial effects

An objection to this set of sentences was that some contrasting pairs differed in syllable count around the target syllable. For example, "The ban(S)danna is green" and "The ban(W)dit is mean" differ in that the word carrying (S) has 3 syllables but the word carrying (W) has two. To find whether this difference affected the results, we extracted from the set of stimuli recorded by Speakers 1 and 2, 15 pairs having the same number of syllables counting both target and following syllables (e.g., mon(W)ey talks, mun(S)dane talks; I tried (W) a dozen, I tried (S) out a dozen) and analyzed them separately. Mean values were calculated for both speakers for both kinds of syllables. For speaker 1, the mean syllable duration in millimeters for (S) syllables was 34.14, while the mean duration for (W) syllables was 30.22 ; for speaker 2, the contrasts were 31.38 for (S) and 27.22 for (W). As there was no main effect of speaker on the previous ANOVA performed on these data, a matched pairs t-test combining values from both speakers was performed. The t-test indicated that the target syllable durations differed significantly depending on whether they were (S) or (W) ($t(29) = 3.26$, $p < 0.01$). Thus for this selected subset, mean values for syllable durations were as predicted for the study.

For the second set of sentences pairs spoken by speakers 3 and 4, a two-way ANOVA showed significant main effects of speaker and of (S) versus (W): (speaker effect: $F(1,74)=8.54$, $p < 0.01$; (S) versus (W): ($F(1,74)=15.34$, $p < 0.01$). The speaker effect is due to the absolute duration differences between speakers, seen in Table 1. Again, no speaker by (S) versus (W) syllable interaction was observed ($F(1,74)=1.48$; NS). This finding matched the results obtained on the first stimulus set.

Given the lack of significant interaction terms in both ANOVAs, data from the two speakers for each stimulus set were combined for matched-pair t-test analyses. For speakers 1 and 2, (S) syllables were reliably longer than (W) syllables: ($t(60)=3.00$, $p < 0.01$). For speakers 3 and 4, the finding was similar: ($t(74) = 3.99$, $p < 0.01$). These findings confirmed the hypothesis of this study.

4. Discussion

The results give statistical backing to the auditory impression created by a sequence like All (S) sam(S)pan(S) hands(S) work(S) long(S) hours--that it has a "syllable-timed" rhythm, differing from the "stress-timed rhythm" of All(W) the sam(W)ple pan(W)eling han(W)dles work(W) alike. Although the -pan of sampan is unstressed, the fact that the vowel is full gives that syllable the same status as all the others in affecting the length of the preceding syllable and being affected in its own length by the following syllable.

4.1 A function for length: contouring

Many factors lead to lengthening. Unlike shortening, which responds more to overall speed, selective lengthening of a syllable is always an option, whether for emphasis, amplification (a l-o-n-g way), emotion (I h-a-t-e you!), or collecting one's thoughts. Why then a nonselective, automatic lengthening conditioned by a following syllable? One plausible explanation is that it is associated with intonation in much the same way that utterance-final lengthening appears to be: utterance-final position usually has to accommodate the most important pitch turn. If two full syllables are juxtaposed and both are to be accented, the first thus needs a certain drawl to be appreciated--one might say that the syllable is divided between an accented part and an unaccented part that trails off; the first is figure and the second ground. When one or more reduced syllables follow, they function as ground and no extra length is needed. It is of course not the case that all successions of full syllables are also successions of accented syllables, but by making the additional length automatic the language provides it for when the pitch turn needs it, and its redundancy the rest of the time does no harm.

4.2 A function for shortening: compactness

As far as contouring is concerned, the language would be served just as well if all syllables were full. What then is the advantage not just of syllable reduction itself but its effect on a preceding full syllable?

A possible answer lies with turning the tables and speaking not of lengthening in (S) but of shortening in (W): a reduced syllable or syllables "borrow time" from a preceding full syllable. As Lehiste (1972: 2018) puts it in summarizing two earlier studies:

> ...the words stead, skid, and skit were compared with steady, skiddy, and skitty. It might have been expected that the latter set would be longer than the former by the average duration of the derivational suffix. It turned out instead that the duration of the base part of the derived word was considerably shortened, so that even with the addition of a fairly long -y, the overall duration of the derived words was not much different from that of the base words.

