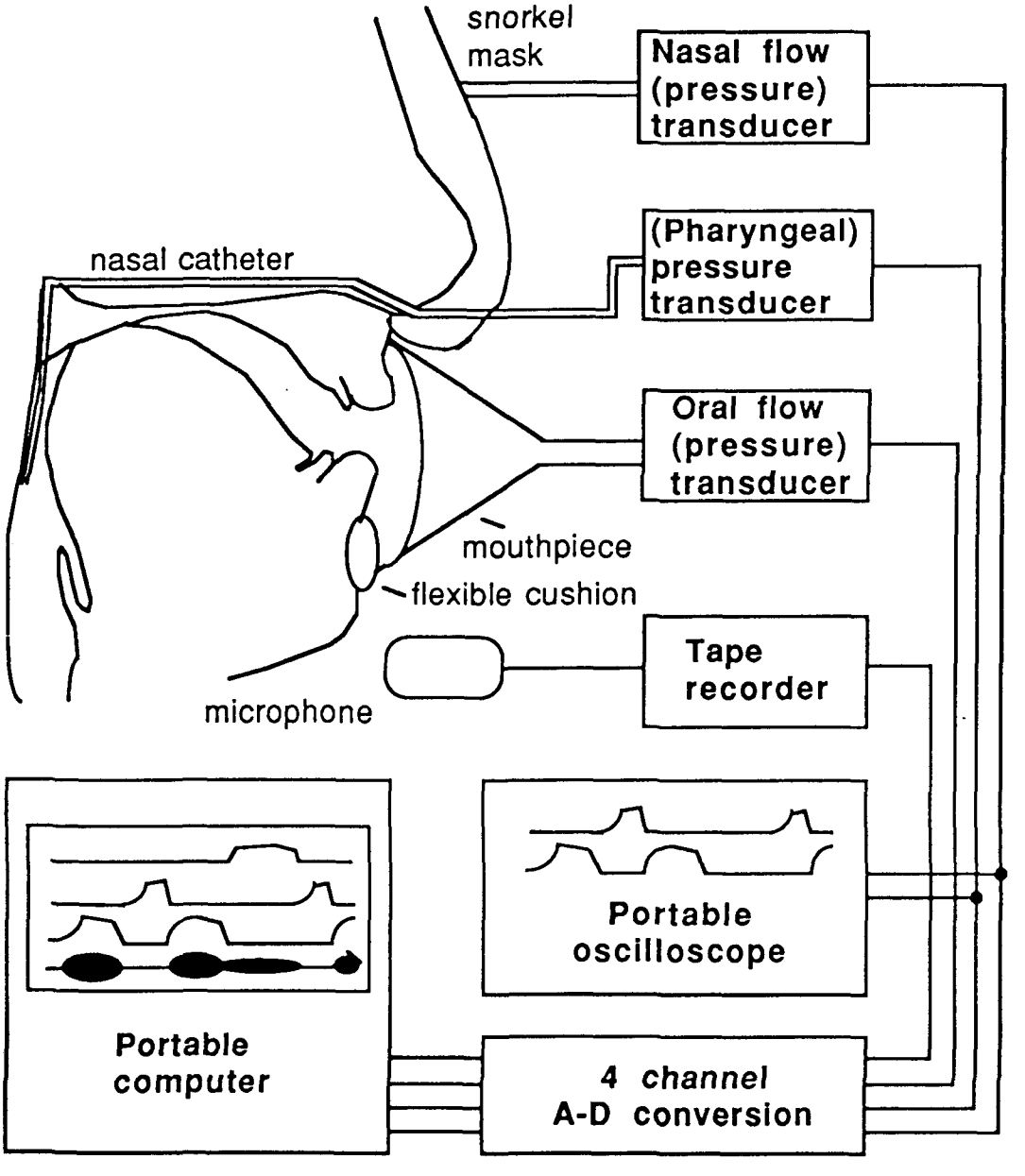
"All the tools of instrumental phonetics can never replace the ears of a skilled phonetician"
(P. Ladefoged 1991)



UCLA WORKING PAPERS IN PHONETICS NO. 78

FEBRUARY 1991

The UCLA Phonetics Laboratory Group

Beatriz Amos
Sue Banner-Inouye
Barbara Blankenship
Dani Byrd
John D. Choi
Sarah Dart
Deborah Davis
Sandy Disner
Edward Fleming
Vicki Fromkin
Robert Hagiwara
Bruce Hayes
Susan Hess
Keith Johnson
Pat Keating
Paul Kirk
Jenny Ladefoged
Peter Ladefoged
Mona Lindau
Ian Maddieson
Paroo Nihalani
Kristin Precoda
Maria Psihountas
Bonny Sands
Stephan Schuetze-Coburn
Andrea Schwartz
Aaron Shryock
Donca Steriade
Henry Teheranizadeh
Camille Teran
Kimberly Thomas
Cynthia Walker

As on previous occasions, the material which is presented in this volume is simply a record for our own use, a report as required by the funding agencies which support the Phonetics Laboratory, and a preliminary account of research in progress for our colleagues in the field.

Funds for the UCLA Phonetics Laboratory are provided through:

NSF grant BNS 87-20098
USHHS grant NS-22726
USHHS grant 1 RO1 DC00642
USHHS grant 1 T32 DC00029
and the UCLA Department of Linguistics.

Correspondence concerning UCLA *Working Papers in Phonetics* should be addressed to:

Phonetics Laboratory
Department of Linguistics
UCLA
Los Angeles, CA 90024-1543
(U.S.A.)

UCLA *Working Papers in Phonetics* is edited by Ian Maddieson.

UCLA Working Papers in Phonetics 78

February 1991

Table of Contents

Peter Ladefoged	Computerized phonetic fieldwork	1
Peter Ladefoged	What do we symbolize? Thoughts prompted by bilabial and labiodental fricatives.	7
Ian Maddieson	Testing the universality of phonological generalizations with a phonetically-specified segment database: results and limitations.	11
Ian Maddieson	Investigating linguistic universals	26
Ian Maddieson and Kristin Precoda	Syllable structure and phonetic models	38
Deborah S. Davison	An acoustic study of so-called creaky voice in Tianjin Mandarin	50
Deborah S. Davison	Stress and tonal targets in Tianjin Mandarin	58
John D. Choi and Patricia Keating	Vowel-to-vowel coarticulation in three Slavic languages	78
Kimberley D. Thomas	Vowel length and pitch in Yavapai	87
Dani Byrd	Perception of assimilation in consonant clusters: a gestural model.	97
Peter Ladefoged	Announcements	127

COMPUTERIZED PHONETIC FIELDWORK*

Peter Ladefoged

The days of Henry Higgins and his notebooks are passed, and battery operated computers and printers are now major assets for the collection of phonetic data in the field. Computers can be used as sophisticated audio editing and listening devices, and they also provide informative acoustic analyses. When combined with air pressure and flow transducers they also enable physiological data analysis to be carried out in the field with language consultants who might not otherwise be available.

Henry Higgins recorded Eliza Doolittle in Covent Garden simply with the aid of pen and paper (Shaw 1914). Right up to World War II phoneticians had to rely almost entirely on their skill in making phonetic transcriptions in order to have any memory of the sounds they were studying in the field. For almost half a century since then the most important instrument available to the phonetician has been the tape recorder. Nowadays few phoneticians would start a fieldwork investigation of a language without having some means of making recordings. However, as computers become more available, the days of dependence on tape recorders may be passing. In many kinds of linguistic fieldwork it may be preferable to record directly onto a portable computer, with the tape recorder being used simply as a backup.

Linguistic fieldwork can take all sorts of different forms. The fieldworker might be, like Henry Higgins, standing behind a church pillar in Covent Garden, or, more probably in these days when secret recording is less favored, sitting in a croft in the Scottish Highlands or in a camp in the Kalahari Desert. But wherever it might be taking place, good fieldwork should always cause the minimum distraction to the people being recorded. There should be no fuss, no busying around setting up equipment, trying to find where to plug it in. Instead one should be able to walk in quietly, produce a microphone from one's pocket, and start work. When more rapport has been established with the community, one can take more trouble about careful placement of the microphone, quieting the goggling children, and shooing the chickens away from the door. As we will discuss later, other equipment can be brought in so that one has much of the power of a full-fledged phonetics laboratory. But it is always true that the less ostentatious one can be, the better the fieldwork situation. Accordingly, in this paper we will be concerned only with portable computers and peripherals which have the advantage that not only can they be used quietly and quickly, but also that they fit in the overhead compartment on an airplane.

Fieldwork computers are, of course, much more than devices for recording and reproducing sounds. But even in these basic tasks they are far more versatile than tape recorders. Typically, phonetic fieldworkers are recording word lists or short paragraphs containing very similar sounding words. They often want to be able to play back selected pieces over and over again, so that they can hear subtle nuances of sounds that are new to them. They also want to be able to hear one sound, and then, immediately afterwards, hear another that may contrast with it. Both tasks can be done somewhat cumbersomely and tediously using tape recorders. But they are trivial, normal operations on any computer equipped with a means for digitizing and editing recorded sounds. We may take as an example the problem of distinguishing between so-called stiff and slack syllables in Bruu, a Mon-Khmer language spoken in Thailand. Suitable words can easily be recorded onto a computer, using the equipment listed in the appendix at the end of this paper. A pair of syllables illustrating the contrast can then be cut out from the surrounding sounds, and displayed on a computer screen as shown in Figure 1. Clicking on the icon representing a loudspeaker in either of the windows on the screen will play the utterance in that window. This makes it easy, or at least easier, to hear the difference between the two syllables. The comparison is also helped by the fact that the waveforms show that the stop at the beginning of the word in the lower window is slightly more aspirated than the stop in the upper window.

* A minimally different version of this paper is in the *Proceedings of the Third Australian International Conference on Speech Science and Technology.* (Ed. Roland Seidl) Australian Speech Science and Technology Association: Canberra. 1990.

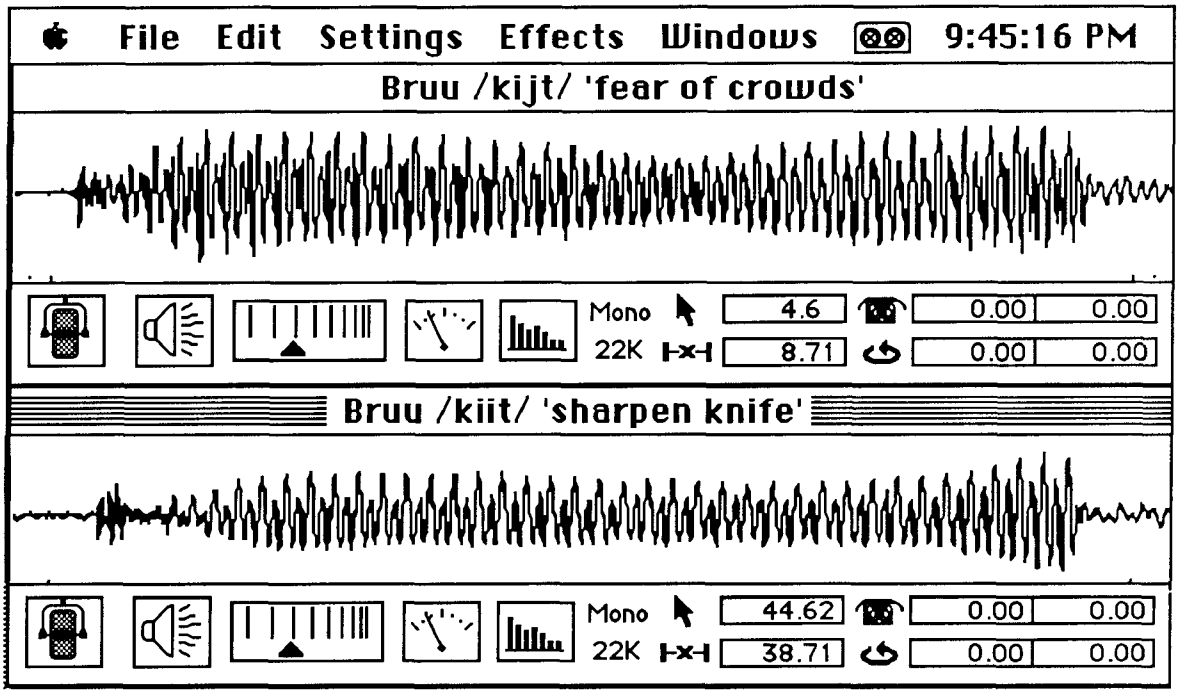


Figure 1. Computer screen showing two words in Bruu, a Mon-Khmer language. Either word can be played by clicking on the speaker icon in its window.

In addition to being useful as a sophisticated playback device, a computer can provide several types of acoustic analysis that a fieldworker might find useful. Consider, for example, the difficulties in analyzing the four contrastive sibilants that occur in Toda, a Dravidian language spoken in the Nilgiri Hills in India. The best way of determining the general acoustic characteristics of these sounds in the field is to make a spectrogram of the kind shown in Figure 2. Each of the words ends in a different sibilant. The laminal dental sibilant at the end of the first word has the highest frequency, and the retroflex sibilant at the end of the last word has the lowest. The apical alveolar and (laminal) palatoalveolar sibilants at the ends of the second and third word have very similar spectral characteristics. (The lowering of the spectral energy peak at the end of the second word is a non-distinctive feature, being simply due to the closure of the lip for the consonant at the beginning of the next word.) These two sibilants are distinguished primarily by their on-glides. The increasing second formant at the end of the third word is due to the raising of the blade and front of the tongue for this laminal sound. In the last word, the lowering of the third formant is probably due to the sublingual cavity that is formed by raising the tip of the tongue for this retroflex sibilant. All these points are readily observable in the recordings of these words; but it is much easier to hear them after one has seen them.

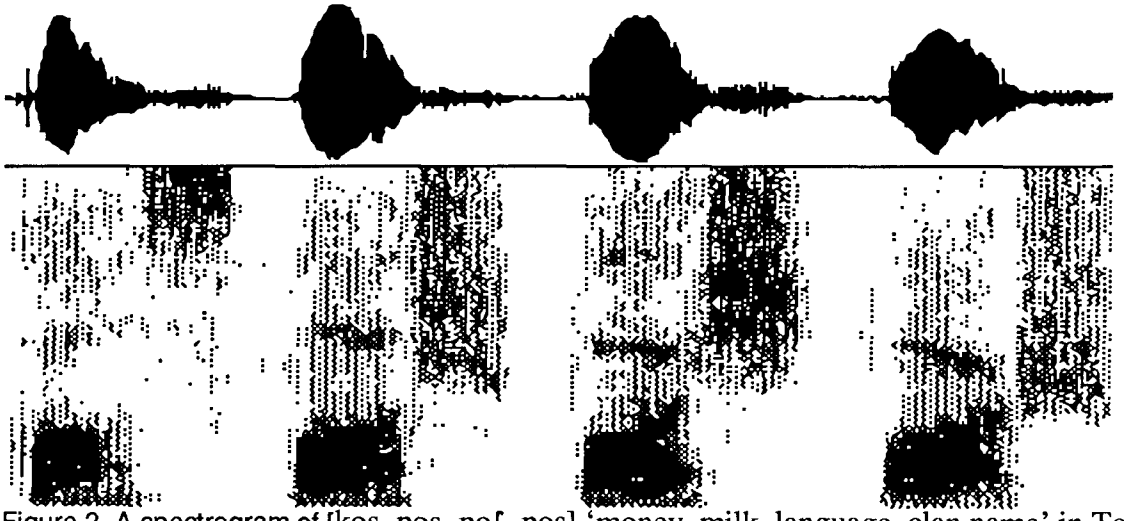


Figure 2. A spectrogram of [koʂ, poʂ, poʃ, poʂ] 'money, milk, language, clan name' in Toda, made under field conditions.

The spectrogram in Figure 2 was produced on a portable computer without a color (gray scale) screen. The figure is an unretouched copy of the printout from a battery operated printer no bigger than my hand, physically cut and pasted into this paper. As portable computer screens and printers improve so that they produce better gray scale or color output, this kind of display will become even more useful. Of course, still more information can be obtained from high quality spectrograms produced by this or another program on a laboratory computer after one has returned from the field.

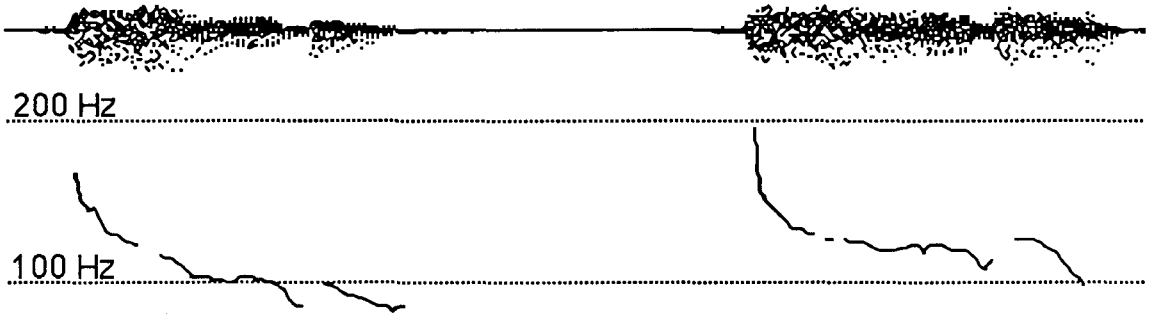


Figure 3. The tonal contrast between /ku!laamba/ 'to lick', and /kulaamba/ 'to be dear, in Sukuma, a Bantu language spoken in Tanzania.

Another kind of analysis that is very useful to the fieldworker is one that indicates the pitch. An excellent way of doing this is to make a narrow band spectrogram, comparable to the wide band spectrogram shown in Figure 2. Displays of this kind are very useful for pitch analysis when a creaky voice quality or other unusual spectral characteristics are involved. But it is often possible to use a simpler kind of display. Figure 3 shows the fundamental frequency in a pair of words with contrasting tones in Sukuma, a Bantu language of Tanzania. The first word, /ku!laamba/ 'to lick', contrasts with the second word, /kulaamba/ 'to be dear', by having a greater decrease in pitch (a downstep) after the the first syllable.

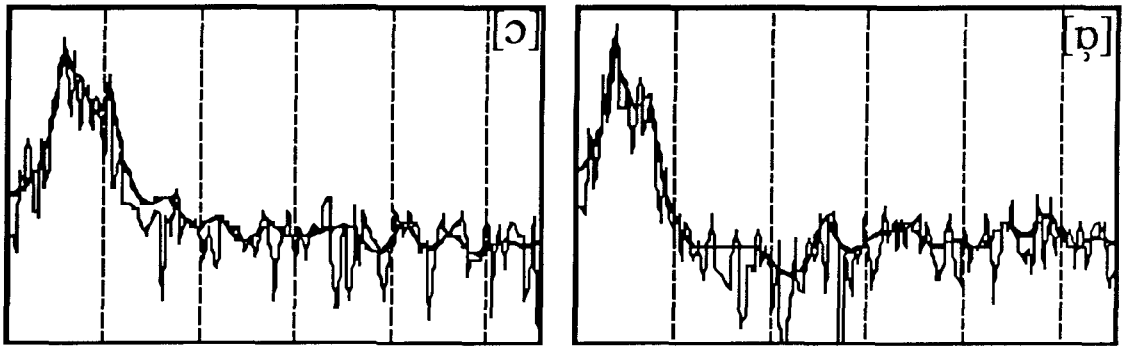


Figure 4. Superimposed FFT and LPC spectra of two Assamese vowels.

The final kind of computer analysis of speech sounds that will be illustrated here is one for determining the formant frequencies, the principal aspects of vowel quality. A common way of obtaining formant frequencies is by inspection and peak picking using superimposed LPC and FFT spectra of the kind shown in Figure 4. The original display has been somewhat reduced in size so as to fit into this paper. The figure shows analyses of two Assamese vowels that are very similar in quality; one is fairly similar to the IPA reference vowels [ɔ], and the other is a vowel with a tongue position like that in [ɑ], but with a lip position more like that in [u]. The analyses suffer from having been made on sounds sampled at 8 bits under field conditions, but the crucial differences are still evident. The first and second formants are at slightly higher frequencies for the higher vowel in the lefthand part of the figure – an unusual situation in that higher vowels normally have lower first formant frequencies. The considerable rounding that occurs in the vowel in the righthand part of the figure has caused both formants to be lower, and it has also caused a very sharp decrease in spectral energy immediately above the second formant.

The value of a portable computer in phonetic fieldwork extends beyond its use in recording and analyzing sounds. It can also be an important part of a system for recording aerodynamic data.

Records of the pressure of the air in the mouth and of the airflow from the nose and the lips have been used for over a century in phonetic research, dating back, in an unquantified way, to the kymograph tracings of the early experimental phoneticians. For many years we have been able to make good, calibrated, records of these variables (Ladefoged 1967). Now this ability is available to the field phonetician. Figure 5 shows a schematic of the instrumentation employed. The language consultant speaks into a specially constructed mouthpiece pressed against the face, which takes the oral air flow through a calibrated resistance so that a pressure transducer provides a signal that is directly proportional to the rate of air flow. The air flow from the nose is similarly transduced after having been captured within a snorkeling mask, which completely covers the upper part of the face; it is difficult to get well-fitting mask that covers just the nose. The air flow mask devised by Rothenburg (1973) does not provide adequate separation between oral and nasal channels, although it is excellent for recording air flow when a high frequency response is required.

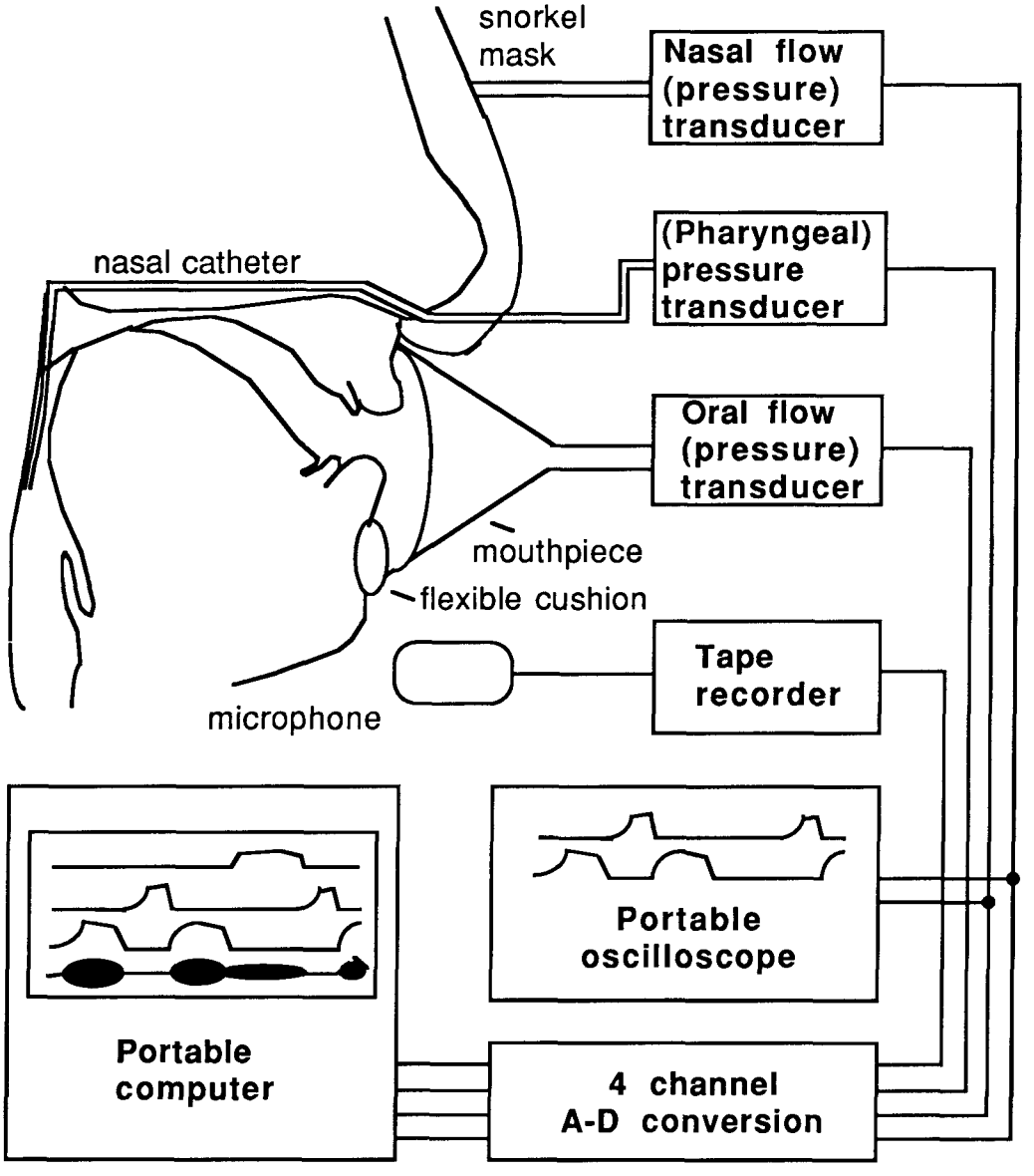


Figure 5. Apparatus for obtaining aerodynamic records in the field.

If one can find a language consultant who is willing to tolerate a nasal catheter, then it is possible to record the pressure build up behind stop closures anywhere in the vocal tract. Alternatively a simple way of obtaining data on just bilabial sounds is to use a small tube inserted between the lips. Another aerodynamic parameter that is useful to record is the subglottal pressure if the speaker is willing, and the system permits recording more than 3 channels. This would involve a second tube through the nose that passed down into the oesophagus so that its pressure sensing end was below the level of the vocal cords.

All these parameters have to be digitized along with the audio signal from a microphone. The sample rate required for digitizing the physiological parameters depends on the kind of information that is being sought. Records of the oral airflow can be filtered so as to remove the formant resonances and reveal the shape of the waveform produced at the glottis If this is the intent, then an appropriate face mask such as that described by Rothenberg (1973) will have to be used, and the oral airflow should be digitized at 5,000 Hz. But if the intent is to use the physiological parameters to show the gross movements of the vocal organs, then a sample rate of 500 Hz is sufficient.

The system in Figure 5 also shows two of the channels being monitored on a portable oscilloscope. There are two reasons for this. Firstly, an oscilloscope provides an easy way to check the base lines and gains of the pressure transducers when setting up the apparatus. Computer systems are usually less tolerant in the range of input values which they will accept, and gross adjustments are more easily handled on a flexible device such as an oscilloscope. Secondly, as all fieldworkers know, things go wrong at crucial moments, and it is always good to have a back up system available. If the computer system does not work, it is still possible to display some of the aerodynamic variables on a storage oscilloscope and photograph the screen with an ordinary camera.

Figure 6 is an example of some aerodynamic data recorded in the Kalahari Desert. The first word begins with an alveolar click with delayed aspiration, in which the air pressure behind the velar closure of the click is released through the nose. The second word begins with a palatal click which is voiced during the closure, but voiceless and aspirated during the release. Computerized phonetic fieldwork is a great asset in recording the details of these complex sounds.

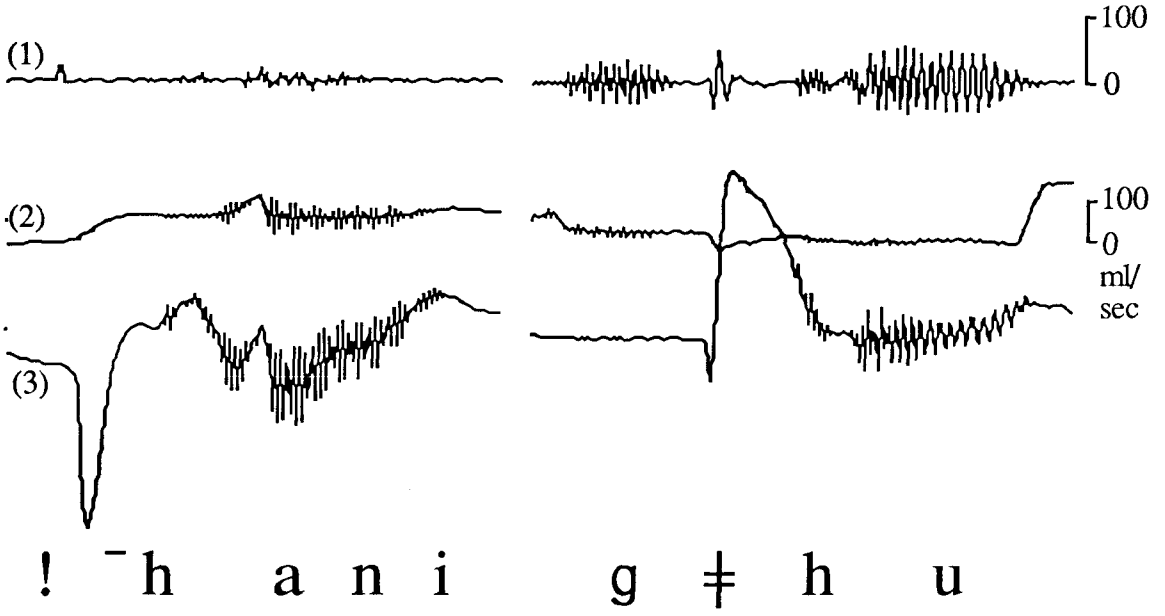


Figure 6. Records of (1) audio, (2) nasal air flow, and (3) oral air flow during the pronunciation of two words in Zhu|oasi, a Khoisan language spoken in Namibia.

Finally, in a technical conference such as this, we should consider the specifications of a computer system that is suitable for fieldwork. Clearly, to be linguistically distinctive the relevant acoustic properties of speech sounds must be well within the auditory limitations of the human ear. People in their prime can hear sounds with frequencies around 20,000 Hz, but this ability declines rapidly with age. No languages relies on distinctions that can be heard only by young adults, and a frequency range up to 12,000 Hz should be sufficient for linguistic purposes. In fact, although some speech sounds (particularly voiceless stops and fricatives) may differ significantly in the amplitude of the frequency components in the 10 to 20,000 Hz range, the major differences (even in voiceless stops and fricatives) are below 8,000 Hz. In order to allow some tolerance in the system, the sample rate should be 2.5 times the highest frequency present. Accordingly, the computer system should be capable of sampling speech at 20-24,000 Hz for high quality listening and analysis.

The frequency range must also be considered when making an FFT. In effect, an FFT provides the amplitudes of the spectral components that are present on the assumption that these components are all multiples of a wave corresponding to the number of points in the FFT. The greater the number of points in the FFT, the longer the wave length, thus the lower the frequency of this wave, and the smaller the interval between calculated components. But any program calculating an FFT will have a certain maximum number of points permissible (usually something like 512 or 1024). Accordingly, the only way to further increase the accuracy in the frequency domain (i.e. to decrease the interval between measured components) is to *decrease* the sample rate. This will have the effect of decreasing the range of frequencies that can be observed. But it will also mean that all the components calculated will be within that range. Given a 512 point FFT and a sample rate of 20,000 Hz, there will be 256 components spaced about 40 Hz apart in the range up to 10,000 Hz. But if the sample rate is reduced to 10,000 Hz, the 256 components will be spaced about 20 Hz apart in the range up to 5,000 Hz, and displayed frequencies can be discriminated more precisely. When vowel formants are being studied it is advisable to use a lower sampling rate. The alternative would be to use an FFT with a larger number of points, but no analysis system will permit the maximum number of points to be increased beyond some fixed limit.

The minimum sample size is more difficult to determine. Compact disks used in high fidelity audio systems specify amplitudes in terms of 32,000 possible levels (16 bits), which allows for the maximum signal level recorded to be 96 dB above the system noise. There is no doubt that a Signal/Noise ratio of this magnitude is highly desirable. But in most fieldwork situations it is difficult to record in an environment in which the background noise is as much as 48 dB below the level of a speaker who is as close as possible to the microphone. There is always a baby crying in the next room, or the wind in the trees, or some other noise that cannot be avoided. Accordingly 256 possible levels (8 bits) providing a 48 dB Signal/Noise ratio may be sufficient, provided that a great deal of care is taken to ensure that the full range is used (i.e. that the signal to be recorded is at the maximum level possible without overloading). Again it should be emphasized that if a battery operated 12 or 16 bit system is available, and if the computer has sufficient memory, then the added margin of safety is highly desirable. But 8 bit systems providing a 48 dB Signal/Noise ratio (such as the MacRecorder used by the SoundEdit program illustrated above) are satisfactory for most purposes. As a point of reference it is worth noting that 48 dB is about the Signal/Noise ratio of a good laboratory tape recorder. Few portable cassette recorders (even so-called professional models) have a Signal/Noise ratio greater than 50 dB.

APPENDIX

Equipment and programs used for producing the data in this paper

Name	Used in figure	Manufacturer
Macintosh Portable	1-6	Apple Computer
WriteMove (printer)	2	GCC Technologies
MacRecorder	1-4	Farallon 2201 Dwight Way, Berkeley, CA 94704
SoundEdit	1	Farallon (supplied with MacRecorder)
Signalyze	2	InfoSignal, 231 Belair E. Rosemere, Quebec J7A 1A9, Canada
Uppsala/UCLA SoundWave	3,4	University of Uppsala, Sweden, and Linguistics, UCLA, Los Angeles, CA 90024
Multi-channel aerodynamic digital recording system	5,6	Linguistics, UCLA, Los Angeles, CA 90024

REFERENCES

Ladefoged, P. (1967) *Three areas of Experimental Phonetics.* London: Oxford University Press.

Rothenberg, M.A. (1973) "A new inverse-filtering technique for deriving the glottal air flow waveform during voicing". J. Acous. Soc. of Am. 53.6, 1632-1645.

What do we symbolize?
Thoughts prompted by bilabial and labiodental fricatives

PETER LADEFOGED

(To appear in *Journal of the International Phonetic Association*)

The contrast between bilabial and labiodental fricatives is comparatively rare in the world's languages. In his survey of 317 languages Maddieson (1984) notes only one that has the full set of voiced and voiceless sounds [ɸ, β, f, v], and only four other languages that contrast just the voiced pair. There is no doubt that these contrasts do occur, and that phonologists such as Sagey (1986) are wrong in neglecting to provide for them. This note will provide further particulars of the contrast in languages in which it has not been so well documented in the general linguistic literature. It will also suggest that this contrast may not be made in the same way in all languages, and discuss the implications for the IPA and for theories of phonetic description.

The best known language containing these contrasts is Ewe, a Kwa language in the Niger-Congo family spoken in Ghana. Some of the neighboring closely related languages or dialects such as Avatime and Logba also have these sounds. Ladefoged (1968) contains a description based on a total of 5 speakers of these West African languages. In these languages, as is shown by the photographs published in Ladefoged (1968), in [β] the lips are narrowed by being compressed in the vertical direction; but there is no sign of any form of rounding through the corners of the lip having been brought forward. In [v] the upper lip is actively raised so as to increase the difference between this sound and the bilabial [β]. All 5 speakers of Kwa languages used this active raising of the upper lip in the formation of the contrasting labiodentals. None of them had any form of lip rounding in either sound.

The contrast between bilabial and labiodental fricatives also occurs in a number of Southern Bantu languages. In these languages there is no upper lip raising in the labiodentals, and the contrast is enhanced in another way, namely by retracting the lower lip when making the labiodentals [f, v], and (sometimes) adding a slight lip protrusion when making the bilabials [ɸ, β]. We investigated contrasts as produced by three speakers of Kwangali, and three speakers of Rugciriku, two Bantu languages in the Kovango group, spoken in Namibia and Southern Angola. Both these languages contrast voiced bilabial and labiodental fricatives, but do not have voiceless sounds of this kind. Both voiced and voiceless bilabial and labiodental fricatives occur in Venda, a major Bantu language of South Africa; we also recorded a single speaker of this language. None of these seven speakers enhanced the bilabial - labiodental contrast by raising the upper lip for the labiodental sound, as did the West African speakers. Instead the contrast was enhanced by drawing the lower lip back over the lower teeth for the labiodental, and (for some speakers) bringing the corners of the lips forward to make a more slightly rounded version for the bilabial (the latter gesture was present in only 3 of our 7 speakers).

Simultaneous full face and side view photographs and sketches based on the originals of these photographs of one of the speakers of Kwangali are shown in Figure 1. The retraction of the lower lip when producing the labiodental fricative can be seen both by the

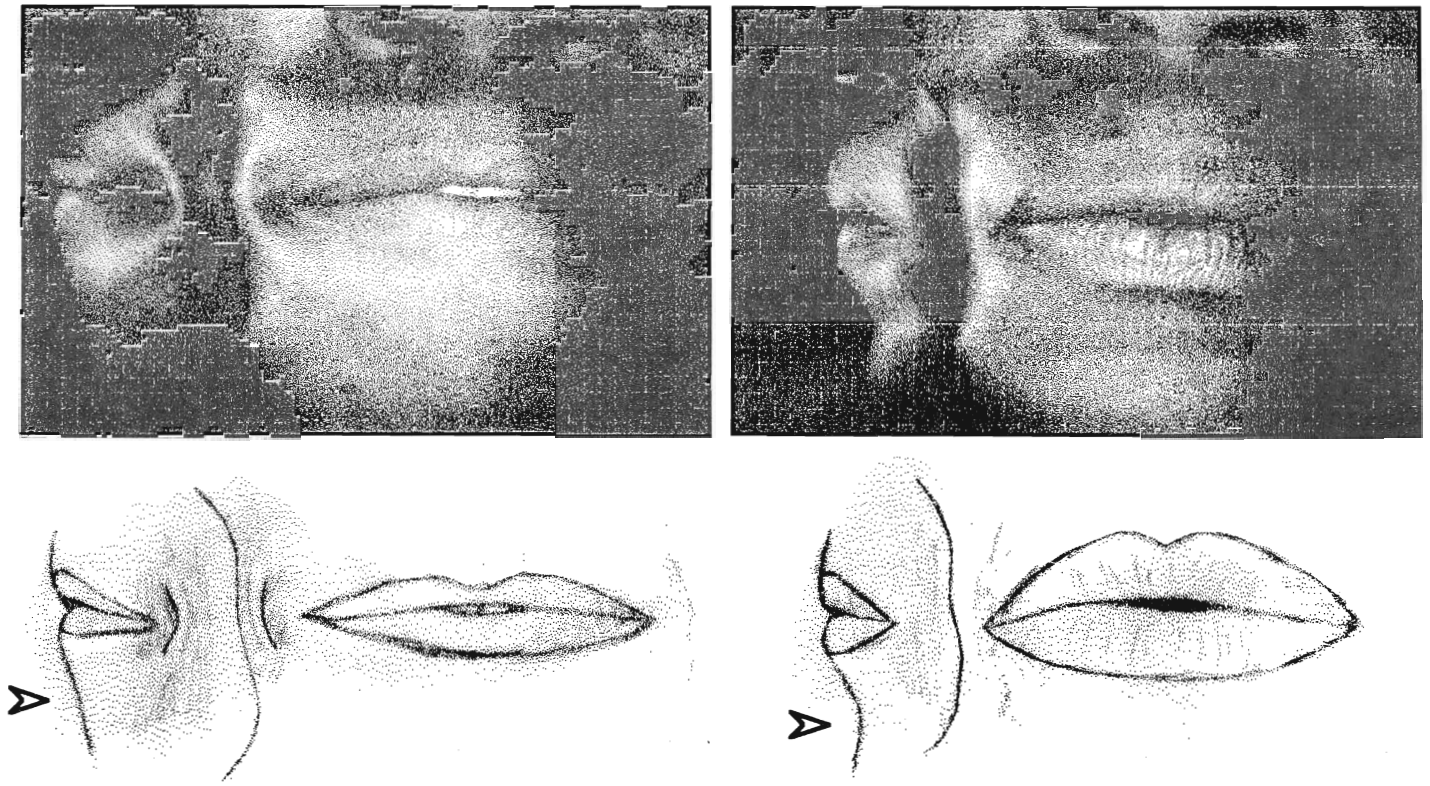


Figure 1. Labiodental (left) and bilabial (right) fricatives in Kwangali. Note profile of chin as indicated by an arrow on each sketch. For explanation, see text.

resulting protrusion of the stretched skin in the region shown by the arrow in the figure, and by the active retraction of the corners of the mouth. When producing the bilabial fricative the skin is not stretched in the region marked by the arrow, the corners of the mouth are further forward, and the upper lip is thicker.

There are interesting implications of this use of different techniques for enhancing the contrast between bilabials and labiodentals. First let us assume that if the contrast is enhanced at all it is done by raising the upper lip for the labiodental in one group of languages (the West African Kwa languages) and by retracting the lower lip in the other group (the Southern Bantu languages), with perhaps also some forward movement of the corners of the lips for the bilabial sound in the latter group of languages. These observations are valid for all the West African and Southern African speakers that we have observed. What does this imply for our linguistic phonetic definitions of the distinctions involved? It would seem to support the view that the linguistic definitions need refer (as now) only to the point of maximum constriction as being bilabial or labiodental. The other differences are properties that distinguish one language from another. These data seem to be one more (small) indication of the fact that our phonetic theories need to make a clear distinction between features that convey linguistic information (the difference between one form and another *within* a language) and those features that convey sociolinguistic information (the difference *between* one language or dialect and another).

The International Phonetic Association has always been somewhat ambivalent about this point. As was noted in the report on the Kiel Convention (International Phonetic Association 1989), the Association's long standing aim as stated in its *Principles* is to show on its chart only those sounds that are known to be contrastive (phonemic) within some language. It has not taken any policy decision as to how two other kinds of differences among speech sounds should be symbolized: (1) allophonic differences within a language; and (2) comparative differences between languages. Abercrombie (1967) offered some valuable guidelines on this topic, noting that more exotic, less romanic, symbols are often used when the aim is to point to a comparative difference between two languages. But a more exotic symbol is available only in those cases when some third language requires it for a contrastive sound, so that it is an officially recognized symbol. Small differences between languages that are not known to be contrastive in any language have to be symbolized by *ad hoc* devices. In considerations of this kind many phoneticians would advocate that the Association should simply add whatever devices are needed for making more precise phonetic descriptions, be they additional rows or columns on the chart, or specially defined symbols.

A better system that is within the IPA tradition is to associate symbols with specific phonetic values, and to use diacritics to show variations occurring only between sounds in different languages. This solution is often used in the case of vowels. The description of vowels is the one area of phonetic theory in which the Association has explicitly recognized the general phonetic, non-linguistic, nature of some of its phonetic categories. By regarding the symbols as defined by cardinal vowel theory, small differences between languages can be readily described.

Any phonetic theory that attempts to assign precise values to symbols rests on two premises: firstly it assumes that the symbols refer to specific points in the appropriate space; and secondly it assumes that we can define the dimensions of this space. The principal dimensions of vowel quality are fairly well agreed in practice (although there is

some disagreement on whether they refer to articulatory or auditory properties of vowels); and the definitions of the points in this space are sufficient for many phoneticians to be able to use them. (However, the Association might well contemplate why it is that only phoneticians trained in the British tradition use this system; scores of highly skilled phoneticians all around the world disregard the possibility of using cardinal vowel theory when describing languages.)

There has been much less work on defining the dimensions of consonants. In some sense we might say that the three principal dimensions are the state of the glottis, and the place and manner of articulation (with, of course, further dimensions being required in the description of many sounds). But none of these three has been at all rigorously defined. When considering the state of the glottis, for example, there is no agreement on how to specify sounds that are slightly breathy voiced or slightly creaky voiced. There are now diacritics for both these situations, but no way of noting that the degree of breathiness in Owerri Igbo voiced aspirated (murmured) stops is different from that in the similarly labeled Marathi sounds. The notions of place and manner of articulation are also not well defined, even (as we now see) in the case of such apparently simple terms as bilabial and labiodental. There is no notion of a cardinal bilabial or labiodental fricative from which the languages we have been considering might or might not depart. Nor are there diacritics for raising the upper lip, or pulling the lower lip back over the upper teeth. Neither of these gestures is part of the IPA theory of phonetic description. It would seem that the Association still has a lot of work ahead of it.

References

ABERCROMBIE, D. (1967). *Elements of General Phonetics*. Edinburgh: Edinburgh University Press.