The shortening does no harm to accentual structure: a contour consisting of full plus reduced syllable(s) is as effective as a contour consisting of one extended full syllable (or, for that matter, consisting of a redundantly extended full syllable followed by one or more full syllables without accent). There is even an advantage, given an established inventory of reduced vowel phonemes, in having certain syllables marked as unaccentable and therefore automatically recognizable as ground for accentual figures.

The upshot is increased compressibility of syllables and much more compact speech. Reduction becomes an ongoing process: an unaccented but full syllable no longer needs its fullness for intonational purposes, and its vowel can join the general drift to shwa.

5. Conclusion

Anticipatory lengthening (or, conversely, regressive shortening) appears to be a constant among the determinants of syllabic duration in English, one that is often overridden but consistently reasserts itself. It is plausibly related on the one hand to intonation structure and on the other hand to compressibility, and probably represents an adjustment between the two.

Further research is needed to determine whether the length effects noted here depend entirely on the nature of the vowel, or whether the nature of the vowel is just one of several contributory factors in the "heaviness" of syllables, in which case one might expect the Bow- of Bowditch to be longer than the Bo- of Bode [bodi-] even though the second syllable in both words is reduced.

*Harvard University and Stanford University (Emeritus)

References

Bolinger, D. (l963) Length, vowel, juncture, Linguistics 1, 5-29.

________. (l965) Forms of English. Cambridge, Massachusetts: Harvard University Press.

________. (l981) Two kinds of vowels, two kinds of rhythm. Bloomington: Indiana University Press.

________. (l986) The English beat: some notes on rhythm. In Problems of standardization and linguistic variation in present-day English (G. Nickel and J. C. Stalker, editors), pp. 36-49. Heidelberg: Gross.

Dasher, R. and Bolinger, D. (1982) On pre-accentual lengthening, Journal of the International Phonetic Association, 12, 58-71.

Lehiste, I. (1972) The timing of utterances and linguistic boundaries, Journal of the Acoustical Society of America, 51, 2018-2024.

Van Draat, P. Fijn (1910) Rhythm in English prose, Heidelberg: Anglistische Forschungen.

Appendix A.

Stimulus sentences, Test One.

They proved a com(S)plex theorem
They proved a com(W)plicated theorem
The ban(S)danna fits a bandit today
The bandanna fits a ban(W)dit today
The ban(W)dit wears a bandanna today
The bandit wears a ban(S)danna today
The ban(S)danna is green
The ban(W)dit is mean
Not all prac(S)titioners are very practical today
Not all practitioners are very prac(W)tical today
Not all prac(W)tical people are practitioners today
Not all practical people are prac(S)titioners today
All thirteen(S) men are ready
All thirteen(W) marines are ready
Is Nim(S)rod as nimble as Nim O'Leary used to be?
Is Nimrod as nim(W)ble as Nim O'Leary used to be?
Is Nimrod as nimble as Nim(W) O'Leary used to be?
Is Nim(W) O'Leary as nimble as Nimrod used to be?
Is Nim O'Leary as nim(W)ble as Nimrod used to be?
Is Nim O'Leary as nimble as Nim(S)rod used to be?
They ro(S)tate both ways

They wrote(W) it both ways
It describes the electron(S) lineup
It describes the electron(W) alignment
They run up Can(W)opy Hill every day
They run up Sun(S)apee Hill every day
I tried(S) out a dozen
I tried(W) a dozen
Wait till you've won(S) over a few
Wait till you've won(W) a few
I didn't say An(W)dy's, I said Andes the other day
I didn't say Andy's, I said An(S)des the other day
I didn't say An(S)des, I said Andy's the other day
I didn't say Andes, I said An(W)dy's the other day
There were ten(S) trustees looking after the affairs of the board
There were ten(S) trusties looking after the other prisoners
There were ten(W) defendants lined up in a court
One(S) Dane talks, and another listens
Mon(W)ey talks, and everyone listens
Mun(S)dane talks are boring
Mon(W)day talks are boring
Those are reprobate(S) justifications
Those are reprobate(W) excuses
Those are substantive(S) justification
Those are substantive(W) excuses
Bring me my man(S)dragon now
Bring me my man(W)dolin now
Who can overcome a handicap(S) like that one?
Who can overcome a handicap(W) as bad as that one?
Who can wear a cap(S) like that one?
Who can wear a cap(W) as big as that one?
It's(S) quite a grotesque idea
It's(W) a quite grotesque idea
She earns(S) starvation wages
She earns(W) a starvation wage
"Vor(W)ta sees" is truth eternal

Vor(W)tices is a plural number
Vor(S)tex stands for one
What are your favorite mountains?-- Oh, I guess the Cascades(W) are the ones I like best
What are your favorite mountains?--Oh, I guess the Cascade(S) Mountains are the ones I like best

Appendix B

Stimulus sentences, Test Two.

No tor(S)ture is worse than that
No to(W)mato is worth all that
They live near Manitou(S) Corners
They live near Manitou(W) Corrals
They wrote(W) it both ways
They ro(S)tate both ways
Because the holiday(S) calendars are here
Because the holiday(W) collection is here
Why don't you give us a lecture on the pontiff?--I'm not INTERESTED in the pon(W)tiff any more
Why don't you give us a lecture on the pontoon?--I'm not INTERESTED in the pon(S)toon any more
I tried(W) a dozen
I tried(S) out a dozen
Who can wear a cap(S) like that one?
Who can wear a cap(W) as big as that one?
Wait till you've won(W) a few
Wait till you've won(S) over a few
They're staying until Sun(W)day only
They're staying until sun(S)down again
They're staying until sundown(W) again
They're staying until sun(S)down only
They're staying until sundown(S) only
Bring me my man(W)dolin now

Bring me my man(S)dragon now
Mon(W)day talks are boring
Mun(S)dane talks are boring
One attack(S) after another
One attack(W) at a time
You've picked the wrong(S) attache
You've picked the wrong(W) design
You've picked the wrong(W) assistant
You've picked the wrong(S) deacon
Was it Pom(W)pey you were talking about?
Was it Pom(S)peii you were talking about?
Mackinaw(S) prime beef
Mackinaw(W) prepared beef
John(S) cracked the walnuts
John(W) corrected the papers
Do you know John(W) O'Toole?
Do you know John(S) Olsen?

Voice perception deficits:
Neuroanatomical correlates of phonagnosia

Diana Roupas Van Lancker, Jody Kreiman, & Jeffrey Cummings

Neuroanatomical substrates of voice perception have been little studied, despite the importance of this ability. To study the perception and cognition of human voices, it is important to distinguish between voice discrimination and voice recognition (Bricker and Pruzansky, l976). Discrimination abilities involve judgments about unfamiliar voices, whereas voice recognition involves identification of a speaker known to the listener.

We wish to report our findings on studies of voice perception deficits, or phonagnosia, in brain-injured subjects, using both kinds of processing (recognition and discrimination). Individuals in four groups-- unilaterally left-brain damaged (LBD), unilaterally right-brain damaged (RBD), bilaterally brain damaged (BBD), and normal control subjects-- were tested on two tasks, familiar voice recognition and unfamiliar voice discrimination. Our purpose was to compare performance on these two tasks and to establish neuroanatomic substrates of these different aspects of voice perception.

In this report, we present data supporting the suggestion that voice discrimination and voice recognition are dissociated neurological functions (Van Lancker and Kreiman, in press; Van Lancker, Cummings, Kreiman and Dobkin, in press), and we explore the specific neuroanatomical correlates of observed deficits in the two abilities.

Methods

Stimuli

The unfamiliar voice stimuli were obtained from recordings of ten male Southern Californians selected from an original set of 22 who were recorded while making telephone survey calls (Kreiman, l987). Speakers ranged from 20-31 years of age, and were matched for regional accent. All were free of vocal pathology.

Using a survey sheet designed for this purpose, speakers were instructed to make telephone survey calls to a prearranged party at each of two recording sessions separated by at least one week. Rather than obtaining the voice recording over telephone lines, the speaker was recorded directly via a high-quality dynamic microphone attached to the mouthpiece of the telephone. This made it possible to obtain good-quality recordings while allowing the speaker to carry on a normal telephone conversation; additionally, only the speaker's voice (and not that of the interviewee) was recorded. One call was selected from each recording session for use in our studies.