MADDIESON, I. (1984). *Patterns of Sounds*. Cambridge: Cambridge University Press.

SAGEY, E. (1986). The Representation of Features and Relations in Nonlinear Phonology. Ph.D. dissertation, MIT.

LADEFOGED, P. (1968). *A Phonetic Study of West African Languages*. Second edition. Cambridge: Cambridge University Press.

To appear in a special issue of *Phonetica* on the role of phonetics in phonology.

Testing the universality of phonological generalizations with a phonetically-specified segment database: results and limitations.

Ian Maddieson

Phonology and Phonetics

As fields of enquiry the relationship between phonology and phonetics is intimate, despite the occasional historical interludes during which the practitioners of these two linguistic subdisciplines conduct hostilities with their colleagues. While both concern the study of sound patterns in language, phonetics primarily reaches out to physics (acoustics and fluid dynamics) and human biological sciences (physiology, neurology and psychology), whereas phonology primarily addresses the interaction of sound patterns within a language and their relation to other levels of analysis internal to linguistics (lexicon, morphology, syntax). Phoneticians may sometimes perceive phonologists as operating too remotely from physical reality; phonologists may feel that phoneticians are insufficiently concerned with the functioning of the entire linguistic system. But the subject matter of phonology and phonetics is a connected whole, linked both by means of direct pathways, such as rules for the phonetic implementation of phonological representations, as well as more general considerations. Phonological theory generally values representations that can be given a phonetic basis over those that can't; phonological rules and processes that can be given an explanation in terms of phonetic naturalness are preferred over those that account for the same data in terms of arbitrary properties. Phonetic research focusses its attention on those matters that are felt to be linguistically significant, meaning primarily phonologically relevant. The study of phonological universals is an area where the two fields closely interact.

Phonological universals

'Phonological universals' are patterns prevalent in the phonological structures of languages that are attributed to natural factors. Some of these are likely to result from higher-level processing requirements related to the way that linguistic knowledge is stored in the brain, but others are due to various factors at the phonetic level (Ohala 1983). In some cases these factors are relatively straightforward biological limitations on our species. Thus, we can be sure that there is no significant linguistic contrast that depends on differences in the signal above 20,000 Hz, since most human beings cannot hear such frequencies well. The limits that are biologically imposed have been very little studied. More attention has been focused on universals that are posited to arise from optimization processes within the boundaries that are set by biology. This viewpoint might be distinguished as 'ecological' rather than 'biological' in that it has to do with adaptation of language to the environment in which it is used rather than with matters that are directly part of the human genetic endowment. Different phonetic patterns have greater or lesser advantages because they differ with respect to communicative effectiveness, economy of articulatory effort, efficiency of

encoding, or in other relevant ways. Languages will tend toward those patterns that have the greater advantage, and the phonology will be shaped accordingly. Universals posited to result from such optimization are not intended to be exceptionless; they are not literally universal, only prevalent. Although, by hypothesis, they result from factors that are immanent in the environment in which natural languages must function, they are evident in differing degrees in different languages. What this means is that such 'universals' are statements that are true of some languages but not of others. Since any correct descriptive statement may be true of some languages but not of others, it is important to show that 'universals' are valid generalizations across a set of relevant languages. In other words, hypothesized universals need to be statistically examined, using a properly constructed sample of languages. Resulting refinements to universals can then feed back into improving models of the nature of the pressures that generate them. In this way, studies of phonological typology and of phonetic models mutually illuminate each other.

Many of the generally affirmed phonological universals concern segment inventories and relate to such matters as the structure of systems of contrasts, and implicational relationships between segment types. These often take the form of statements with the structure of one of the following:

i) A plurality of (or most, or all) languages have the (set of) segment(s) *x* (*y, z,* ...).
ii) A language with the (set of) segment(s) *x* (*y, z,* ...) will also have the (set of) segment(s) *a* (*b, c,* ...).
iii) A language with a set of *n* segments of class *y* will show property *x* in the structure of the set.

Familiar examples of each of these types of statements would be the following:

i) All languages have coronal stops.
ii) A language with voiceless nasals will have voiced nasals with the same places of articulation.
iii) A language with only five vowels will have the set /i, e, a, o, u/.

Statements of type (i) simply describe a property of languages and are most obviously 'statistical' observations, in the lay sense of the term. But statements of the other two types, describing a contingent relationship between two properties, also require a statistical interpretation. A type (ii) statement, a familiar implicational hierarchy, is intended as the equivalent of saying that occurrence of segment(s) *x* (*y, z,* ...) is restricted to languages that also have the segment(s) *a* (*b, c,* ...) to a degree that is significantly higher than chance. Similarly, type (iii) statements amount to saying that property *x* occurs in languages with *n* segments of class *y* to a degree that is significantly higher than chance. Note that statements in the form of types (ii) and (iii) in the way they are usually phrased can be true without there being a significant association between the properties named. For example, if *all* languages have both *x* and *a* then there is no significant restriction of occurrence of *x* to languages with *a*.

Statements of these types can be statistically evaluated with a database containing segment inventories of a sample of languages, though it is important that the sample be representative. General models of the nature of some of the phonetic pressures that may be affecting languages can also be studied with the use of such samples. For example, work associated with the ideas of Lindblom

(Liljencrants & Lindblom 1972, Lindblom 1986, Lindblom & Maddieson 1986, Vallée 1989) has pursued the theory that the structure of segment inventories is governed (at least in part) by a principle that dispersion of segments within an auditory space represents an optimization. The predictions of this theory have been compared with observed patterns in segment inventory databases and the theory progressively modified. Other studies (e.g. Abry et al. 1989) have used a segment inventory database to investigate aspects of the quantal theory developed by Stevens (1972, 1989). This model proposes that selection of segments located in areas where there is a nonlinearity in the relationship between articulatory and acoustic change represents an optimization. Since Lindblom's model, now known as the Theory of Adaptive Dispersion, and Quantal Theory assume that optimal solutions involve different factors, they differ to some degree in the predictions they make concerning expected phonological universals. By comparing their success in predicting universals the models can themselves be tested.

UPSID

A possible source of data on universals of segment inventories is the UPSID database, now enlarged from that described in Maddieson (1984). The current version of this database contains data from 435 languages.[1] The languages are chosen to represent a properly structured quota sample of the genetic diversity of extant languages. One and only one language is included from each cluster of related languages judged to be separated from its nearest relative to a degree similar to the separation of North and West Germanic (taken to be equivalent to about 1500 years of separate development). Each language is represented by a listing of the phonologically contrastive segments it uses. The segments are in turn given a feature specification in terms of 64 phonetic attributes, consisting of an elaborated set of the categories of standard phonetic theory, such as places and manners of articulation for consonants, height, backness and rounding values for vowels, and so on. Segments are positively specified for those attributes possessed by the most basic allophone of the segment in question. In most cases this is the most frequent allophone, but sometimes there are reasons for thinking that another phonetic form is more basic, particularly when the more common form seems like a relaxed variant of the other. For example, in a number of languages there is variation in the realization of a phoneme /r/ between a trill [r] and a tap [ɾ], with the tap being more frequent in casual speech. UPSID will code this as a trill for the following reason: lingual trills are observed to have a natural frequency of 25-30 Hz, so a faster rate of speech will readily shorten the trill duration to the point where only one contact is made. Thus it is expected for a trill to be reduced to a tap ("a one-tap trill"), whereas there is no parallel reason for an articulatory gesture that is programmed as a tap to be expanded into a trill.

UPSID allows validation to be supplied for many generalizations concerning segment inventories. We will illustrate this with a fairly straightforward example. As part of a set of "assumptions about nasals", Ferguson (1963) proposed that no language will have more nasalized than oral vowels. The extent to which this holds for the languages in the UPSID database can easily be verified by examining the vowel inventories of those languages that have any nasalized

vowels. In the current sample of 435 languages there are 102 (23%) with nasalized vowels, and in each case the number of nasalized vowels is equal to or less than the number of oral vowels. If the UPSID sample successfully represents the full range of the world's languages, the probability of a language violating Ferguson's generalization may be said to be very low indeed. But we may go further than just confirming the claim and use the sample to investigate the possible basis in the phonetic world on which it rests.

As noted, when nasalization occurs with vowels it often occurs with a smaller number of vowel qualities than are used for oral vowels in that language. This pattern has been attributed to the fact that nasalization reduces the auditory distinctiveness between other properties of vowels (Wright 1986). For example, nasalization typically broadens the bandwidth of formants, making it harder to discriminate differences in their frequencies. Thus, although natural coarticulatory processes lead to the evolution of nasalized vowels, particularly from syllable-final vowel+nasal sequences, sustaining the distinction between all of the different vowels when they are nasalized requires perceptual distinctions that are too fine to be convenient. Hence there is a prevailing pressure to limit the number of distinctions between nasalized vowels. The plausibility of this account is enhanced by considering the pattern of vowel nasalization in the light of overall vowel inventory size. A striking fact that emerges from UPSID is that almost half (48, 47%) of the languages with nasalized vowels have as many nasalized vowels as they have oral vowels. Since these languages are concentrated among those with a small number of vocalic distinctions, this pattern actually supports and strengthens the hypothesis that inventories of nasalized vowels may be shaped by difficulty of discriminating between them. Of the 48 languages with equal numbers of oral and nasalized vowels, 36 (75%) have 6 or fewer oral vowels. On the other hand, a language with 7 oral vowels is most likely to have only 5 nasalized ones (50% of cases). We can plot the relationship between number of oral vowels and number of nasalized vowels as in figure 1 and see that the number of nasalized vowels is not linearly related to the number of oral vowels, but falls off as oral vowel inventory size increases. Thus we can refine the statement of the universal, and of the hypothesis that is aimed at explaining it. It seems that there is a threshold, at about 6 nasalized vowels, above which the problems of distinguishing between them become more acute. The hypothesis of a threshold is open to empirical testing: perceptual experiments should show that successful discrimination between the members of a set of nasalized vowels falls off much more sharply as the number in the set rises above 6 than is the case with sets of oral vowels.

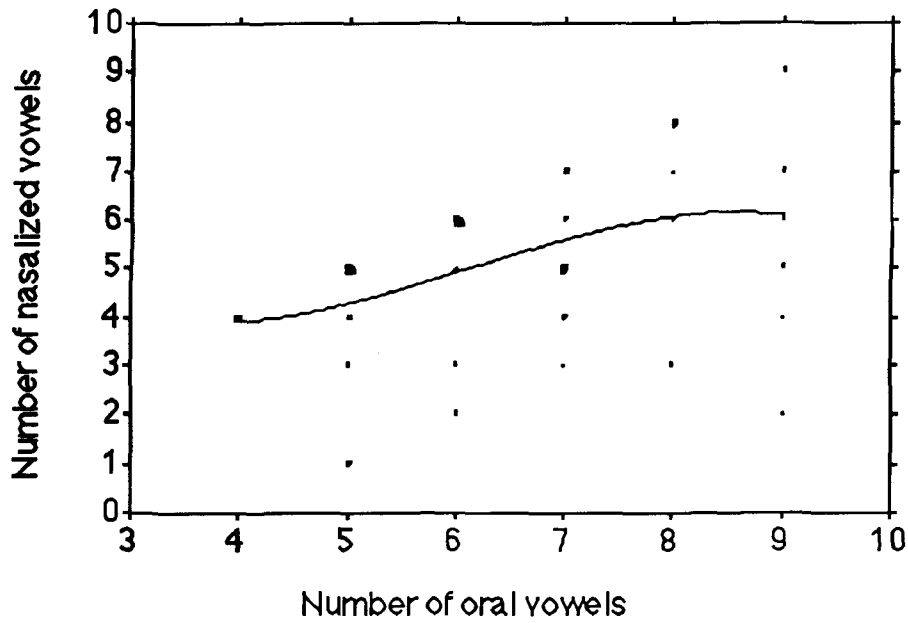


Figure 1. Scatterplot of oral vs nasal vowel inventory sizes in languages with both. The size of the points at each intersection of oral and nasal vowel numbers indicates the number of cases. A third-order polynominal regression function is fitted to the data. Languages with more than 9 oral vowels are dropped from this plot since the numbers of cases per cell are too few.

Data problems

However, in using the database, allowance must be made for problems arising from lack of an agreed descriptive framework and from other deficiencies in the data. The questions examined must be tailored to match the level of detail that is available. The UPSID format requires that segments be assigned to a particular character code, and assigned a positive or negative value for all features in the UPSID list. This is done so that the compiler, or a user who chooses to add to the data, is forced to consider whether a given segment in any one language is comparable to one that occurs elsewhere in the set of sampled languages and should be assigned the same representation. There are recurrent challenges in performing this task. Sources for some languages targeted for inclusion are vague or ambiguous in describing their segments. Different linguistic traditions and individual linguists use non-commensurate terminology. Some detective work is sometimes required to translate these as far as possible into a uniform notation. Problem areas serve to draw attention to where greater precision in segmental descriptions would be appropriate in phonological statements.

For consonants, there are frequently difficulties in deciding cross-language equivalences between places of articulation. These difficulties can be particularly acute with respect to places that would be grouped together as coronal (or lingual)

articulations. Some languages, including the majority of those spoken in Australia, have a contrast between dental and alveolar stops and nasals, so these places must be distinguished. But for many linguists the term 'dental', or its equivalent in the language they are writing in, encompasses both the dental and alveolar areas (e.g. Le Bris & Prost 1981, Ozanne-Rivierre 1982). The Russian term 'peredjazychnyj' can encompass an even larger articulatory zone (e.g. Chispijakova 1983). In some French-language sources the term 'alvéolaire' refers to a postalveolar region (i.e. [ʃ] would be referred to as 'alvéolaire') (e.g. Monod-Becquelin 1975), and both dental and alveolar regions are described as 'dentale'; however, sometimes (e.g. Sauvageot 1965) the usage parallels that in most recent English-language phonetics texts (e.g. Ladefoged 1982). Some sources provide only a symbol, such as [t], with no explicit guidance as to the particular articulation involved (e.g. Krauss 1975, Katz 1975).

Ladefoged & Maddieson (1988) argue that within-language contrasts provide evidence that linguo-labial, interdental, dental, alveolar, postalveolar and palatal are all validly distinguishable places of articulation, but that there is also a cross-cutting parameter of tongue profile (tongue-tip up or tongue-tip down — usually referred to as 'apical' vs 'laminal'). At dental, alveolar, postalveolar and palatal places of articulation both apical and laminal articulations are known to occur in different languages. However, relatively few descriptions comment on the tongue profile except in selecting the term 'retroflex' (also cerebral, cacuminal) to refer to tip-up articulations in the postalveolar or palatal regions, or 'palato-alveolar' for tip-down articulations in the postalveolar region. The set of places provided in UPSID therefore includes palato-alveolar (i.e. laminal postalveolar), retroflex (i.e. either apical postalveolar or sublaminal palatal) and palatal (always laminal). A 'postalveolar' segment must thus be classified as either retroflex or palato-alveolar. However, at dental and alveolar places, especially when these do not contrast, data is only rarely available on the tongue profile employed. Often it is not even clear whether segments are dental or alveolar in place.[2] Because of such limitations UPSID can only be used to give a crude approximation in answer to questions concerning the relative frequency of dental versus alveolar articulations and cannot be used to verify the hypothesis that dentals are usually laminal and alveolars usually apical (Stevens, Keyser & Kawasaki 1986).

For vowels, it is often impossible to tell from a given language description whether there is an independent parameter of tongue root position in the vowel system. Prior to the availability of x-ray studies the articulatory basis of the contrast between vowels belonging to the vowel harmony sets in languages such as Igbo and Akan was not recognized. The differences were described as if they were simply aspects of the vowel height parameter (beginning with Christaller 1875), or the vowel harmony sets were given labels such as 'tense' and 'lax' which correspond to no well-defined phonetic parameter (e.g. Swift et al 1962). Similar descriptions are found in recent publications. After x-ray studies showed the role of tongue root position, beginning with Ladefoged's work on Igbo (Ladefoged 1964), it became common to describe African vowel harmony systems in terms of advanced and retracted tongue root positions. The terminology was also used to describe register systems of Mon-Khmer languages (Gregerson 1976) and certain other situations in languages where words also fall into two

categories differentiated by properties of the vowels (Cook 1983, 1989, Trigo 1989). Unfortunately, judging whether vowels have distinct tongue root positions is not that readily done on the basis of auditory data. Only a handful of African languages with vowel harmony systems have been studied using x-rays, including Igbo, Akan, Ijo and Ateso (Lindau 1974, 1975), DhoLuo (Jacobson 1978), Anyi (Retord 1972), and Ndut (Gueye 1986). Although both Niger-Kordofanian and Nilo-Saharan languages are included in this set and all of these have tongue root harmony, it is quite certain that not all African vowel harmony systems are based on tongue root position. Svantesson (1985) argues from acoustic studies that Khalkha Mongolian vowel harmony depends on contrast in tongue root position, but so far neither acoustic nor x-ray studies on Mon-Khmer languages have been able to confirm any use of this parameter (Lee 1983, Thongkum 1987a,b). It seems certain that tongue root position contrasts have been attributed to languages that do not have them and have been overlooked in languages that do have them. A consistent treatment of tongue root position is therefore impossible in UPSID.[3] Hence the frequency with which this feature takes part in phonological contrasts and its relationship to other properties of inventories cannot be reliably estimated.

Sources of uniformities and group anomalies

In evaluating whether a given generalization should be considered a universal, various possible sources of observed cross-language uniformities must be considered. Prevalent patterns may well result from the ecological circumstances of human language use; these are the ones that most researchers are interested in looking for. But it is not given that a common pattern, even one found in a majority of languages, is due to a factor that is basic in human language. This is one possibility. But is also possible that some patterns result from areal contact phenomena or from the accidents of language survival and proliferation. Considerable numbers of languages spoken today belong to close-knit families containing hundreds of very closely related languages; the Bantu branch of Niger-Kordofanian (Bouquiaux 1980) and the Oceanic branch of Austro-Tai (Ross 1988) are two examples. These particular language families have proliferated because the speakers were successful at expanding into new areas while maintaining political structures of small size, encouraging language differentiation. The typological similarities between languages within such groups are undoubtedly due much more to shared inheritance, further reinforced by geographical proximity, than to environmental factors shaping the individual languages in the same way. The set of living languages is the set of survivors from a series of such economic and political actions, and there are many chance factors involved. All the languages in UPSID with clicks have relatively small inventories of vowel qualities. There are 6 languages involved and none has more than five basic vowel qualities. (This observation appears also true of languages with clicks not included in UPSID). Since there are few surviving languages with clicks it is easy to understand that such a pattern may reflect coincidence of two unrelated properties.

There are two tests that should be applied before an observation is accepted as a universal. The first we may call the test of generality, and the second the test of motivation. By generality what is meant is that a pattern should

be shown to occur separately in each major areal or genetic group of languages. This is a minimal check that the phenomenon is in fact universal in scope, and is similar to a procedure proposed by Dryer (1989) in his work on universals of word-order. Possible groupings would include purely geographical ones, such as the four inhabited continents and their associated islands, or purely genetic ones, such as the 11 major language families, each containing hundreds of members, into which the majority of the world's languages fall, plus a remnant group for the smaller families. I have used a combination of geographical and genetic factors in an analysis of vowel systems, described below. Of course, the nature of the particular proposed universal will determine if it is possible to demonstrate its presence in each major grouping. The strength of implicational hierarchies, type (ii) statements above, can only be tested with languages that contain relevant segments, and their distribution may itself be restricted to particular families or regions, as is the case with clicks. On the other hand, the test of generality is straightforward to apply a statement of type (i) above, such as "all languages have coronal consonants" or "all languages have fricatives". In fact, languages in all parts of the world have at least one coronal consonant—there are no exceptions in UPSID, and also none are known in languages outside the sample (cf Paradis & Prunet 1990). However, there is a serious exception to the generalization that all languages have fricatives, since this is not true of the great majority of Australian languages. Twenty of the 23 Australian languages in UPSID have no fricatives, and two of the exceptions have only an approximant-like voiced velar fricative. If absence of fricatives can be a local deviation, then how can we be sure that their almost invariable presence in languages of the other parts of the world is not also a kind of local deviation?

The only check is the test of motivation. For a proposed universal it should be possible to provide a hypothesis concerning the design factors at work. Thus we may argue that the widespread presence of fricatives in the world's languages can be predicted from the fact that the common fricatives are among the most salient of sounds, e.g. they are rarely confused with other sounds (Singh & Black 1966). Hence their presence follows from a broader hypothesis, namely, that languages will tend to select more salient over less salient sounds. Although the notion of salience requires further research, it may eventually be possible to show that other salient sounds are no less broadly distributed than fricatives, and that languages everywhere averagely include similar numbers of salient sounds in their inventories. In this way Australian and non-Australian languages can be subsumed under a broader generalization. Likewise, even though it may initially appear that the occurrence of small vowel inventories in languages with clicks is chance, if a good reason to expect a relationship between these two factors can be found the conclusion shoulf be re-examined. In fact, such a motivation can be found. For the production of lingual clicks, both the coronal and dorsal articulators are required to be in specific positions, and there is little freedom to vary the position of the tongue body. The result is that co-articulation with following vowels is very limited (see Sands 1990a, b for documentation of this point for Xhosa). Since vowel discrimination is aided by co-articulation with neighboring consonants (Repp 1892, Tomiak et al 1987), it will follow that languages with a large number of clicks in their inventories might have a design preference for a small number of vowels, since, lacking coarticulation with a significant number of

their consonants, it is harder to distinguish a given number of vowels than is the case in a language that does not have clicks.

We should nonetheless continue to be wary of the 'universality' of a pattern that is not replicated in each major grouping of languages. For example, it has been suggested that the common 5 vowel system consisting of /i, e, a, o, u/ represents a perceptually optimal exploitation of the two most basic acoustic dimensions of the vowel space (e.g. Ladefoged 1989: 30-32). Crothers (1978) likewise suggested that five is the optimal size for vowel quality systems and wrote that:

> "this optimal size should be predictable from such factors as the limits on the human ability to distinguish vowels, the average amount of noise in the speech situation, the average information content of spoken communication, the relative contribution of consonants and vowels to the total information content, and so on."

As in the sample of languages used by Crothers, a system with 5 vowel qualities is more common than any other in the languages of the UPSID sample (cf. Vallée 1989). The number of languages with vowel inventories of different sizes is as shown in figure 2. Across the whole sample, the pattern of vowel system sizes approaches a normal distribution with a peak at 5.

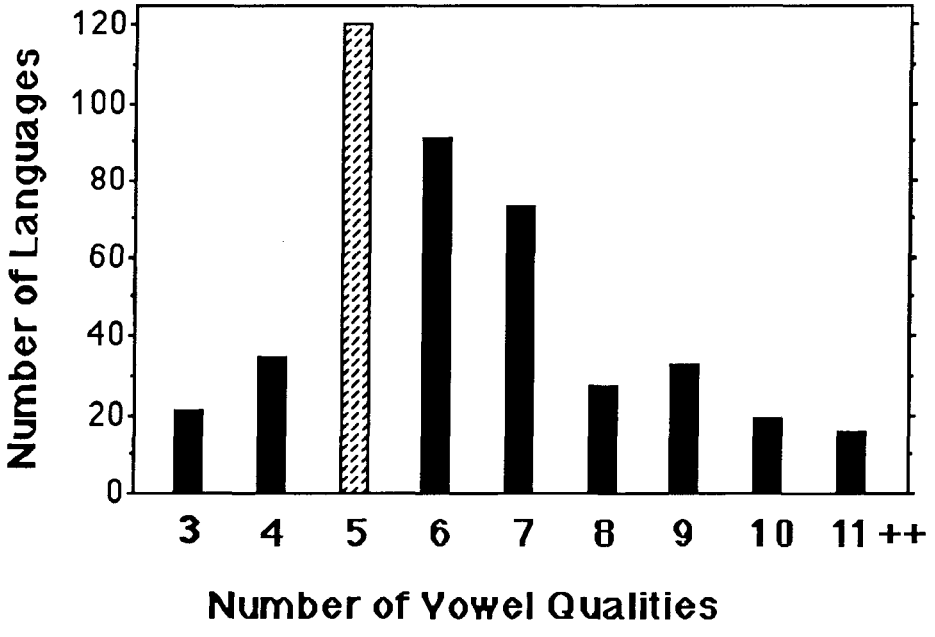


Figure 2 . Histogram showing percentage of languages in a sample of 435 with vowel quality inventories of different sizes. Vowels that share the same basic height, backness and lip position features but differ in length, phonation type, etc are said to have the same quality. Vowel inventories containing 11 or more members are pooled together.

However, if the groups of language families in the four major continental landmasses are examined separately, an interesting discrepancy arises. Whereas systems with 5 vowel qualities are the most common in the languages of the families centered on the Eurasian, American and Australasian landmasses, a 7 vowel system is the most common in the languages belonging to the 4 major language families centered on the African continent. The differences can be seen in the histograms in Figure 3. Pooling the data from all languages, as in Figure 2, swamps this local variation.

If the kinds of factors mentioned by Crothers and Ladefoged are working to favor five vowel systems, we have every reason to expect they should apply to African languages as much as to others. The peak in the African distribution at 7 is as well-defined as the peak at 5 in the other large geographical/genetic groupings. The African sub-sample contains a substantial number of languages drawn from four major language families (Afro-Asiatic, Niger-Kordofanian, Nilo-Saharan, Khoisan). This local deviation is therefore not easily dismissed and constitutes a clear exception to the preference elsewhere for five vowels. How is this exception to be interpreted? It might result from some wide-spread areal similarity between African language families. That factor could be a sharing of vowel harmony systems using an additional dimension of the vowel space, although a system of this type containing only seven vowel qualities must be considered a 'degenerate' one (Lindau 1975). Antell et al. (1974) speculate that tongue-root based vowel harmony did spread from Niger-Kordofanian to Nilo-Saharan languages, and with it, expanded vowel inventories. Vocalic harmony is also found in some Afro-Asiatic languages (Ebert 1974, Lydall 1976), though I am not aware of any Khoisan languages with such rules.

Does this mean that African languages show a preference for a non-optimal vowel system? Not necessarily. Retention or adoption of vowel harmony may be facilitated by the fact that it usually limits the vowels within a (phonological) word to being drawn from a set that contains no more than five members. Thus the expansion of the total vowel inventory resulting from the added parameter of vowel contrast is mitigated. The permitted lexicon is doubled but the value of the vowel parameter underlying the vowel harmony needs to be recognized only once per word. The 'benefit' of added lexical possibilities is set against only a small 'cost' in more complex representations, since the added segments are not freely distributed. The result is a net advantage, or at least an offsetting balance. If this is the appropriate interpretation, our models need to predict that there may be several distinct local optima in any optimization function concerning vowel inventories, depending on how the vowels relate to other phonological features of the language. We note that this would be more readily reconciled with models of the dispersion type than with a quantal theory, which seems to predict a single optimum.

This interpretation would be supported by evidence that languages elsewhere have found the same local optimum — perhaps it could be shown that a significantly higher than expected proportion of all languages with 7 (or more) vowel qualities have a vowel harmony restriction on their distribution. Of course, distributional restrictions are not represented by a listing of the vowel inventory,

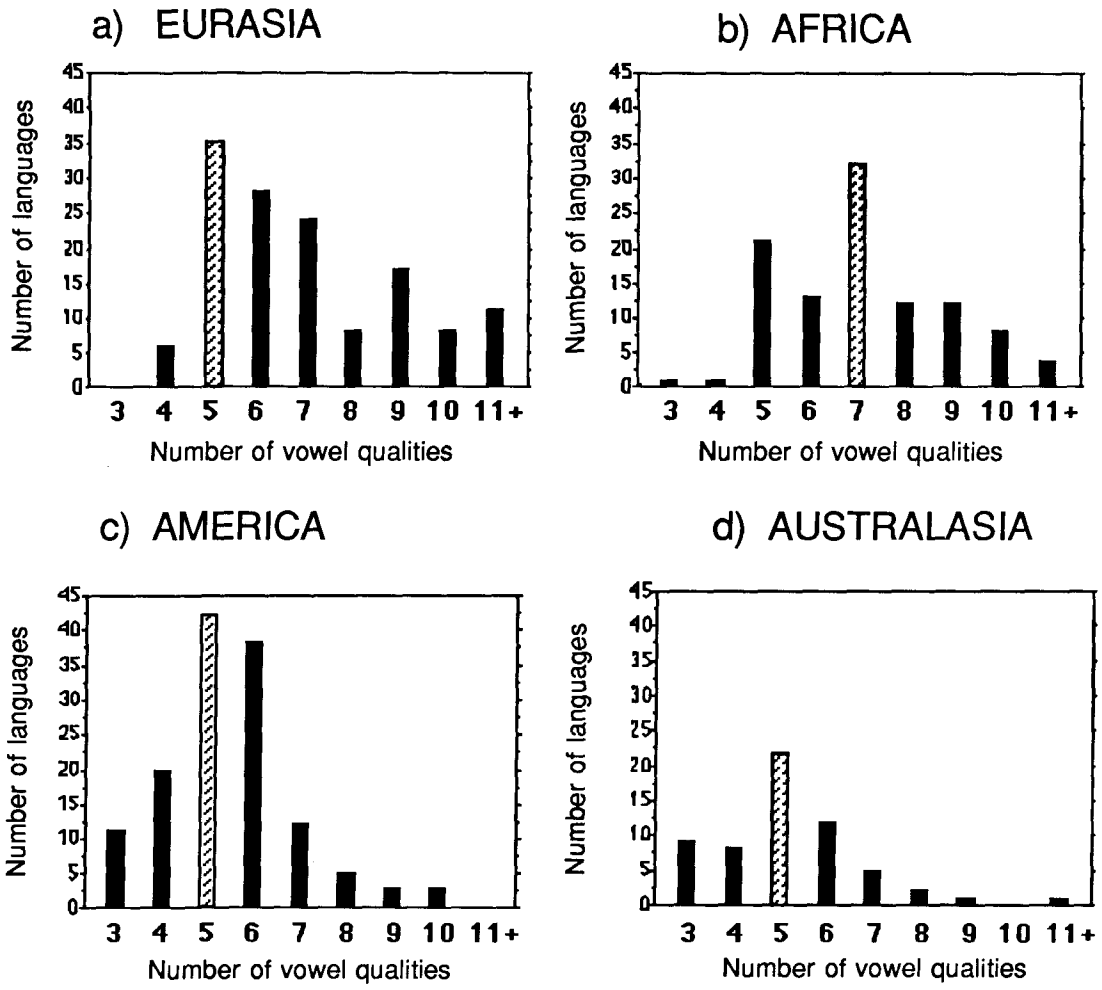


Figure 3. Histograms showing percentage of languages with vowel quality inventories of different sizes in four groupings. The groupings are (a) **Eurasia**, Indo-European, Ural-Altaic, Sino-Tibetan, Austro-Tai, Austro-Asiatic, Dravidian, Caucasian and Chukchi-Kamchatkan languages in the UPSID sample together with a few single languages; (b) **Africa**, Afro-Asiatic, Niger-Kordofanian, Nilo-Saharan, and Khoisan languages; (c) **America**, Amerindian, Na-Dene and Eskimo-Aleut languages; (d) **Australasia**, Australian and Papuan languages.

and so are not represented in the standard UPSID format. Alternatives are rather more threatening to our belief that we can use the relative frequency of typological patterns to understand the constraints and pressures on the development of human language. Major typological divergences between large groupings of languages might imply that none of the patterns concerned are optimal, but that certain patterns become prevalent simply as a result of areal influence and/or proliferation of languages that are genetically related, if remotely. 'Universals' of this type result from contact and inheritance rather than from design limitations. Given that individual languages display great variability in phonological structure, and all are at least adequate to the tasks they must serve, it is hard to rule out the possibility that some prevailing patterns result from accidental survival and spread. We might consider this in the light of the much-discussed possibility that all modern humans have a relatively recent single African ancestor (Brown 1980, *Cann et al. 1987, Stringer & Andrews 1988), even though this idea, based on a* dating technique using mitochondrial DNA, is robustly rejected by many scholars (Avise et al. 1984, Spuhler 1988, Wolpoff 1988). If it is true, then it is overwhelmingly probable that all living human languages likewise have a comparatively recent single common ancestor, also with an African origin. This follows from the fact that the human genetic data, if interpreted correctly, requires that all pre-existing human groups were exterminated and did not interbreed with the newcomers as they migrated across the continents. This scenario makes it improbable that any pre-existing languages survived or contributed anything to the language of the newcomers. And it is equally unlikely that the newcomers did not carry language with them but invented it independently in different locations later. The complexity of social organization implicit in pre-newcomer archaeological records in both Europe and Asia (America and Australasia were settled later; West 1981, Flood 1989) is usually interpreted as requiring the inference that these humans had linguistic communication. Of course, the hypothetical single parent language has since diversified greatly, just as modern human populations show considerable intra-specific diversity.

A simple (naïve?) hypothesis consistent with this view of human evolution might be that all language families outside Africa can be traced back to a dispersion from a location reached by the newcomers soon after leaving Africa (as proposed by Stringer & Andrews 1988). Hence non-African languages are more closely related to each other than they are to those of Africa, and the prevalence of five vowel systems in these languages might reflect a typological characteristic of the common parent, retained in many of its offspring. Are African languages preserving an ancient property of human language that has been lost elsewhere? The seven vowel system might represent a stage that is older in the phylogeny of human language. In this case, presumably it is this pattern that theories of language ecology should explain.

Acknowledgment
The work reported in this paper was supported by grant ROI DC00642 from the National Institutes of Health.

Footnotes

1. This represents an interim version of the database as it stood in August 1990. Approximately 30 additional languages are targeted for inclusion, and review of the previously included languages needs to be completed. This final version will be distributed with the program described in Maddieson & Precoda (1989).

2. The solution adopted in UPSID is to add "unspecified dental or alveolar" to the list of place features. Segments identified only by symbols such as /t,d,n,s,l/ are assigned this place feature, as well as segments described as "dental" where this seems to be a cover term covering both dental and alveolar.

3. Where descriptions refer to tongue root position, vowel differences are coded in UPSID according to auditorily-based impressions of height and backness. The relationship between tongue root constriction and pharyngealization are discussed in Ladefoged & Maddieson (1990).

References

Abry, C., L-J. Boë, & J-L. Schwartz. 1989. Plateaus, catastrophes and the structuring of vowel systems. *Journal of Phonetics* 17: 47-54.

Antell, S.A., G.K. Cherono, B. L. Hall, R. M. R. Hall, A. Myers & M. D. Pam. 1974. Nilo-Saharan vowel harmony from the vantage point of Kalenjin. *Afrika und Übersee* 57: 241-267.

Brown, W. M. 1980. Polymorphism of mitochondrial DNA of humans as revealed by restriction endonuclease analysis. *Proccedings of the National Academy of Sciences of the USA* 77: 3605-3609.

Bouquiaux, L. (ed.) 1980. *L'expansion bantoue (actes du colloque internationa du CNRS, Viviers), Vol 2.* SELAF, Paris.

Cann, R. L., M. Stoneking & A.C. Wilson. 1987. Mitochondrial DNA and human evolution. *Nature* 325: 31-36.

Chispijakova, F. G. 1983. Soglasnye Kondomskogo dialekta Shorskogo jazyka. In V.M. Nadeljaev (ed.)*Sibirskij Foneticheskij Sbornik: Sbornik Nauchnyx Trydov.* Akademija Nauk SSSR, Sibirskoe Otdelenie, Novosibirsk: 16-31.

Cook, E-D. 1983. Chilcotin flattening. *Canadian Journal of Linguistics* 28: 123-32.

Cook, E-D. 1989. Articulatory and acoustic correlates of pharyngealization: evidence from Athaspaskan. In D.B. Gerdts & K. Michelson (eds.) *Theoretical Perspectives on Native American Languages.* State University of New York Press, Albany: 133-145.

Crothers, J. 1987. Typology and universals of vowel systems. In J.H. Greenberg et al (eds.) *Universals of Human Languages, Volume 2, Phonology.* University of Stanford Press, Stanford: 93-152.

Dryer, M. 1989. Large linguistic areas and language sampling. *Studies in Language* 13: 257-292.

Ebert, K. 1974. Partial vowel harmony in Kera. *Studies in African Linguistics, Supplement 5*: 75-80.

Ferguson, C. A. 1963. Assumptions about nasals: a sample study in phonological universals. In J.H. Greenberg (ed.) *Universals of Language.* M.I.T. Press, Cambridge, MA: 53-60

Flood, J. 1989. *Archaeology of the Dreamtime, revised ed.* Collins, Sydney.

Gregerson, K. J. 1976. Tongue root and register in Mon-Khmer. In *Proceedings of the First International Congress on Austroasiatic Linguistics (Oceanic Linguistics, Special Publication 13).* University of Hawaii Press, Honolulu: 323-369.
Gueye, G. 1986. Les correlats articulatoires et acoustiques de la distinction ± ATR en ndut. *Travaux de l'institut de phonétique de Strasbourg* 18: 137-249.
Jacobson, L. C. 1978. DhoLuo vowel harmony. *UCLA Working Papers in Linguistics* 43.
Katz, H. 1975. *Selcupica I: Materialen von Tym (Veröffentlichungen des Finnisch-Ugrischen Seminars an der Universität München, Serie C, Band I).* Universität München, Munich.
Krauss, M.E. 1975. St. Lawrence Island Eskimo phonology and orthography. *Linguistics* 152: 39-72.
Ladefoged, P. 1964. *A Phonetic Study of West African Languages.* Cambridge University Press, Cambridge.
Ladefoged, P. 1982. *A Course in Phonetics (2nd ed).* Harcourt Brace Jovanovitch, New York.
Ladefoged, P. 1989. Representing Phonetic Structure. *UCLA Working Papers in Phonetics* 73.
Ladefoged, P. & I. Maddieson. 1988. Places of articulation for stops. In L.M. Hyman & C.L. Li (eds) *Language, Speech and Mind.* Routledge, London and New York:
Ladefoged, P. & I. Maddieson. 1990. Vowels of the world's languages. *Journal of Phonetics* 18: 93-122.
Le Bris, P. & A. Prost. 1981. *Dictionnaire Bobo-Fing (Précedée d'une introduction grammaticale et suivi d'un lexique français).* SELAF, Paris.
Lee, T. 1983. An acoustical study of the register distinction in Mon. *UCLA Working Papers in Phonetics* 57: 79-96.
Liljencrants, J. & B. Lindblom. 1972. Numerical simulation of vowel quality systems: the role of perceptual contrast. *Language* 48:839-862.
Lindblom, B. 1986. Phonetic universals in vowel systems. In J.J. Ohala & J.J. Jaeger, eds., *Experimental Phonology.* Academic Press, Orlando: 13-44.
Lindblom, B. & I. Maddieson. 1988. Phonetic universals in consonant systems. In L.M. Hyman & C.L. Li (eds) *Language, Speech and Mind.* Routledge, London and New York: 62-78.
Lindau, M. 1974. The feature Advanced Tongue Root. In E. Voeltz (ed.)*Third Annual Conference on African Linguistics.* Indiana University, Bloomington: 127-134.
Lydall, J. 1976. Hamer. In M. L. Bender (ed.) *The Non-Semitic Languages of Ethiopia.* African Studies Center, Michigan State University, East Lansing: 393-438.
Maddieson, I. 1984. *Patterns of Sounds.* Cambridge University Press, Cambridge.
Maddieson, I. & K. Precoda. 1989. Updating Upsid. *UCLA Working Papers in Phonetics* 74: 104-111.
Monod-Becquelin, A. *La Pratique Linguistique des Indiens Trumai (Haut Xingu, Mato Grosso, Brésil, Tome I.* SELAF, Paris.
Ohala, J. J. 1983. The origin of sound patterns in vocal tract constraints. In P. F. MacNeilage (ed.) *The Production of Speech.* Springer-Verlag, New York: 189-216.

Ozanne-Rivierre, F. 1982. Langues de Hienghène et Proto-Océanian: Phonologie comparée. In A-G. Haudricourt & F Ozanne-Rivierre *Dictionnaire Thématique des Langues de la Région de Hienghène.* SELAF, Paris.

Paradis C. & J.-F Prunet (eds.). 1990. *The Special Status of Coronals: Internal and External Evidence (Phonetics and Phonology 3).* Academic Press, San Diego.

Retord, G. L. A. 1972. L'agni, variété dialecte sanvi: phonologie, analyses tomogrophiques, documents. *Annales de l'université d'Abidjan, Série H Linguistique, 5.1:* 1-206.

Ross, M. D.. 1988. *Proto-Oceanic and the Austronesian Languages of Western Melanesia (Pacific Linguistics C-98).* Australian National University, Canberra.

Sands, B. 1990a. Some of the acoustic characteristics of Xhosa clicks. *UCLA Working Papers in Phonetics* 74: 96-103. Abstract in *Journal of the Acoustical Society of America* 86: S123-124 (1989).

Sands, B. 1990b. Effect of vowel on preceding click. Abstract in *Journal of the Acoustical Society of America* 88: Suppl 1, S54.

Sauvageot, S. 1965. *Déscription synchronique d'un dialecte Wolof: Le parler de Dyolof.* IFAN, Dakar.

Singh, S. & J.W. Black. 1966. Study of twenty-six intervocalic consonants as spoken and recognized by four language groups. *Journal of the Acoustical Society of America* 39: 372-387.

Spuhler, J. N. 1988. Evolution of mitochondrial DNA in monkeys, apes and humans. *Yearbook of Physical Anthropology* 31: 15-48.

Stevens, K. N. 1972. The quantal nature of speech: evidence from articulatory-acoustic data. In E. E. David & P. B. Denes (eds.) *Human Communication: A Unified View.* McGraw-Hill, New York: 51-66.

Stevens, K.N. 1989. On the quantal nature of speech. *Journal of Phonetics* 17:3-45.

Stevens K. N., J. Keyser & H. Kawasaki. 1986. Toward a phonetic and phonological theory of redundant features. In J.S. Perkell & D. H. Klatt (eds.) *Invariance and Variability in Speech.* Lawrence Erlbaum, Hillsdale, N.J.: 426-449.

Stringer, C B. & P. Andrews. 1988. Genetic and fossil evidence for the origin of modern humans. *Science* 239: 1263-1268.

Svantesson, J-O. 1985. Vowel harmony shift in Mongolian. *Lingua* 67: 283-327.

Thongkum, T. 1987a. Phonation types in Mon-Khmer languages. *UCLA Working Papers In Phonetics* 67: 29-48.

Thongkum, T. 1987b. Another look at the register distinction in Mon. *UCLA Working Papers In Phonetics* 67: 132-165.

Trigo, L. 1989. Back-of -the-mouth phonology. Manuscript. Department of Modern Foreign Languages, Boston University.

Vallée, N. 1989. *Typologie des Systèmes Vocaliques,* Travail d'Études et de Recherche de Maîtrise, Institut de la Communication Parlée, Université Stendhal, Grenoble.

West, F. H. 1981.*The Archaeology of Beringia.* Columbia University Press, New York.

Wolpoff, M. H. 1988. Multiregional evolution: the fossil alternatives to Eden. In C.B. Stringer & P Mellers (eds.) *The Origins and Dispersal of Modern Humans: Behavioural and Biological Perpectives.* University of Chicago Press, Chicago:

Wright, J. T. 1986. The behavior of nasalized vowels in the perceptual vowel space. In J.J. Ohala & J.J. Jaeger, eds., *Experimental Phonology.* Academic Press, Orlando: 45-67.

Draft of a paper to be presented at the International Congress for the Phonetic Sciences, Aix-en-Provence, August 1991. Session : Typology and Universals

Investigating Linguistic Universals

Ian Maddieson

For centuries linguists have tried to understand what is essential to the nature of Language, as opposed to something that is particular to a given language or group of languages, by studying language universals. The reasoning is essentially as follows: if a feature is distributed widely enough in the world's languages to be labeled universal then it is either a necessary property of human *language, or it is in some sense a desirable one.* A *similar argument applies* when clusters of co-occurring linguistic properties are described under the heading of linguistic typology. If certain sets of properties repeatedly occur together in languages, then it can be argued that their co-occurrence is a necessary or a desirable property. That is, typology includes the study of contingent universals.

It is usually because of this line of thinking that people are interested in universals of language. The study of prevalent patterns in languages, of universals, is a window to examine the question of why language is the way it is. By this hypothesis, universals arise because of biological limits and environmental pressures that are at work on all languages simply by virtue of the fact that they are in use by members of the same species of mammal. Despite the great geographic dispersion of our species and a good deal of individual and group variability between its members, all humans make use of basically the same equipment of brain, vocal tract and auditory system. Studying universals is therefore not so much a goal in its own right as a challenge to the linguist to come up with explanatory accounts of what these pressures are and how they affect human language in general. The goal is to produce models of these pressures that predict the universals that have been observed.

However, if we are interested in universals we have two great problems to face. One is the problem of obtaining knowledge about them. How do we decide what is universal? That is, how do we go about finding what is prevalent enough in the languages of the world to count as a possible universal? The second problem is how to distinguish those properties that we wish to consider 'universal' in the particular sense that they arise from design considerations that apply to human language in general from prevalent patterns that arise from other sources of uniformity? It is a hypothesis that important properties of human languages are common because they are based on inherent characteristics of the human species and of the environment in which we as a species employ our linguistic abilities. This hypothesis must be compared with alternative hypotheses that might explain the data in better or equally satisfactory ways.

Since the concern of this conference is with the phonetic sciences, the discussion of these issues which follows will be directed to and illustrated with examples from the domains of phonological and phonetic universals, based in part

on my own work with the UCLA Phonological Segment Inventory Database or UPSID (Maddieson 1984, Maddieson & Precoda 1990), but there is nothing in the general principles concerned that would be any different if the field of enquiry was in some other area of linguistics.

A. How do we find meaningful prevalent patterns?

As has been pointed out before, but is worth stressing again, some kind of structured systematic sampling of the universe of known languages is essential if we want to know what linguistic patterns are prevalent (Hurford 1977, Bell 1978, Maddieson 1984, Dryer 1989). Prevalence is an essentially statistical concept. We need to be able to say with some confidence that the set of languages within which some property is said to be prevalent (or more common than some other pattern) represents the larger universe that we are really interested in studying, ultimately that of all possible human languages. Above all, if we are looking at patterns of co-occurrence of properties, at typological patterns, we must be able to evaluate the independent distribution of these properties, in order to be able to say if they are significantly associated with each other.

An obvious way to know how widely distributed a particular feature is would be to count the frequency of that feature in all human languages. Even if we limit ourselves to languages still spoken at this time, there are two straightforward practical objections to attempting this, and two theoretical ones. The practical objections are, first, that linguists have not yet got around to examining all of the world's living languages, and, second, that, even if they had, surveying descriptions of all languages would be impossibly time-consuming.

The theoretical objections concern the need to survey the data in a way that gives appropriate weight to each language. First, there is no unambiguous principle to define the borderline between the degree of difference between two speech varieties that warrants assigning them to different languages and the degree of difference which can be accomodated within the construct of a single language. Different linguists will give various classifications of the same speech varieties. Without an answer to this problem we might include only one dialect of one language but many varieties of another, giving it undue weight in the survey. This makes it impossible to be certain that we have assigned equal weight to each language.

Secondly, we know that where we find close-knit families of languages existing today this reflects an evolution from an earlier stage at which the precursors of these now-distinct languages were dialects of a single language. Separately counting all members of such close-knit groups transparently gives undue influence to the group, just as separately counting all dialects of a language does for that language. This is because members of the group will have so many features in common that are simply inherited. Few of their shared features are likely to be due to *independent* response to the pressures shaping human language that it is our ultimate objective to investigate. We may take the North Germanic languages as an example. There are perhaps five living languages in this group, Icelandic, Faroese, Swedish, Norwegian and Danish, and we know that they go back to a common Norse parent language that was quite uniform as

recently as five or six hundred years ago. The descendent languages, unlike their next nearest relatives in the West Germanic group, share some elements of a pattern relating consonant quantity to vowel quality and quantity features. We might overestimate the global prevalence of such a pattern, that is, the number of independent occurrences of the pattern, by counting each of these languages separately. In contrast, a language such as Albanian in the same period of time has not fragmented into a number of daughter languages. We would underrepresent features that might have been shared by the daughter languages it never had. The problem is just the same as if we were to count each of a large number of modern dialects of English but only to count one variety of modern French. In that case, our survey might show an inflated number of interdental fricatives, and a correspondingly depressed number of front rounded vowels.

It is therefore necessary for both practical and theoretical reasons to develop some strategy so as to make a selection of languages such that each contributes an appropriately equal weight to the sample. One needs to create a sample that can be trusted to represent in a fair way the overall frequency of the properties of interest in the world's languages. For the UPSID project the decision made was to aim for a sample that includes one and only one language from each group of languages that is separated from its closest relative by a genetic distance similar to that separating the North Gemanic languages from the West Germanic languages. In terms of time depth this might translate into about 1500 years of separation, a long enough period for substantial independent developments to occur in the phonological patterns of any two languages belonging to the same larger family. Related languages will, of course, have certain elements of their phonological patterns in common, or we would hardly be able to recognize their relatedness. But at the same time they will have a degree of independence. Languages with *no* closer relatives are also included, as they too represent the outcome of certain lines of independent development. The current UPSID sample size is *(about)* 460 languages, probably between 5% and 8% of the world's existing languages.

However, despite the restriction built in to constructing the UPSID sample, problems concerning whether the selected languages can be considered truly independent samples do not go away. This remains true even if the sample is drawn up in more restrictive ways. I will return to this question when it comes to discussing the interpretation of prevalent patterns. But first I will provide a simple illustration of the use of this database to derive estimates of the frequency of phonological patterns.

It is generally agreed that there are more languages with a voicing contrast in stops than languages with a voicing contrast in fricatives (Hockett 1955) . But let us suppose that we want to investigate the claim that voicing contrasts in fricatives *preferentially* occur in languages which have a voicing contrast in stops, that is, there is an implicational universal involved. To do this, it is not good enough simply to point to a large number of languages that do indeed share both types of voicing contrasts and then list a number of languages that have a voicing contrast in stops but not in fricatives. It has to be shown that there are *fewer than expected* cases of languages that have a fricative voicing contrasts but lack stop voicing contrasts. The frequency of fricative voicing and stop voicing independently can

be estimated from our language sample, and each number can be viewed as the probability that a given language will have the property in question. We can then multiply these independent probabilities together to obtain estimates of how frequently fricative and stop voicing might be expected to co-occur if there was no contingent relationship between them. The expected value can then be compared with observed frequencies of co-occurrence and singular occurrence, and the significance of the association of the voicing contrasts with each other can be statistically evaluated.

In our UPSID database, about 66% of the languages included have voicing contrasts in stops and about 31% have voicing contrasts in fricatives (*Note: these are preliminary values based on an incomplete expanded data file. Final text will have accurate numbers for the completed database).* That is, the probability of one of the individual languages in the database having a stop voicing contrast is .66, and the probability of an individual language having a fricative voicing contrast is .31. If we multiply these two probabilities together, the result is about .21. That is, if there is no connection between the occurrence of these two things, we may expect 21% of these languages to show both stop and fricative voicing, leaving about 10% that have fricative voicing without stop voicing, . The observed figures are in fact *(about)* 27% and 4% respectively. A simple χ-square test can then be applied to compare the observed with the expected distributions, yielding the answer that there is about a one in ten chance that these results are accidental. Since there is only one degree of freedom in this problem the level of significance should perhaps not be taken too seriously, but for what it is worth, the result suggests that the connection between the ocurrence of fricative and stop voicing is not all that strong.

As a final note in this section, a word should perhaps be said about the care required in drawing conclusions from any assemblage of data about a set of languages. Typically, when a large number of descriptions of languages are brought together to get a view of the variety of language, a wide range of explicit or implicit linguistic theories are represented. Scholars of different language families and from different parts of the world are trained in different traditions, so that different facts are observed and the same facts will be reported in different ways. Also as theoretical models in any given tradition evolve, what are considered to be the significant properties of a language change. We need to be sure that descriptions are commensurate before generalizations are drawn, and to be sure that the inferences made are responsive to the particular nature of the data that is represented.

B. How do we interpret prevalent patterns?

As remarked above, universals in themselves are not objects of ultimate interest. It is the theory that will account for universals that is the focus. Compiling a sample such as UPSID provides a basis for stating which types of patterns might be justly interpreted as prevalent. For example, since over 98% of languages in the UPSID sample have stops at bilabial, anterior coronal (dental or alveolar) and velar places of articulation, we can say that it is a valid generalization about languages that they are overwhelmingly likely to have stops at these three places. What this means is that we would expect to find this to be true of some different but

representative sample of extant languages or if we could travel 2000 years forward or backward in time and sample the languages spoken at that time. In this case, as in any other, once it has been established that some pattern is prevalent or that there is a certain set of properties that tend to co-occur in the world's languages, we are challenged to look for the explanation that might be responsible for that pattern.

There are several types of explanations that may be entertained. They fall into two basic groups. The first type posits that prevalent patterns reflect necessary or desirable properties of language. The second group takes more account of the extent to which prevalent patterns might be due to inherited similarities between languages or to the spread of traits due to contact. These two types of explanations reflect on the one hand the fact that the faculty of language is a basic part of our human make-up and on the other hand the fact that the particular languages that survive and spread result from accidents of history shaped by many socio-political and environmental factors.

The first type of explanation includes the possibility that certain universals are inevitable. Some universals may be due to species-specific biological constraints, which would at least set limits to the range of variation that languages may exhibit. However, the absolute biological constraints that can be stated at this time do not seem to be very interesting. This is perhaps because we know relatively little about what our language-related biological limitations actually are, and hence are restricted mostly to stating the obvious, such as that in their speech mode languages must use articulations that are possible human gestures that leave some acoustic signature of their presence. Thus, although languages make use of various gestures involving the lower lip, such as bilabial, labiodental and linguo-labial articulations, labio-uvulars are universally absent. This is so for the rather uninteresting reason that the lower lip and the uvula cannot meet. This tells us why labial-uvulars are absent but does not tell us why bilabials are universal and linguo-labials very rare.

Aside from articulatory impossibilities, we can also point to certain inevitabilities in speech production of the sort that have been teh focus of research by John Ohala and some of his associates (e.g. Ohala 1983). These are effects that arise from the operation of physical laws applicable to the functioning of the vocal apparatus. They are not species-specific in any sense, but since the physical laws apply to all individuals, these effects are also inevitable. Ohala points out how physical laws produce assymetrical results. For example, given the higher resistance to air-flow in high vowels, there is a certain level of subglottal driving pressure at which voicing of high vowels will fail to occur but voicing of low vowels will be sustained. The consequence is that voiceless high vowels are a little more likely to occur than voiceless low vowels. This addresses the observation that there are languages in which all vowels devoice and languages in which only high vowels devoice, but no language is known that has devoicing of low vowels only.

The possibility that there are innate categorical classifications of certain sound types due to the way the perceptual system works remains uncertain, but further biologically-determined limits could arise from such a cause.

Other universals may reflect desirable design attributes of languages rather than inevitable properties. Let us think about one class of desirable properties. For a variety of reasons, humans do not wish to operate *near* the limits of their capabilities. In any mode of activity, errors increase when performance is pushed towards the limits. The nearer an approach is made to an operating limit the greater the difficulty of learning becomes, the more variable individual levels of success become, the greater the degradation of performance under stressful conditions, the greater the difficulties resulting from effects of age, tiredness, etc, and so on. For spoken language, the relevant limits would include limits on the range and speed of movement of the constituent parts of the vocal tract mechanism, limits on the ability of the auditory system to resolve distinctions between sounds, and limits related to the capacity for storage of linguistic knowledge in the brain. Even without knowing exactly where any of these limits lie, we can understand what represents movement *towards* these limits. It seems safe to assert that it is a desirable property of language that it should avoid any approach to the performance limits. This is at once a more inclusive and more cautious formulation of old observations that are usually phrased in terms of languages maximizing ease of articulation and auditory distinctiveness. These two principles have been appealed to in selective ways to account for particular synchronic or diachronic patterns in languages, but the implications of proposing these principles as ones that affect languages across the board have rarely been taken seriously.

An exception is the ambitious phonetic model of phonological origins being developed by Björn Lindblom. The aims of this theory, the Theory of Adaptive Dispersion (summarized in Lindblom 1990), include being able to account for the ontogeny of segments and the structure of segment inventories. Lindblom's presentations of his theory include a model of the way we might envisage a language developing phonological patterns through selecting an optimal set of syllables. The optimal set is the one that minimizes the value of aggregate articulatory effort, expressed as the sum of deviations from a neutral vocal tract position plus the magnitude of articulatory trajectories in transitions between onset and offset of syllables, and, at the same time maximizes the value of overall auditory contrastiveness, expressed as the sum of differences over time in the auditory spectrum across the set of syllables. This model has at present been developed more as a demonstration that it is possible to predict the optimal set of syllables from any set of input candidates using the very general principles described, and it is not intended that the particular set of selected syllables has any special standing. So it is not appropriate to analyze the set of selected syllables to see if they reflect the preference patterns seen in actual languages. However, we can see in principle how such a model might explain the relative frequencies of bilabials and linguo-labials. Linguo-labial contact requires a tongue gesture of much greater magnitude than the lip-rasing gesture used for a bilabial.

But it is possible that this part of Lindblom's theory is too deterministic. The articulatory and auditory components produce a single optimal solution for a given input.[1] Our impression is that languages are more variable than this. Considering just segment inventories, the evidence from surveys such as UPSID provides no clear evidence that languages are tending towards unique solutions. Consonantal and vocalic systems show certain similarities in their common core but the ways that they are elaborated beyond this common core are quite variable, and reduced systems with less than the common core are not unusual. A cross-linguistic study of syllable patterns currently under way at UCLA (Maddieson 1991b) shows that most of the languages studied do not have the strong dependencies between adjacent consonants and vowels that might be expected if ease of articulation and auditory distinctiveness, evaluated at the syllable level, play dominant roles in selecting preferred sound patterns. And, after all, languages with linguo-labials do occur (Maddieson 1990).

Rather than leaving this variability to be accounted for by social factors, as Lindblom provides, two directions for developing the more strictly phonetic elements of the model seem to merit consideration. One is to add further parameters, reflecting costs and benefits of other aspects of sound patterning, such as rules of word formation and phonological alternations. The 'cost' of the higher degree of articulatory difficulty of, say, consonant clusters may be mitigated when these result from morphological processes such as affixation (as English *move, moved*). Typically, affixes form a closed set and articulatory precision can be relaxed. Similarly, the cost of reduced auditory distinctiveness associated with an increased number of vowel contrasts might be mitigated by the presence of a rule of vowel harmony that limits the free distribution of these vowels at the word level. Recognition of whatever phonetic parameter forms the basis of the vowel harmony distinction is only required once per word, rather than for each syllable. The possibility of an association between larger vowel inventories and vowel harmony is suggested by the fact that for languages in Africa the modal size of the vowel inventory is 7, whereas on the other continents it is 5. Vowel harmony systems are more prevalent in the language families of Africa than in most other areas (Maddieson 1991a).

The second, perhaps complementary, direction for taking the development of such a model is to relax the constraint that it seeks a single, optimal, solution, so that it produces a variety of possible solutions that cluster around the optimum. That is, accepting our restatement of the desirable design requirements and modelling avoidance of extremes rather than maximization of ease of articulation and auditory distinctiveness. Of course, here the problem would be to determine how close an approach to the limits should be modeled as acceptable.

If we cannot be satisfied that universals arise from inevitable causes or result from shared pressures towards desirability, our other alternative is to consider that they may result from inherited similarities (or at least transmitted similarities). That is, we may see prevalent patterns that are not the result of innate limits or pressures to select desirable traits independently applying to many separate languages, but are the result of preservation of traits, possibly quite accidental ones, of a parent language which is ancestral to many or even all of the surviving languages (or a

language which influenced surviving languages at an early stage). For this reason, it is important for universalists to be very concerned with the issue of how closely related the surviving languages are. Otherwise our conclusions may be little better than the ones we would draw from a sample consisting only of modern English dialects. At one time it seemed that our understanding of the story of human evolution might have allowed for the likelihood that language evolved in parallel in several different areas and over a long period of time. Early diffusion of hominids through the Old World seemed to be followed by a long period of somewhat separate but parallel development (see, e.g. Campbell 1966). Present-day populations in East Asia, Africa and Europe were believed to reveal traces of characteristics seen in ancient fossils found in those areas. This would have allowed for the interpretation that when two language families were said to be unrelated, it meant more than that the relationship could not at present be demonstrated by traditional historical-comparative methods. They could actually be of independent origin. The picture now seems more confused.

First of all, there seems to be increasing evidence that many of the groupings of languages that linguists were once content to say were "not related" can be shown to have genetic relationships demonstrable by traditional methods (Kaiser and Shevoroshkin 1987). Reorganizations of the familiar major language families are disruptive to the scholarly communities involved and tend to be met with resistance, or ignored. But even conservative scholars are beginning to concede that the data being assembled in favor of relating (at least) Indo-European, Dravidian, Ural-Altaic, Afro-Asiatic and Kartvelian together has merit (Nichols 1990). Sagart (1990) has recently provided strong evidence that Chinese may be more closely related to Austronesian than to Tibeto-Burman. Since Benedict (1975) has shown Austronesian and Thai-Kadai languages to be related, and Sino-Tibetan comparisons still seem valid, a macro-grouping of languages in Asia seems to be emerging. Benedict has further claimed Japanese as a relative of Austronesian (Benedict 1988), whereas Miller (1971, 1980) has shown strong reasons for linking it with Ural-Altaic. If both connections are valid, then a huge number of the languages of the Old World are genetically linked. Missing so far from this agglomeration are the three other major language families of Africa. While there is no shortage of fanciful speculation on their wider relationships (Homberger 1941, Stopa 1972, Diop 1988), at least the data assembled by Gregerson (1972) and Boyd (1978) seems to indicate the serious possibility that the Niger-Kordofanian and Nilo-Saharan families are related. As for the New World, many Americanists reject Greenberg's (1987) grouping of most American languages into a single Amerind family (Kaufman 1989, Campbell 1989) as being, at best, premature. However, cautious scholars continue to show how parts of the picture relate together (e.g. Payne 1989) and eventual demonstration that many of these languages are related seems probable. The late twentieth century is thus a period during which we are recognizing more and more of the world's languages as related to each other, and pushing back the time depth at which relationships can be recognized.

Secondly, our picture of human origins is shifting as modeling of the past based on studies of genetic markers in present-day populations is added to the tools of paleontology. A plausible account has been offered that the surviving

human population may trace back to a single African origin of a considerably more recent period than earlier models suggested (Brown 1980, Cann, Stoneking & Wilson 1987, Stringer & Andrews 1988; but see Spuhler 1988 for a more cautious view). This would suggest that all languages also have a single origin of the same time depth (not more than 200, 000 years, or about 4 to 6 times as long as humans have colonized areas such as Australia and the Americas, and perhaps as little as 100,000 years). The recognition of language relatedness among larger groupings tends to support this possibility of a single parent language at a not impossibly remote time period.[2] Since this language doubtless had its share of arbitrary and idiosyncratic features, we must be concerned that some at least of the properties that we see as prevalent in the world's languages trace back to the idiosyncratic features of this postulated parent language. Such features would be misleading testimony concerning which properties are necessary or desirable in human languages.

Of course, we know that languages change their phonetic and phonological structures over time and much diversity would have evolved from any ancient parent language. Historical studies show that, for example, vowel systems tend to be quite changeable. But there are certain other properties that tend to remain quite stable (Hagège & Haudricourt 1978). Nasals tend to remain nasals in syllable-initial position, for example. Another diachronic pattern is that stops tend to remain stops, at least in pre-stress syllable-onset position, and to retain their place of articulation, especially in low vowel environments. As noted above, stop systems including three major places (bilabial, anterior coronal, and velar) are nearly universal in languages. This seems a candidate for a trait that might be a conservative, inherited, feature. All reconstructed languages at the greatest time-depth that linguists go back to have stops at these places and the great majority of the daughter languages have retained them. There seems to be no necessity for languages to have a stop system with this many places. Some languages, such as Ahtna (Kari 1989), get along quite well with no bilabials, and it is easy to imagine languages that would have no contrast between front and back tongue articulations, with a rule-governed distribution like that of the second articulation in the so-called labial-velar stops of Nzema and Dagbani—alveolar with front vowels and velar with back vowels.

If the minimal three-place structure of stop systems is not necessary, can we show that it is universal because it is desirable? The answer, at least at present, is that we probably can't. This is because our most effective tools for attempting to understand the issue of desirability depend on having variability to analyze and on being able to look at the co-occurrences of particular properties. The many fruitful investigations of the structure of vowel systems in the last several decades—including Liljencrants & Lindblom 1972, Crothers 1978, Lindblom 1986, Abry, Boë, & Schwartz 1989, and Vallée 1989—illustrate this point. All of these studies analyze the covariation between the size and the content of vowel inventories, and draw their conclusions in the main from comparing changes in the modal structure of vowel inventories as the number of contrasting vowels varies. Without such variability, our ability to create models is impaired, and there is a shortage of data points with which to test the success of the predictions of any model.

The perhaps paradoxical conclusion is that we study the effect and nature of the ambient pressures on language with the most confidence when studying those aspects in which languages display the greatest variability, rather than in studying aspects in which they show the most conformity. Where universal or near-universal conformity is found, and we cannot explain it as due to biological factors or physical laws, it is difficult to reject the hypothesis that the trait in question is inherited.

Footnotes

1. Lindblom's model provides for cross-language variability in two ways; the number of distinct syllables can vary and the output of the articulatory and auditory components can be modified by a matrix of sociolinguistically determined functions. These are not specified in any detail but would presumably include such things as the role of linguistic markers in identifying group membership.

2. Thomason and Kaufman (1988) argue for multiple language origins, making the point that creole languages have no single parent and hence their descendents are unrelated to other languages. They also argue that it is impossible to know how often the social circumstances that lead to formation of a creole may have occurred in the distant past. These may be valid points, although it is uncertain how often conditions for long-term survival of creole languages are likely to arise. However, in the creole languages of which we know the recent histories, the sound patterns are constructed out of material that is present in one or more of the 'input' languages. There is no reason to believe this would have been different at earlier times. Ancient creole languages would not represent independent language development in the sense we are concerned with here. They would reflect continuity of traits such as three-place stop systems.

Acknowledgments

The work reported in this paper was supported in part by grant ROI DC00642 from the National Institutes of Health and grant BNS 8720098 from the National Science Foundation.

References

Abry, C., L-J. Boë, & J-L. Schwartz. 1989. Plateaus, catastrophes and the structuring of vowel systems. *Journal of Phonetics* 17: 47-54.

Bell, Alan. 1978. Language samples. In J. H. Greenberg et al (eds.) *Universals of Human Language, Vol 1, Method and Theory.* Stanford University Press, Stanford: 123-156.

Benedict, Paul. 1975. *Austro-Thai Language and Culture.* HRAF Press, New Haven.

Benedict, Paul. 1988. *Japanese/Austro-Tai.* Karoma, Ann Arbor.

Boyd, Raymond G. 1978. À propos des ressemblances lexicales entre langues Niger-Congo et Nilo-Sahariennes. In *Études Comparatives (Bulletin de la SELAF* 65): 43-94.

Brown, W. M. 1980. Polymorphism of mitochondrial DNA of humans as revealed by restriction endonuclease analysis. *Proceedings of the National Academy of Sciences of the USA* 77: 3605-3609.

Campbell, Bernard. 1966. *Human Evolution.* Aldine Press, Chicago.

Campbell, Lyle. 1988. Review article: Language in the Americas, by Joseph H. Greeenberg. *Language* 64: 591-615.
Cann, R. L., M. Stoneking & A.C. Wilson. 1987. Mitochondrial DNA and human evolution. *Nature* 325: 31-36.
Crothers, John. 1978. Typology and universals of vowel systems. In J.H. Greenberg et al (eds.) *Universals of Human Languages, Volume 2, Phonology.* University of Stanford Press, Stanford: 93-152.
Diop, Cheik Anta. 1988. Nouvelles recherches sur l'egyptien ancien et les langues négro-africaines modernes.
Dryer, Matthew. 1989. Large linguistic areas and language sampling. *Studies in Language* 13: 257-292.
Greenberg, Joseph H. 1987. *Language in the Americas.* Stanford University Press, Stanford.
Gregarson, Edgar. 1972. Kongo-Saharan. *Journal of African Languages* 4: 46-56.
Hagège, Claude & A.G. Haudricourt. 1978. *La phonologie panchronique.* Presses Universitaires de France, Paris.
Hockett, C. F. 1955. *A Manual of Phonology.* Indiana University, Bloomington.
Homberger, L. 1941. *Les Langues Négro-Africaines.* Payot, Paris.
Hurford, James R. The significance of linguistic generalizations. *Language* 53: 574-620.
Kaiser, Mark & Vitaly Sheveroskin. 1987. On recent comparisons between language families. *General Linguistics* 27: 34-46.
Kari, James. 1990. *Ahtna Athabaskan Dictionary.* Alaska Native Language Center, Fairbanks.
Kaufman, Terrence. 1989. Language history in South America: What we know and how to know more. In Doris L. Payne (ed.) *Amazonian Linguistics: Studies in Lowland South American Languages*, University of Texas Press, Austin: 13-73.
Liljencrants, Johan & Björn Lindblom. 1972. Numerical simulation of vowel quality systems: the role of perceptual contrast. *Language* 48: 839-862.
Lindblom, Björn. 1986. Phonetic universals in vowel systems. In J.J. Ohala & J.J. Jaeger, eds., *Experimental Phonology.* Academic Press, Orlando: 13-44.
Lindblom, Björn. 1990. Models of phonetic variation and selection (Paper presented at the conference on Language Change and Biological Evolution, Institute for Scientific Interchange, Turin, May 1988). *PERILUS (Phonetic Experimental Research, Institute of Linguistics, University of Stockholm)*, 11: 65-100.
Maddieson, Ian. 1984. *Patterns of Sounds.* Cambridge University Press, Cambridge.
Maddieson, Ian. 1990. Linguo-labials. In R. Harlow & R. Hooper (eds) *VICAL (Papers form the Fifth International Conference on Austronesian Linguistics, Auckland), Volume I, Oceanic Languages.* Linguistic Society of New Zealand, Auckland: 349-375.
Maddieson, Ian. 1991a. Testing the universality of phonological generalizations with a phonetically-specified segment database: results and limitations. To appear in special issue of *Phonetica* on the role of phonetics in phonology.
Maddieson, Ian. 1991b. Syllable structure and phonetic models. Paper presented at the Linguistic Society of America, Chicago.

Maddieson, Ian & Kristin Precoda 1990. Updating UPSID. *UCLA Working Papers in Phonetics* 74: 104-111.
Miller, Roy A. 1971. *Japanese and the Other Altaic Languages.* Chicago University Press, Chicago.
Miller, Roy A. 1980. *Origins of the Japanese Language.* University of Washington, Seattle.
Nichols, Johanna. 1990. Linguistic diversity and the first settlement of the New World. *Language* 66:475-521.
Ohala, John J. 1983. The origin of sound patterns in vocal tract constraints. In Peter McNeilage (ed.) *The Production of Speech* Springer, New York: 189-216.
Payne, David L. 1989. Some widespread grammatical forms inSouth American languages. In Doris L. Payne (ed.) *Amazonian Linguistics: Studies in Lowland South American Languages*, University of Texas Press, Austin:75-87.
Sagart, Laurent. 1990. Chinese and Austronesian are genetically related. Paper presented at the International Conference on Sino-Tibetan Languages and Linguistics, Arlington, Texas.
Spuhler, J. N. 1988. Evolution of mitochondrial DNA in monkeys, apes and humans. *Yearbook of Physical Anthropology* 31: 15-48.
Stopa, Roman. 1972. *Structure of Bushman and its traces in Indo-European.* Zaklad Narodowy, Wroclaw.
Stringer, C. B. & P. Andrews. 1988. Genetic and fossil evidence for the origin of modern humans. *Science* 239: 1263-1268.
Thomason, Sarah G. & Terrence Kaufman. 1988. *Language Contact, Creolization and Genetic Linguistics.* University of California Press, Berkeley and Los Angeles.
Vallée, N. 1989. *Typologie des Systèmes Vocaliques,* Travail d'Études et de Recherche de Maîtrise, Institut de la Communication Parlée, Université Stendhal, Grenoble.

Syllable structure and phonetic models[1]

Ian Maddieson and Kristin Precoda

Through construction of predictive models phoneticians, are making increasingly sophisticated attempts to account for certain aspects of the phonological structure of languages from very general principles (Stevens 1989, Lindblom 1990a, Ohala 1990). One of the most elaborated of these models is the theory that Lindblom has named the Theory of Adaptive Dispersion, abbreviated TAD. TAD incorporates the often-mentioned principles of ease of articulation and perceptual distinctiveness. Support for TAD has been drawn from predicting the composition of phonological inventories, especially systems of vowel contrasts. Lindblom (1986) has shown that his theoretical predictions concerning the structure of vowel systems match well with the most typical patterns reported in cross-language surveys such as Crothers (1978) and Maddieson (1984). In these simulations of the dispersion of vowels within a phonetic space, segments have been viewed in isolation, not as part of larger units. Yet TAD proposes that segments arise from principles that, among other things, favor *sequences* of articulatory gestures which involve less articulatory movement in the transition from one segment to the next, and acoustic *sequences* with sufficient auditory contrast. To give a simplified example, the 3-vowel system consisting of /i,a,u/ might arise because, say, the syllables [di], [ga] and [bu] constitute a good contrastive set and are articulatorily economical. Figure 1, redrawn from Lindblom (1990a), shows the results of a rather well-known simulation of Lindblom's, demonstrating how the sequential considerations act to pick out certain preferred syllables, marked as black squares, from a set of possible combinations of syllable onsets and endpoints. The simulation shows how, in principle, an inventory of vowels or consonants might be selected. Note that of the 6 consonant onsets considered in this case, only three are included in any of the set of 24 optimal syllables selected, but these three consonants each appear with a different range of vowels. Because it weighs the importance of articulatory and acoustic trajectories, TAD therefore predicts relative frequencies of particular sequences more directly than relative frequencies of particular segments or inventory structure.

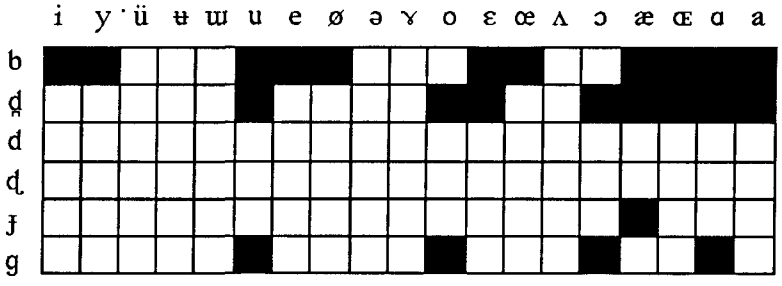


Figure 1. Matrix of derived phonetic syllables from a set of 6 onsets and 19 endpoints.

In contrast to TAD, the Quantal Theory proposed by Stevens (1972, 1989) gives no overt role to ease of articulation. Stevens' theory proposes that language exploits nonlinear relationships between changes in an articulatory parameter and the resulting acoustic/auditory response, as modeled idealistically in figure 2, redrawn from Stevens (1989: 4). Here, region II is an area where there are large changes in the acoustics for small shifts in articulation. Regions I and III show relatively little difference in the acoustics for equivalent articulatory shifts, but the difference between Region I and Region III is large. Stevens suggests that segment sequences are selected so that they cross regions such as Region II in this figure, producing rapid changes that serve as landmarks in the acoustic stream. Stevens writes:

> "In the acoustic signal, therefore, there will be an alternation between temporal regions where the acoustic parameters remain relatively steady, and narrow regions marked by acoustic events where there are rapid changes. These somewhat discontinuous attributes of the acoustic signal occur in spite of rather continuous movements or changes in the articulatory parameters." (Stevens 1989: 5)

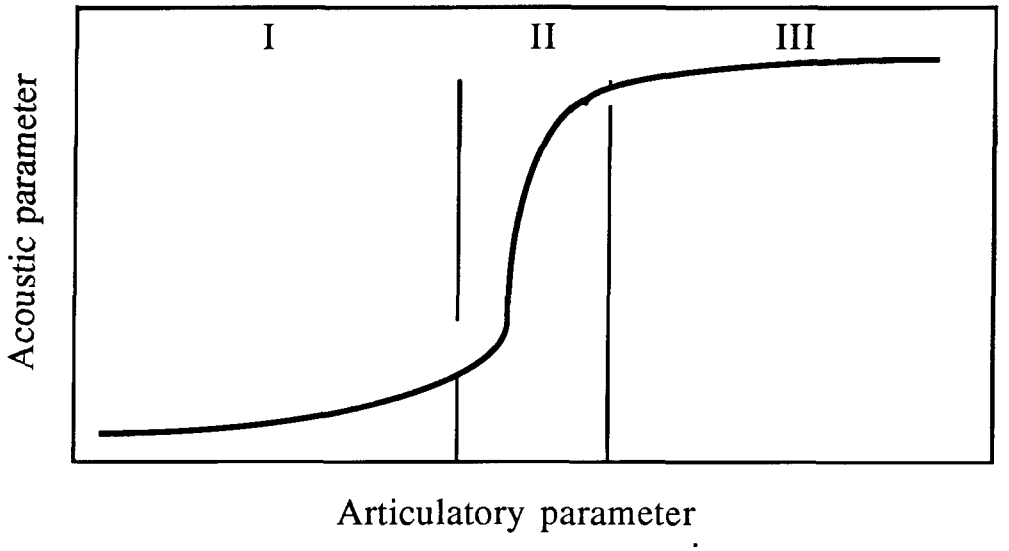


Figure 2. Model of a "quantal" change in acoustic response as values of some articulatory parameter are adjusted.

The primacy of acoustic distinctiveness implicit in Stevens (1989) was more directly advocated by Kawasaki in her 1982 dissertation and in Ohala & Kawasaki (1984). Kawasaki argued that universal constraints on sound sequences arise because segment sequences which contain salient acoustic change are preferred to those that do not have large acoustic changes. Both Stevens and Kawasaki predict that we should be able to see cross-linguistic preferences for sound sequences that have large acoustic change.

However, comparatively little is known about the cross-linguistic frequency of particular sound sequences. We therefore know relatively little about the linguistic

basis for inferring the role of principles of conservation of articulatory effort and optimization of acoustic or auditory contrast in their construction. A significant beginning was made by Janson (1986) in an article written mainly in response to Kawasaki's work. He reasoned that more favored sequences will tend to occur more frequently in texts. By looking across a sample of distantly related or unrelated languages, it should be possible to extract universal tendencies for favored CV sequences.

Janson reached a strong conclusion, suggesting that CV's requiring smaller articulatory movements were preferred, rather than CV transitions that show strong auditory/acoustic contrast. As he put it:

> "The combinations that are favored are mainly the ones in which the articulators do not have to make extensive movements from the consonant gesture to the vowel gesture."

In particular, he claims to have shown tendencies for rounded vowels to be preferred with labial consonants, front vowels with coronal (i.e. dental or alveolar) consonants and back vowels with velar consonants. Because, to a very rough approximation, CV syllables with small articulatory movements will tend to have smaller acoustic transitions than those with large movements, he adds that:

> "[T]he trend to make things easier for the speaker clearly overrides the considerations of the needs of the listener."

However, Janson stated his conclusions rather more strongly than his results actually warrant, as we shall presently show in a reanalysis of his data.[2]

For pragmatic reasons concerned with data availability, Janson's language sample consisted of Finnish, Latin, Latvian, SeTswana, and Turkish. Syllables containing consonant clusters, including affricates, were discarded and the onset CV sequences of the remaining syllables in texts of varying lengths were counted. Janson collapsed across classes of vowels, so that all languages are represented as having a set of front unrounded vowels, a set of back rounded vowels, and a low central vowel /a/ alone or as the first element of a diphthong. To simplify our presentation, we will omit vowels in Finnish and Turkish that do not fit into these groups.

For each language Janson compared the observed frequency of any CV combination with its expected frequency. The expected frequency is simply derived from the overall frequency of the individual segments which comprise the sequence. Percentage deviations from the expected score were then calculated for each consonant with each vowel class. These deviation scores indicate the percentage by which a vowel class is more or less frequent than expected following a given consonant. Janson discusses these results consonant by consonant on the basis of tables such as table 1, reporting on occurrence of the nasal /n/ with the different classes of vowels. The table indicates, for example, that there were 12.86% more CV's with /n/ followed by front unrounded vowels in Finnish than expected.

Table 1. Classes of vowels after /n/ from Janson (1986).

	I (front unround)	**A** (low central)	**U** (back round)
Finnish	12.86	-29.86	-5.19
SeTswana	9.46	36.04	-71.68
Latin	3.69	-22.11	6.69
Latvian	1.34	-26.74	54.43
Turkish	0.10	-27.24	4.58

Janson points to the fact that all values in the first column are positive, that is, in each language there are more instances of coronal nasals with front vowels than expected. This is taken as indicating that this type of sequence—coronal consonant with front vowel—is favored. However, note that three of the positive values in the first column are quite low, indicating no real preference, and that except for Finnish, every language has a higher positive value in another column. It is obvious that this nasal is unexpectedly frequent with low vowels in Setswana, and with back vowels in Latvian, and to a lesser degree in Latin and Turkish too. Taken by itself this table does not provide very strong evidence that front consonants are preferentially paired with front vowels. Results in some of the tables for other consonants are even more varied than this.

What is missing from his presentation is any demonstration that there is an overall cross-linguistic trend for the consonants at a given place to associate with particular classes of vowels. We therefore reanalyzed his results in a way that looks at the magnitude and the consistency of the deviations in the entire data set. Deviation scores for consonant/vowel pairs were recalculated from raw values provided in Janson's article, and the results grouped by place of consonant articulation according to the categories shown in table 2.

Table 2. Consonants included in Janson's data

	labial	coronal	velar
Finnish	p, v, m	t, s, n, l, r	k
SeTswana	p, b, f, m	t, s, n, r	k, x
Latin	p, b, f, m	t, d, s, n, l, r	k, g
Latvian	p, b, v, m	t, d, s, n, l, r	—
Turkish	p, b, f, m	t, d, s, n, l, r	—
	19	27	5

The means of the deviation scores for each vowel class with each consonant place are shown in table 3. If there are overall effects associating particular vowel classes with particular consonant places, then the set of deviation scores for those consonant place/vowel class pairs will have a mean that is significantly different from zero. Note first that most of these mean values are relatively low, especially in the case of labial and coronal consonants, where a larger number of cases are present. In order to isolate effects within individual cells of this table, the deviation scores were

analyzed using t-tests. The significance level is shown below the mean for each cell. There is a weakly significant association between coronal place and front vowels, where the significance level is .03, and a hint in the mean values that velars may combine more naturally with back than with front vowels, but the results as a whole cannot be said to support the strong assertion that Janson makes about preference for small articulatory trajectories; neither do they clearly suggest that preferences are based on auditory considerations.

Table 3. Reanalysis of Janson's data by consonant group. Mean deviation from expected, and significance level (t-test).

	front unround	low central	back round
labial	-3.98	2.83	1.11
n=19	*.131*	*.234*	*.634*
coronal	4.34	-1.44	-2.95
n=27	*.031*	*.315*	*.053*
velar	-8.76	1.02	5.53
n=5	*.181*	*.785*	*.520*

The failure of such trends to emerge could mean that the phonological (syllabic) structure of languages is not in fact shaped by phonetic factors of the kind proposed by Kawasaki, Stevens, Ohala, Lindblom and others, but it may have other explanations. The language sample may have been too small or too skewed in some way, or the chosen method of counting (frequency in texts) may have been inappropriate; the categories to which Janson assigned segments may have obscured important phonetic differences between the languages, including the different numbers of vowels grouped together and the role of coda consonants and consonant clustering.

It is obvious that the issue is not resolved, and further study of patterns of segment sequences is needed. The remainder of this paper reports part of a larger study using lexical frequency counts of syllables rather than text frequency counts, which will include a much larger and geographically and genetically more diverse sample than that used by Janson. By using lexical counts it is hoped that undue influence in text counts of frequently occurring grammatical elements, such as the noun class markers of SeTswana, can be reduced. One subset of the languages chosen are ones with particularly limited segment inventories and simple phonotactics so that all possible syllables can be readily included in the counts. The three primary languages in this set are shown in table 4, which also displays the complete segment inventories assumed for this study. Hawaiian is a Polynesian language of the Central Pacific, Rotokas is a Papuan language spoken on Bougainville Island, Pirahã is a language of Brazil that is sometimes classified as a member of an Andean language family. These languages have between 6 and 8 consonants and between 3 and 5 vowels. Only CV and CVV syllables occur. The total number of syllables counted for each language is also shown in this table.

Table 4. Phoneme inventories of "small inventory" languages

Hawaiian p, m, w; n, l; k; ʔ, h; i, e, a, o, u.
11090 syllables
Rotokas p, ß; t, r; k, g; i, e, a, o, u.
9400 syllables
Pirahã p, b; t, s; k, g; ʔ, h; i, a, o.
10567 syllables

Two additional languages, Eastern (Labuk) Kadazan and Shipibo, with slightly less limited inventories will also be included in our report. Kadazan is an Western Austronesian language spoken in Sabah, Shipibo is a Panoan language spoken in Peru. These two languages have only four vowels and about 15 consonants; both allow a very limited number of CVC sequences. The consonants to be reported on in these languages and their vowel inventories are shown in table 5. The five languages shown in tables 4 and 5 have a relatively high level of independence from each other. Kadazan and Hawaiian are both Austronesian, but are not particularly closely related, and presumably Pirahã and Shipibo are related at some deep level, but overall few shared inherited or areal similarities are likely across this set of languages.

Table 5. Kadazan and Shipibo vowels and partial list of consonants

Kadazan p, b, ß, m; t, d, n, l, r, s; k, g, ŋ; i, ə, a, u.
5339 syllables
Shipibo p, ß, m; t, n, r, s, ts; k; i, ə, a, u.
15520 syllables

Lexical data for each language were obtained from published and unpublished sources. For Hawaiian we excerpted every fifth headword from the dictionary by Pukui and Elbert (1986). For the other languages all words in the sources used were counted. For Rotokas the source was the trilingual vocabulary of Rotokas by Firchow, Firchow & Akoitai (1973), interpreted according to the analysis provided in Firchow and Firchow (1969). For Pirahã it was a vocabulary of verb roots compiled by Keren Everett and made available to us by Dan and Keren Everett. This contains many example sentences and all words that occur in the examples were extracted, not just the headwords. Interpretations of Pirahã phonology which differ with respect to the relationship posited between [h] and [k] are offered in Everett (1982) and Everett (1986); our data retain these as distinct segments. The data on Kadazan and Shipibo are from SIL dictionary files compiled by Hope Hurlbut and Dwight Day respectively, and made available through the good offices of Eugene Loos.

The lexical data were analyzed into syllables, and frequency counts of each consonant-vowel pair were made for each language, counting only the CV portion of CVV and CVC syllables where these occur. These raw scores were then converted into deviation scores in the following way. The total number of occurrences of each consonant was calculated and the frequency of each vowel with that consonant was

expressed as a percentage of the total for the consonant. The total number of occurrences of each vowel was calculated and expressed as a percentage of all vowels. The deviation score for a given vowel with a given consonant is then obtained by subtracting the percent occurrence overall for that vowel from the percent occurrence of that vowel with the given consonant. This procedure is equivalent to that employed by Janson and the results can be interpreted in the same way.

Results will be shown separately for the three major consonant place groupings in the next three tables, starting with labials in table 6. A mean deviation score for the labial consonants in each language is shown for each of the six vowels that occur in the set of languages. Of course, vowels represented by the same symbols in these languages may not be precisely similar in pronunciation, or have the same range of variation: but it is clear that each symbol represents segments that belong in at least the same broad category. The number of labial consonants in each language is shown in parentheses after the language name.