Sentences were excerpted from each speaker using a wave-form editing computer program. A stimulus tape was constructed consisting of 26 pairs of voice samples. For 13 pairs, the two samples represented the same speaker; in the other 13 pairs, they represented two different speakers. Within a pair of voices, the speakers always said the same thing; when speakers were the same, one utterance was taken from two survey calls, so listeners never compared two identical stimulus tokens. Each of the ten voices appeared equally frequently; each voice occurred as both the first and second member of a "voices different" pair and in at least one "voices same" pair. Mean performance scores were known for the voice pairs used. As far as possible, "voices different" and "voices same" trials were matched for difficulty: 20 normal listeners aged 45-72 (mean age = 58.8, SD = 7.88) made an average of 4.08 errors on the "voices different" trials and 4.15 errors on the "voices same" trials (Kreiman, l987).

Response sheets for this task consisted of a column of 26 same/different choices. Clinical subjects were asked to say "same" or "different" (or to point to the written words "same" or"different" if a speech difficulty interfered with a spoken response) on hearing each stimulus pair. Normal subjects circled "S" or "D" on an answer sheet.

The 25 <u>familiar voice</u> stimuli consisted of brief excerpts of texts spoken by famous males politicians and entertainers (e.g., John F. Kennedy, Johnny Carson). Using a computer wave-form editing program, the voice samples were then edited to create 4-second stimuli free of pauses, background noises, and identifying content. A written pretest administered to a group of normal subjects demonstrated that no targets were identified at above chance from linguistic content alone.

For the clinical group, response sheets consisted of vertically aligned photographs of the target speaker and three foils, randomly ordered, with typed names next to each photograph. Linguistic content of the samples was neutral. Foils were matched to the target speaker for speaking style and perceived voice quality to challenge the listener to actually recognize the target voice and not to use cues from content, rhetorical style (i.e., politician versus comedian) or other deductive strategies.

For each test item, brain-damaged subjects were presented with the 4-choice response sheet; the four names were then read aloud and the stimulus voice played. Response alternatives were thus made available to these subjects in visual, written, and spoken forms (to compensate for any specific impairments in language comprehension, facial recognition or reading). Normal subjects circled one of 5 written names on an answer sheet.

At the end of this test session, for each target speaker, subjects were asked whether they felt they would normally recognize the voice. Responses were scored only for those voices a given listener claimed were familiar. This ensured that performance scores measured familiar voice recognition, and not merely familiarity with the set of test voices.

Subjects

In all, 56 brain-damaged patients were tested. One subgroup consisted of 23 LBD, 15 RBD patients, and 6 BBD subjects who were tested on both the voice recognition and the discrimination protocols. Another subgroup consisted of 12 unilaterally brain-damaged patients (2 LBD, 10 RBD) who were tested with familiar voice recognition only.

All the BD subjects had cerebral infarctions; all except two were right handed; all, without exception, were native speakers of American English, and were educated in the United States. In most cases, site of lesion was determined by computerized tomography (CT) supported by clinical and neurological data.

The LBD patients were tested using standardized language assessment instruments (Boston Diagnostic Aphasia Examination and Western Aphasia Battery). All were aphasic (7 fluent, 12 nonfluent, 6 anomic). In the LBD group, ages ranged from 41 to 80, with a mean of 61 years (SD=8.67) . RBD subjects ranged in age from 34 to 82, with a mean of 62.5 years (SD=10.54). The BBD subjects were ages 57 to 82, with a mean of 71 years (SD=10.38). In all three clinical groups, years of schooling ranged from 2 years of high school to postgraduate work. The normal-control group consisted of 48 subjects aged 51 to 85 years, with a mean of 64 years (SD=8.73). Their education ranged from 4 years of high school to postgraduate work.

Two kinds of analyses were conducted. First, a statistical analysis of group performance data compared normals (n = 48) and the unilaterally damaged patients (n = 38) who had performed both the recognition and discrimination protocols. Group means were obtained and an analysis of variance (ANOVA) performed to determine the effect of hemispheric side of lesion on discrimination and recognition abilities.