Table 6. Vowel deviation scores with labials in 5 languages.

	i	e	a	o	u	ə
Hawaiian (3C)	-5.4	-1.1	13.8	-4.2	-3.2	—
Rotokas (2C)	7.6	-0.1	3.6	-10.5	-0.2	—
Pirahã (2C)	6.9	—	**-2.3**	-4.6	—	—
Kadazan (4C)	-1.5	—	**3.1**	—	-3.7	0.3
Shipibo (3C)	-14.8	—	6.8	0.2	—	7.8
mean across lgs	*-1.4*	*-0.8*	*5.0*	*-4.8*	*-2.3*	*4.1*
mean across segs	*-2.7*	*-0.9*	*5.5*	*-4.2*	*-2.7*	*3.8*
p (2-tail t-test)	*.27*	*.74*	*.06*	*.14*	*.26*	*.21*

Two ways of summarizing the results for each vowel column are also shown, a mean of the deviation scores across the languages, and a mean across all of the individual segments. Comparing these two provides a check on the influence of the varying number of segments in the different languages. The significance of the difference from zero of the segment mean is indicated in the final line. Again, if there are no consistent, significant effects pairing these consonants with particular vowels, these means will be close to zero.

Articulatory convenience might be held to predict a preference for labial consonants to precede rounded vowels, that is for deviation scores to be positive with such vowels. These languages do not show such a preference. In fact mean scores of labials with rounded vowels are negative, though not significantly so. If any vowel is

preferred it is low central /a/. These results would be slightly more consistent with predictions from acoustic salience, although positive scores with front vowels might have been expected in that case.

Occurrences of the different vowels with coronals is shown in Table 7. Here articulatory convenience predicts positive deviation scores with front vowels. These mean scores are in fact positive but only by small margins. The mean score for the back vowel /u/ is similarly positive, but only because Kadazan has positive values. Positive values with back rounded vowels might be predicted from acoustic salience, since transitions would be large. But neither of these factors is demonstrating a dominant effect. The strongest tendency is for the low vowel /a/ to occur less often than expected after coronals, although even this result is not statistically significant.

Table 7. Vowel deviation scores with coronals in 5 languages

	i	e	a	o	u	ə
Hawaiian (2C)	**2.9**	4.0	-2.7	-1.1	-3.1	—
Rotokas (2C)	**3.1**	-0.2	-4.1	**2.1**	-0.8	—
Pirahã (2C)	**3.7**	—	-3.0	**-0.6**	—	—
Kadazan (6C)	**2.5**	—	-2.6	—	**4.8**	**8.2**
Shipibo (5C)	**1.4**	—	0.7	-0.2	—	-1.9
mean by lg	***3.0***	*1.9*	*--2.3*	*0.1*	*0.3*	*3.2*
mean by seg	*2.7*	*1.9*	*-1.9*	*0.0*	*2.1*	*3.6*
p	***.40***	*.61*	*.30*	*.99*	*.23*	*.16*

Occurrences of the different vowels with velars is shown in Table 8. Articulatory convenience predicts a preference for velars to occur with back vowels, whereas acoustic salience would predict a preference for non-back vowels. The results are quite variable, and are perhaps distorted by some unresolved questions about the actual distribution of /k/ in Pirahã, which results in very high deviation scores for velars in this language. It is not apparent that there is any statistically significant trend across these languages to confirm either predicted preference. Though no language prefers high front vowels with velars, two of the three languages with /u/ show no preference for pairing it with a velar.

Thus, to summarize, over the set of consonant place/vowel pairs examined in this study, our results show no major effect of the predicted speaker preference for articulatory convenience. Equally, there are no indications that preference for acoustically distinct transitions influences the occurrence of particular places with particular vowels in a consistent way.

Table 8. Vowel deviation scores with velars in 5 languages.

	i	e	a	o	u	ə
Hawaiian (1C)	-0.4	-2.2	-2.3	-1.7	6.6	—
Rotokas (2C)	-11.6	0.8	1.0	10.9	-1.2	—
Pirahã (2C)	-18.7	—	10.9	7.8	—	—
Kadazan (3C)	-0.9	—	3.1	—	0.0	5.0
Shipibo (1C)	-18.4	—	2.3	7.8	—	8.2
mean by lg	*-10.0*	*-0.7*	*3.0*	*6.2*	*1.2*	*6.6*
mean by seg	*-9.1*	*-0.2*	*3.7*	*7.3*	*0.7*	*5.8*
p	*.04*	*.88*	*.08*	*.20*	*.74*	*.10*

Our data, especially when additional languages which are not otherwise discussed in this paper are included, do show two types of particular salient deviation patterns. First, glides and secondary articulations tend to be avoided before the cognate vowels. For example, syllables like [wu], [kwu] and [ji] are disfavored. Second, there is evidence that dorsal articulations show coarticulatory adjustments to the location of the following vowel in the front-back dimension. The relatively high negative score for /i/ in table 8 is a partial reflection of this pattern. Note that of these two patterns the first is primarily attributable to auditory factors and the second to articulatory ones.

However, in general, the deviation scores in the languages reported here are quite low, meaning that across this set of languages, predicting the pattern of vowel occurrence after consonants at different places from the overall frequencies of the vowels in the given language comes quite close to the mark. This may suggest that rather than operating to favor particular *sequences* of sounds over others, any speaker- or listener-based preferences affect the frequency of individual *segments*. Having made a choice of segments, and determined their relative frequency, these languages for the most part do not further restrict the combinations into which they may enter.

Ohala (1980) has commented on the tendency for languages to make "maximum utilization of the available distinctive features" at the segmental level by using all compatible combinations of a small set of feature values. For example, a language with voiced and voiceless stops and nasals tends to have all of these at the same place places of articulation. What we have observed here seems somewhat akin to this, but operating at the syllabic level. It could be labeled a principle of "maximum utilization of the available segments", but it operates within the limits imposed by the overall frequency of individual segments. A grid showing the type of syllabic combination expected under this principle would look like figure 3. It has no empty

cells such as those that would remain in a grid such like that in figure 1 after all empty rows and columns had been deleted. If we imagine a language with (at least) the 6 vowels and 7 consonants shown, and these have the frequencies shown as percentages at the margins of this figure, then the percentage of the total syllables of the language that each CV combination would represent is that shown in the cell.

		i	e	a	o	u	ə	...
	%	20	10	25	10	15	5	...
p	10	2	1	2.5	1	1.5	0.5	···
t	20	4	2	5	2	3	1	···
k	15	3	1.5	3.75	1.5	2.25	0.75	···
m	5	1	0.5	1.25	0.5	0.75	0.25	···
n	20	4	2	5	2	3	1	···
ŋ	10	2	1	2.5	1	1.5	0.5	···
s	15	3	1.5	3.75	1.5	2.25	0.75	···
...	...	...	...	...	...	...	...	...

Figure 3. A grid of syllable frequencies derived from segment frequencies.

The relatively full utilization of syllabic potential that we find in these languages might well be attributable to cognitive factors to do with efficiency of storage of, and access to linguistic knowledge. By using the full range of combinations of a more limited number of articulatory routines, a given number of syllables has a more parsimonious representation than would be the case if different subsets of gestures were used in different sets of syllables. Moreover, many psycholinguists envisage the process of lexical access to involve 'activation' of a large number of candidate words from which final identification is made as more information becomes available, as in the cohort model of Marslen-Wilson (1987, 1989), or the TRACE II model of McClelland & Elman (1986). In the cohort model, the process is viewed as one which includes considering a cohort of possible words with a given onset and progressively eliminating candidates as more information about later segments is processed. Even if such bottom-up processing is only a part of accessing lexical information, a more equally populated syllable space enhances the efficiency of the search by tending to equalize the size of the cohorts.

Note that we are not saying that syllables of different types are equally frequent: They are not, because segments are not equally frequent. However, syllables of different phonetic shapes are more nearly frequent than might have been expected from considering articulatory and auditory factors in sequencing sounds. We suggest that some evening out of the inequalities of syllable frequencies is a desirable trait in language design for what might be called cognitive reasons. [3]

Ackowledgments
This research is supported by grant ROI DC00642 from the of the National Institutes of Health. We acknowledge the assistance of Eugene Loos, Dan and Keren Everett, Dwight Day et al.

Footnotes

1. This paper reflects the combined content of two conference papers, one presented at the 112th Meeting of the Acoustical Society of America, San Diego (Maddieson and Precoda 1990), and one presented at the Linguistic Society of America Annual Meeting, Chicago, January 3-6 1991 (Maddieson 1991).

2. Janson also specifically critiques the validity of Kawasaki's conclusions, which were based on looking at co-occurrence restrictions in a number of languages and drawing out general patterns of occurrence and avoidance. She concluded, inter alia, that labial consonants disfavor following (back) rounded vowels and that coronal consonants disfavor following front vowels. Janson pointed out that Kawasaki had in some cases stated her conclusions in broader terms than were warranted from the data she had gathered. For instance, the conclusion that labial consonants are avoided with following rounded vowels is based on data that actually largely shows restrictions on consonants with a secondary articulation of labialization or the labial-velar approximant /w/ itself, i.e. segments in which there are vocalic features in the consonants. Janson finds only five languages in Kawasaki's survey in which plain labial consonants are involved in co-occurrence restrictions of the relevant sort.

3. Future research will be addressed to the question of whether similar patterns obtain in languages with much richer syllabic inventories, or if this is a tendency that is particularly strong in languages with relatively small numbers of distinct syllables.

References

Everett, Daniel. 1982. Phonetic rarities in Pirahã. *Journal of the International Phonetic Association* 12: 94-96.

Everett, Daniel. 1986. Pirahã. In *Handbook of Amazonian Languages* (ed. D Derbyshire & G. Pullum). Mouton, Berlin: 200-326.

Firchow, I., J. Firchow & D. Akoitai. 1973. *Vocabulary of Rotokas-Pidgin-English.* Summer Institute of Linguistics, Papua New Guinea Branch, Ukarumpa.

Firchow, Irwin & Jacqueline Firchow (1969). An abbreviated phoneme inventory. *Anthropological Linguistics* 11: 271-6.

Janson, Tore. 1986. Cross-linguistic trends in CV sequences. *Phonology Yearbook* 3: 179-196.

Kawasaki, Haruko. 1982. An acoustical basis for universal constraints on sound sequences. Ph. D. dissertation, University of California, Berkeley.

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

Lindblom, Björn. 1986. Phonetic universal in vowel systems. In *Experimental Phonetics* (ed. J.J. Ohala & J.J. Jaeger). Academic Press, Orlando: 13-44.

Lindblom, Björn. 1990a. Models of phonetic variation and selection. In *Language Change and Biological Evolution* (ed. L. Cavalli-Sforza & A. Piazza). Stanford University Press, Stanford CA.
Lindblom, Björn. 1990b, ms. Phonological units as adaptive emergents of lexical development. To appear in *Phonological Development*. York Press, Parkton MD.
Maddieson, Ian. 1991. Syllable structure and phonetic models. Paper presented at the Annual Meeting of the Linguistic Society of America, Chicago, January 3-6, 1991.
Maddieson, Ian & Kristin Precoda. 1990. Preferred syllables. *Journal of the Acoustical Society of America*: 88, Suppl 1: S82-S83 (Abstract).
Marslen-Wilson, William. 1987. Functional parallelism in spoken word-recognition. *Cognition* 25: 71-102.
Marslen-Wilson, William. 1989. Access and integration: projecting sound onto meaning. In *Lexical Representation and Process* (ed. W. Marslen-Wilson). MIT Press, Cambridge MA: 3-24.
McClelland, J. L. and J. L. Elman. 1986. The TRACE model of speech perception. *Cognitive Psychology* 18: 1-86.
Ohala, John J. 1980. Chairman's Introduction. *Proceedings of the Ninth International Congress of Phonetic Sciences.* University of Copenhagen: 184-5.
Ohala, John J. 1990. There is no interface between phonology and phonetics: a personal view. *Journal of Phonetics* 18: 153-171.
Ohala, John J. & Haruko Kawasaki. 1984. Prosodic phonology and phonetics. *Phonology Yearbook* 1: 113-128.
Pukui, Mary K. & Samuel H. Elbert. 1986. Hawaiian Dictionary, revised and expanded edition. University of Hawaii Press, Honolulu.
Stevens, K.N. 1972. The quantal nature of speech: evidence from articulatory-acoustic data. In *Human Communication: A Unified View* (ed. E.E. David & P.B. Denes). Academic Press, London: 51-66.
Stevens, K.N. 1989. On the quantal nature of speech. *Journal of Phonetics* 17: 3-45.

An Acoustic Study of So-Called Creaky Voice in Tianjin Mandarin

Deborah S. Davison

0. Abstract: the relation of pitch to other laryngeal properties in Mandarin Chinese dialects

Chao (1968) observes that the "creaky voice" articulation observed in citation pronunciation of Beijing Mandarin Chinese dialect low dipping tone three always co-occurs with low pitch in Chinese dialects. Thus it is not distinctive. Beijing Mandarin has only one low tone of four lexically contrastive ones. In this paper the distribution and acoustics of "creaky voice" in Tianjin Mandarin is examined, a closely related dialect with two of its four tones articulated in the lower register. It is found that in Tianjinese only tone three occurs systematically with a particular phonation type, and it does so in a manner orthogonal to its occurrence on relatively higher or lower pitch. Nevertheless, the two low tones are distinguished as well by pitch contour, in most environments. Of the possible contrastive features, then, pitch contour is assumed to be the more salient.

1. Data: the language, its tonal contours and history

Tianjin Mandarin is spoken by some three million natives of China's third largest city, located 100 km. southeast of Beijing. Average f0 contours of the four etymological citation tones in Beijing Mandarin and Tianjin Mandarin are compared in Figure 1. Notice that they differ significantly only in the contour of tone one (T1): Beijing's T1 is high level and Tianjin's is low falling. Thus Tianjinese has two tones, numbers one and three (T1 and T3), articulated entirely in the lower pitch register, whereas Beijing has only one, viz., tone three (T3).

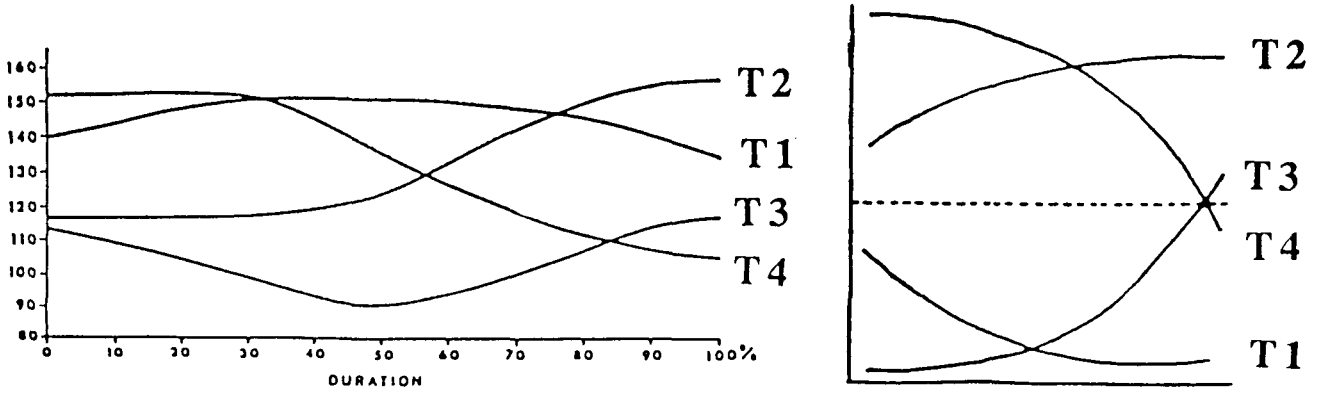


Figure 1. Average f0 contours of the four lexically distinctive tones in Beijing, left (after Howie 1976) , and Tianjin, right, (after Shi 1988).

Beijing Mandarin T3 in citation form is described synchronically as "a dipping pitch contour mildly falling in the first two thirds of the vowel, reaching almost the bottom limit of the range, and sharply rising in the last third; the loudness is generally slightly falling with a small rise at the very end of the vowel, and the duration is far above average" (Kratochvil 1968:37). A similar description is appropriate for Tianjin T3 in citation pronunciation.

On the traditional distinctive f0 contour analysis (Chao 1968), creaky voiced phonation optionally accompanying the articulation of this tone is assumed to be due to and coincide with reaching the bottom limit of the pitch range. Yet Ladefoged (1973:75) observes about the creaky voice phonation mechanism that "because the arytenoid cartilages move forward as they come together, the vocal cords tend to be less stretched in creaky voiced sounds; they are therefore likely to vibrate at a lower frequency. But the coming together of the arytenoids and the movements of the thyroid cartilage that stretch the vocal cords are independent laryngeal gestures, so that it is quite possible for creaky voiced sounds to occur on any pitch."

The four phonologically distinctive tones on monomorphemic syllables in Mandarin Chinese dialects are described in the phonetic and phonological literature primarily in terms of f0 contour and secondarily in terms of amplitude and duration, without reference to phonation differences. This is in contrast to other tone languages, including in particular the genetically related Wu Chinese dialects (Cao & Maddieson 1989) as well as many Southeast Asian languages, for which breathy voice and creaky voice registers associated to "tonal" categories are widely acknowledged.

Haudricourt (1956), Egerod (1971), Pulleyblank (1978), Baxter (1984) and others writing on Chinese historical phonology argue that, on the model of development of the tone systems of Southeast Asian languages, systematic phonetic features of modern-day Mandarin tones are reflexes of distinctions in phonation type from which the tones arose. One candidate, mentioned specifically in Egerod (1971), is the creak associated with Mandarin T3. Whether or not the tone system of Mandarin retains acoustically discernible traces of phonation contrasts largely remains uninvestigated but cf. the Beijing Mandarin electromyographic study by Sagart et al. (1988).

Sagart et al. show a high degree of correlation between peaks of activity of the sternohyoid (SH) muscle in both segmental articulation and control of pitch (f0). A strong peak of SH activity "aligned with the initial consonant or with vowel onset...is believed[to]...reflect the segmental component of SH activity." They also observe that "the highest SH peak occurring during initial consonant is seen with T2, suggesting that this peak results from the superposition of a sgment-oriented component and a f0 oriented component, the latter being related to the f0 trough at the beginning of T2...Otherwise, peaks of SH activity occur during T3...and in the second part of T2...SH activity in these two locations appears to be related to f0 decrease (in the case of T3) or deceleration (in the case of T2). Note that the f0 fall in T3 does not require a participation of the SH...Participation of the SH in [T1] is dubious." (p. 8)

T2 is known to derive historically from voiced initial T1 words. Its initial depression followed by a pitch rise, correlated, according to Sagart's data, with SH muscle activity, is conceivably a direct reflex of the earlier primarily consonantal contrast. How SH involvement relates to the history of T3, as opposed to T2, is less clear, though the historical argument in favor of its representing a reflex of the earlier T3 final consonant could be made.

2. Experimental design: pitch and spectral measurements of elicited natural speech

Natural, connected speech as well as the citation pronunciation of a variety of words in Tianjin Mandarin were elicited from three monolingual natives of Tianjin, one female and two males, and recordings made *in situ*, in 1981. The data were analyzed using a Kay DSP Sona-Graph Model 5500. A corpus of 222 tokens of fully stressed syllables containing the low central vowel /a/ was selected for analysis. Three equidistant values of the highest well-defined harmonic across the vocalized portion of the syllable were measured from narrow-band spectrograms, and the f0 at each point was calculated. The mean pitch values at onset, middle, and offset and standard deviation about the mean for the four tones across all tokens were determined.

Measures of spectral tilt expressed as the amplitude difference between f0 and H2 were calculated from narrow-band power spectra of a 50 ms window moved across the vowel steady state portions of the same tokens. To ensure that the points measured for each token were representative, three to six points were measured across each vowel. In case of variation the highest and lowest values were rejected, as were peripheral points, after which the lowest value represented by two or more points was selected, where possible. The spectral tilt calculation, involved subtraction of the peak amplitude value of f0 from that for H2. High positive values (where the amplitude of the fundamental is much higher than H2) correlate with creaky voice, mid values with modal voice, and high negative values (f0 amplitude weaker than H2) with breathy voice, after Ladefoged et al. (1988). Thus for example breathy phonation is correlated with high amplitude of f0 in comparison to the second harmonic. The resulting measurements from all three speakers were pooled. Sample tokens and points measured are illustrated in Figure 2. F0 and spectral tilt values were plotted against each other for all tones and for each of the four phonemic tones, as displayed in Figure 3.

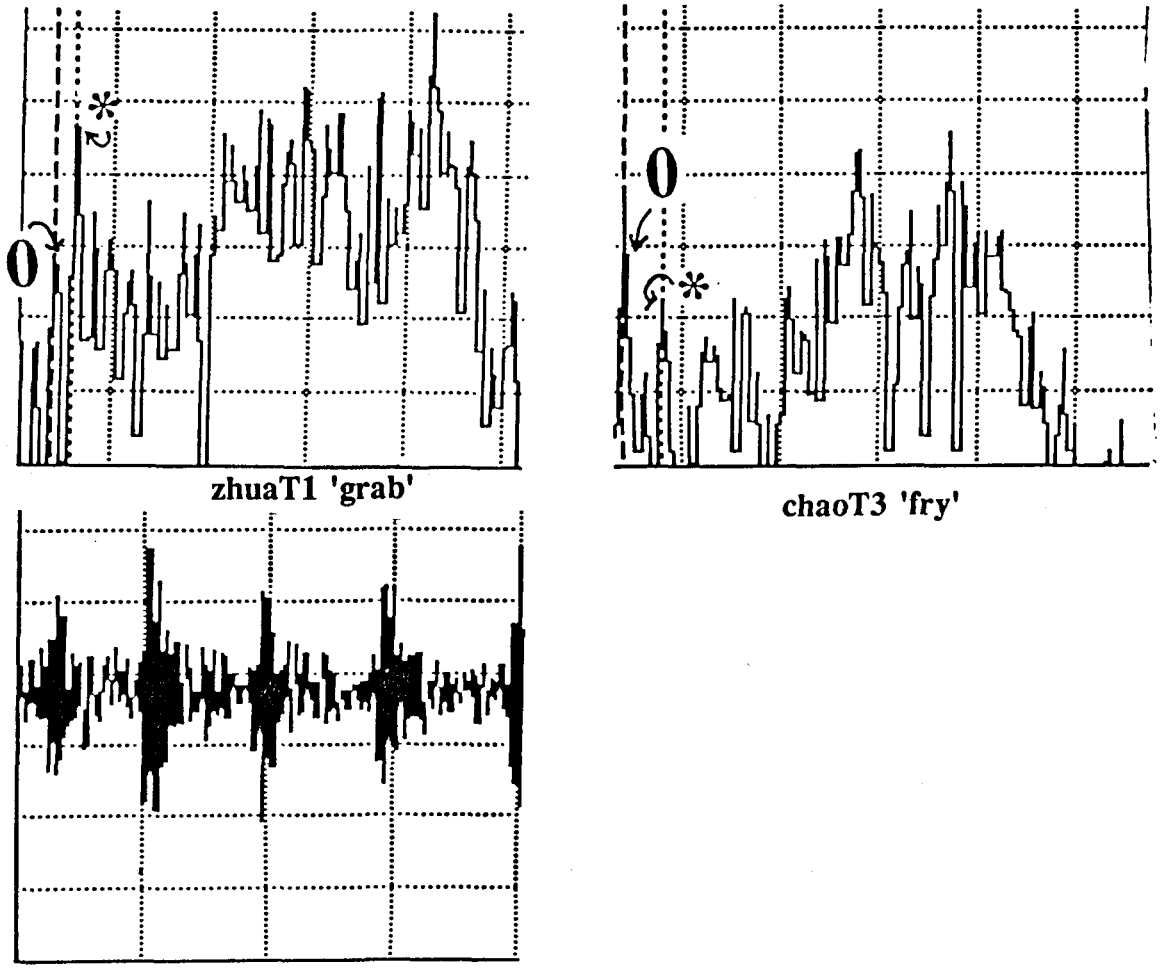
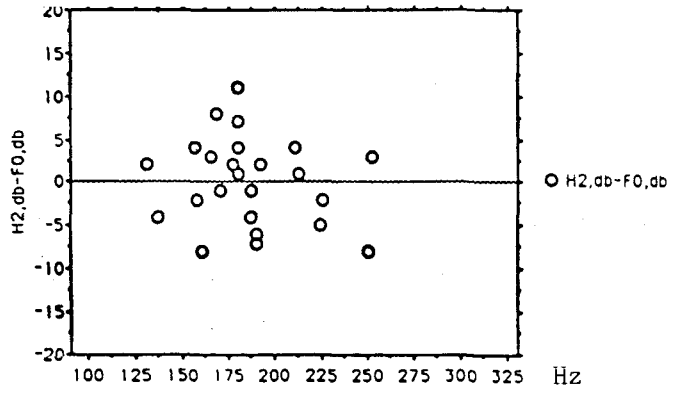
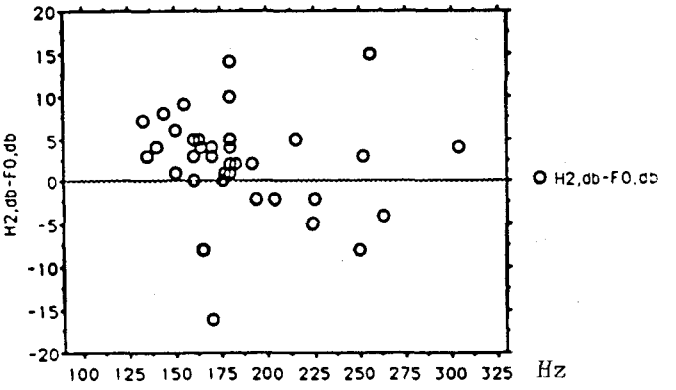


Figure 2. Sample tokens of the power spectra used for measurement. Spectra are from the steady state portions of the vowel /a/ in the disyllabic word [tʃwa ˩ tʃaw ˧˥] *zhua chao* "grab - fry" (T1 + T3). The waveform of the vowel in the first syllable is also illustrated. The fundamental component of the spectrum is indicated by **0** and the second harmonic by *.

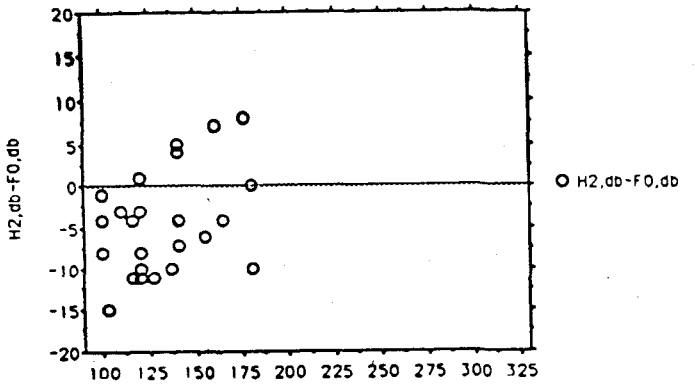
Tone 1 (low falling pitch)



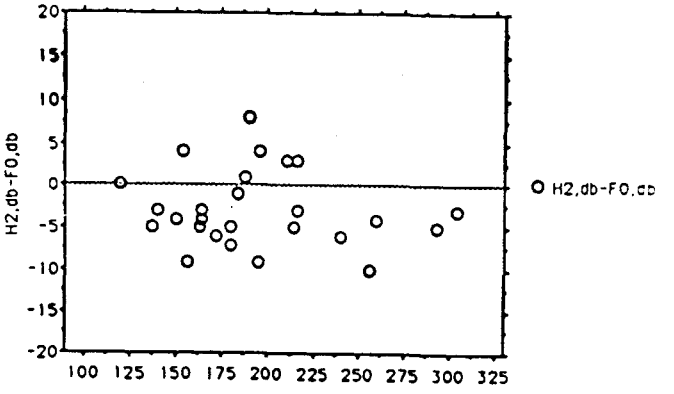
Tone 2 (high rising pitch)



Tone 3 (low dipping pitch)



Tone 4 (high falling pitch)



Composite (all tones)

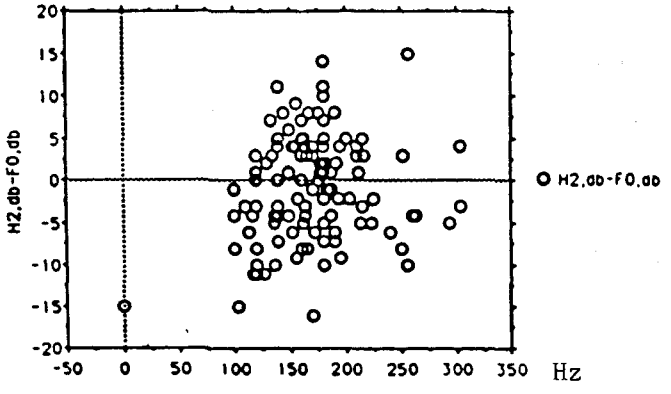
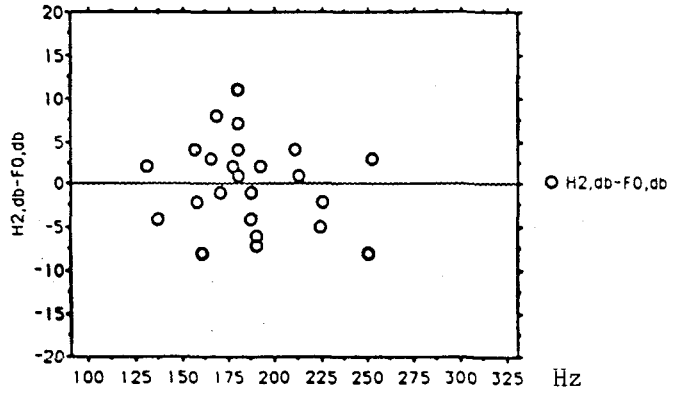
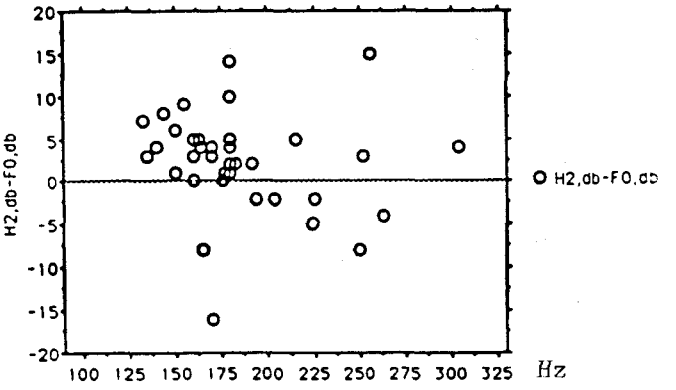


Figure 3. Scattergrams of x1:spectral tilt to y1:f0

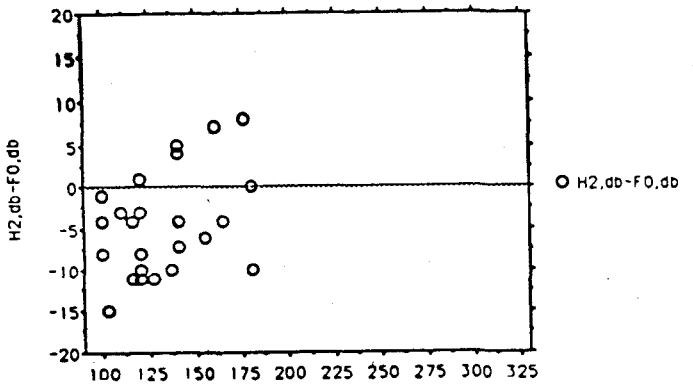
Tone 1 (low falling pitch)



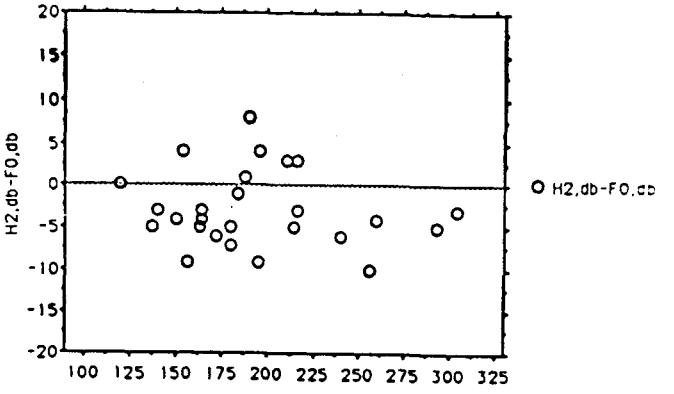
Tone 2 (high rising pitch)



Tone 3 (low dipping pitch)



Tone 4 (high falling pitch)



Composite (all tones)

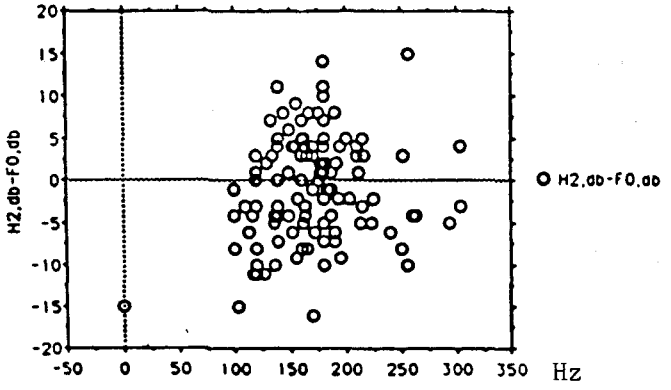


Figure 3. Scattergrams of x1:spectral tilt to y1:f0

3. Results

From the scattergrams in figure 3 it is clear that pitch and spectral tilt do not co-vary, either for individual tones or in aggregate. Thus the claim that the so-called creaky voice phonation depends on low f0 is not supported by the Tianjin Mandarin data summarized here. On the other hand, analysis of variance shows that spectral tilt serves to distinguish between tonal categories. The results are shown in table 1. A scattergram displaying the range of spectral tilt values for each tone is shown in figure 4.Post hoc evaluation of the difference between the means shows that T3 is significantly different from T1 and T2 at the 95% level. The level of significance is slightly less for T1 vs. T3 than for T2 vs T3, possibly indicating a weak interaction of low pitch with spectral values. Thus these data indicate that in Tianjinese the phonation type characterizing pronunciation of T3 may be a redundant but systematic, language-specific enhancement of the T3 pitch contour.

Of the three speakers, two have T4 values similar to those of T3, while one does not. Further examination of T3/T4 spectra across speakers is needed to determine whether the spectral tilt of these two tones consistently pattern together.

The electromyographic data on Beijing Mandarin, grouping tones T2 and T3 against T1 and T4 with respect to firing of the sternohyoid muscle are orthogonal to the Tianjin Mandarin acoustic data, which group T1 and T2 against T3 and T4, with respect to spectral tilt.

Although T1/T2 and T3/T4 may differ significantly in values for spectral tilt, the pitch contours remain independently distinctive (except in tone sandhi environments, see below.) Thus whether the phonation contrast is phonological in Tianjinese remains to be determined. Gårding et al. (1986)'s perception test results distinguishing simulated Beijing Mandarin T3 and T4 which varied in gradient fashion across tokens according to pitch peak location and steepness were not affected by the addition of simulated creak.

Table 1. Analysis of variance of spectral tilt measures for main effect of tone, and analysis of means (1=T1, 2=T2, 3=T3, 4=T4)

Analysis of Variance Table

Source:	DF:	Sum Squares:	Mean Square:	F-test:
Between groups	3	840.886	280.295	7.935
Within groups	110	3885.641	35.324	p = .0001
Total	113	4726.526		

Model II estimate of between component variance = 81.657

Group:	Count:	Mean:	Std. Dev.:	Std. Error:
1	26	1.154	6.473	1.269
2	27	2.815	6.587	1.268
3	27	-4.481	5.8	1.116
4	34	-1.676	5.032	.863

Comparison:	Mean Diff.:	Fisher PLSD:	Scheffe F-test:	Dunnett t:
1 vs. 2	-1.661	3.237	.345	1.017
1 vs. 3	5.635	3.237*	3.969*	3.451
1 vs. 4	2.83	3.069	1.114	1.828
2 vs. 3	7.296	3.206*	6.782*	4.511
2 vs. 4	4.491	3.037*	2.865*	2.932
3 vs. 4	-2.805	3.037	1.117	1.831

* Significant at 95%

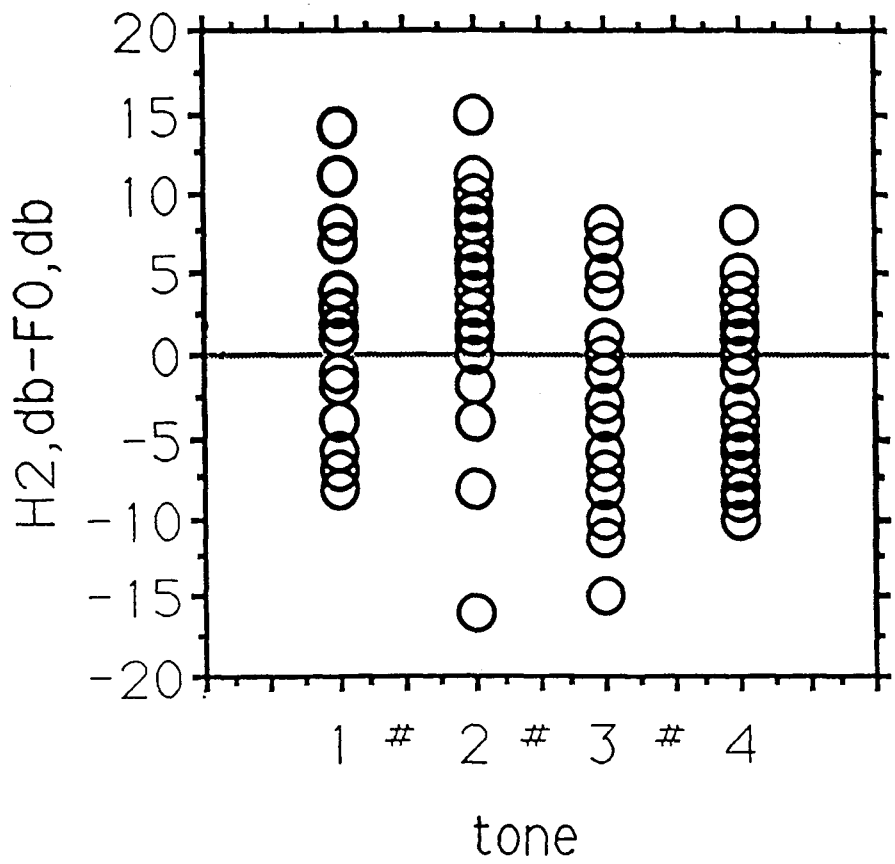


Figure 4. Scattergram showing range of spectral tilt values by tone category

4. Discussion: the non-modal phonation is auditorily breathy but spectrally creaky

A distinctive phonation type, impressionistically identified in the literature as "creaky voice", is found to co-occur systematically with one more than the other of the two low phonemic tones of Tianjin Mandarin Chinese dialect, consistent with Ladefoged's claim of the potential independence of the articulatory mechanisms producing low pitch and creak. This is observable by examining the scattergram in Figure 4. T3 has the lowest negative values of spectral tilt overall, T2 the highest positive ones. Values for T1 and T4 lie somewhere in between. Nevertheless, the statistical analysis show that mean spectral tilt values for the low falling T1 and low rising T3 are significantly different at 95% by Fisher PLSD and Scheffe F-test and at 99% by Fisher PLSD but not Scheffe F-test.

The data support the interpretation from Chinese historical phonology that "creaky voiced" phonation paired with its particular tone category three is a reflex of an earlier distinctive phonation type which later developed into a contrast marked primarily by pitch contour. Low rising and high falling pitches may have originated as redundant features of distinctive, non-modal laryngeal gestures, as in the historical reconstructions.

Power spectra for the non-modal phonation type in modern-day Tianjin Mandarin resemble those for breathy voice, not creaky voice, in other languages: high f0 amplitude in comparison to H2. Breathy voice is assumed to be characterized by low frequency sinusoidal vocal cord movement, as opposed to abrupt, sharp contact, typical of creaky voice. Since the auditory impression of T3 is not of breathiness, the explanation for T3's spectral profile remains unclear. Curiously, the historical prediction for T4, which is not significantly distinct from T3 in spectral tilt, is that it derives from breathy voiced phonation. We will attempt to examine Beijing Mandarin and other closely related dialects, in future work.

5. Postscript: tone sandhi and neutralization

Beijing and Tianjin Mandarin dialects undergo tone sandhi. A rule shared by both dialects changes T3 to T2 before a following T3. Tianjin also has three additional rules, described as changing a sequence of T4s to T1 plus T4, a sequence of T1s to T3 plus T1, and a sequence of T4 plus T1 to T2 plus T1. These rules are presented in table 2.

Table 2. Contextually conditioned pitch contour changes: tone sandhi in Beijing and Tianjin Mandarin

(a) low dipping + low dipping -> high rising + low dipping T3 + T3 -> T2 + T3 (Beijing & Tianjin)	˨˩˦ + ˨˩˦ -> ˧˥ + ˨˩˦
(b) low falling + low falling -> low dipping + low falling T1 + T1 -> T3 + T1 (Tianjin)	˩ + ˩ -> ˨˩˦ + ˩
(c) high falling + high falling -> low falling + high falling T4 + T4 -> T1 + T4 (Tianjin)	˥˩ + ˥˩ -> ˩ + ˥˩
(d) high falling + low falling -> high rising + low falling T4 + T1 -> T2 + T1 (Tianjin)	˥˩ + ˩ -> ˧˥ + ˩

The relevant data in this corpus are very few, but nevertheless limited observations about them may be made. According to my limited data, T1, supposed to neutralize with T3 before a following T3, does not change to both the pitch and phonation type characteristic of T3, as complete neutralization might predict. The four tokens of T1->T3 resulting from the application of (b) have spectral tilt values (H2-f0) of 14, 3, 4, and 1 respectively. The two tokens of an original T3 before T1 have -4 and 8. These results are consistent with Davison's (1984) observation that rule (b) more closely resembles one of the form T1 + T1 -> T2 + T1 auditorily. As for rule (a), the only token of T3->T2 has value 0, consistent with Egerod's (1971) analysis of the Beijing Mandarin rule as involving phonation dissimilation.

Acknowledgements

Fieldwork in Tianjin in 1980-82 was supported by a grant from the Committee on Scholarly Communication with the People's Republic of China. This research was supported by NIH Phonetics Training Grant No. 1T32 DC 00029-01.

References

Baxter, William H., III. 1984. Reconstructing Old Chinese: the Bodman-Baxter system. 17th International Conference on Sino-Tibetan Languages and Linguistics, Eugene.

Chao, Yuen Ren. 1968. *A Grammar of Spoken Chinese.* Berkeley, University of California Press.

Cao, Jianfen & Ian Maddieson. 1989. An exploration of phonation types in Wu dialects of Chinese. *UCLA Working Papers in Phonetics* 72:139-160 & *J. Phonetics*, forthcoming.
Davison, Deborah S. 1987. Tone sandhi in Tianjin dialect. Paper presented at the XXIth International Conference on Sino-Tibetan Languages and Linguistics.
Egerod, Soren. 1971. Phonation types in Chinese and South East Asian languages. *Acta Linguistica Hafniensia* (Copenhagen) XIII, 2:159-171.
Gårding, Eva, Paul Kratochvil, Jan-Olof Svantesson & Jialu Zhang. Tone 4 and tone 3 discrimination in Modern Standard Chinese. *Language and Speech* 29:3:281-293.
Haudricourt, A. G. 1954. Comment reconstruire le Chinois archaique. *Word* 10:351-364.
Howie, John Marshall. 1976. *Acoustical studies of Mandarin vowels and tones*. Cambridge: Cambridge University Press.
Kratochvil, Paul. 1968. *The Chinese language today: features of an emerging standard.* London: Hutchinson University Library.
Ladefoged, Peter. 1973. The features of the larynx. *J. Phonetics* 73:73-84.
Ladefoged, Peter, Ian Maddieson, & Michel Jackson. 1988. Investigating phonation types in different languages. In Fujimura, Osamu, ed., *Vocal Physiology: Voice Production, Mechanisms and Functions,* New York: Raven Press, Ltd.
Pulleyblank, Edwin G. 1978. The nature of the Middle Chinese tones and their development to Early Mandarin. *Journal of Chinese Linguistics* 6.2:173-203.
Sagart, Laurent *et al.*. 1988. Electromyographic investigation of the tones of Modern Standard Mandarin. Paper presented at the XXIth International Conference on Sino-Tibetan Languages and Linguistics.
Shi, Feng. 1988. Shilun Tianjinhuade shengdiao jiqi bianhua [Analysis of Tianjin dialect tones and their transformations]. *Zhongguo Yuwen* 5:351-360.

Stress and Tonal Targets in Tianjin Mandarin

Deborah S. Davison

0. Abstract

A target and interpolation model of acoustic data is used to represent aspects of the linguistic coding of lexical and phrasal stress and phonological tone sandhi (TS) rules in Tianjin Mandarin dialect. It is argued that the phonology makes reference to high and low tones, while register and scalar effects belong to the phonetic, not phonological representation. The TS rules are interpreted as serving to mark prominence relations within an iambic foot.

1. The problem: describing the interaction of the realization of tones, the application of tone sandhi, and the realization of prominence in Tianjin Mandarin

As described by Li & Liu (1985), Chen (1986), and Chen et al. (1987), Tianjin Mandarin dialect has four tone sandhi rules (TSRs, see table 3) whose application, especially to multisyllabic domains, is incompletely understood. Chen (1986) and (1987) claims to show that the TSRs do not apply uniformly either in a particular direction, across a particular size domain, or at a particular level. Authors have attempted to rationalize aspects of the problem with reference to various phonological principles. Yip (1989) analyzes some of the rules as involving dissimilation of identical sequences of contour tones. Milliken (1990) analyzes the TS effects as phonological responses to the more standard types of OCP violations or to what he calls the "maximum association condition" (MAC) in the cyclic phonology, and as involving floating tone attachment in the post-cyclic phonology (see (7)).

In this paper I argue that the Tianjinese tone sandhi rules serve rather to locate or isolate a single high (H) accent either near the leftmost boundary of or on the second syllable of a disyllabic word, just in case the second syllable of the disyllable is prosodically strong. My analysis is consistent with the fact for Tianjinese that TSRs generally apply only to disyllabic words or phrases bearing a weak-strong (iambic) stress pattern.

Consistent with this interpretation, systematic pitch-based correlates of lexical and phrasal prominence are also found in environments other than the tone sandhi ones. For example, an expanded pitch range resulting in phonetic super-high highs and super-low lows may characterize the f0 values of the first syllable of a disyllabic word bearing a strong-weak stress pattern. To support the claim of the existence of a close relationship between the realization of tonal allotony and prominence, examples of assorted means of phonetic realization of prominence in Tianjinese are given. These include lexically H tone becoming super high, contour tones shifting to high register, and systematic delay of the occurrence of a tonemic H peak to late in the syllable, just in case the syllable is heavily stressed.

A complete account of the phonetics of tone and stress in Tianjinese is beyond the scope of this study. Here the focus is to show the dependency between prominence and both the realization of basic tones and the conditions of application of TS to the basic tones. It is then argued that TS application is not merely conditioned by prominence relations but is in fact a means by which they are linguistically encoded.

The structure of the paper is as follows. First the facts of Tianjin tones and tone sandhi as presented in the existing literature are reviewed. Then an example of a problematic case of TS application to a trisyllabic structure is presented. A reanalysis is offered in terms of stress induced tone reduction and subsequent tone shift. Next, one example each is presented of stress-induced register shift, change of tonal target from high to super-high, and tonal peak delay. Having

shown that prominence affects tone realization on the phonetic level in Tianjinese, spectrographic evidence is then given of its effect on TS application. I conclude with the proposal that the reason Tianjin tone sandhi is restricted to applying within an iambic foot is because it operates to create a LHL tune (duh dUUH-like), in which the H is located at the intermediate syllable boundary, which is characteristic of and linguistically marks iambic feet in particular.

The new Tianjinese data presented here are from samples of natural, connected speech of four monolingual speakers of Tianjin city dialect, collected by the author in Tianjin, People's Republic of China, in 1980-82. Pitch contours are traced from narrow-band spectrograms produced from tape recordings using the Kay DSP 5500 spectrograph. Figures based on these materials will be discussed at relevant points below. Most of my data on Tianjin tones is in the form of natural rather than elicited speech, and several points of interest about the dialect currently remain unclear to me. Here I focus only on the more robust, relatively unambiguous effects.

2. Tianjin Mandarin tones and tone sandhi

Tianjin Mandarin dialect is spoken in Tianjin, China's third largest city, located 100 kilometers southeast of Beijing. Like Beijing Mandarin, Tianjin Mandarin has four lexically contrastive tones, which are traditionally labeled A, B, C, and D. In Beijing Mandarin the tones can be illustrated with the examples *ma*A ˥ mother, *ma*B ˧˥ numb, *ma*C ˨˩˦ horse, *ma*D ˥˩ angry. The cognate Tianjinese citation tone values are somewhat similar: *ma*A ˨˩ mother, *ma*B ˧˥ numb, maC ˨˦ horse, *ma*D ˥˩ angry. The most obvious auditory difference between the citation tones of Beijing and Tianjin Mandarin is that Beijing's tone A is high, while Tianjin's tone A is low. Roman numerals from one to five are used to represent the phonetic contour of the pitches over time, one being low pitch, five high pitch, as in Table 1. The representations given by Li & Liu (1985) and Shi (1986) differ slightly in the assigned pitch values.

Table 1. Citation tones of Beijing and Tianjin Mandarin compared

	Beijing	Tianjin (Li & Liu, 1985)	Tianjin (Shi, 1986)
Tone A	55	21	211
Tone B	35	45	455
Tone C	214	213	113
Tone D	51	53	553

These tonal contours have been given a number of different representations in terms of tone features in recent literature. Five of these representations are compared in Table 2.

Table 2. Underlying phonological representations of Tianjin tones

Note: L=[-high], H=[+high], -R=[-rising], R=[+rising], -C=[-convex], C=[+convex]. Floating tones in bold type.

	Shi (1986)	Davison (1987)	Davison (1988/9)	Yip (1990)	Milliken (1990)
Tone A	L, -R, -C	L	**L**	L	**H**L
Tone B	H, +R, +C	LH	**H**	H	H
Tone C	L, +R, -C	L**H**	L**H**	LH	LH
Tone D	H, -R, +C	HL	**H**L	HL	HL

There are four tone sandhi rules (TSR's) in Tianjinese whose phonetic effects are shown in Table 3. They may be interpreted as changes of tone categories, in the way shown on the right in this table. In other words, the TSRs operate to create environments of tonological neutralization in the leftmost of a pair of syllables. Thus the rule affecting a sequence of two A tones (the AAR) changes two low falling tones to a low rising plus low falling tone, or CA sequence; the CCR

changes two low rising tones to a high plus low rising tone, or BC; the DDR changes two high falling tones to a low plus high falling tone pattern, or AD; and the DAR changes a high falling plus low tone to high plus low pattern, or BA.

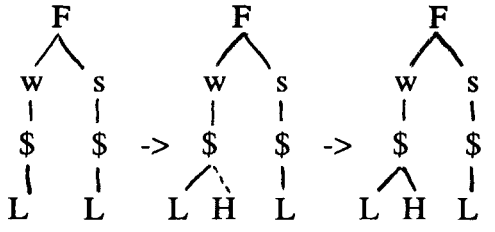
Table 3. Tianjin tone sandhi rules

	phonetic form	categorical shift
AA Rule:	21 + 21 -> 13 21	AA -> CA
CC Rule:	13 + 13 -> 45 13	CC -> BC (also in Beijing)
DD Rule:	53 + 53 -> 21 53	DD -> AD
DA Rule:	53 + 21 -> 45 21	DA -> BA

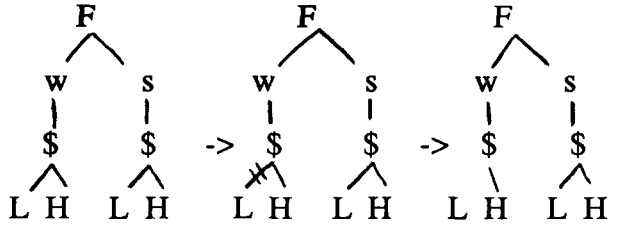
Table 4 summarizes this author's interpretation of the autosegmental processes involved in the application of the TSRs in Table 3 to iambic disyllables in Tianjinese. The underlying representations used here as elsewhere in this paper are those found in Table 2 under Davison (1988/9). (These are identical to those of Yip (1990).) Later they are compared to the representations of Milliken (1990) (also given in Table 2), in particular his representation of tone A as underlyingly a floating **H** plus a L tone. Notice that tone A, while phonetically low falling 21 in citation form, is phonologically simply L, and tone B, phonetically high rising 45, is phonologically simply H. This choice of representation is defended in the discussion of Shi (1986)'s dynamic tone analysis in section 2.3 and the analysis in the concluding section.

Table 4. Autosegmental analysis of Tianjin TS rules. Tone A=L, B=H, C=LH, D=HL.
F=foot, w=weak, s=strong, $=syllable.

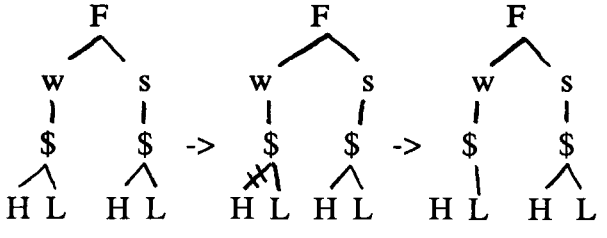
AAR: High tone insertion



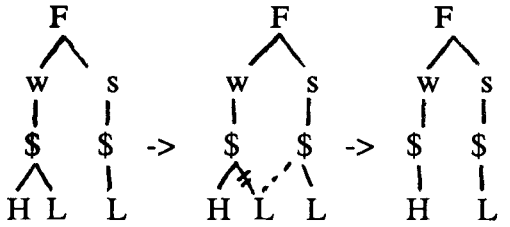
CCR: Contour simplification



DDR: Contour simplification



DAR: Low tone deletion



2.1 The Tianjin tone sandhi paradox

The "paradox of Tianjin tone sandhi", discussed in Chen (1986) and Chen, Hung, Tan, and Zhang (1987), is that no single, unified manner of application of Tianjin dialect's four tone sandhi rules as cyclic (5a) or post-cyclic (5b), right-to-left (5b) or left-to-right (5c), uniformly accounts for the pattern of application of each of the four TSRs to multisyllabic words observed in Li & Liu (1985). Table 5 contains examples from the literature purported to illustrate that the TSRs do not apply uniformly to prosodic units larger than disyllables. There is disagreement about the facts of application to trisyllables and general acknowledgement that the four rules vary among themselves in the degree of obligatoriness with which they apply even to disyllables. An approximate consensus view is summarized in Field (1990). It maintains that the AAR and DDR apply right to left, the CCR left to right; and that the DAR applies last/post-cyclically. Trisyllables thus are derived as in (5d).

Table 5. The "paradox" of Tianjin TS exemplified

(a)*Cyclic application:* [*chou*A->C [[*zhan*D->A *dou*D]->B *yan*A]] 'smoke Struggle cigarettes'

DDR		[A]	
DAR		[B]	
AAR	[C		]
	'smoke'	'struggle'	'cigarette'

DDR applied on the first cycle feeds AAR on the second cycle (post-cyclic R-to-L application would give *ADBA). However,

(b) Post-cyclic application, R-to-L:

[[*jian*D *zhu*D]->A *wu*D] 'building'

[	A]	DDR
'building'	'thing, instance'	

[[*ji*A *guan*A]->C *qiang*A] 'machine gun'

[	C]	AAR
'machine'	'gun'	

Neither DDR nor AAR applies on the innermost cycle, therefore they are post-cyclic. They apply to the rightmost pair of syllables first, therefore they must apply R-to-L. Whereas,

(c) Post-cyclic application, L-to-R:

[*chang*C->B [*dang*C->B *wei*C]] 'Party committee of the factory'

[	B	B]	CCR
'factory'	'party'	'committee'	

If CCR were cyclic or post-cyclic R-to-L (rather than post-cyclic L-to-R) it would apply first on the innermost cycle to give *CBC.

(d) TSR application to trisyllables (Chen et al. 1987)

ADD -> AAD -> CAD	*xin [dian shi]* 'new television'
CAA -> CCA -> BCA	*[bao wen] bei* 'thermos cup'
CCC -> BCC -> BBC	*l[ing dao] jiang* 'leader said...'
DAA -> DCA	*[dian che] xin* 'trolley (is) new'
DDA -> ADA -> ABA	*[da mai] zhou* 'oat porridge'
DDD -> DAD	*zuo [zuo ye*] 'do homework'

Here I do not review the derivations of multisyllabic words in detail, since Field (1990)'s study reveal the data to be seriously flawed, see 2.4. It is rather assumed that the portion of multisyllabic data discussed below is more accurately analyzed as involving reduction to tonelessness of the middle syllable under conditions of weak stress, a process which is optionally accompanied by compensatory tone shift in some contexts.

2.2 Dynamic phonology: phonetic evidence

Before proceeding to the autosegmental analysis adopted in this paper, the usefulness of the dynamic tone analysis proposed by Shi (1986) following Wang (1967) is briefly considered. The issue arises in particular regarding Tianjinese because in citation pronunciation Tianjinese has no phonetic level tones. This is apparent in the numerical notation: tone A is 21, hence low-falling; tone B is 45, hence high rising; tone C is low dipping in isolation and low rising in most other contexts; and tone D is high-falling. Thus in principle a dynamic tone analysis is possible involving two rising and two falling tones which contrast in a high or low register feature. Accordingly, Shi describes the Tianjin basic tones (as well as TSRs) using contrasting values for the features high register (H/L), rising (R), and convex (C), as in Table 2. The TSRs so described can be represented as changes in one feature: CCR and DDR change the first syllable from low to high and high to low register respectively; and AAR and DAR change the first syllable from falling to rising.

However, there are phonetic reasons to reject the dynamic tone solution. In contexts other than isolation citation pronunciation, phonetically all four tones occur bearing opposite or neutral values for the dynamic feature Rising. This is apparent in the pitch traces exemplified throughout the text and in the composite data on tonal co-articulation in Figure 2 below. We see that low falling A tone is often level word internally; low rising C tone may occur as level before tones B, D, and toneless syllables; high rising tone B is found to fall in some contexts (see Figure 3c); and high falling tone D demonstrably rises, then falls, when stressed (Figure 3b). It is not apparent on examination of the pitch traces that these changes would be described more efficiently by a dynamic model than by (at least a simple) model which interpolates between tonal targets.

As to the question whether tone or register is involved in the sandhi rules, the DD->AD data serve as an interesting test case, since on a gross auditory basis both the CCR and DDR could be conceived as involving register shifts (cf. Shi 1988). In fact, on closer examination of the pitch traces we find for the DD->AD case that the first syllable is neither always low nor always falling. In the former case production errors or intonational influence could be invoked. But in the latter, the fact that the derived low tone sometimes drifts upward as it approaches the D syllable boundary seems most compatible with a tonal target and interpolation rather than a dynamic registral treatment of the DDR process. The AAR and DAR in any case are uncontroversially able to be represented as tonemic phonological processes.

Additionally, independent arguments have been made against treating citation contours as basic in many Chinese dialects (Chan 1985, Davison 1989). Since the low tones A and C are often level in context, the variation noted above arguably should be described with reference to level rather than contour underlying forms. Lastly, the rules in Table 4 are not excessively more complicated than the dynamic ones, so for the above reasons and the general one of choosing a language universal tonal paradigm where possible, a treatment in terms of dynamic tones is rejected for purposes of the current study.

2.3 Variable application of tone sandhi: the role of prominence

Apart from Chen et al. (1987)'s lexical phonological and Milliken (1990)'s autosegmental treatments of the Li & Liu (1985) data, another approach is found in Field (1990). Field elicited from fourteen Tianjin native speakers the reading pronunciation of thirteen disyllabic and sixty-seven trisyllabic strings, for a total of 1,103 tokens. Each syllable was assigned a tonemic value of A, B, C, or D based on Hz values. The results are summarized in Table 6. Trisyllabic patterns occurring with less than 6% of the total number of tokens for a particular input are omitted from the table. Taking for example (6b), of 69 tokens of trisyllabic phrases with underlying CAA as input, 74% were pronounced with a CBA, i.e., presumably, LH-H-L pattern; 12% with BCA, i.e., H-LH-L, and 14% other.

Table 6. Variable TS patterns produced on 483 trisyllabic tokens by 14 native Tianjin speakers (Field 1990)

(a) ADD as input (n = 153).	Output AAD 58%, ADD 24%, CAD 16%, other 2%
(b) CAA as input (n = 69).	Output CBA 74% [!], BCA 12%, other 14%
(c) DDA as input (n = 109).	Output DBA 39%, ABA 30%, DDA 20%, other 11%
(d) DDD as input (n = 152).	Output DAD 69%, DDD 16%, other 15%.

Field (1990)'s study confirms impressionistic reports in the earlier literature of significant intra- and inter-speaker disagreement on acceptability judgements on applications of the TSRs to trisyllabic forms. Field's data also show virtually no correlation between tone pattern and phrase-internal direction of syntactic branching, possibly excepting the boundary between subject and verb+object. This fact and the amount of variability reported by Field suggest that the speakers' choices are influenced by higher level more than word-level syntactic and prosodic factors.

2.4.1 Tone sandhi and stress I: primary stress on the target blocks TS

Other anomalies in Field's data, from the point of view of earlier work, include that some of his output tone combinations show TS not to have applied where expected. Other outputs are not generatable by the TSRs. According to my interpretation, both of these anomalous results are explicable with reference to the fact that tone sandhi application reflects most of all stress placement. Thus note that in three strings a significant percentage of tokens had no application of sandhi rules, giving input = output: ADD (24%), DDA (20%), and DDD (16%). According to Shen (1989:56), non-application of sandhi occurs in Beijing Mandarin only when the syllable targeted for the sandhi change receives primary stress. Since Field's subjects were engaged in a reading task, a 16-24% likelihood that the speakers elected to give full stress to each syllable, causing the suspension of TSR application, seems entirely plausible.

Some of the examples in my recorded data clearly indicate retention of primary stress has blocked TS. Figure (1a) shows the pitch contour of an example of a four-syllable 'elaborate expression', *hun sang jia qu* 'wedding+funeral+marry out a daughter+marry in a daughter-in-law = family celebrations', the first two syllables of whose underlying tone pattern, AADC are eligible for AAR TS application. However, as described above, the TSR does not apply, in this case apparently because the speaker elects to stress each syllable fully, in a list fashion (while

semantically also intending the compounded meaning, as is apparent from context). One indication in this example that primary stress on each syllable is intended, apart from the four syllables' equal duration, is that the second syllable's A tone falls, characteristic of its isolation pronunciation, rather than leveling off or rising, as expected in word-internal context (cf. the discussion in 2.2).

Primary stress assigned to the target by the metrical structure can also serve to *block* TS in environments where it would otherwise be expected to apply. Assuming that Mandarin has binary branching right headed feet and that stress is assigned from right to left, the normal four syllable expression will have a W-S-W-S pattern. The often observed fact that TS rarely/never applies between the middle two syllables of a four syllable pattern thus falls out automatically. Cf. Figure 1(b) *guojia fenpei* BA AD 'national job distribution', where AAR has not applied to the middle two syllables.

2.4.2 Tone sandhi and stress II: zero stress on the target blocks TS

The complete absence of stress may also affect sandhi application. Thus consider again Table 6(b); the surface contour produced 74% of the time on an underlying CAA trisyllable is CBA, according to Field's method of analysis. This pattern cannot be generated at all by the TS rules in Table 3. (The rule-generated output, CAA -> BCA, in contrast, was produced on just 12% of the tokens.) What I would suggest is that the apparent CBA result reflects rather the reduction to tonelessness of the middle syllable. The middle syllable of Mandarin trisyllabic phrases in particular is often weakly stressed (see below) and may receive a pitch value by tone autosegmentalization, as perhaps in this pattern, or by interpolation between those of its neighbors, as in the example in Figure 1(d) below. TS application to CAA thus is bled by stress reduction to tonelessness.

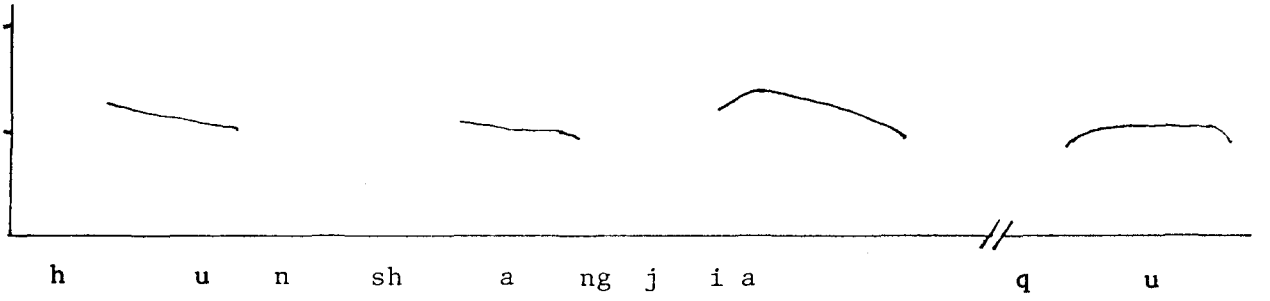
Lacking an exactly comparable example, consider instead the two pitch traces made from narrowband spectrograms in Figure 1(c) *xihuan chi* C0A 'like to eat' and 1(d) *ba ta neige mawar* C0 D0 D0 '[do] to it that watchamacallit'. The first three syllables of each example are C0A and C0D respectively, where 0 is a neutral tone, that is, a toneless syllable. The middle syllables, which I have interpreted as having been reduced by stress to the point of lacking their own contrastive pitch contour, theoretically have lexical tones D (high falling) and A (low falling) respectively. However, the second syllable of *xihuan* 'like' in 1(c) is regularly reduced in Mandarin, to the extent that it has a C0, not CD, designation in the dictionary. In the second syllable of 1(d) the pronoun *ta* 'it', though occurring under tone A when used as a subject pronoun, is often cliticized and hence toneless as an object pronoun.

Note that the second syllable of Figure 1(c) *xihuan* 'like' and middle syllable of 1(d) *ta* 'it' could be treated as extrametrical when pronounced in isolation, assuming underlying right strong binary branching stress feet. However, the standard expectation, sometimes claimed for Beijing Mandarin, that extrametrical toneless elements surface with their underlying tones when embedded in a non-peripheral context, is not met. Thus the contours on the first three syllables in Figure 1(c) and (d) are not describable as LH-HL-L and LH-L-HL respectively. Rather for the C0A sequence in 1(c), recall that C is LH and A is L. The spectrogram shows that the first syllable has low pitch, although underlyingly LH. The middle syllable is rising in pitch, although underlyingly toneless. And the third syllable is an A, that is to say, low tone. Thus an analysis in which Tianjin Mandarin has both right and left headed binary branching feet seems to be required.

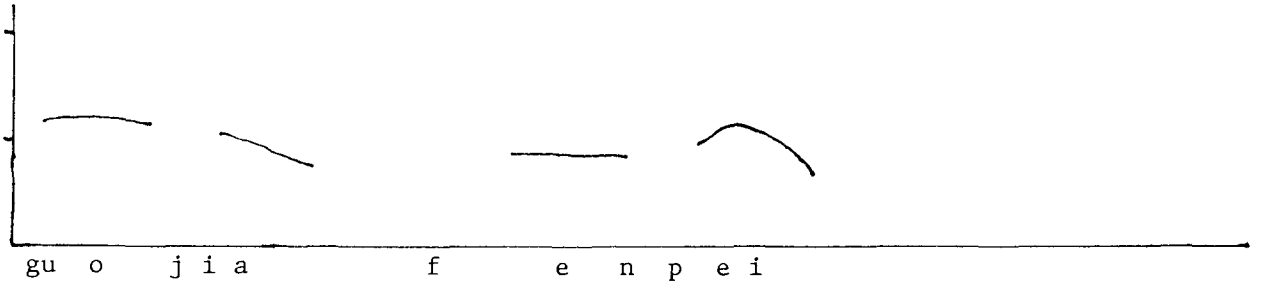
To account for the changes in pitch on the first two syllables, I assume the second syllable to have acquired a H tone by tone shift from the LH contour of the first syllable, giving L-H-L. This results in a surface tonal contour on the second syllable closely resembling a H tone B. Notice that the CAA -> CBA (LH-L-L -> LH-H-L) pattern, unaccounted for by Field, gives virtually the same tune as that for my C0A token. Assuming thus that Field's CBA is actually better

Figure 1. Pitch contours, traced from the fourth harmonic of narrow-band spectrograms of natural, connected Tianjinese speech

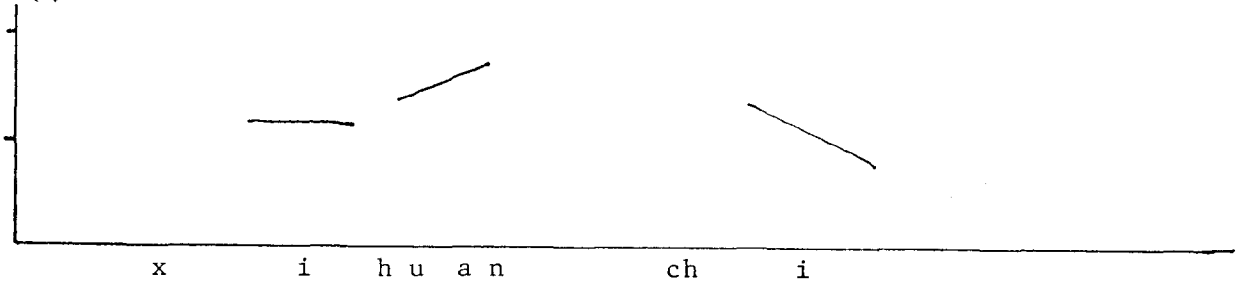
(a) *hun sang jia qu* AADC 'wedding+funeral+marry out a daughter+marry in a daughter-in-law= family celebrations'. L-L-HL-LH.



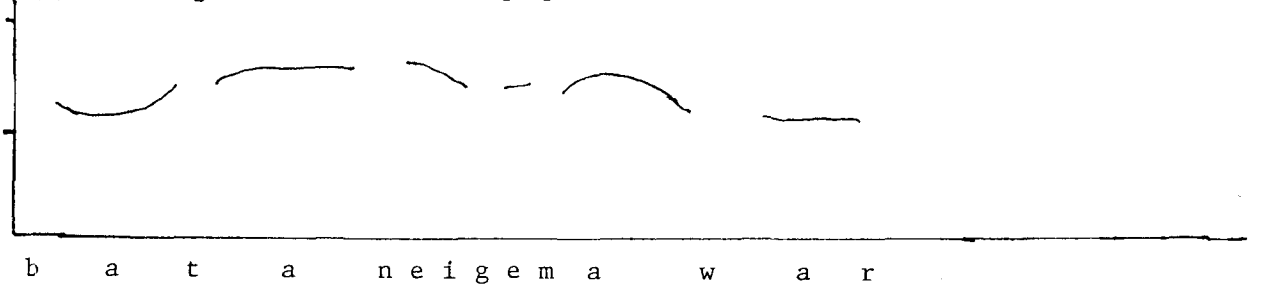
(b) *guo jia fen pei* BAAD 'national job distribution'. H-L-L-HL



(c) *xi huan chi* C0A `like to eat'. LH-0-L -> L-H-L

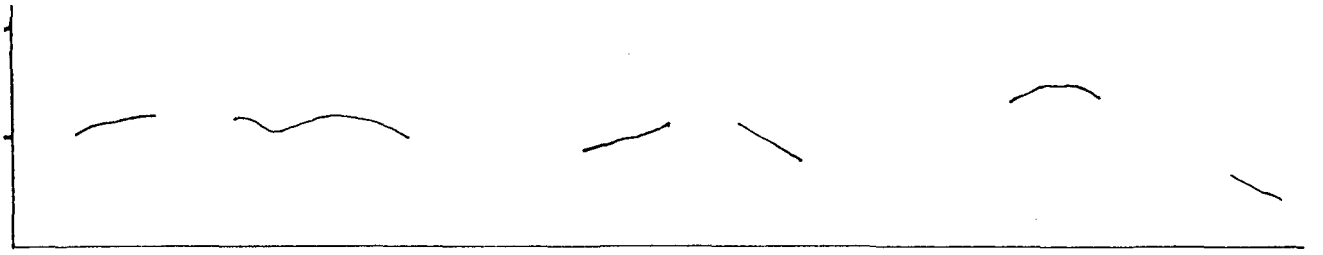


(d) *ba ta nei ge ma war* C0 D0 D0 `[do] to it that whatchamacallit'. LH-0-HL-0 -> LH-0-HL-0

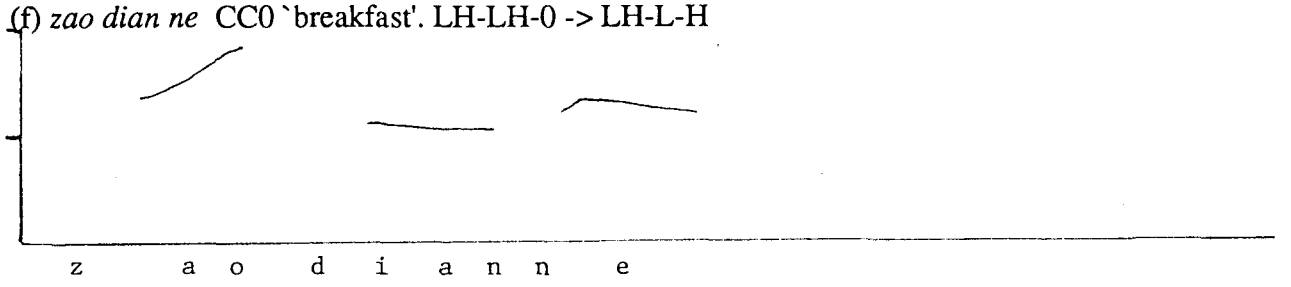


/continued

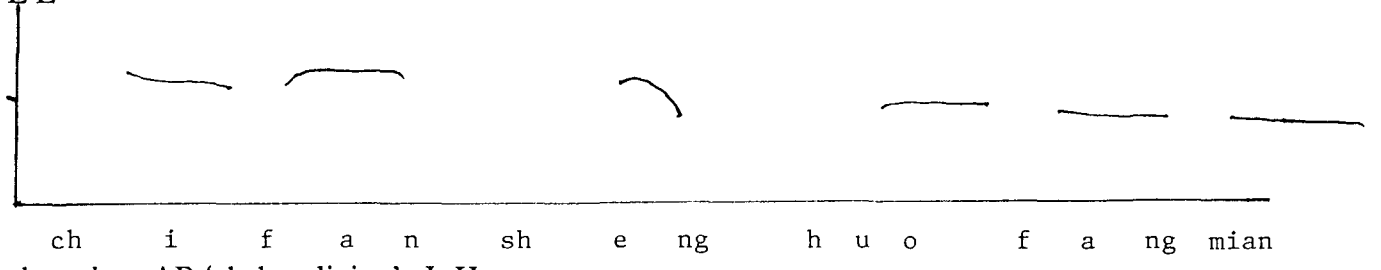
(e) *yuan lai xi huan ji ju* BB C0 BD 'originally liked to congregate'. H-H-LH-0-H-HL



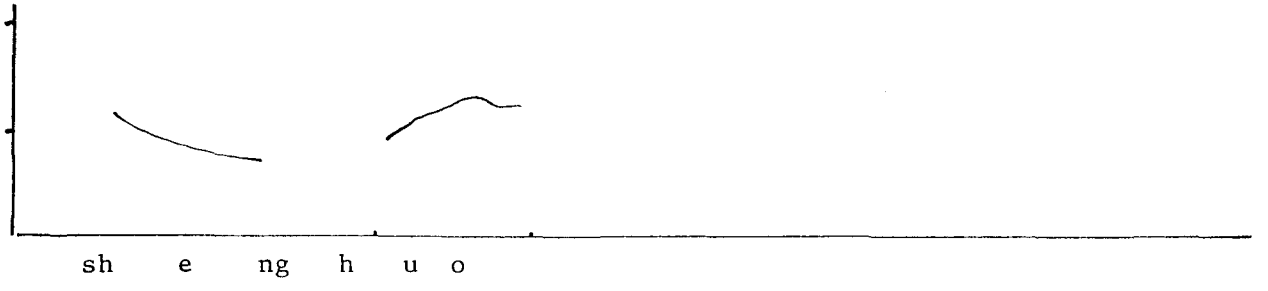
(f) *zao dian ne* CC0 `breakfast'. LH-LH-0 -> LH-L-H



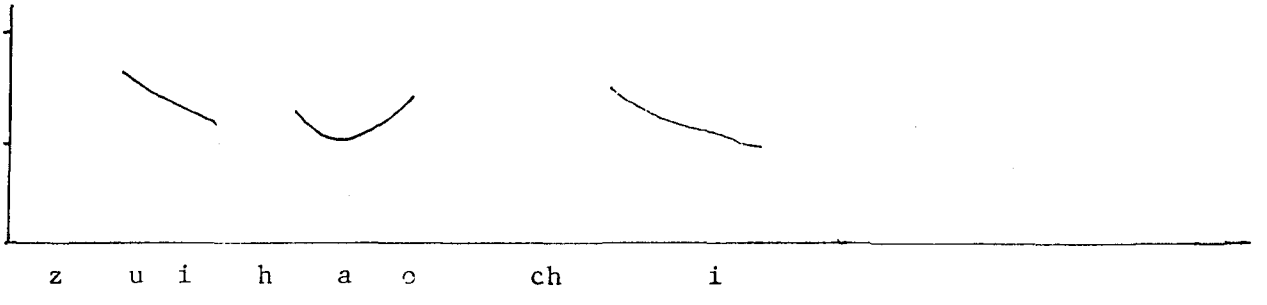
(g) *chi fan sheng huo fang mian* ADA000 'food & shelter concerns'. L-HL-L-L-L-L->L-H-L-L-L-L



sheng huo AB 'shelter; living'. L-H



(h) *zui hao chi* DCA 'most-good-eat = best-eating' HL-LH-L



represented as C0A, I have motivated it with reference to stress induced tone reduction and subsequent tonal shift.

In 1(d) interpolation between the first syllable's final H tone of tone C and the third syllable's initial H tone of tone D gives the high level contour across the middle syllable. Thus the LH-0-L underlying trisyllabic tone sequence for (1c) has redistributed its tonemes one to a syllable, surfacing as L-H-L, while the LH-0-HL underlying sequence of (1d) surfaces with assigned tonemic pitch values unchanged, but the phonetics interpolates between the H tones of the first and third syllables to produce the middle syllable's high level tone.

2.4.3 Conditions favoring H tone shift: attraction to primary stress position, boundary marking

It is of interest to note that the two other occurrences of the word *xihuan* C0 'to like', along with most other C0 tonal combinations in the corpus, surface with a (tonally unchanged) LH-0 tonal distribution, as is typical of Tianjin Mandarin. (Beijing Mandarin, on the other hand, regularly has L-H for C0.) The only noticeable difference in context for Figure 1(c) and the couple of other C0 combinations in the data undergoing LH-0 -> L-H, vs. the majority cases, including 1(d) (cf. also 1(e) *yuanlai xihuan jiju* BB C0 BD 'originally liked to congregate'), appears to be that, in 1(c), tone C's H tone is the only H in the string, and since the middle syllable is toneless, the H, attracted to the primary stress of the third syllable, shifts toward it as far as it can go, namely, to the second syllable's right edge. (This follows from Mandarin metrics' having right-headed binary feet (unless lexically marked to be left-headed), right-to-left application, and rightmost primary stress.) Thus, informally, 1(c) has a 2-0-1 stress pattern, and the first syllable's H tone has shifted toward the primary stress across a toneless syllable.

In the other two cases of LH-0 -> L-H, the final H of tone C docks onto a phrase-final stressless suffix, or as in Figure 1(f), *zaodian ne* CC0 'For breakfast,' onto a phrase-final particle, thus appearing to have migrated to serve rather a boundary marking function. Notice in 1(f) that while the H peak of the second and third syllables is clearly on the rightmost particle, not the second syllable C tone, it is not all the way to the right.

At this juncture it is relevant to mention a possible alternate analysis, Milliken (1990)'s floating **H**-L underlying representation of Tianjin tone A, given in Figure 4. On the normal interpretation of floating tones, the floating H tone of Milliken's tone A might be expected to surface in Figure (1c) on the middle syllable, or more specifically, according to Milliken's treatment, on the second TBU of the middle syllable, i.e., exactly where the H tone is in fact found. However, the disappearance of the H tone of tone C LH remains to be explained. Milliken's input of LH-0-**H**L would be expected to result in a 1(d) type contour, with a level high pitch interpolated across the middle syllable. In contrast, in my treatment there is no extra H tone to remove.

These data can be compared also to Figure 1(g), *chifan shenghuo fangmian* [[[AD] [AB]] AD] 'food [and] shelter concerns'. Notice first that a DAR context exists, though only across a syntactic boundary. The pitch traces show that the DA combination in 1(g) is realized pretty clearly as a sequence of high plus low tones, as expected if DA -> BA (HL-L -> H-L) has applied. It is also noteworthy that the third syllable, *sheng*, has the longest duration, and the pitch of the following four syllables steadily falls from it. This indicates that primary stress has fallen on *sheng*, with the string of syllables following it all toneless and stressless (extrametrical? a problem for the future). The pitch trace of *shenghuo* AB 'shelter, living' in isolation pronounced by the same speaker in a different context shows that it normally has low plus rising intonation, i.e. its underlying AB tones both surface, thus further showing that destressing of all the syllables after *sheng* happens at the phrasal, not lexical level. Thus it is also clear that the DAR has applied post-cyclically, in this case. The underlying syntactic bracketing which blocked TSR from applying in

Figure 1(b) has been superseded by the assignment of phrase-level main stress on the third syllable of a four syllable phrase, in this example. It is thus possible to interpret the function of the DAR here as being to replace tone D HL's final L tone with a H tone, near the onset of the final low A tone, signalling that the following A tone syllable has primary stress. Only the metrical account can explain the variation in application of the various TSRs conditional on stress placement, as above, rather than merely specifying the stress conditions. For this and other reasons given below the metrical treatment is to be preferred to Milliken's purely autosegmental TS analysis.

The question arises whether the examples in Figure 1(c) and (d) are typical, i.e. whether the middle of three syllables always reduces. Shen (1990:50) observes for Beijing Mandarin that almost all speakers read the Beijing Mandarin three-syllable phrase *lao gudong* CCC 'old conservative man', embedded in a nine-syllable sentence of C tone words, as B-neutral tone-C rather than BBC. According to her rules (p. 41) neutral tone after both tones B and C is higher than the B or C H tone itself. However, as argued above, the S-W-S pattern is subject to phrase level intonational effects. As noted previously, syntactic bracketing is not at issue, as exemplified by Figure 1(h) 'most good [to] eat', which is right-branching like *lao*C *gu*C*dong*C yet for semantic reasons has none of the three syllables completely reduced. Thus in 1(h) the middle C tone syllable is both long and shows a full LH contour, the H co-articulating into the third syllable.

These data show that stress plays an important role in the realization of basic and sandhi tones. However, the status in the grammar of the application of the optional and apparently at least partially lexically conditioned sandhi rules to disyllabic words with the requisite iambic metrics remains to be explicated. Shen (1990:61) says of Beijing Mandarin that "tone sandhi is completely a phonetic phenomenon: in a [C] tone ditoneme *hao jiu* CC 'good sake', when the preceding C tone is heavily stressed, no sandhi takes place...it may be concluded that stress does not have its own carrier; its existence is signaled through change in tones...Tone and stress are inseparably interlaced, and the manifestation of stress is also at the same time the effect of stress on tones." In the present paper it is argued that the phonetic basis of TS in stress-induced modification of pitch range is relatively transparent in certain contexts, consistent with Shen's view. Be that as it may, within the domain of the disyllabic word, the TSRs in Tianjinese have essentially become phonologized, and operate on L and H in autosegmental fashion. In the following section, stress is shown to affect phonetic aspects of tone realization such as register and location of tonal targets within the syllable boundary in Tianjinese. In the last section the relation of stress to the TS rules is described.

3. Phonetic study of Tianjin tones

Figure 2 displays mean onset, midpoint and offset values for Tianjin tones from one male speaker, "Li", based on measurements made from spectrograms. Each trace represents the average of 4-10 tokens. The "Units" on the vertical axis are Hz. Tone shapes in isolation are indicated by the black continuous lines. The different phonetic contours for the same tone resulting from differences in tonal context are indicated by providing separate traces for each following tone (lefthand side of the figure) and each preceding tone (righthand side of the figure).

The only previous acoustic study of Tianjinese of which I am aware is Shi (1986). Shi analyzes spectrograms of two male speakers pronouncing 10 words each of the 16 possible disyllabic tone combinations, a total of 160 disyllabic tokens. He displays average pitch curves for each speaker, for each tone in each context.

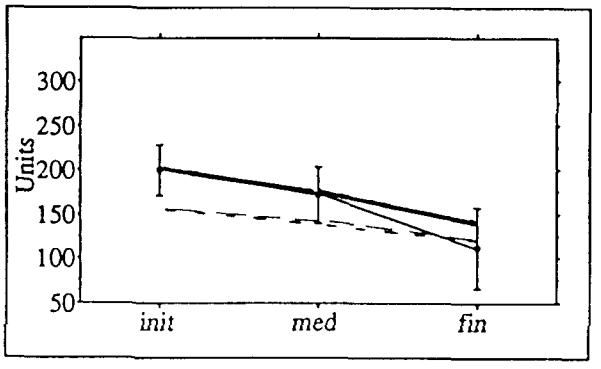
3.1 Stress induced register shift, tone C (LH)

One of Shi's findings is that Tianjin tone C has a particularly high register, often occurring as a high rising tone, before "neutral" stressless, toneless syllables (tone 0) and before tone A. In Shi's terms, then, CA tones have values 35-21. (Notice that AA tone sequences undergo the AAR

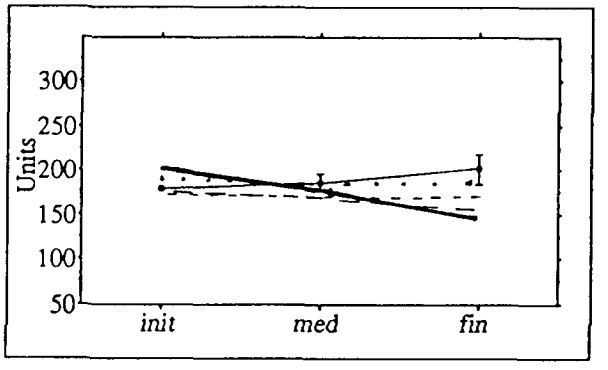
Figure 2. Mean pitch contours of each tone in isolation and in different tonal environments. Data represent means measurements at three time points made from spectrograms of Tianjin natural speech tokens spoken by speaker Li. Isolation contours and contextual variants are indicated according to the coding below. Effects of preceding tonal environment are shown in the left hand panels; effects of the following tonal environment in the right hand panels.