Second, the larger group (n = 56) was used to determine intrahemispheric lesion sites correlated with deficits in voice recognition and discrimination. A descriptive study is presented, followed by a quantitative analysis of observed clinicopathological correlations.

		Group			
		LBD	**RBD**	**Normal**	
Task					
Recognition	x	80.10	58.57	82.10	
	SD	13.79		23.01	11.39
Discrimination	x	77.07		68.21	87.19
	SD	13.30	17.54	9.96	

Table 1. Mean percentage correct scores for left-brain damaged (LBD), right-brain damaged (RBD), and normal-control groups on discrimination and recognition of voices. Standard deviations (SD) are also given.

Results

Statistical Analyses

Mean scores for each group (LBD, RBD and normal) on the recognition and discrimination tasks are given in Table 1. A two-way (group x task) ANOVA with repeated measures on task showed a significant effect of group on score (F (2,83) = 20.40, $p < 0.001$), a significant effect of task (F (1,83) = 4.58, $p < 0.04$), and a significant group x task interaction (F(2,83) = 3.64, $p < .03$). Post-hoc comparisons showed that, on the recognition task, LBD subjects did not differ from normals (F(1,69)=0.42, n.s.), while RBD subjects performed significantly worse than either normals (F (1.61) = 28.58, $p < 0.001$) or LBD subjects (F(1,36) = 13.07, p. < 0.001). Both LBD and RBD subjects performed significantly worse than normals on the discrimination task (LBD: F(1,69) = 12.84, $p <0.01$; RBD: F(1,61) = 28.01, $p < 0.01$). The two clinical groups did not differ in performance on this task.

In summary, an impairment in familiar voice recognition was found in association with RBD, whereas a deficit in unfamiliar voice discrimination was observed to be associated with damage to either hemisphere. LBD patients were impaired relative to normal subjects only in the voice discrimination task; RBD patients were impaired in both familiar voice recognition and unfamiliar voice discrimination. This suggests, in agreement with the preliminary study (Van Lancker and Kreiman, in press), that recognition and discrimination of voices are separate and independent abilities with different neuroanatomical substrates.

Clinical and Radiographic Analyses

More precise neuroanatomic substrates of phonagnosia were also sought. Two types of analysis were performed. First, a descriptive study was undertaken, to establish which lesions sites were associated with voice perception deficits. Second, a quantification of the data was performed, testing hypotheses derived from the descriptive analysis.

For the descriptive analysis, we examined records of 56 patients (25 LBD, 24 RBD and 6 BBD patients) tested on one or both voice protocols. Radiographic data were used to determine where in the right hemisphere (RH) damage was most often associated with a familiar voice recognition deficit, and which sites in either hemisphere were correlated with deficient unfamiliar voice discrimination.

CT scans of the head were obtained for 30 of the patients (12 RBD, 13 LBD, 5 BDD). The regions of infarction were mapped onto anatomical templates (Damasio, 1983; Matsui and Hirano, 1978). Lesion sites were classified by determining the lobe with the largest extent of the lesion (e.g., parietal, temporal, frontal, occipital). If the lesion involved two or three lobes equally, a combined lobar assignment was made (e.g., fronto-parietal, temporo-parieto-occipital). Subcortical lesions were indicated according to whether the lesion involved primarily the basal ganglia or the thalamus.

For this analysis, stringent criteria were utilized for classifying patients. A CT-scan must be available, and the patient was counted as deficient on one or the other task if he/she scored more than 2 standard deviations (SD) from the normal mean on either task. Of 29 unilaterally BD patients with available CT-scans who took the voice recognition test (18 LBD, 11 RBD), 9 were deficient in voice recognition. Seven of the patients with voice recognition deficits had lesions that included (n = 6) or undercut (n = 1) the right parietal lobe. Exceptions were two patients scoring in the deficient range who had large left frontal lesions. Conversely, none of the patients scoring in the normal range on voice recognition had lesions in the right parietal lobe.

Analysis of the 5 BBD patients agreed with findings in the unilaterally BD group. Two of these patients had right parietal lobe damage; both were severely impaired on voice recognition. The other three BBD patients scored in the high normal range on familiar voice recognition, and none had any observable damage in the right parietal lobe.