▬▬ = #__#; —— = A; - - - - - = B; ····· = C; — — — = D; · · · · · = 0

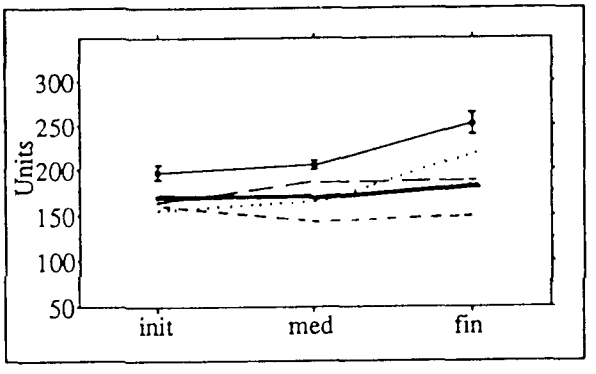
Tone A / X__



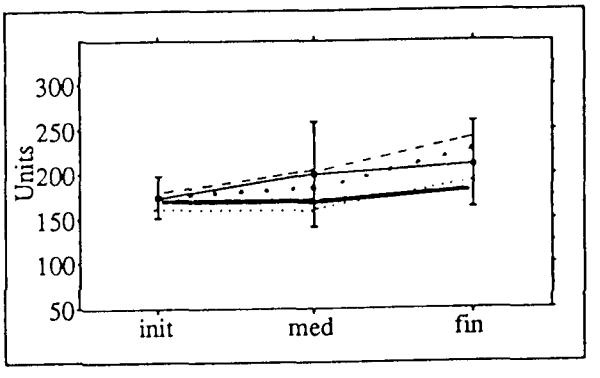
Tone A / __X



Tone B / X__

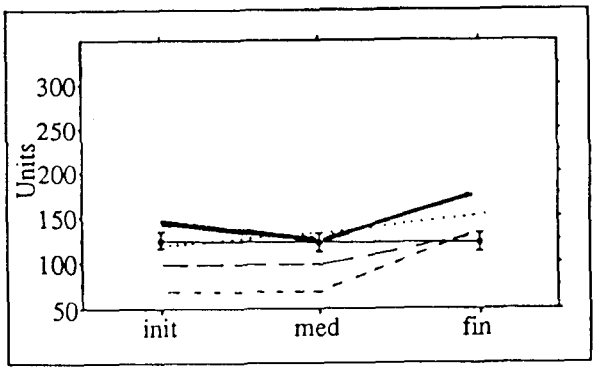


Tone B / __X

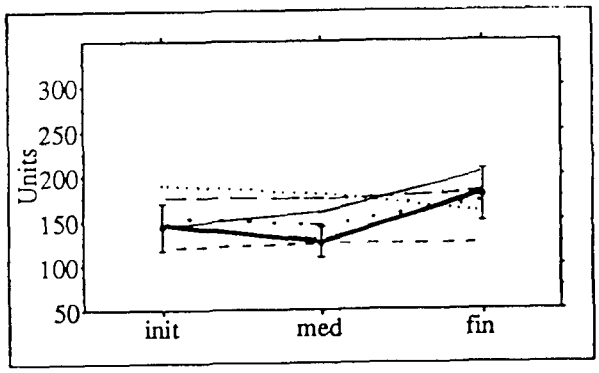


/continued

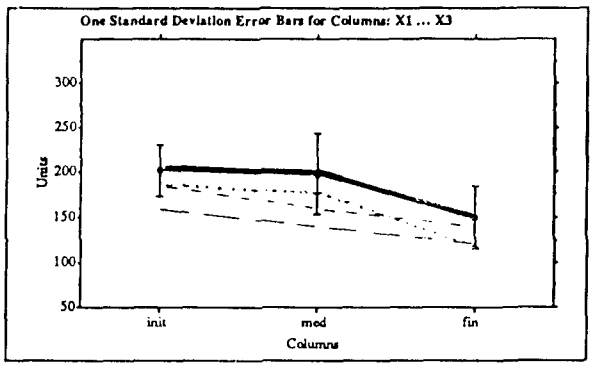
Tone C / X__



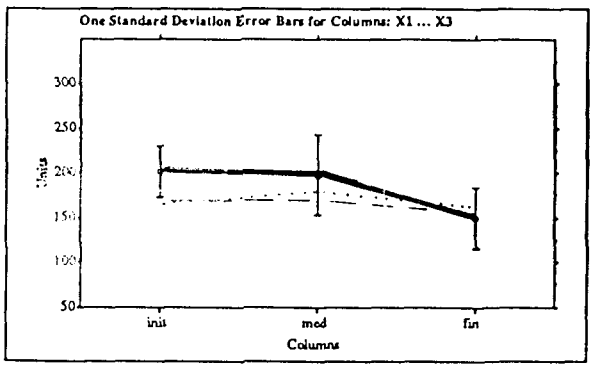
Tone C / __X



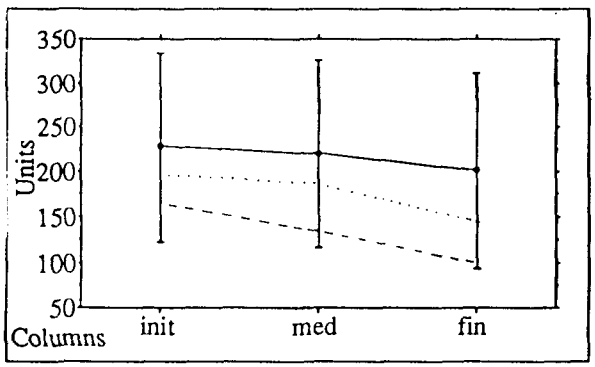
Tone D / X__



Tone D / __X



Tone 0/X__



to merge with the allotonic 35-21 CA set, though somehow remaining distinguished from BA 45-21. The appropriate perception tests remain to be done to determine the salience of this distinction.) This raised variant is described by Shi as tone C's allotone.

My data are generally consistent with Shi's, though as is evident in Figure 2 in the righthand panel under tone C, my speaker's tone C before tone A (represented by the narrow continuous line) has a higher terminus than tone C before neutral tone (represented by the widely spaced dots) as well as the other alternants. For tone C to rise more before a prosodically strong A tone, rather than also before the neutral tone, is consistent with my hypothesis.

In addition, tone D lowers allotonically from 53 to 42 before the two rising tones B and C. These two conditioned register shifts of tone A up and tone D down are non-distinctive, applying equally to underlying and sandhi-derived tones. Thus they belong in the phonetic component.

Shi also observes that in second position in the disyllable, tones A and C occur on higher than average pitch while tones B and D have lower than average pitch, across the length of the contour. I.e., the register value in second position is inversely related to its value in citation form. Shi considers tones A and C to be low register tones, tones B and D to be high register tones. Moreover, tone A begins higher both in second position and before neutral tone. If true such `register inversion' would be an interesting variation on the claim that stress was marked by an expanded pitch range. However, the co-articulation data on Tianjin natural speech from my informant Li, given in Figure 2, does not entirely corroborate this claim. More research is needed on this point.

Notice that, regarding neutral tone, Shi claims that it is mid-falling after rising tones B and C and low-falling after falling tones A and D. My own data agree with Shi except for tone A, after which I find both higher and lower pitches on the neutral tone, apparently depending at least in part on the intonation contour. Cf. Shen (1990:40) for Beijing Mandarin: "...the instrumental evidence of my data shows that it may be higher or lower than the preceding Tone A, depending on the syllable position that the neutral tone occupies in the sentence." Apart for conditioning of neutral tone realization after tone A, which requires further study, my data are thus generally consistent with Shen (1990) for Beijing Mandarin neutral tone realization and support her claim for this being a (possibly universal) phonetic co-articulatory effect. Therefore, pitch assignment to "neutral tone", that is, underlyingly toneless, syllables is relegated to the phonetic component.

Lastly, it is observable in Figure 3(a) *qiling shiqi* AB B0 'the mixed up period' that for a sequence of two H B tones, the highest pitch is achieved on the second, primary stressed one. The rising contour remains the same.

3.2 Stress-induced tone peak delay, tone D (HL)

Another example of phonetically conditioned tonal variation involves the location of the H toneme peak of tone D. Shih (1988) observes that her isolation pronunciation tokens of Beijing Mandarin tone D generally locate the H of the HL contour slightly after onset of voicing, what she describes as a "slightly delayed peak". She notes that the delayed peak was sometimes masked by voiceless initial consonants and enhanced by voiced initial consonants and/or glides. The occurrence of delayed peaks for tone D in data from three speakers was qualitatively assessed and the results are given in table 7 below. The variation in my data on the peak location of the HL contour of tone D in Tianjinese does not significantly correlate with initial consonant voicing (in any case the great majority of the Tianjinese tokens begin with phonologically voiceless initial consonants). However, the pooled data from three speakers were suggestive of other correlations.

Table 7. Occurrence of tone D with delayed H peaks in Tianjin natural speech data: in all contexts; in first syllable and second syllables of a disyllabic word; after tone C; after tone A; and in monosyllabic isolation context.

Speaker	All contexts total	delayed	First syllable total	delayed	Second syllable total	delayed	CD	AD	#D#
Li	32	10	14	2	12	4	2/2	3/2	6/4
Liu	29	10	14	2	14	8	3/3	7/4	1/0
Yang	28	5	13	1	8	4	4/3	0/0	7/0
All speakers	89	25	41	5	34	16	9/8	10/6	13/4

Before discussing these data, it may be useful to characterize what is meant by a delayed paek more precisely. Compare the pitch traces in Figure 3(b) and 3(c). In 3 (b) the second syllable, *fan* of *chifanle* AD0 'have eaten', has its peak delayed until after the middle of the vocalism. In contrast, when tone D is in the first position of a disyllable having weak-strong stress it looks spectrographically like the pitch trace for the first syllable *cong* of *congxiao* BC 'since childhood', in Figure 3(c). I.e., tone D in first unstressed position falls straight away from a H (both for pretonic unstressed D and, optionally, [sic] B); the H tonal target is not delayed when in weakly stressed first position of a disyllable.

This pattern is born out statistically. See Table 7, which gives the numbers of delayed and non-delayed tokens in different environments. Delayed peaks occur in almost half of the second syllables of disyllabic words (16 of 34 tokens), but in only 5 of 41 tokens of first syllables of disyllables, and only 4 of 13 monosyllables. Of the 5 on the first syllable of a disyllable, 3 had voiced initials, and 4 were on trochees. Of the 4 on monosyllables, three were pronounced in complete isolation, and the fourth was emphatically stressed. Of the second syllable of disyllables with delayed peaks, 8 of 9 were in CD patterns, 6 of 10 in AD patterns, 1 of 3 in DD patterns, and 1 of 8 in BD patterns.

Thus my data indicate both that prominence on a tone D syllable may be signalled by delayed peaks on tone D, and that delayed peaks are especially to be observed after tone C, and to a lesser extent tone A. The latter fact could conceivably be accounted for by the need to distinguish CD from the sequences CA or C0, which as noted above both have a high target at the end of the first syllable and in the case of C0 sometimes delay the H target into the beginning of the second syllable. Thus notice the extremely late H target in the second syllable in Figure 3(b). If maximizing distinctiveness is to be invoked the relevant adjustment rule may also be expected to be part of the phonetic component.

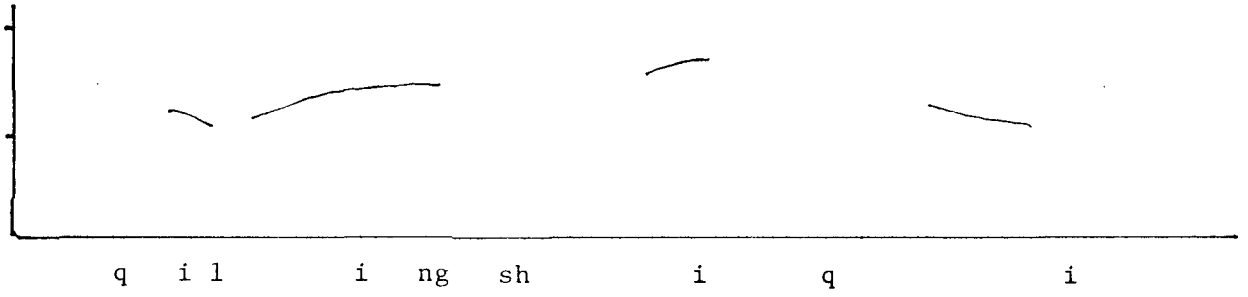
Notice finally the example in Figure 3(c), in which tone B, high or high-rising in citation form, falls before a stressed tone C. We might speculate that whereas the rightmost stressed syllable in 3(a) has a delayed peak, the leftmost unstressed syllable in Figure (3c) has an early peak. The frequency of tokens of the sort represented in Figure 3, having systematically early or delayed peaks in certain contexts; the possibility that rising and falling tones neutralize phonetically under the appropriate prominence conditions; and the acoustic cues distinguishing them all merit further study.

4. Prominence effects and tone sandhi

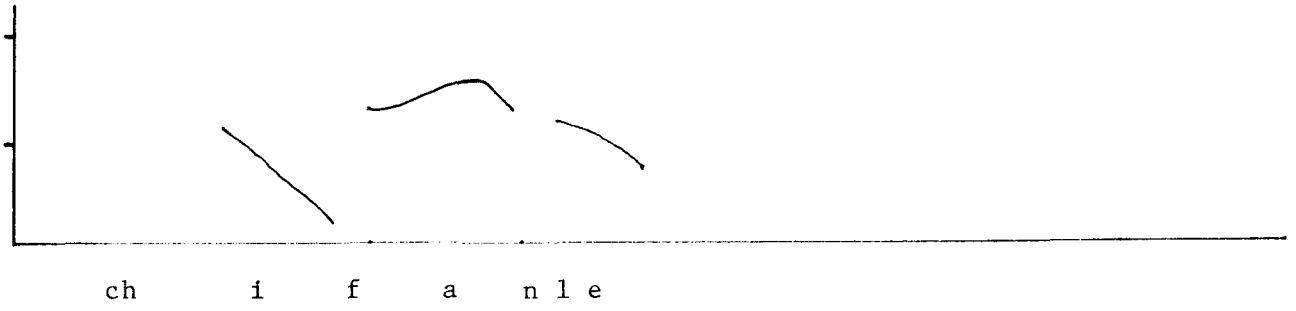
Above we have seen contextually conditioned allotones of Tianjin tones C and D, the rising and falling contour tones, which differ from the basic tonal forms only in occurring in a higher or lower register, the contour remaining the same. We have also seen the realization of the high tonal target for tones C and D (and possibly also B) delayed into the following syllable or late in the

Figure 3. Pitch contours illustrating the phonetic effect of prominence on H tone peaks in Tianjinese.

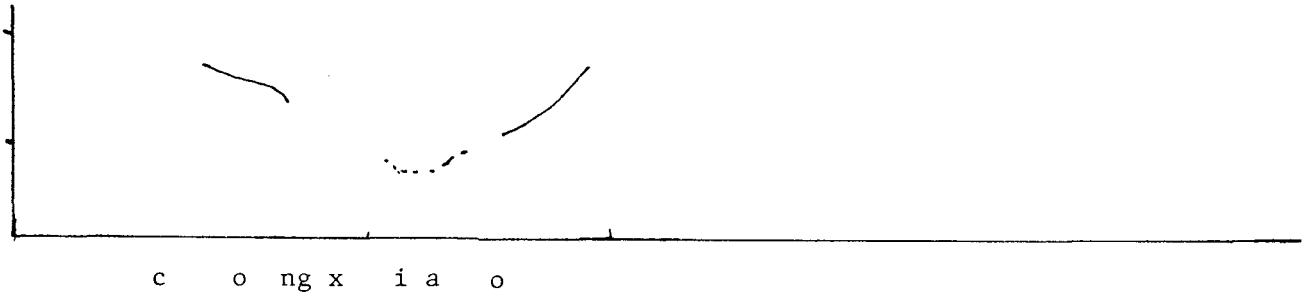
(a) *qi ling shi qi* ABB0 'the mixed up time period'. L-H-H-0



(b) *chi fan le* AD0 `ate'. L-HL-0



(c) *cong xiao* BC `since childhood'. H-LH



syllable, depending on stress. Lastly we have seen evidence that f0 values for underspecified syllables may be derived by interpolation from one tonal target to the next. Thus the phonetic evidence is consistent with Yip (1990)'s proposal, that Chinese type contour tones behave both as indivisible, dynamic units and as decomposable tonemes. She characterizes this type of structure as involving the branching of a tonal node rather than the clustering of tonemes to the vowel or syllable node.

Yip (1990)'s phonological analysis of Tianjin tones (see Table 2 above) assumes the same sort of dualistic structure. Shi (1988) is somewhat similar, following Wang in contrasting height, slope, and contouricity features. In contrast, Milliken (1990) gives an autosegmental account of Tianjin tone sandhi in which branching tonal nodes do not have special status, as in Figures 4 and 5. In Milliken's proposal tone A is represented as a cluster of a floating high tone plus a low tone attached to a single root node; tone C is a low plus high tone attached to separate root nodes; and tone D is a high plus low tone attached to separate root nodes. High and low tones are on separate tiers (Cf. Davison 1987, 1988, 1989, and Table 4).

The various analyses nevertheless have superficial similarities. Davison (1987) treats the H of tone C as floating, while Milliken (1990) has a floating H for tone A which docks postcyclically. Davison's contour simplification rule is functionally equivalent to Milliken's Maximum Association Condition (MAC) (see Figure (5b)). Both Yip and Milliken invoke the Obligatory Contour Principle (OCP) to motivate the TS rules. However, Yip claims that violation of the OCP causes dissimilation of sequences of identical branching tone contours, not only sequences of identical level tones, whereas Milliken places H and L tones on separate tiers to create tier-adjacent violations of the OCP for contour tone sequences.

Milliken provides an autosegmental phonological treatment of the sandhi forms on trisyllabic as well as disyllabic words. Specifically, he generates Field's most common pattern for underlying DDA sequences, DDA -> DBA, and the two most numerous token types for DDD, DDD -> AAD, DAD. In the latter case Milliken claims AAD occurs with left-branching structures, DAD with right-branching syntactic structures. As noted above, Field's data do not support the correlation of tone sandhi application with syntactic branching. Milliken does not discuss the ADD and CAA patterns.

Milliken's distinction between cyclic and post-cyclic rules is intended to coincide with the facts of differentially branching structures receiving distinct tone patterns in the cyclic phonology. Of course, for cyclic rules to generate the same outputs by different derivational routes, cf. Milliken's right vs. left branching DDA->DBA and AAA->ACA, begs the question.

Another problem with Milliken's treatment is the floating tone for A which links to the rightmost tonal node of the preceding syllable, post-cyclically. A similarly workable solution arises from associating an underlying floating tone with C, one which has the added advantage of more exactly paralleling the floating tone analysis of Beijing Mandarin, a closely related dialect, in Yip (1980), to account for the high tone occurring also on post-tonic toneless syllables as described above for both Tianjin and Beijing dialects. Aside from CC tone sandhi and high tone spreading to toneless suffixes as motivations for a C tone associated floating H tone, C also has a low level allotone in both Tianjin and Beijing dialects, though its distribution is restricted in Tianjin to occurrence before tones B and D. Given Milliken's account of tone A we might expects its floating H tone to associate to pretonic toneless syllables, on analogy with tone C, yet this does not in fact happen.

The question remains as to which solution, when sufficiently patched up, is most likely to correctly represent the sorts of phonological processes these TS rules are. In light of the preceding I would argued on the one hand in favor of what in Table 4 is referred to as contour simplification and by Milliken in Figure (5b) as the MAC. This interpretation has precedents in the literature (see

Figure 4. Milliken (1990)'s underlying representational tone structure

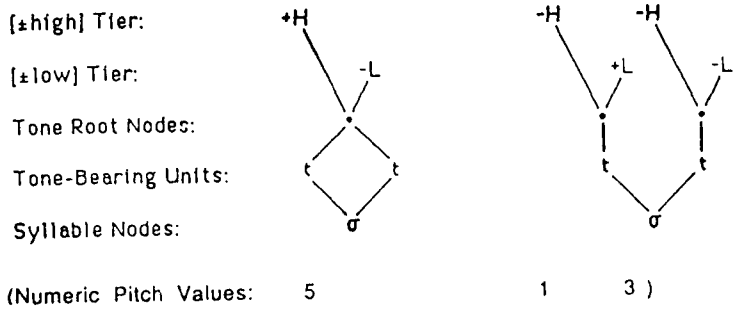
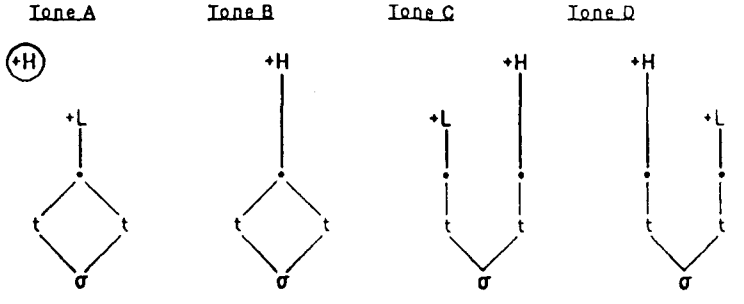
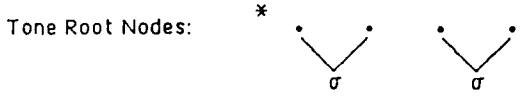


Figure 5. Milliken (1990)'s underlying representation of the four basic tones of Tianjinese

(a) Representation of the four tones



(b) Maximum Association Condition (MAC)



(c) Conditions

--Floating tones are only permitted in the cyclic phonology. (Language specific)

--Adjacent identical autosegments are prohibited. (OCP)

--Representations are always scanned right to left. (Language specific)

Chan 1985) and is supported by the phonetics of Tianjinese, as noted earlier regarding the C tone plus toneless syllables. Here then we are arguing on the basis that natural phonetic processes attested elsewhere in the language are good candidates for evolving into phonological ones. The production of tunes in Tianjinese in relation to the MAC will be discussed in a later paper.

Secondly, I would like to suggest on the basis of the preceding that the source of the H tone in both the AAR and DAR, which Milliken represents as A tone's floating H, is rather the marking of the second syllable of the disyllable as prominent, i.e. stressed. One reason to prefer this analysis to Milliken's is that tone A's floating H must be arbitrarily restricted to apply only in the cyclic phonology, to account for the fact that it does not associate to a neighboring (preceding) toneless syllable as we saw C tone's H tone do above in Figure 1(c). However, as formulated, Milliken requires floating **H** to dock only post-cyclically. The resulting contradiction suggests that the floating **H** is not the correct representation.

Arguments against Milliken's floating H analysis of tone A appeared above. The main reason to consider the TS rules as phonologizations of a H tone accent on the second syllable of a disyllabic word is the following. As noted above, the AAR only applies to weak-strong disyllabic patterns. I.e., if the first syllable of a disyllabic AA tone word is stressed, as in the place name *tianjin* A0 'Tianjin' the resulting form does not undergo TS, so the standard pronunciation is a steadily falling tone across two disyllables, rather than the sandhi-produced low rising-falling pattern. Word initial stress can be lexicalized, as in the case of the place name, or may be due to phrase level emphatic stress. (I note parenthetically that Tianjinese is unlike Beijing dialect in disallowing TS application to precede lexical stress shift, as far as I have been able to determine.)

I am claiming that the TS rules conspire to create a tune having a single high pitch peak at or near the rightmost edge of the prosodic domain. A systematic study of all the relevant environments will be attempted in future. Here, for selected TSRs we have seen the H peak target to be either the second syllable itself or near the left edge of the second syllable. The DAR and CCR effect this sort of change inasmuch as the tone which is phonologically H, tone B, is phonetically high rising, that is, its highest pitch peak is at the terminus. Since the A tone is the only tone in Tianjinese lacking an underlying high tone altogether, H accent tone insertion necessarily changes its percept, creating the conditions for phonological neutralization.

To conclude, an advantage of this analysis is that it motivates the general observation made by Cheng (1966) and others that in Mandarin Chinese dialects only low register tones appear to dissimilate. Each dialect must of course be carefully analyzed on a case by case basis. However, at least for Beijing and Tianjin Mandarin dialects, we may now have an explanation for the distribution of tone sandhi, in particular that it applies to low register tones, but not sequences of phonological H tones.

Thus we may assume that a lexical H tone can be made prominent by expanding the pitch range and making it super-high. In contrast, accent on a L tone is realized most saliently on an adjacent H tone, even an inserted one, if necessary.

Acknowledgments

Fieldwork in Tianjin was supported by a grant from the Committee on Scholarly Communication with the People's Republic of China, 1980-1982. This research was supported by NIH grant no. 1T32 DC 00029-01.

References

Chan, M. K.-M. 1985. Fuzhou Phonology: a Non-Linear Analysis of Tone and Stress. Ph.D. dissertation. Seattle: University of Washington.

Chao, Y. R. 1968. A Grammar of Spoken Chinese. Berkeley: University of California Press.

Chen, M. Y. 1986. The paradox of Tianjin tone sandhi. *Proceedings of the Chicago Linguistic Society Annual Meeting 22:98-114.*

Chen, M. Y. et al. 1987. A symposium on Tianjin tone sandhi. *Journal of Chinese Linguistics* 15:203-26.

Cheng C. C. 1966. Guanhua fangyande shengdiao zhengxing gen liandiao bianhua [Tone features and tone sandhi in the Mandarin dialects]. *Dalu Zazhi* 33:102-8.

Davison, D. S. 1987. The tonology of Tianjin Mandarin. International Conference on Sino-Tibetan Languages and Linguistics, Vancouver.

Davison, D. S.1988. A laryngeal source for the Mandarin tone sandhi rule. New Orleans: Annual Meeting of the Linguistic Society of America.

Davison, D. S.1989. Lexical prosodies of Mandarin: Comparative evidence from Northern Chinese dialects. University of California at Berkeley, Ph.D. dissertation.

Field, K. L. 1990. Tianjin tone sandhi revisited: an instrumental analysis approach. International Conference on Sino-Tibetan Languages and Linguistics, Austin.

Li, X-J & Liu S-S. 1985. Tianjin fangyande liandu biandiao [Tone sandhi in Tianjin dialect]. Beijing: *Zhongguo yuwen* 1:76-80.

Milliken, Stuart. 1990. Resolving the paradox of Tianjin Chinese tone sandhi. International Conference on Sino-Tibetan Languages and Linguistics, Austin.

Shen, X.-N. S. 1990. The Prosody of Mandarin Chinese. University of California Publications in Linguistics, Vol. 118. Berkeley: University of California Press.

Shen, X.-N. S. 1990. Tonal coarticulation in Mandarin. *Journal of Phonetics* 18:281-295.

Shi, F. 1986. Tianjin fangyan shuangzizu shengdiao fenxi [Analysis of Tianjin dialect tones in bisyllables]. *Yuyan Yanjiu* 1:10:77-90.

Shih, C-L. 1988. Tone and intonation in Mandarin. *Working Papers of the Cornell Phonetics Laboratory* No. 3, pp. 83-108.

Yip, M. J. W. 1980. The tonal phonology of Chinese. Ph.D. dissertation. Cambridge MA: MIT

Yip, M. J. W. 1989. Contour tones. *Phonology* 6:149-174.

Vowel-to-vowel coarticulation in three Slavic languages

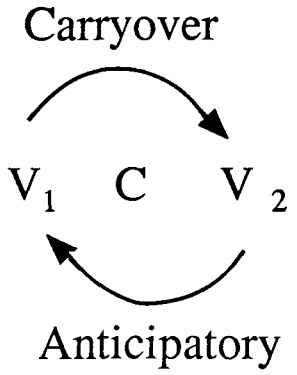
John D. Choi & Patricia Keating

Paper presented at the Acoustical Society of America Meeting
San Diego, CA November 26-30, 1990

Introduction

Coarticulatory effects of one vowel on another across an intervening consonant, called vowel-to-vowel coarticulation, have been the focus of much phonetic research. Presence of vowel-to-vowel coarticulation has been argued to provide evidence that speech does not consist of a purely local concatenation of discrete segments.

Figure 1. Carryover and anticipatory vowel-to-vowel coarticulation.



Much of the work on vowel-to-vowel coarticulation has been concerned with differences across languages. Among the languages which have been studied, Russian has been reported to exhibit little or no transconsonantal coarticulatory effects. Öhman (1966) provided preliminary indication that Russian showed much less anticipatory vowel-to-vowel coarticulation, as measured on F2, than did Swedish and English. More extensive work by Purcell (1979) found no effects on F2 in either direction. This contrasts with the numerous other languages studied so far, e.g. Swedish (Öhman 1966), English and Japanese (Magen 1984; 1986), Shona, Sotho and Ndebele (Manuel 1990), Swahili (Manuel & Krakow 1984), Catalan (Recasens 1984), Spanish (Recasens 1987), and Italian (Farnetani 1990), which all exhibit some vowel-to-vowel coarticulatory effects.

One explanation for the lack of coarticulation in Russian, originally proposed by Öhman, is the presence of contrastive secondary articulation in this language. Russian contrasts palatalized versus nonpalatalized consonants. The tongue body assumes a particular position during the consonant, in addition to the consonant's primary articulation. Thus the Russian consonants have a more complex articulation that, like vowels, involves the tongue body, and this articulation is hypothesized to pre-empt any interactions between the vowels. Nonetheless, the previous studies of Russian have not used statistical analyses sensitive to the small differences often involved with vowel-to-vowel coarticulation. Furthermore, Russian is only one language with palatalization, so other examples would be needed to show that the difference is systematic.

The goal of this study is to further investigate the occurrence of vowel-to-vowel coarticulation across consonants with and without secondary articulations. The hypothesis being tested is that secondary articulations on consonants should block or strongly inhibit vowel-to-vowel coarticulation.

Method

Three Slavic languages, Russian, Bulgarian and Polish, were examined. Palatalization is contrastive in Russian and Bulgarian - in Polish, palatalized consonants are derived before high front /i/. In addition to these three Slavic languages, data from California English were also collected to serve as a point of reference for comparison.

Three native speakers, all adult males, were recorded for each language. Each subject read from a list of VCV utterances in which the first syllable was always stressed. The vowel phonemes in the VCV tokens were restricted to /i/ and /a/ in all the languages, producing four possible combinations. A third high front vowel /ɪ/ which does not trigger consonantal palatalization was also elicited in Polish to allow for all the relevant phonetic comparisons. The consonant in each of the languages varied across 3 places of articulation and voicing, with the exception of English in which only voiced consonants were examined. Lastly, the contrast between plain and palatalized consonants was examined in the Slavic languages.

The effects of each vowel on the other was assessed at the closest transition edge. Measurements of the second formant frequency were made at the offset of the first vowel and the onset of the second vowel from a wide band FFT power spectrum examined in tandem with a spectrogram and waveform display, as shown in Figure 2. The FFT spectrum were based on a 5msec window, carefully positioned so as to enclose a single glottal pulse. While this resulted in reduced frequency resolution, the method provided a more accurate measure of F2 at the VC and CV boundary than would a larger analysis window in which averaging would smear the measured F2 value across time. This especially holds true in fast moving transitions. The F2 measurements were then analyzed with a series of repeated measures ANOVAs.

Results

Both within-language and cross-language tests were conducted. Beginning with English, our results show significant main effects for vowel-to-vowel coarticulation in both directions. Figure 3 shows the difference in both the anticipatory and carryover effects of /i/ vs /a/ on V1 and V2, respectively. The values plotted here represent the mean F2 values and standard deviations for V1 and V2 collapsed across all consonants and all the relevant vowels. The first pair of values shows the difference of F2 between vowels preceding /i/ and vowels preceding /a/, showing that the vowels have a higher F2 before /i/. The second pair of values shows the same effects on the second vowel - again, the vowels have a higher F2 after /i/. Were there no effects, we would expect to see no difference within each pair. These results concur with previous studies, although the effects are somewhat larger in magnitude than those reported, for example, by Magen (1986).

Our results for Russian differ from previous studies. The data show small, but significant carryover coarticulation. Figure 4 illustrates the overall effect of V1 on V2, where as before, the values represent the standard deviations and mean F2 values of the second vowel collapsed across all consonants and V2s. Again, the vowels have higher F2s after /i/. Although it cannot be seen on this graph, such coarticulation is found across both plain and palatalized consonants. It should also be noted that this difference is more robust on the low vowel /a/ than on high vowel /i/. Anticipatory vowel-to-vowel coarticulation is also found, but is confined to certain occurrences of the low vowel and is smaller in magnitude than the carryover effects.

Bulgarian also has contrastive palatalization and like Russian, exhibits carryover effects. The magnitude is even smaller than in Russian, so that it is seen largely on /a/. This is illustrated in figure 5 which shows that /a/ has a higher F2 after /i/. Anticipatory effects in Bulgarian could only be tested across palatalized consonants due to constraints on possible C-V combinations. Such an effect can be seen only in one particular VCV combination, again on the low vowel.

Polish differs from Russian and Bulgarian in that palatalization on consonants is derived. Vowel-to-vowel coarticulation is found, but only in particular VCV combinations. The carryover effects are the most general and are illustrated in figure 6. The first pair of values shows that /i/ has a higher F2 after the high vowel than the low vowel, even though this /i/ only occurs after palatalized consonants. The second pair shows the effects on the high vowel /ɪ/ which occurs after nonpalatalized consonants. As we can see, this vowel also has a higher F2 after /i/. The last pair

Figure 2. Sample acoustic display of Russian /ádi/.

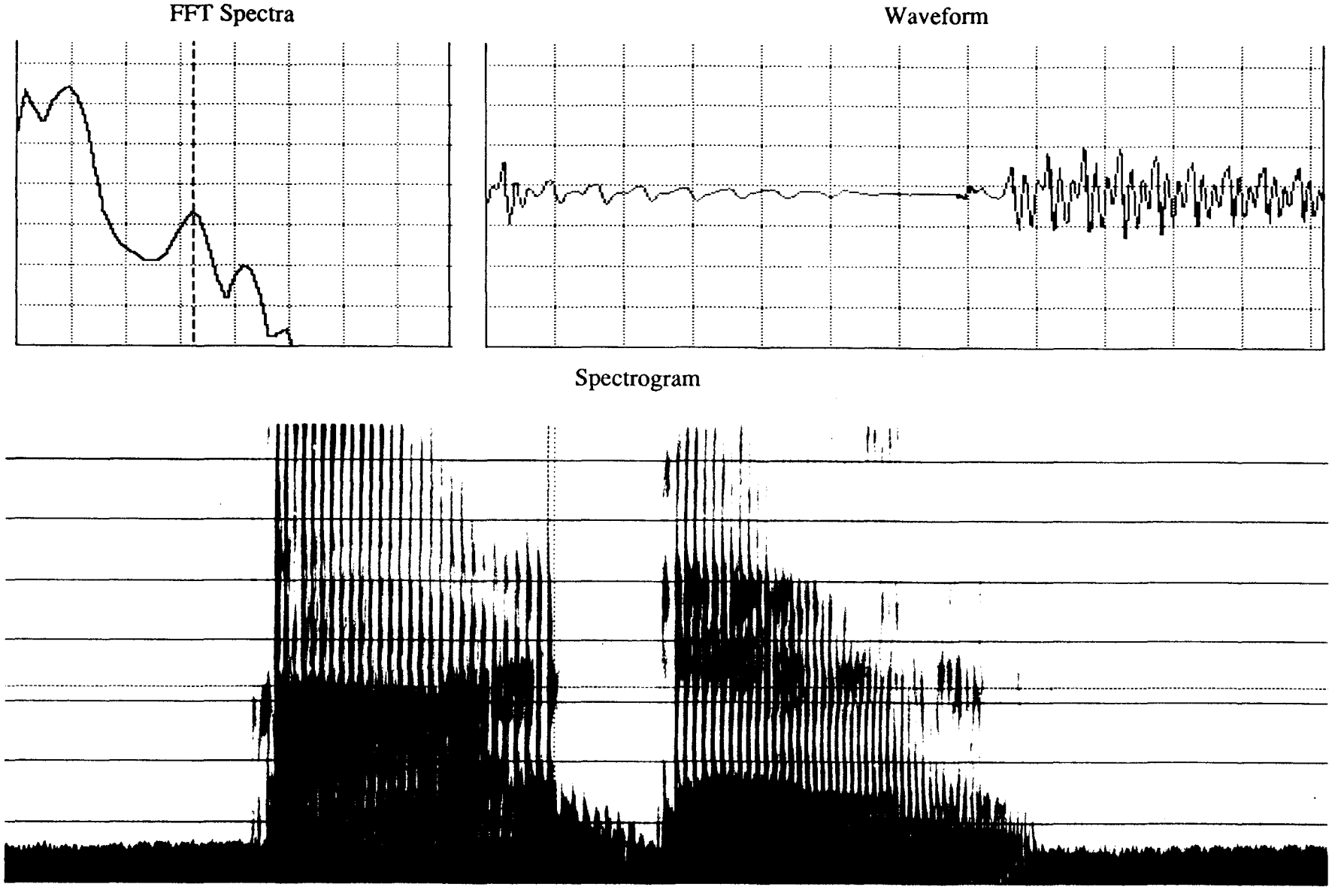


Figure 3. English main effects.

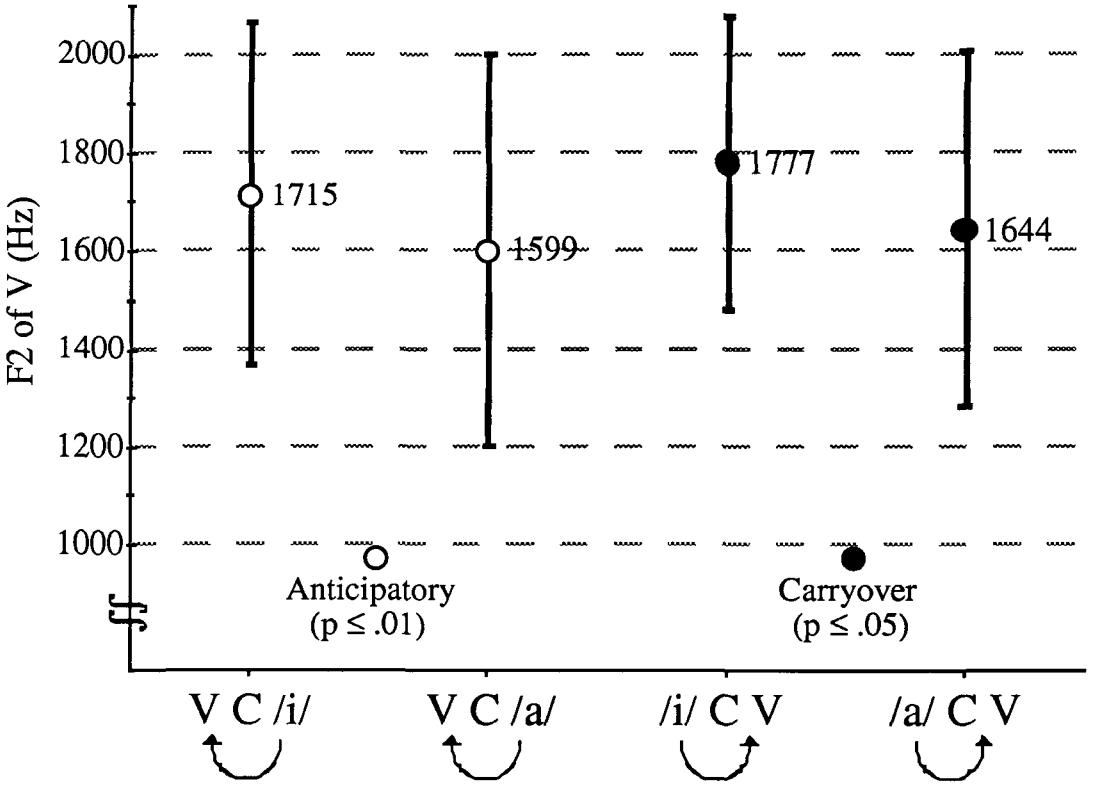


Figure 4. Russian main effects on V2.

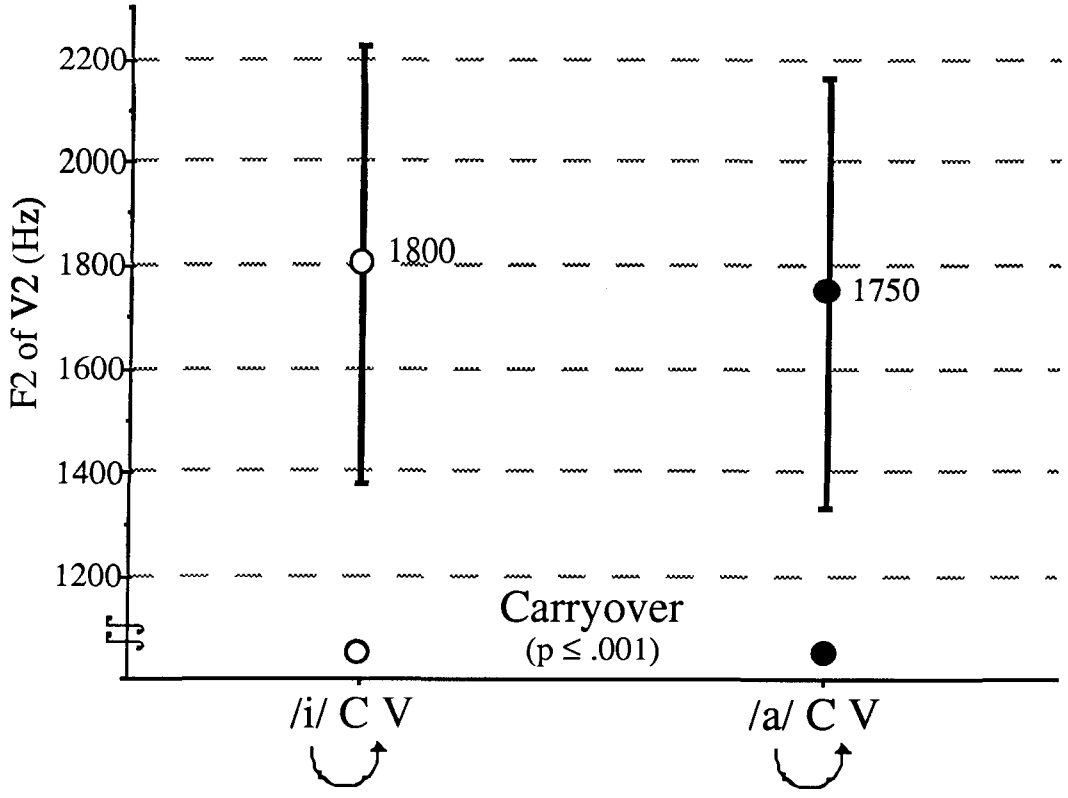


Figure 5. Bulgarian carryover effects on V2.