Of 20 (6 RBD, 14 LBD) patients who had taken the discrimination test and for whom CT-scans were available, those with voice discrimination deficits had lesions involving either the right or left temporal lobe, and the lesions tended to be large (extending beyond the temporal lobe in 4/6 cases). Three patients with a left frontal lobe lesions apparently sparing the temporal lobe also had an impairment of voice discrimination. Six patients with left temporal lobe lesions had normal voice discrimination; in one case the lesion involved only the anterior temporal pole, one involved the posterior and superior temporal regions, and four had large anterior left-sided lesions.

The 5 BBD patients conformed to this pattern of a temporal lobe involvement in voice discrimination abilities. The 4 with BBD sparing both temporal lobes scored in the high normal range on voice discrimination, and the fifth, having temporal lobe damage on both sides, performed at chance on the voice discrimination task.

To facilitate quantification of the neuroanatomical information, 9 patients (5 LBD, 4 RBD) for whom radiologists' reports were available were added and the criterion for having a deficit in either voice discrimination or voice recognition was liberalized to one SD from the normal mean. Using these revised criteria, patient information regarding lesion location and performance on the two voice protocols was used to obtain a Chi-Square statistic. The total group submitted to the Chi-Square analysis consisted of 17 RBD, 21 LBD, and 5 BBD patients. From the available neurological data we could reliably classify patients into cells to test two hypotheses derived from the descriptive analyses described above.

The first hypothesis is that abnormal performance in familiar voice recognition is associated with injury to the right parietal lobe; and the second hypothesis is that abnormal performance on unfamiliar voice discrimination is associated with damage to either right or left temporal lobes.

Subjects were sorted according to whether they attained high scores or low scores (more than one SD away from the normal mean) on either task, and whether or not they had documented damage to the right parietal (for the recognition task) or to either temporal lobe (for the discrimination task). A separate matrix was prepared for each task. That is, for the recognition task, each subject was identified as "high" or "low" on performance, and as having "right parietal lobe damage" or "damage elsewhere"; for the discrimination task, each subject was identified as "high" or "low" on performance and as having "left or right temporal damage" or "damage elsewhere."

		RIGHT PARIETAL	ELSEWHERE
Score	HI	0	27
	LO	9	7

Figure 1. Matrix for voice recognition scores (Chi-square = 15.96, $p < 0.01$).

Figure 1 shows the results for voice recognition. As was seen in the descriptive study, this analysis also shows that no subjects having right parietal lobe damage performed normally on voice recognition. Twenty-seven patients with lesions elsewhere than the right parietal lobe (LBD=19, RBD = 6, BBD = 2) did perform normally on the voice recognition task. Nine patients (7 RBD and 2 BBD) having damage to the right parietal lobe showed impaired performance on the task. Exceptions to the pattern are 6 patients with damage elsewhere than the right parietal lobe who also showed impaired performance on familiar voice recognition. The Chi-square (with 1 degree of freedom, with Yates' correction for continuity) = 15.96 ($p < 0.01$).

	Damage	
	L OR R TEMPORAL LOBE	ELSEWHERE
HI	4	9
LO	13	4

Figure 2. Matrix for voice discrimination scores (Chi-square = 4.54, $p < 0.05$).

The matrix for unfamiliar voice discrimination is shown in Figure 2. Patients were sorted by their high or low scores on the task and by the presence or absence of a lesion in either left or right temporal lobe, or elsewhere in the brain. CT-scans or radiologists reports were available for 25 patients on whom voice discrimination scores had been obtained. The presence of a temporal lesion on either side was significantly associated with low performance on voice discrimination. Of the 25 patients examined, 13 with temporal lobe damage had deficient performance on the voice discrimination task, while 9 without temporal lobe damage (damage elsewhere) performed normally on the task. Eight patients (of 25) did not conform to this pattern: Four patients with temporal lobe damage performed in the normal range; all had left hemisphere lesions. In addition, four patients with damage elsewhere performed poorly on the task. Two had large, right-sided subcortical damage undercutting the temporal lobe, a third had a large left frontal lesion apparently sparing the temporal lobe, and the fourth had anterior-parietal damage on both sides, possibly involving superior-posterior portions of the temporal lobes. Chi-square (with 1 degree of freedom, including Yates' correction for continuity) = 4.54, $p < 0.05$.