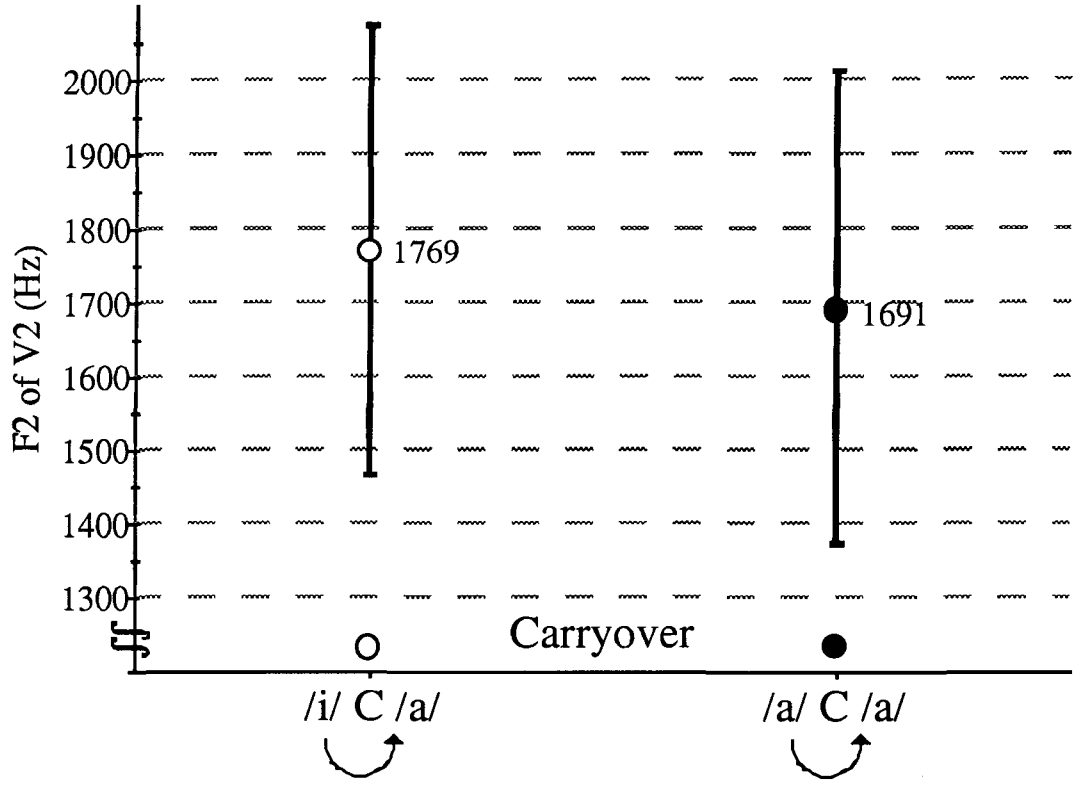
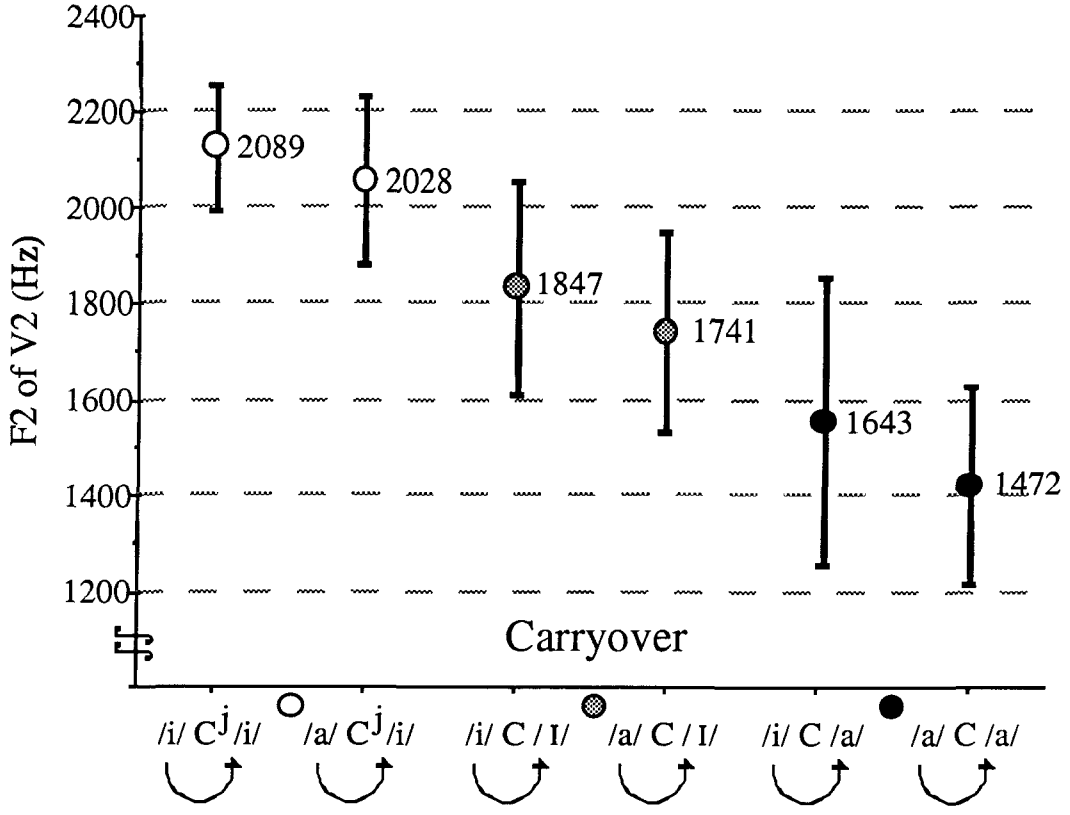


Figure 6. Polish carryover effects on V2 across voiceless consonants.

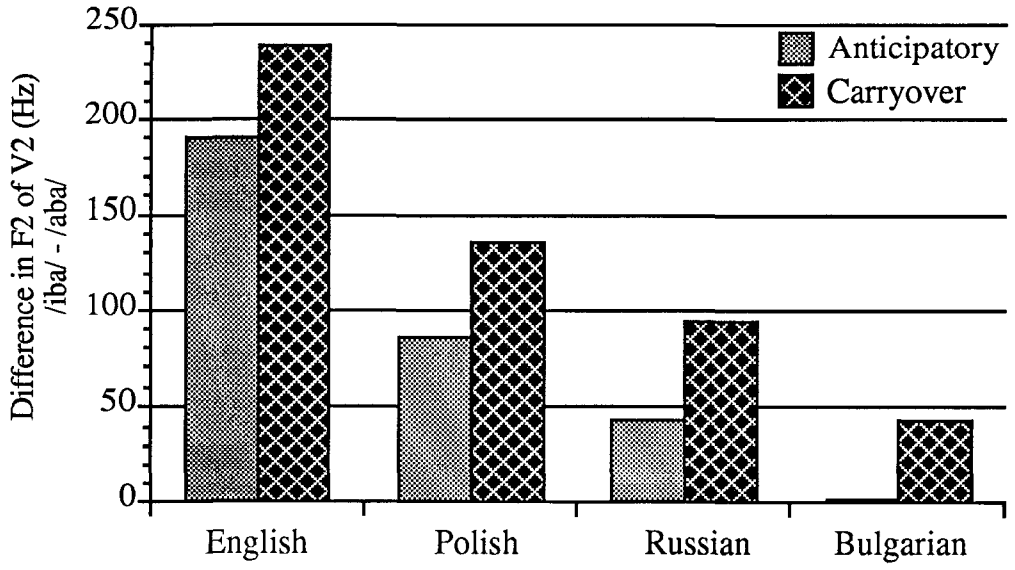


shows the same effect on the low vowel /a/. As in the other languages, the low vowel shows the largest magnitude effect. Again, we see effects across both plain and palatalized consonants, as we did in Russian and Bulgarian. Anticipatory vowel-to-vowel coarticulation is also found, but only on /a/, and is smaller in magnitude than the carryover effects.

Based just on these separate language analyses, the following generalizations emerge. In all the languages, some statistically significant vowel-to-vowel coarticulation is found. Except for English, carryover coarticulation is more general than anticipatory. Lastly, coarticulation is more generally found on /a/.

Statistical comparisons across all four languages do not reveal any further differences, but rather simply confirm the generalizations evident in the individual within-language tests. Figure 7 shows the size of the difference produced by a contextual /i/ versus a contextual /a/ on all /a/ tokens across the voiced bilabial stop /b/ (recall that coarticulation is more robust on the low vowel). Anticipatory coarticulation is weaker than carryover in all four languages, presumably because the first vowel was always stressed. By contrast, differences in magnitude of coarticulation can be seen across the languages. Generally speaking, English exhibits the strongest coarticulatory effects, and Bulgarian the weakest. These magnitudes for the Slavic languages are smaller not only compared to English, but also to other languages studied by other researchers. But these differences are not strong or consistent enough with only three speakers per language for the statistical comparisons to show significant language differences.

Figure 7. Magnitude differences across languages.

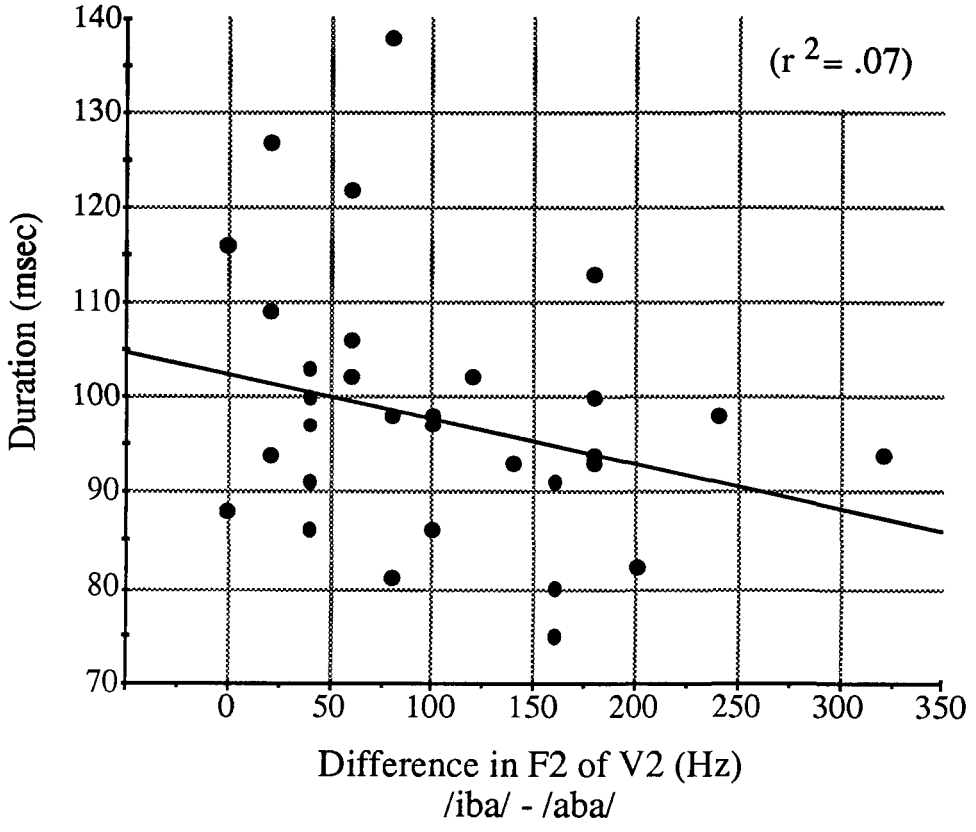


Given that these effects are smaller than those found for other languages, it follows that tests for their occurrence must be fairly sensitive. This, we believe, explains why our results for Russian differ from previous studies which showed little or no coarticulatory effects on F2. Differences in experimental design, measurement technique and statistical analysis combine to make our study more sensitive to these effects. Figure 8 compares the relevant factors in our research to those in Öhman and Purcell's research. Öhman's preliminary study of Russian was based on one speaker, looking only at anticipatory coarticulation on /ɛ/. With respect to measurement technique, F2 measurements were made from spectrograms. It is unclear from the published text whether or not he ran statistical tests. Purcell's study, on the other hand, incorporated 4 speakers. F2 measurements were made from LPC spectra which, we assume, used a 25.6 msec window. In regard to statistical analysis, Purcell ran a large series of regression tests comparing effects across categories of vowels.

Figure 8. Comparison of carryover techniques.

	Öhman	*Purcell*	*Choi & Keating*
Experimental Design	1 speaker (anticipatory effects on /ɛ/ only)	4 speakers	3 speakers
Measurement Technique	spectrograms	LPC spectra (25.6 msec window)	FFT spectra (5 msec window)
Statistical Analysis	?	series of regression tests (vowel categories against F2 values)	repeated measures ANOVA

Figure 9. Correlation between consonant duration and magnitude of coarticulation.



Öhman's study, then, differed from ours in looking at a very limited set of data. In particular, we would expect anticipatory effects on /ɛ/ to be extremely weak, given our finding that the strongest effects are the carryover effects on /a/. Purcell's study differed from ours in using a large analysis window which was likely to miss effects occurring in fast changing transitions. In Russian, it is indeed the case that the effects occur only at the very edges of the transitions. The statistical technique that he used, moreover, served only to compound this insensitivity.

Summary

In summary, the Slavic languages have statistically significant vowel-to-vowel coarticulation. These effects, however, are very small and most likely imperceptible. It appears, therefore, that secondary articulations do not, by themselves, block the occurrence of vowel-to-vowel coarticulation. However, something about these languages seems to inhibit or weaken these effects. Two possible explanations for this difference in magnitude are briefly considered.

One possibility is that these languages have longer consonants, either because of the secondary articulations or because of differences in overall speaking rate - and with longer consonants, there is less overlap of vowel gestures, resulting in weakened coarticulation. To explore this possibility, we tested the correlation of consonant duration with magnitude of coarticulation on the low vowel /a/ across the voiced labials. The results of this test, illustrated in figure 9, shows that there is no relationship between consonant duration and magnitude of coarticulation. Hence, whatever the cause of the weakened vowel-to-vowel coarticulation, it is probably not that vowels overlap less in Slavic languages.

Another possible explanation for small coarticulatory effects relates to the perceptual distinctiveness required by a language to maintain linguistically contrastive vowels. Manuel and Krakow (1984) propose that, all else being equal, languages with more vowel distinctions will show smaller coarticulatory effects. However, this kind of explanation will not work in the cases at hand since these languages all have only 5 or 6 vowel phonemes. On the most literal interpretation of this hypothesis, the Slavic languages should exhibit the same magnitude of coarticulation as that seen for languages like Ndebele, Shona and Japanese. We have seen, however, that this is not the case. As Manuel and Krakow also recognized, tongue body contrasts within consonant systems crucially interact with those for vowels in determining how much the vowels may vary. The precise way in which this interaction occurs remains to be explored.

Acknowledgements

This research was partially supported by NSF Grant BNS 8418580 and NSF Grant BNS 8720098.

References

Farnetani, Edda. 1990. V-C-V lingual coarticulation and its spatiotemporal domain. In W. J. Harcastle and A. Marchal (eds.) *Speech Production and Speech Modelling*. Kluwer Academic Publishers, Dordrecht: 93-130.

Magen, Harriet. 1984. Vowel-to-vowel coarticulation in English and Japanese. *Journal of the Acoustical Society of America* 75: S41.

Magen, Harriet. 1986. Coarticulatory effects between vowels of non adjacent syllables. Paper presented at the 1986 Linguistic Society of America Meeting.

Manuel, Sharon. 1990. The role of contrast in limiting vowel-to-vowel coarticulation in different languages. *Journal of the Acoustical Society of America* 88: 1286-1298.

Manuel, Sharon and Rena Krakow. 1984. Universal and language particular aspects of vowel-to-vowel coarticulation. *Haskins Laboratories Status Report on Speech Research* SR77/78: 69-78.

Öhman, Sven. 1966. Coarticulation in VCV utterances: spectrographic measurements. *Journal of the Acoustical Society of America* 30: 149-154.
Purcell, Edward. 1979. Formant frequency patterns in Russian VCV utterances. *Journal of the Acoustical Society of America* 66: 1691-1702.
Recasens, Daniel. 1984. Vowel-to-vowel coarticulation in Catalan VCV sequences. *Journal of the Acoustical Society of America* 76: 1624-1635.
Recasens, Daniel. 1987. An acoustic analysis of V-to-C and V-to-V coarticulatory effects in Catalan and Spanish VCV sequences. *Journal of Phonetics* 15: 299-312.

Vowel Length and Pitch in Yavapai

Kimberly D. Thomas

Introduction

As early as 1970 (published in *Hokan Studies*, 1976), Shaterian claimed that there are three distinctive vowel lengths in Yavapai, a Native American Indian language of the Yuman family. He reiterated these claims in his 1983 dissertation. Shaterian does not limit his remarks to Yavapai, but rather suggests a three-vowel-length distinction as a phenomenon within the Pai (Northern Yuman) subgroup. He cites Joel's work (1966) in further support of this claim.

> Joel (1966) describes a similar situation in Paipai, a Yuman language of Lower California with suggested strong affinities to Northern Yuman [Havasupai, Hualapai, and Yavapai]: She finds three vowel lengths, one short, one long, and one of varying quantity. A cursory comparison of her recordings with those of Yavapai reveals definite correspondences in the three vowel lengths, although the matter of how systematic these correspondences may be has yet to be thoroughly explored. (Shaterian, 1976:88)

Shaterian also claimed that Yavapai, a predominantly intonational language, makes limited use of tone:

> I have found, in addition, distinctive pitch in both . . . [Northeastern Yavapai] and . . . [Western Yavapai], the two dialects with which I have worked most closely, although I have not yet been able to take precise acoustic measurements of the relationship between pitch and length; nevertheless, this is, I am certain, going to prove a very interesting area of research; and it is quite likely that the results of these investigations will...shed new light on Proto-Yuman vocalism. (Shaterian, 1976: 88, 89)

The following study will address two primary points. First it will determine whether the existence of three lengths is statistically verifiable. Secondly, it will address the relationship between vowel length, pitch, and syntactic category—the last in an attempt to find a morphosyntactic connection.

Shaterian's claim is of general interest because few languages of the world boast three contrastive vowel lengths. In an article entitled "Vowels of the World's Languages," Ladefoged and Maddieson (1990) discuss three languages—Estonian, Mixe, and Kamba—that use three and four contrastive vowel lengths. Estonian has been shown to have three vowel lengths (Lehiste, 1970). However, distribution of these three degrees of length is limited to some degree by syllable structure and word patterning. Hoogshagen (1959) describes Mixe as a language which uses three contrastive vowel lengths. Unlike Estonian, Mixe does not seem to be influenced by

word patterning or syllable structure. Whiteley and Muli distinguish four contrastive vowel lengths in the Bantu language Kamba (Ladefoged and Maddieson, 1990). The third and fourth lengths, however, may be morphologically derived.

Methods

The data in this study are from tape recordings Shaterian made in 1981 and 1989 of a Northeastern Yavapai speaker named Clara Starr. The two sets of minimal triplets that constitute the main data set are shown in Table 1. The lengths will be referred to as short, long, and extra-long. Long vowels and extra-long vowels are transcribed /aˑ/ and /aː/ respectively. The term *duration* will only be used to refer to measured quantities of length in milliseconds. There are two sets of words designated Set 1 and Set 2. They begin with initial clusters /ʔh/ and /ʔɲ/ respectively; in actual pronunciation a predictable vowel is inserted between these consonants. Note that Set 2 contains a fourth word with a short vowel and a glottal stop. This point will become relevant when examining vowel durations.

TABLE 1: Yavapai Length Triplets

Set 1		Set 2	
/ʔhà/	'water'	/ʔɲá/	'road'
/ʔhâˑ/	'be bitter'	/ʔɲàˑ/	'be black'
/ʔháː/	'cottonwood'	/ʔɲáː/	'sun'
		/ʔɲáʔ/	'me'

All forms were elicited several times during the recordings. Samples in which Ms. Starr was cued improperly or in which she hesitated or seemed to be thinking out loud were excluded. Once the proper samples were identified, vowel durations and their corresponding pitch measurements were made using a digital spectrograph machine (Kay Speech, DSP, Model #5500). The number of tokens of each word measured ranged from 3 to 8.

The general difficulty in measuring duration is consistency. The measurement techniques described below were used in an attempt to control as many variables as possible. The technique involves locating the vowel, demarcating it between two time cursors, and noting the duration in milliseconds. The vowel is measured from its onset, defined as the end of the preceding consonant to the end of the voice-excited formant for the vowel. Because word-final vowels trail off into voicelessness, the major difficulty is demarcating the end of the vowel. The end of voicing for the vowel is determined by visually inspecting the spectrograph display as well as by a listening method, described below.

The listening method involves positioning one cursor near the end of the vowel and a second cursor just beyond the end of the word. The portion between

the cursors was then played to detect whether it included perceptible voicing. If some voiced vowel was heard, then the first cursor was moved further toward the second cursor (i.e., toward the end of the vowel). The process continued until no voicing was heard. The position of the first cursor was then judged to be the end of the vowel, and it always corresponded closely with the location chosen by examining formants.

In Set 2, forms with nasals, the end of the characteristic nasal formant structure, which is also the onset of the vowel in question, is identifiable by a rapid change in amplitude and formant pattern. The amplitude is defined by the intensity of color: the nasal formant is lighter and the vowel is much darker. This difference in intensity provides a reliable method of separating the nasal from the vowel. The end of the vowel was determined as in Set 1 by the listening method as well as examination of formant structure.

Results

The mean durations for the final vowels in the individual words in Sets 1 and 2 are shown in Figure 1. Note that the duration values in Set 1 appear to reflect three different vowel lengths, with a mean difference of 93 msec between short and long, and 153 msec between long and extra-long.

Note that in Set 2, the minimal pair /ʔŋa/ and /ʔŋaʔ/ have very similar durations. The final glottal stop has very little effect on the vowel. In the following analysis the two are combined. The difference between short and long is 161 msec; the difference between long and extra-long is 115 msec. Note that for both Sets 1 and 2, the difference between lengths is about 100-150 msec.

In addition, the absolute measure for each length of one set closely corresponds to the length of the other set. For example, the mean value for the short duration in Set 1 is 217 msec. In Set 2, it is 197 msec. Long and extra-long lengths across the two sets also have similar durations. These observations allow us to pool the short, long, and extra-long samples. This data is represented in Figure 2. The difference between the short (including /ʔŋaʔ/) and long duration is 119 msec; between long and extra-long it is 142 msec.

At this point, we would like to know whether the differences observed in length represent a reliable distinction between different categories or whether they are due to random variation. Analysis of variance (ANOVA) was carried out on these measurements. There was a highly significant main effect of the hypothesized length categories ($p < .0001$). Post hoc analysis on the means using the Scheffé F-test for multiple simultaneous comparison of means showed that the differences between all pairs of lengths are significant at better than the .01 level.

The data clearly show that there are three phonetic vowel lengths. It must next be determined whether these phonetic length differences in isolated words are related to other conditioning factors.

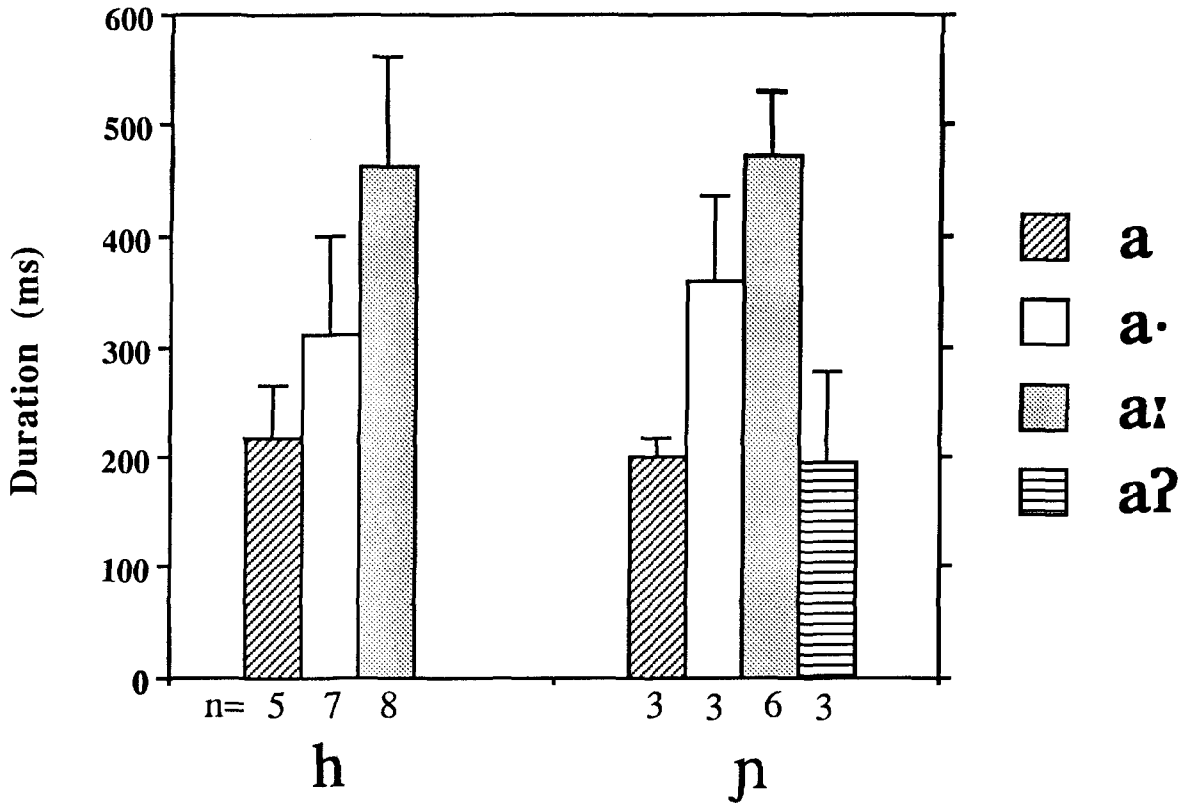


Figure 1 - Duration values for Set 1 (h) and Set 2 (ɲ).

Vowel Length and Pitch Interaction

Lehiste (1970) notes that the third degree of vowel length in Estonian is accompanied by a falling Fo contour. Woo(1969), working with Mandarin, discovered that more complex pitch patterns were correlated with extra-long vowel lengths. Given that in both tonal and non-tonal languages vowel length and pitch may be correlated, and the fact that Shaterian has recorded varying pitch patterns for some of the the forms cited in Sets 1 and 2, it is necessary to determine how pitch and vowel length interact in Yavapai. In order to establish the relationship between pitch and vowel length, pitch contours of the same sets of words were also examined.

Pitch contours were measured from narrow-band spectrograms. Fundamental frequency was calculated from the harmonic most clearly visible

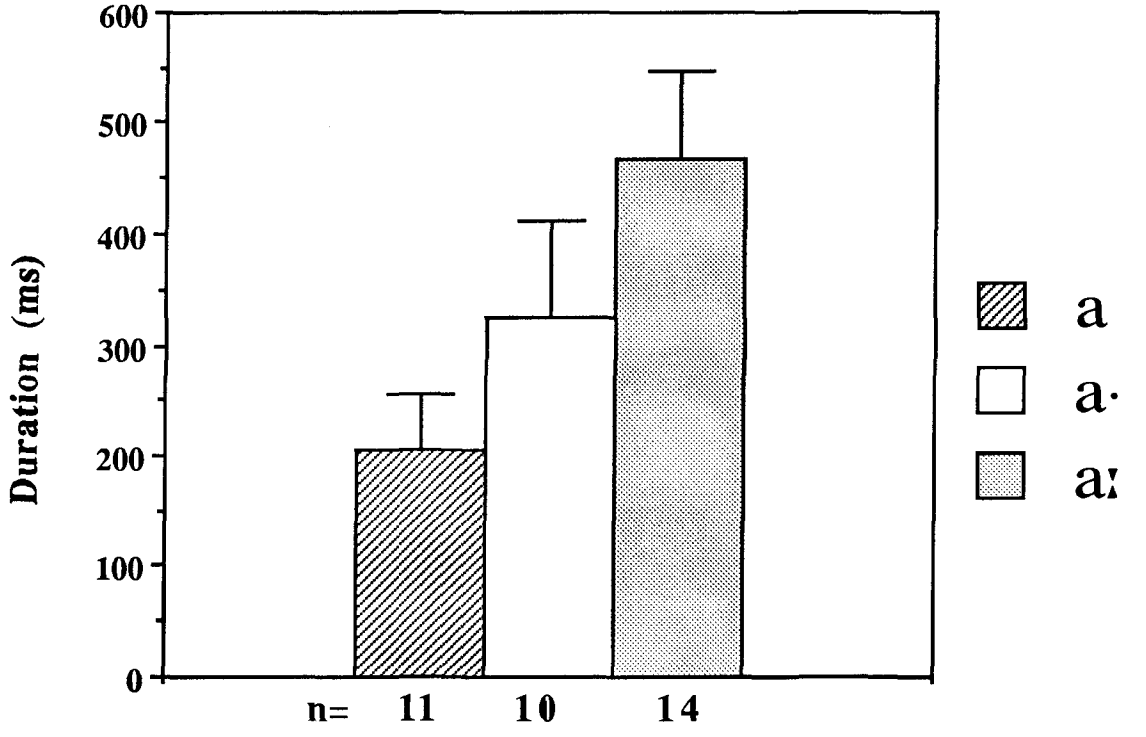


Figure 2 - Pooled duration values.

throughout the vowel. To represent the moving pitch contour, measurements were taken at the onset, the mid point, and the offset of the harmonic. The mean F_0 values for these three points for Sets 1 and 2 respectively are shown in Figures 3a and 3b.

In Set 1, the onset ranges from 185 to 187 Hertz, the midpoint ranges from 195 to 199, and the offset ranges from 194 to 230. Set 2 shows an onset range from 180 to 211 Hertz, with a midpoint range of 188 to 224 and an offset pitch range from 202 to 217. The pitch contours look very similar in all sets.

The numbers were pooled to determine the overall average of the pitch contours for each length category. Figure 4 represents these data. By comparing the vertical height of the bars, it becomes evident that the general contour for each length category are virtually identical. Analysis of variance confirms this observation. No significant main effect of length was observed. All comparisons of short/long and long/extra-long were not significant. Note that the general pitch contour for words of each group is a rising pattern, and that the long and extra-long contours are virtually identical. The shape of the short syllable differs slightly from the long and extra-long contours, but this difference is not statistically significant.

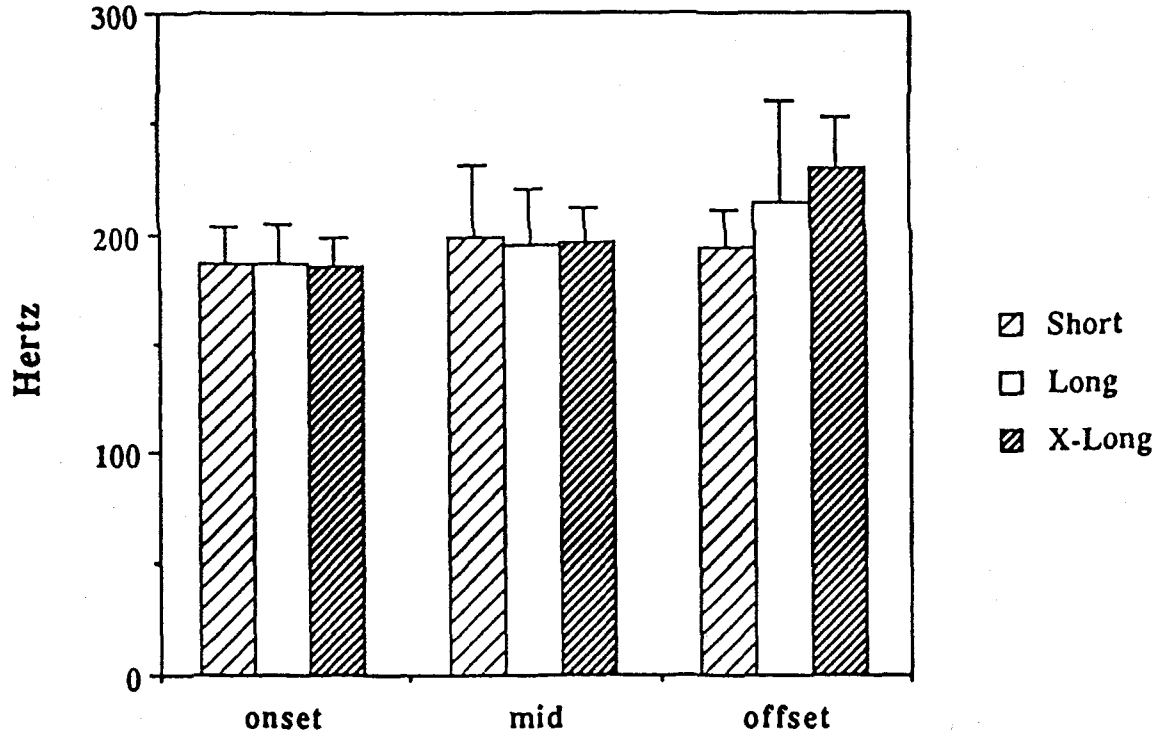


Figure 3a - Pitch contour data showing similarity of pitch points for each length category for Set 1.

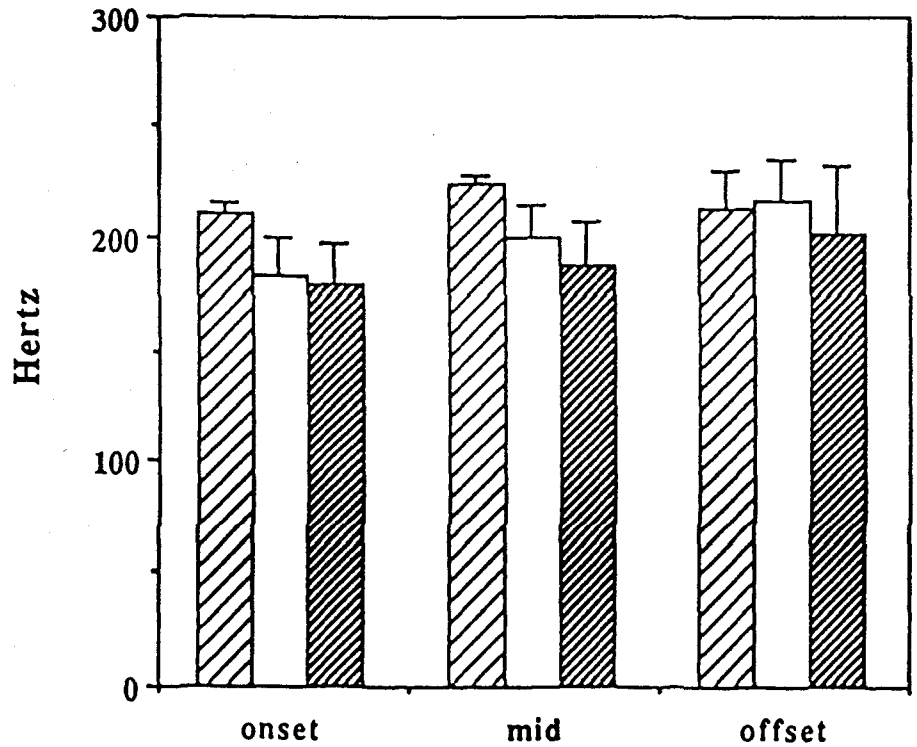


Figure 3b - Pitch contour data showing similarity of pitch points for each length category for Set 2.

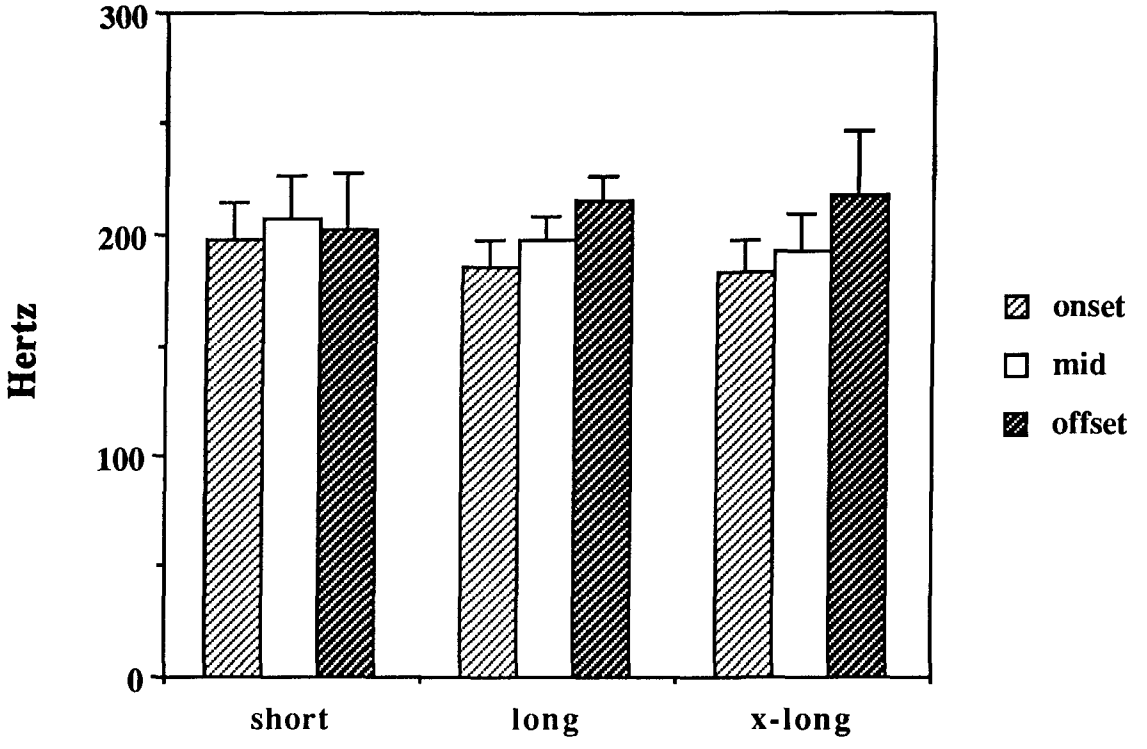


Figure 4 - Pooled pitch contour data by length category.

We noted that typically pitch rises throughout the syllable. We might therefore expect that as a vowel becomes longer, the offset of the pitch becomes higher. In order to establish the relationship between pitch and vowel length, pitch contours of the same sets of words were also examined. Comparing Figure 3a and 3b shows that this is true in Set 1 but not in Set 2. The offset value is more variable in the data in general. Its relationship to the mid point is less consistent across different lengths than the relationship between the onset and the mid point.

We can conclude that these three phonetic vowel lengths are not related to pitch, at least not in the context of isolated utterances.

Vowel Length and Syntactic Category

Another hypothesis, expressed in personal communication by Pam Munro (1990) suggests that syntactic category plays a role in predicting vowel length. Looking at the main data set again, we see that in both Sets 1 and 2, the long-length syllable is verbal (as in 'be bitter' and 'be black') and the short and extra-long lengths are nouns (as in 'water,' 'cottonwood,' 'road,' 'sun,' and 'me'). However, this hypothesis does not hold in an additional minimal triplet measured.

Set 3, in Table 2 shown below, is an example of a near-minimal length triplet in which all of the words are nouns.

TABLE 2: Yavapai nouns with /u/

Set 3

/ʔyu/	'my eye'
/ʔyuˑ/	'owl'
/huː/	'nose'

Statistics cannot be calculated on this triplet because only one token was available. However, we can still look at duration and pitch as a preliminary examination of a minimal triplet of only nouns. Figure 5 is a comparison of the duration of /u/ in Set 3 with the means for /a/ in Sets 1 and 2.

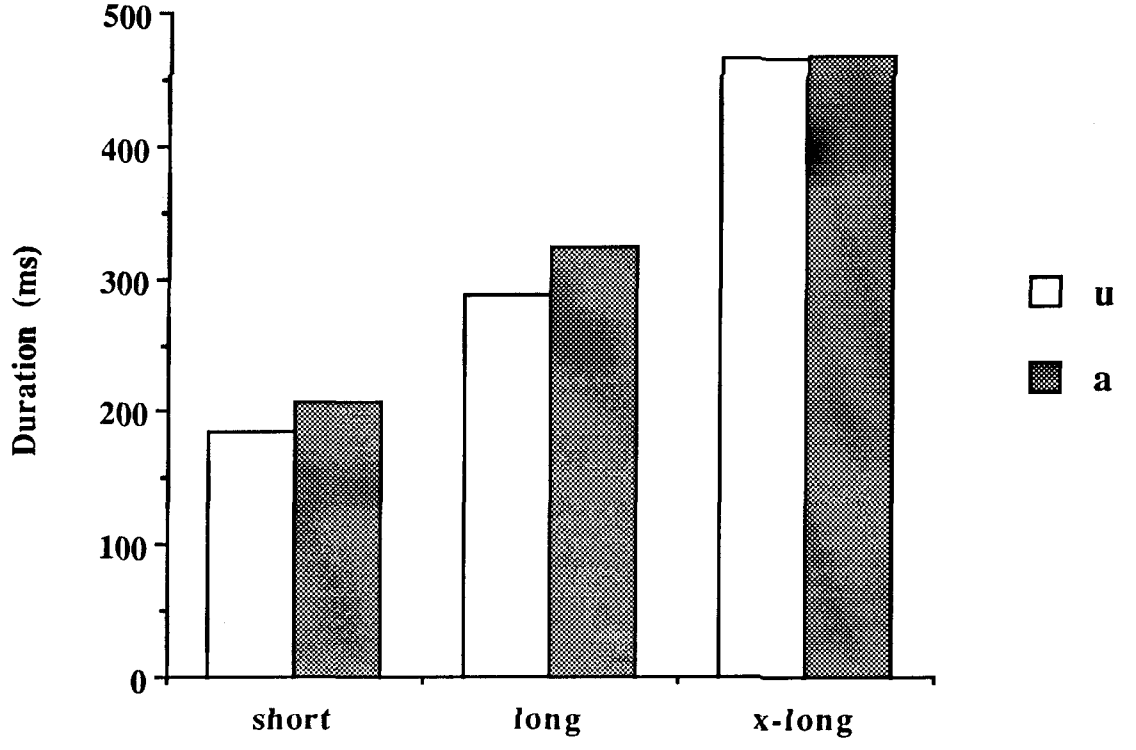


Figure 5 - Comparison of the duration values of /u/ and /a/.

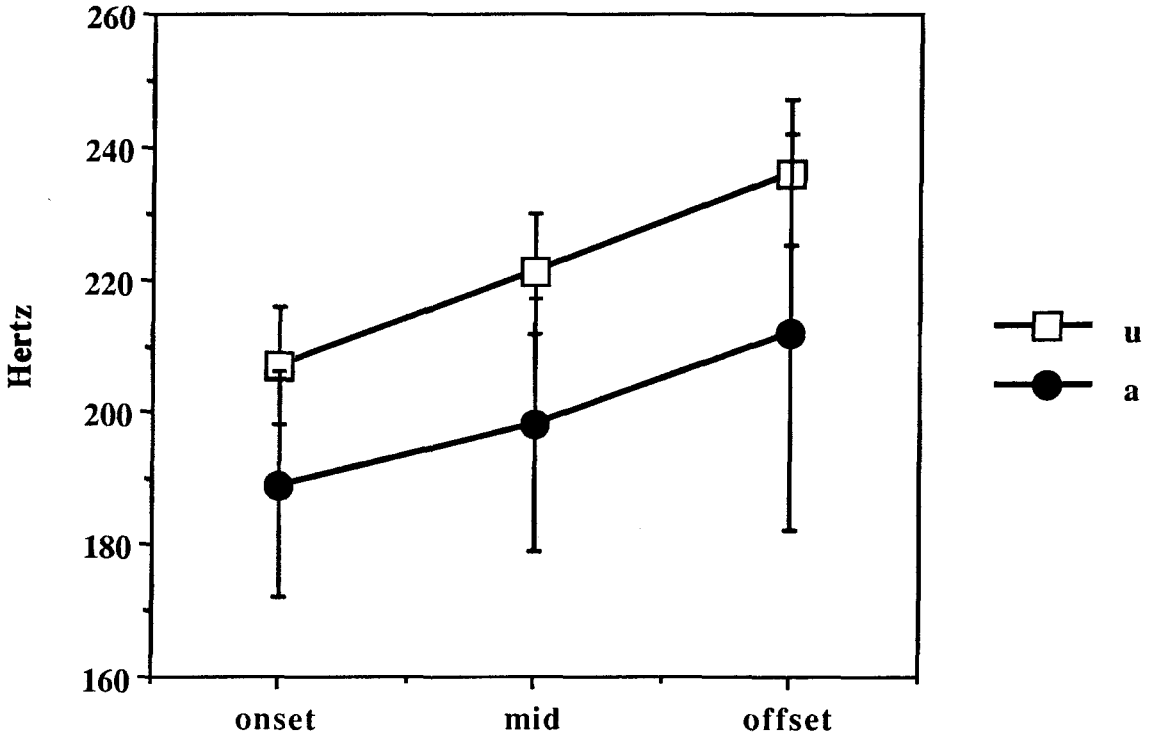


Figure 6 - Comparison of pitch contours for /u/ and /a/.

The difference in duration between the short /u/ and short /a/ is 22 msec, between long /u/ and /a/ it is 36 msec, and between extra-long /u/ and /a/ it is 1 msec. The durations of /u/ correlate well with the durations of /a/.

The pitch data correlate as well. The comparison can best be seen in a line graph, as shown in Figure 6 below. Since we saw in Sets 1 and 2 that the contours have a similar shape for each length, we have pooled the three length categories to obtain one pitch contour for /a/. The onsets, midpoints, and offsets were also pooled to obtain one representative pitch contour for /u/. The contours have more or less the same shape, with /u/ showing the same contour as /a/ but transposed a little higher, as is expected for a high vowel (Hombert et al. , 1979).

In this minimal triplet of three nouns, the vowel lengths and pitch contours correlate well with minimal triplets of two nouns and a verb. From this we conclude that syntactic category does not enable us to predict the difference between the two longer vowel lengths. More minimal triplets of various noun and verb combinations must be measured to substantiate this claim.

Conclusion

We can conclude that there are three phonetic vowel lengths operating in Yavapai. These lengths are not conditioned by pitch factors or syntactic category in the data we have examined. Therefore, the bolder conclusion is that Yavapai has three distinctive vowel lengths.