Discussion

Several conclusions can be drawn from the current study. A RH specialization for familiar voice recognition abilities was found, supporting similar findings of Van

Lancker and Canter (l982) and Van Lancker and Kreiman (in press). Assal, Zander, Kremin and Buttet (l976) reported only a trend toward an association of RH damage with voice perception deficits, but the lack of a strong RH finding in that study may reflect the use of unfamiliar voices in a discrimination task. This study gave further evidence of a previously reported double dissociation for the two aspects of voice perception, recognition and discrimination (Van Lancker, Cummings, Kreiman and Dobkin, in press). Of 43 brain-damaged patients performing both tasks, there were 7 who could discriminate well but recognized voices very poorly (below 2 SD from the normal mean), and 9 patients who could recognize well but not discriminate voices (below 2 SD). The double dissociation, along with the different neuroanatomical substrates described in this paper, indicates that the two skills are independent. Dissociation of recognition and discrimination abilities has been reported for a similar faculty in the visual modality, face perception (Malone, Morris, Kay and Levin, l982; Benton and Van Allen, l972; De Renzi, Faglioni, and Spinnler, l968).

The correlation of familiar voice recognition deficits with right parietal lesions is compatible with current theories about RH function. First, voice identity is a type of nonverbal information carried in speech prosody, and several studies have found nonlinguistic-prosodic comprehension to be mediated by the RH (Ross, l981; Kent and Rosenbek, l982; Heilman, Scholes and Watson, l975). Further, the voice is a complex auditory pattern. A current theory of hemispheric specialization describes the right hemisphere as a holistic pattern recognizer, whereas the left hemisphere excels at sequential and featural analysis (Bradshaw and Nettleton, l983; Bogen, l969; Bever, l975). This model ascribing different modes of processing to left and right cerebral hemispheres might be invoked to explain the differences seen in this study between performance of familiar-recognition and unfamiliar-discrimination voice tasks. Familiar voice recognition requires the association of an auditory pattern with a mental "trace" of a person, as has been postulated for face recognition by Damasio, Damasio and Van Hoesen (l982). Such cognitive processes occur in the posterior "heteromodal" association areas (Mesulam, l985). In contrast, evaluating unfamiliar voices is more purely a "unimodal" auditory function, and requires comparison of selected features in the voice pattern without multimodal processing or reference to another stored entity. This activity engages the temporal lobe, with its specialized auditory abilities and, for optimum performance, requires both hemispheres, as both featural and holistic strategies are needed. These neuroanatomical findings give further support to the notion that recognizing a familiar voice is neuropsychologically different from distinguishing among unfamiliar voices.

This study represents, in part, an effort to "map" the RH, in a tradition of localizing functions, and suggests that voice recognition is disturbed by right parietal lesions, whereas voice discrimination may be disrupted by lesions of either temporal lobe. This lesion approach leads to the conclusion that the right parietal is principally responsible for mediating voice recognition, and functions of both temporal lobes are necessary for competent voice discrimination. The few observed exceptions to these rules may be attributed to plastic central nervous system properties, individual differences in brain organization, alternate strategies for performing tasks, neural reorganization, and recovery of neural function after partial brain damage.

Voice recognition plays a key role in phylogenetic history, infant development, and in communicative interaction. Recent field work with nonhuman primates has shown that maternal Vervet monkeys recognize the voices of their own offspring (Cheney and Seyfarth , I980), and that this information is used in maintaining family contact. Human research has shown that 3-day-old infants recognize the voices of their own mothers from a set of maternal voices (DeCasper and Fifer, I980), and that young children can recognize familiar voices as well as can adults (Mann, Diamond and Carey, I979). For daily communicative interaction, it is obvious that voice recognition is crucial, and that it occurs in tandom with language comprehension. The identity of the speaker and the linguistic-phonetic information are extracted simultaneously from the acoustic material of the speech signal. The data reported here indicate that while the left hemisphere processes the linguistic-phonetic information in speech, the right hemisphere, from that same signal, establishes the identity of the speaker.

Acknowledgments

We wish to express our appreciation to Bruce Dobkin, M.D., Wayne Hanson, Ph.D., Gail Monahan, M.S, Bette Hadler, M.S, E. Jeffrey Metter, M.D., and Douglas Noffsinger, Ph.D. for assistance in obtaining patient information; we thank D. Wesley Burgess, Ph.D., for help in preparing stimulus materials. This project was supported by the Veterans Administration.

References

Assal, G., Zander, E., Kremin, H., and Buttet, J. (l976). "Discrimination des voix lors des lesions du cortex cerebrale." *Archives Suisses de Neurologie, Neurochirurgie, et de Psychiatrie,* ll9, 307-315.

Benton, A. L. and Van Allen, M.W. (l972). "Prosopagnosia and facial discrimination. " *Journal of Neurological Science,* 15, 167-72.

Bever, T. C. (l975). "Cerebral asymmetries in humans are due to the differentiation of two incompatible processes: holistic and analytic." *Annals of the New York Academy of Science*, l9, 251-62.

Bogen, J. (l969). The other side of the brain. *Bulletin of the Los Angeles Neurological Societies,* 34, 73-105.

Bradshaw, J.L. and Nettleton, N. C. (l983). *Human Cerebral Asymmetry.* Englewood Cliffs, New Jersey: Prentice-Hall.

Bricker, P. D. and Pruzansky, S. (l976). "Speaker recognition." In *Contemporary Issues in Experimental Phonetics.* N. J. Lass, ed., pp. 295-326. New York:Academic Press.

Cheney, D. L. and Seyfarth, R. M. (l980). "Vocal recognition in free ranging Vervet monkeys." *Animal Behavior* 28, 362-367.

Damasio, A.R., Damasio, H., and van Hoesen, G.W. (l982). "Prosopagnosia: Anatomic basis and behavioral mechanisms." *Neurology*, 32, 331-341.

Damasio, H. (l983). "A computed tomographic guide to the identification of cerebral vascular territories". *Archives of Neurology*, 40, 38-l42.

DeCasper, A. and Fifer, W. (l980). "Of human bonding: Newborns prefer their mother's voices." *Science* 208, 1174-1176.

De Renzi, E., Faglioni, P., and Spinnler, H. (l968). "The performance of patients with unilateral brain damage on facial recognition tests." *Cortex*, 4,17-34.

Heilman, K., Scholes, R., and Watson, R. (l975). "Auditory affective agnosia." *Journal of Neurology, Neurosurgery, and Psychiatry*, 38, 69-72.

Kent, R. D. and Rosenbek, J. C. (l982). "Prosodic disturbance and neurologic lesion." *Brain and Language,* 15, 259-291.

Kreiman, J. (l987). Human memory for unfamiliar voices. (Doctoral dissertation, University of Chicago), to appear in *UCLA Working Papers in Phonetics.*

Malone, D. R., Morris, H. H., Kay, M. C., and Levin, H.S. (l982). "Prosopagnosia: a dissociation between recognition of familiar and unfamiliar faces." *Journal of Neurology, Neurosurgery, and Psychiatry,* 45, 820-822.

Mann, V. A., Diamond, R., and Carey, S. (l979). "Development of voice recognition: parallels with face recognition." *Journal of Experimental Child Psychology,* 27, 33-40.

Matsui, T. and Hirano, A. (l978). *An Atlas of the Human Brain for Computerized Tomograp*hy. Tokyo, Igaku-Shoinu, Ltd.

Mesulam, M-Marsel. *Principles of Behavioral Neurology.* Philadelphia: F.A. Davis, l985.

Ross, E. (l981). "The aprosodias: functional-anatomic organization of the affective components of language in the right hemisphere." *Archives of Neurology,* 38, 561-569.

Van Lancker, D. and Canter, J. l982. "Impairment of voice and face recognition in patients with hemispheric damage". *Brain and Cognition*, 1, l85-l95.

Van Lancker, D. and Kreiman, J. (in press). "Unfamiliar voice discrimination and familiar voice recognition are independent and unordered abilities."

Neuropsychologia.

Van Lancker, D., Cummings, J., Kreiman, J. and Dobkin, B.H. (in press). "Phonagnosia: A dissociation between familiar and unfamiliar voices." *Cortex.*