Unlike Estonian and Kamba, Yavapai length does not seem to be predictable from other phenomena present in the language. In this respect it is more like Mixe. Thus, Yavapai can be added to the sparse list of languages which utilize three contrastive vowel lengths.

Acknowledgments

I would like to thank Ian Maddieson, whose assistance on this project proved invaluable. Special thanks goes to Alan Shaterian for making his data readily available to me upon request. I would also like to thank Pamela Munro for making valuable comments on this work as well as providing her insights on Yuman languages in general. A similar version of this paper was presented at the 120th meeting of the Acoustical Society of America in San Diego, California. This research was supported by NSF Grant BNS 87-20098.

References

Hombert, J-M., J. Ohala., and W. Ewan. (1979). "Phonetic explanations for the development of tones." *Language*. 55,1. 37-58.

Hoogshagen, S. (1959). "Three contrastive vowel lengths in Mixe."*Zeitschrift fur Phonetik und allgemeine Sprachwissenschaft*, 12, 111-115.

Joel, Judith. (1966.) *Paipai Phonology and Morphology*. Ph.D. dissertation. University of California, Los Angeles.

Ladefoged, Peter and Ian Maddieson. (1990). "Vowels of the world's languages." *Journal of Phonetics*, 18, 93-122.

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

Shaterian, A. (1976). "Yavapai [+sonorant] segments. " *Hokan Studies*, ed. by Margaret Langdon and Shirley Silver (*Janua Linguarum, Series Practica*, 181). The Hague: Mouton.

_____. (1983). *Yavapai Phonology and Dictionary*. Ph.D. dissertation, University of California, Berkeley.

Perception of assimilation in consonant clusters: A gestural model

Dani Byrd[1]

1. Introduction

In recent years, researchers at Haskins Laboratories and elsewhere have begun to approach speech production from a theoretical standpoint novel to traditional linguistics. Browman and Goldstein of Haskins Laboratories have proposed a linguistic gestural model which describes human speech in units of articulator movements called "gestures" temporally organized in a "gestural score." For example, a gesture might be the coordinative movement of the jaw and both lips to produce the first or last sound in the syllable [bab] or the motion of the tongue body required for the first sound in the syllable [kæt]. The ges-tural score would specify the temporal relationship between the movement of the articulators in a sequence like [babkæt]--"bobcat." One basic motivation for this framework has been that "much is missed when the line between phonological patterning and physical processes is drawn too finely." (Browman and Goldstein, 1990, in press-b, p. 1) A phonological theory in which "dynamically-defined articulatory gestures are the basic...units" (Browman and Goldstein, 1988) allows for an explicit account of a variety of phonological processes for which a satisfactory description has been lacking in previous phonological theory. Articulatory phonology proposes that variations in fluent speech such as deletion, assimilation, insertion, and reduction can be described as changes in gestural magnitude, temporal relations between gestures, and/or other aspects of gestural specification.

The topic of this investigation is the effect of temporal differences in gestural organization. Gestural overlap may cause the articulation of one sound to interact with the articulation of another, thereby changing its acoustic make-up. The acoustic result of these co-produced units reflects their combined influence. While coarticulation occurs in all natural speech, in casual or fast speech, gestural overlap between neighboring consonant gestures may increase enough to result in the perception of assimilation or deletion.

In order to determine the ability of this theory to accurately describe speech, researchers at Haskins Laboratories are developing a computational model based on these theoretical assumptions. With this model, one can represent utterances in terms of an input specifying an abstract characterization of coordinated movements of articulators within the vocal tract, the gestural specification, and then observe their acoustic consequences. A gesture involves the closure and release of a vocal tract constriction and is modeled as a dynamical system that regulates the movement of the articulators. (Browman & Goldstein, in press-a; Saltzman, 1986)

[1]*This research was conducted in full at Haskins Laboratories, New Haven, Connecticut, supported by NIH grant HD-01994, and was made possible only through the generosity of the researchers and staff at Haskins in allowing this student to use their facilities and benefit from their experience and knowledge. I thank Peter Ladefoged and Patricia Keating for insightful comments on an earlier version of this paper. It is with much gratitude and fondness that I acknowledge the guidance of my teacher, Dr. Louis Goldstein of Haskins Laboratories and Yale University.*

> The computational system (Browman & Goldstein, 1987; Saltzman, Goldstein, Browman & Rubin, 1988) provides a representation for arbitrary...input utterances in terms of such gestural units and their organization over time, called the *gestural score.* The layout of the gestural score is based on the principles of intergestural phasing (Browman & Goldstein, 1987) specified in the *linguistic gestural model.* The gestural score is input to the *task dynamic model* (Saltzman, 1986; Saltzman & Kelso, 1987), which calculates the patterns of articulator motion that result from the set of active gestural units. (Browman & Goldstein, in press-a)

Note that the task dynamic model is not language specific, in fact having been originally designed to model physical motor movements unrelated to speech. These articulatory movements are input to the articulatory synthesizer (Rubin, Baer & Mermelstein, 1981) which calculates the output waveform. This speech synthesizer is a model of the human vocal tract from larynx to lips and is constrained by the physical limitations of the vocal tract. Speech is synthesized by the mathematical calculation of the resonances resulting from different vocal tract lengths and shapes over time.

Speech samples synthesized in this way are to be used here as stimuli for studying the perception of co-produced consonants. It is hoped that investigating co-production within the framework of the gestural model will allow insight into the articulatory events which give rise to the percept of assimilation. In this experiment, the gestural model described above will be employed to examine how the organization of gestures over time affects the perception of assimilation in consonant clusters. Assimilation is a phonological event that increases in frequency in casual speech. The gestural model will help explain why assimilation is, for example, more frequent in informal speech than in carefully articulated speech (Barry, 1985; Kerswill, 1985; Nolan, 1989; Zsiga and Byrd, 1990). Because assimilation is commonly accepted as a casual speech process, it is worth considering how various theoretical frameworks have characterized both casual speech processes in general and coarticulation and assimilation in particular. In this way, we can better evaluate the gestural model of assimilation.

The motivation of casual speech processes has been discussed at length by many researchers. All agree that these processes are the result of the physiological nature of the articulatory system in response to extra-linguistic cues such as informality. The framework of articulatory phonology which derives from gestural theory sees many assimilation processes of casual speech as instances of overlap in which a gesture (or feature set(s) in standard phonology) is not deleted, delinked, or replaced but is rather obscured by an overlapping gesture. The fact that the assimilated gesture is not lost is supported by phonetic evidence in the form of X-ray microbeam and electropalatography studies which demonstrate that while acoustic evidence of a gesture may not be available to the listener, the physical gesture itself remains present (Browman and Goldstein, 1987; Barry, 1985; Hardcastle and Morgan, 1982, Nolan, 1989). Barry, using electropalatography, did conclude that the alveolar gesture disappeared in some cases of assimilation. However, Browman and Goldstein suggest that these instances may be described as a reduction in gestural magnitude and that the gestural "deletion" observed by Barry is not always confirmed by other types of measurements (such as X-ray microbeam analysis) which do not rely on on actual active articulator to passive articulator contact (Browman and Goldstein, 1987). Nolan also suggests something of this sort saying, "...the underlying

alveolar specification is still leaving a trace in the overall articulatory gesture, even though the target normally thought of as primary for an alveolar is not being achieved."(Nolan, 1989, p. 8) In the framework of articulatory phonology, casual speech processes are explained as changes in temporal overlap between gestures as well as their spatial magnitude. A large enough change in gestural magnitude may result in the articulatory disappearance of a gesture. In the experiment which will be described below only variation in gestural overlap within a consonant cluster will used to model the coarticulation causing the percept of assimilation; gestural magnitude will not be a variable.

Phonologists and phoneticians have described various characteristics of casual speech phenomena. In the framework of lexical phonology, for example, many casual speech processes fall under the heading of post-lexical rules which Kaisse and Shaw (1985) describe as having the following properties: producing gradient (non-binary) output, optionality, rate sensitivity, and yielding segments not present in the underlying segment inventory. Barry, (1985:2) also, cites optionality and gradience of output as typical of casual or connected speech processes. The processes are assumed to take place due to "ease of articulation," a poorly defined concept (see Ohala, in press-a, p. 2).

However, researchers seem to disagree as to whether this optimization in efficiency is accompanied by a perceptually costly increase in complexity. That is, there is no consensus on whether co-production, for example, aids a listener in "predicting" an upcoming piece of acoustically important information or whether the "compacting" of gestures obscures the acoustic information the listener is trying to obtain. Liberman (1988) proposes that coarticulation "is coordinated to gain high rates of transmission" yet the complexity of such phenomena requires him to posit a complex model of the perceptual system, necessitating, along with other motivating factors, a perceptual module specific to speech.

> [listeners] rely on a phonetic module, specifically adapted for processing the speech signal so as to recover the coarticulated gestures that produced it (Liberman & Mattingly, 1985; Mattingly & Liberman, in press-a, in press-b). This module is, of course, merely the other face of the one that governs the processes of coarticulation. Thus, there is but a single module with two complementary and specifically linguistic processes, one for producing phonetic gestures, the other for perceiving them. (Liberman, 1988, p. 147)

While the correctness of Liberman's motor theory of speech perception is not the topic here, it is described so as to illustrate a viewpoint regarding the relationship between coarticulation and perception. Some researchers support the view that gestural overlap and the consequent coarticulation aids listeners' perception. Fant (1981) suggests that "the variability and seeming complexity of speech may be resolved only when viewed from the receiving end," and he continues, "one could argue that speech perception relies on the sensing of temporal gestures or contrasts rather than on the identification of targets per se."(pp. 274, 275) This would point to the perceptual importance of temporal organization of gestures appealed to by Browman and Goldstein as the source of many casual speech variations. Repp describes this perceptual importance of parallel transmission:

> Because of anticipatory coarticulation, the speech signal frequently carries advance information about phonetic segments whose primary acoustic correlates occur later in time. This advance information may be helpful to

> listeners, especially when the principal cues for a segment are ambiguous or when a fast decision is required. (Repp, 1983, p. 420)

However, other researchers take the position that coarticulation is a product of the motor system which confuses the listener and must be filtered out or ignored. Lindblom says:

> Since coarticulation and reduction introduce variability of acoustic cues their effect is to increase the complexity of the acoustic coding, that is to create departures from acoustic invariance. This should mean that the explanations for coarticulation and reduction ought to be primarily production-based rather than listener-oriented. (Lindblom, 1981, p. 12)

This view seems to hold unlikely the possibility that the acoustic consequences of coarticulation may provide any advantage to the listener in identifying ambiguous utterances. Fowler (1981) suggests two accounts of the listeners' treatment of coarticulation that she calls the contrast account and the articulatory account.

> ...the contrast account presumes that coarticulatory influences on the acoustic signal for a vowel will always be erased by perceptual contrast effects. The articulatory account presumes that the acoustic consequences of gestures for the coarticulatory context constitute a frame perceptually, and the coarticulated segment is perceived relative to the frame. The effect of this is to "subtract" the coarticulatory influence of the frame from the signal for the coarticulated segment. (1981, p. 138)

Again, neither of these accounts would point to any perceptual usefulness of the acoustic consequences of coarticulation.

Clearly such a variety of views on the listeners' potential use of acoustic information provided by coarticulation results partly from the fact that co-production in natural speech is a gradient phenomena which may refer to variations ranging from the influence of a place of closure on a preceding vowel offset to "total" assimilation which, in articulatory phonology, may be described as gestural overlap resulting in the perceptual loss of an entire segment. The variety of positions sampled above may not be entirely incompatible in light of this range of variation.

One can design an experiment within the gestural framework to examine the role of gestural overlap on the perception of assimilation. That is, assimilation can be considered to be an extreme of coarticulation produced at specific degrees of temporal and spatial concurrence of articulatory gestures. Under this supposition, the gestures for all coarticulated consonants remain. The percept will be dependent on factors including, but perhaps not limited to, degree of gestural overlap (and corresponding temporal length of closure for consonant clusters) and the relative constriction locations of the gestures involved. In response to these factors, the listener may "filter out" effects of co-production created by a relatively small amount of overlap. In such instances, no assimilation or deletion is noticed. This minimal coarticulation occurs in all speech (for example, vowel onset and offset transitions in CVC environments). However with a greater amount of overlap, the resulting acoustic cues will be motivating enough to create the perception of an assimilated segment. That is, the *same* mechanism causing "segment-to-segment" acoustic

transitions in all speech is also responsible for assimilation and the perceptual loss of segments. This mechanism is co-production or gestural overlap. The specific nature of the listener's percept will be determined by the combined effects of the degree of temporal overlap between gestures, constriction location, and the gestural magnitudes.

One environment in which consonant assimilation typically may occur is a VC1C2 context. In such a case, C1 may be perceptually lost in favor of C2. That is, in connected speech, the C2 closing gesture may overlap the C1 closing gesture to such a degree that only the place of articulation of C2 is perceived; even though articulatory gestures for both consonants remain fully intact. The experiment described below will examine coarticulation in VC#CV sequences in order to determine the interaction of gestural overlap with the perception of assimilation.

Browman and Goldstein (1987) define gestures as consisting of an abstract underlying 360 degree cycle which can be represented using a *critically damped*, mass-spring equation. As unit mass and critical damping are assumed, only the stiffness and equilibrium position are parameters which can vary for different vocal tract variables. The stiffer a gesture is, the more rapid its motion. This in turn means that, all else being equal, the associated articulator(s) will reach its equilibrium position faster. The equilibrium position approaches the peak displacement or "target." Browman and Goldstein (1987) specify the achievement of target to be 240 degrees in the cycle based on observations from American English microbeam data. The explanatory usefulness of a gestural model relies largely on its access to information about the internal durations of gestures and the temporal relations between gestures. Gestures are organized temporally by their intrinsic stiffnesses and the relation between gestural activation periods in terms of a specified *phasing* which coordinates one gesture with another. A phasing association exists between two (tier-) adjacent vowels, between vowels and preceding and following consonants, and between consonants in sequence. Gestures are initiated with respect to the internal states of other gestures in a "relativistic phasing" (Kelso and Tuller, 1985; Browman and Goldstein, 1987). This organization of articulation allows invariant gestures to *overlap* spatially and temporally yielding an acoustic output which will vary depending on the behavior of all concurrently active gestures. The specific characteristics of an utterance are a direct result of the phasing relations and gestural details represented in the gestural score. It is the interaction in gestural magnitude and temporal overlap which enable the gestural model to describe phonological events such as assimilation. I will demonstrate that phonological alternations such as assimilation need not be the result of operations or conditions altering the phonological representation but rather can be the direct output of the representation itself.

Research described by Zsiga and Byrd (1990) has shown that the co-production of two consonants across a word boundary (a VC1#C2 pattern) produces formant movements in the preceding vowel which are distinctive for the various places of constriction of C2. These formant movements interact with the normal vowel offset created by the place of constriction of C1. Because the vowel quality is distinct for each consonant environment, the three cases of [alveolar#alveolar], [alveolar#bilabial], and [alveolar#velar] can be statistically discriminated using the vowel-final F2 and F3 values. This distinction was evident in both natural speech and speech synthesized according to the designated rules of intergestural phasing. This research was necessary to determine the acoustic effect of a second consonant in a sequence spanning a word boundary. After it was shown that a

consonant in this position did distinctively effect the offset formants of a preceding vowel, the next task was to assess the effect of the second consonant on the perception of the first.

Motivated by these findings, the perceptual experiment to be described was undertaken to determine whether listeners will use the acoustic information resulting from the co-production of two neighboring consonants in identifying C1 or whether perception will fail to show an absolute correlation with the acoustic information that we have determined to be present in the vowel offset. Understanding to what degree listeners utilize such acoustic information is vital to the formation of a satisfactory theory of speech perception. This type of inquiry is valuable in a larger sense in that its purpose is to delineate the degree of conjunction and disjunction between the productive and perceptual systems, a question in many fields of psychology and linguistics.

There are two hypotheses being tested here. The first is that the effect of gestural overlap will be more readily perceived in an environment including both the pre- and post-consonant cluster vowels rather than in an environment in which the pre-consonant sequence is spliced off and presented separately. It might be considered that the listener would be more likely to filter out the effects of coarticulation in the full presentation condition because he or she "knew" that the environment was a "potential assimilation site" (Nolan, 1989) and was perceptually *counteracting* this possibility. However, it is rather hypothesized here that the perception of assimilation will be *more* likely in the two-word presentation condition because the listener will be able to utilize more cues indicating that gestural overlap has occurred.

The second hypothesis is that there will be an asymmetry with respect to place of articulation due to acoustic peculiarities arising when a tongue tip gesture overlaps another consonant gesture. The exceptional behavior of coronals in assimilation processes has long been discussed in the literature, (see for example: Avery and Rice, 1989 and Paradis and Prunet, 1989), and the experimental results achieved with the gestural model will be discussed with respect to phenomena observed in assimilation processes in natural speech.

The acoustic effects of varying amounts of gestural overlap can be discovered and perceptually evaluated with the use of speech stimuli synthesized by an articulatory synthesizer implementing the linguistic gestural model. Small degrees of gestural overlap can be precisely controlled and quantified using the computational model. According to the phasing rules of Browman and Goldstein (1987), the second consonant is phased with respect to the offset of the first. The determination of the canonical consonant cluster phasing was based on X-ray films of the movement of articulators in real speech (Browman and Goldstein, 1987). The offset of C1 was determined to be at 340 degrees in its underlying cycle. Here the phasing relationship between two consonants is determined by the manipulation of the point in the second consonant which is specified in the gestural score to coincide with the temporal point occurring at 340 degrees in the cycle of the first consonant. For example, if the temporal point occurring at 340 degrees in the abstract cycle of the gestural activation of the first consonant is phased to coincide with the corresponding 340 degree point in the second consonant's gestural activation period, the two consonants will be modeled as completely (100%) overlapped. In the following discussion, this will be represented in the abbreviated form: [C_1=340 C_2=340]. Note that due to the inherent inertia of this physical model, phasing of the gestural activation periods is not synonymous with phasing of the actual articulator movements. The theoretical structure of this model of articulation demands that activation periods of gestures be the

subject of the phasing rules. As a result, a phasing creating a consonant cluster overlap of 100% or 107% does not mean that the articulator movements are exactly simultaneous or, in the case of 107% overlap, that the closure gesture for C2 is beginning before that of C1. The articulator movement begins as a function of the inertia inherent in the system or, conceivably, the particular articulator in question.

Specifically, this experiment is designed to determine the interactions of 1) the effect of bilabial versus alveolar place of articulation on assimilation in a consonant cluster ; 2) the effect of the degree of gestural overlap or phasing relations in these consonant clusters; and 3) the effect of context on the perception of assimilation, ie. will assimilation of consonants across a word boundary be perceived more readily in a two-word condition or in a truncated condition where the second word has been excised leaving the acoustic effect of C2 on the offset of the isolated word. Presentation in the truncated environment is required to isolate the perceptual effects of coarticulation in the vowel offset from those of closure interval and VC2 onset. This presentation variable is aimed at determining if hearing speech in context, ie. as part of a larger utterance, encourages or discourages the perception of assimilation.

2.0 Method

Phrases were synthesized with consonant clusters having the following structures: [b#b], [b#d], [d#b], and [d#d]. The phrases are shown in Table I.

Table I

The synthesized stimuli

[bæb # bæn]	bilabial#bilabial	control--no assimilation
[bæb # dæn]	bilabial#alveolar	
[bæd # bæn]	alveolar#bilabial	
[bæd # dæn]	alveolar#alveolar	control--no assimilation

Again, only the phasing of the two consonants are manipulated in this experiment. Neither movement velocity (ie. stiffnesses) nor closure times are manipulated for the two gestures. These qualities derive inherently from the articulatory model and the specified gestural overlap. The phasings at which the consonant clusters in the above phrases were synthesized are in 25 degree steps down from a 107% [total] overlap, (gestural activation periods of [C_1=340 C_2=365]), through 34% overlap at [C1=340 C2=115]. The respective phase and percent overlap for the complete set are shown in Table II.

Table II

C_1=340 C_2=	365	340	315	290	265	240	215	190	165	140	115
percent overlap=	107	100	93	85	78	71	63	56	49	41	34

The influence of the overlapped C2 on the vowel offset of VC1 naturally increases with progressively greater amounts of overlap beginning at the point where the articulator movement for C2 starts to co-occur with the vowel. When there is so little overlap that a C1 release occurs, no acoustic influence of C2 results. Note, however, that even if there is no release between the consonants and the gestural activation periods do overlap, there may still be points at which the **movement** for C2 has not yet begun when the C1 closure is first reached. If this is so, no formant influence of C2 could be expected in the preceding vowel. Only when the articulator movement producing the second consonant overlaps with the vowel production will this second constriction in the vocal tract influence the vowel formants.

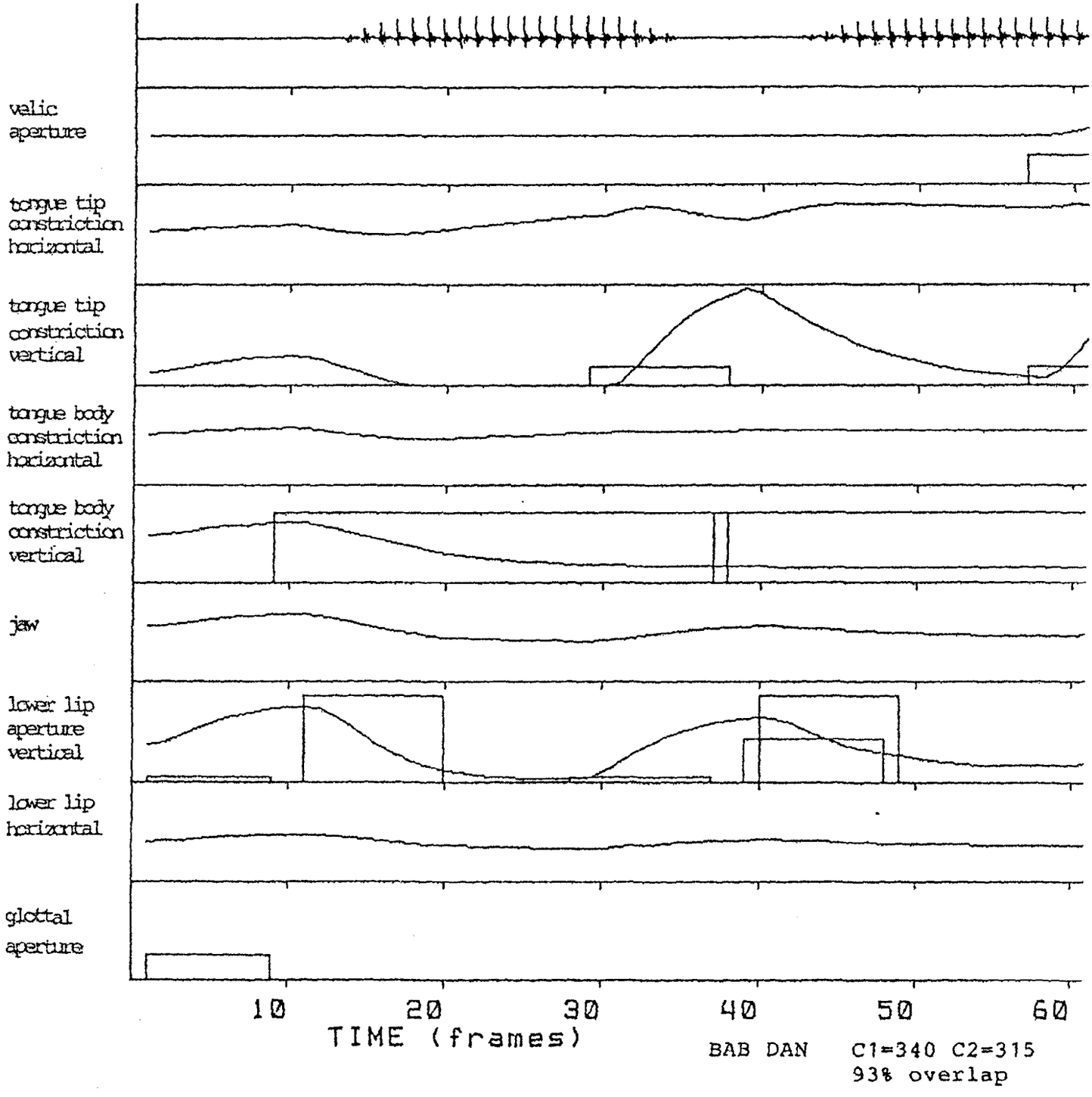
This format yielded four phrases each at 11 grades of overlap, a total of 44 stimuli. Twenty-two of these (all [bæb bæn] and [bæd dæn] phrases) were controls in that no assimilation in place of articulation could result from overlap.

These phasings were overlapped enough to prevent a release in all but six of the synthesized phrases. These were the two least overlapped tokens of [bæb dæn] and the four least overlapped tokens of [bæd bæn]. The fact that C1 is released at different phasings for different places of articulation is reflective of the fact that different gestures entail different speeds of release, or different effective stiffnesses in movement away from their targets. Because the tongue tip articulator moves more rapidly than the lip articulator, the [alveolar#bilabial] sequence yielded a released C1 with a greater amount of overlap than the [bilabial#alveolar] sequence did.

Note that the synthesized stimuli used here did not include release bursts for either C1 or C2. The acoustic cues of interest here were only the formant transitions and length of closure interval.

A sample gestural score is included as Figure 1. In this figure, the boxes represent the period of gestural activation, and the curves represent the movement in space of the articulators. The corresponding waveform is shown across the top.

FIGURE 1


velic aperture
tongue tip constriction horizontal
tongue tip constriction vertical
tongue body constriction horizontal
tongue body constriction vertical
jaw
lower lip aperture vertical
lower lip horizontal
glottal aperture
10
20
30
40
50
60
TIME (frames)
BAB DAN C1=340 C2=315
93% overlap

For the second presentation environment, a truncated version of each stimuli was also created to isolate the perceptual effects of the vowel offset formants. These stimuli were created from the stimuli of the two-word condition, the crucial difference being that the second word, [C2æn], was excised leaving only the first, [bæC1]. Note that any release that may have occurred after C1 was *not* included. Recall that the presentation of the truncated stimuli was required to isolate the perceptual effects of gestural overlap on the vowel offset from that of closure interval and VC2 onset. Of course, the complete accessibility of acoustic cues (ie. formant transitions *and* closure interval) is crucially dependent on the presentation of the non-truncated (two-word) condition.

This environment again yielded four phrases at 11 phasings. Thus the total number of synthesized stimuli was 88 utterances.

2.1 Stimulus Presentation

A pilot evaluation was made to ensure that the synthesized phrases at relatively low levels of overlap (63% and lower) were perceived as having the correct places of articulation. A group of ten introductory phonetics students were asked to transcribe this subset of the synthesized phrases. Place of articulation was misidentified only in approximately 15% of the stimuli, and the lack of releases or bursts did not appear to cause any misidentification of place of articulation. This error level was considered acceptable as the quality of the synthesized speech was unfamiliar, and subjects in the final experiment were to be given a forced-choice paradigm, not any transcription-related tasks.

Five repetitions of each set of the 44 tokens of each condition were randomized in blocks for a total of 220 tokens in the reported experiment. The same random order was used in both the two-word and truncated conditions. In the two-word condition, these tokens were separated by a period of four seconds, with a nine second interval after every tenth token. In the truncated condition an inter-stimuli interval of three seconds was used, with a six second pause after every tenth token.

The synthesized stimuli were recorded onto a chromium dioxide cassette and replayed from a Nakamishi cassette recorder over headphones. The stimuli were presented binaurally to seven subjects. The seven paid-volunteers included four men and three women, all with no background in phonetics or synthetic speech. These subjects heard both the set of two-word utterances and the set of truncated stimuli.

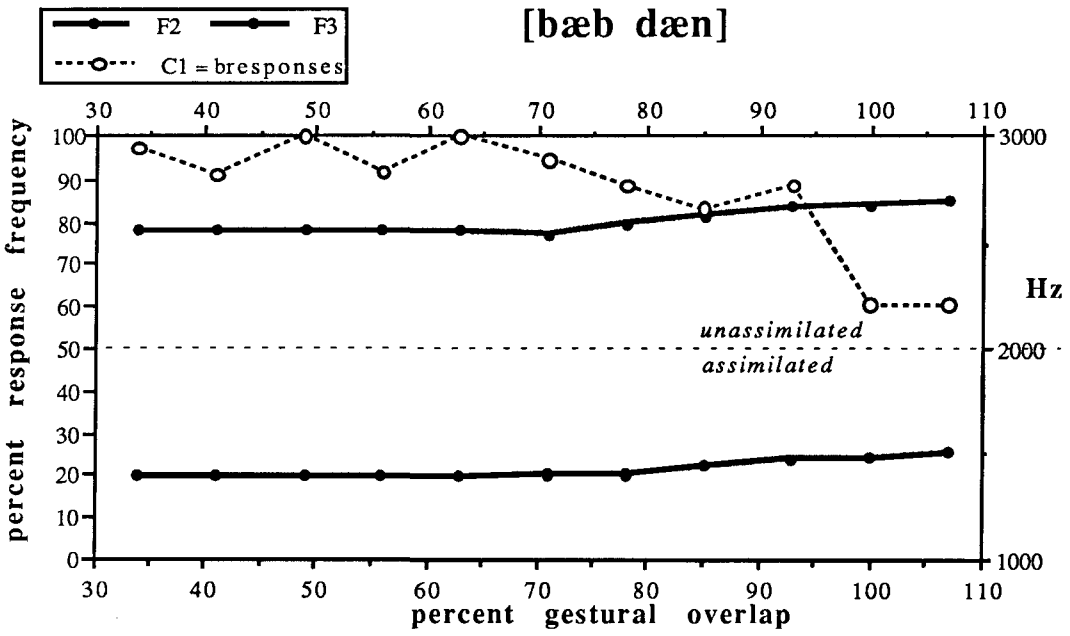
The subjects were faced with a forced-choice paradigm. They were presented with an answer sheet listing the relevant possibilities (four phrases in the full presentation condition, two words in the truncated condition) and were asked to circle or underline their answers.

In both conditions, the least overlapped phasing (ie. 34% overlap/gestural activation phasing of [C1=340 C2=115]) of each possible answer was played five times as a training token on which the subject was informed of the corresponding correct answer. This was to ensure that the somewhat unnatural quality of the synthesized speech was not too distracting for the listener.

The heavy solid lines in Figures 2 and 3 show changes in the stimuli's VC offset formant frequencies with respect to consonant cluster phasing and place of articulation for

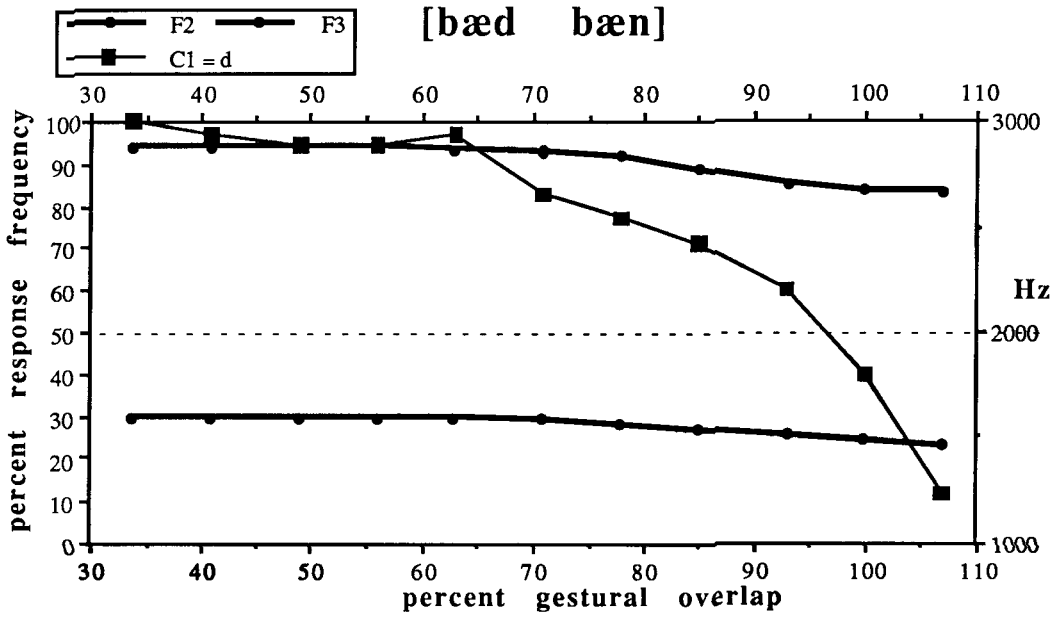
the [bd] and [db] cluster. (The formant values for the [bb] and [dd] clusters were the same across all stimuli, ie. they did not change with respect to the consonant cluster phasing.) The values represented are the vowel F2 and F3 formant values immediately before closure, as the effects of the overlapped C2 gesture were presumed to be greatest here.

Figure 2



Two-word presentation--Results for [bæb dæn]

Figure 3



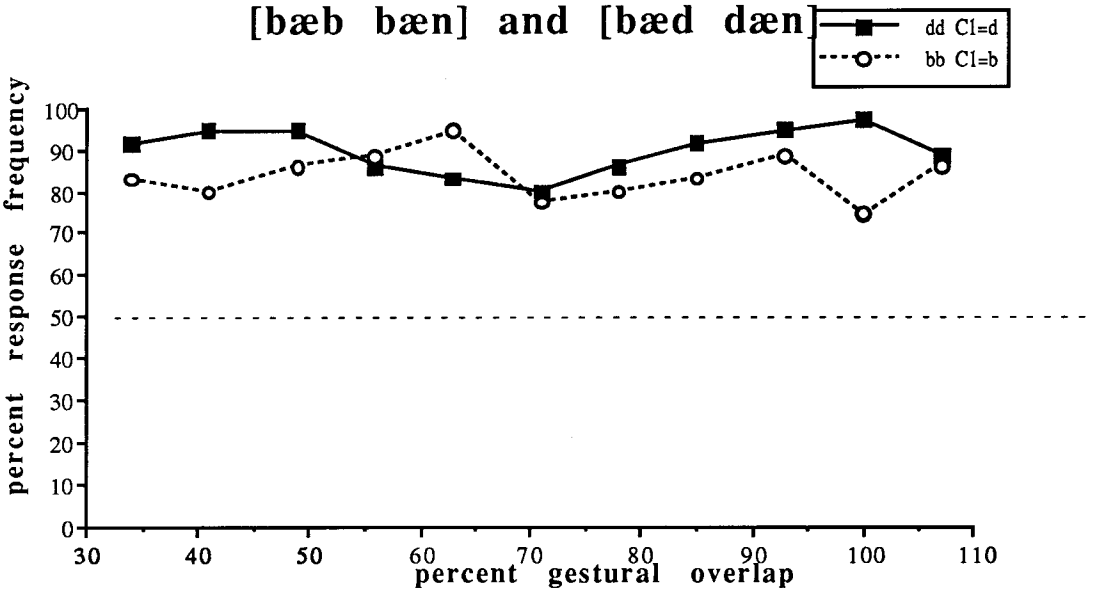
Two-word presentation--Results for [bæd bæn]

3. Analysis and Results

The results are described in terms of totalled subject identifications of C1 with varying degree of consonant gestural overlap for each of the four consonant cluster environments. The results for the two-word condition are shown graphically in Figures 2 through 4.

The two control conditions composed of homorganic stops show only small variations in perception as gestural overlap increased. (Figure 4)

Figure 4



Two-word presentation--Results for [bæb bæn] and [bæd dæn]

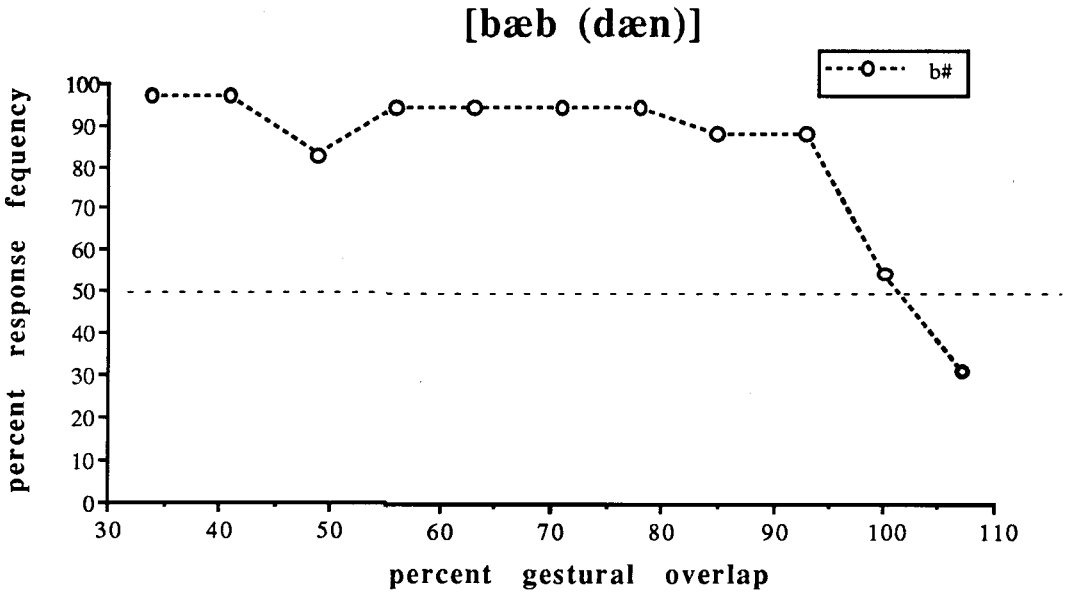
They were generally perceived with about an 80% accuracy which was maintained through all phasings. There is a slight increase in "wrong" responses with both homorganic clusters at the intermediate phasings; a result not seen in the truncated condition. Presumably, this is because at this point the closure interval is too long for a single gesture but too short for two gestures (such as required for two successive homorganic stops). This would also explain the asymmetry of [dd] having this intermediate area beginning at slightly longer closure intervals than [bb], as the labial gesture is slower. The [dd] tokens might have been considered ambiguous at shorter closure intervals than the [bb] tokens for this reason.

The two nonhomorganic phrases (Figures 2 and 3) both demonstrated a change in percept dependent on gestural overlap. For [bd] the percent correct, ie. C1=b, responses first falls below 90% at a gestural overlap of 78%. Note that this is where correct responses also decline slightly in the [bb] cluster, suggesting the slight ambiguity due to closure interval noted above. However, the assimilated responses do not go above 20% until there is a 100% overlap in activation periods. With this consonant order where the bilabial closure precedes the alveolar, there is never a complete crossover in the percept of C1. That is, the perception of C1 as alveolar never becomes more frequent than its perception as bilabial, although the frequency of response does become close.

This is in contrast to the [db] consonant cluster order, where the perception of an assimilated C1 becomes *more* frequent than the unassimilated perception at approximately 95% overlap of the gestures. The frequency of the assimilated response also rises rapidly from the phasing of 63% overlap. At the most overlapped phasing, this [db] consonant cluster had an 88.5% chance of being perceived as having a bilabial C1. In contrast, although some perception of assimilation occurred with the [bd] cluster, at this same phasing there was only a 40% response frequency identifiying an alveolar C1.

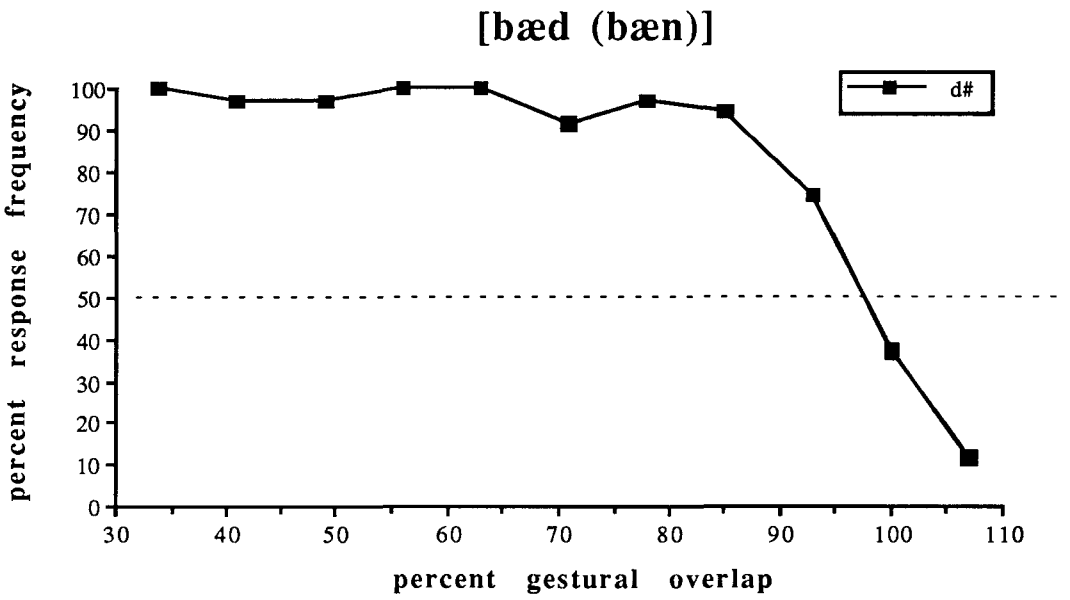
In the truncated condition, where subjects were only presented with the offset formant frequencies of the VC1 closure, the results were similar. They are presented in Figures 5 through 7.

Figure 5



Truncated Presentation--Results for [bæb (dæn)]

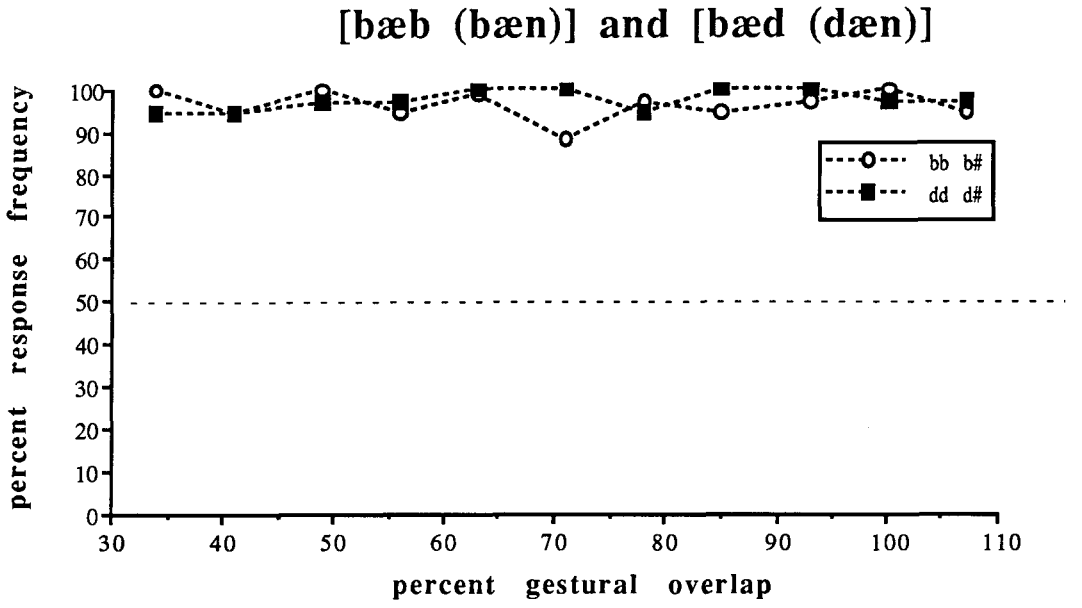
Figure 6



Truncated Presentation--Results for [bæd (bæn)]

The only difference, aside from more consistent overall scores, is that the truncated nonhomorganic clusters required a greater amount of overlap for the percept of the final consonant (C1) to begin to change due to the coarticulatory influence of (the removed) C2 (see Figures 5 and 6). The homorganic clusters underwent no change in perception with respect to gestural overlap (see Figure 7).

Figure 7



Truncated Presentation--Results for [bæb (bæn)] and [bæd (dæn)]

In the [bd] consonant cluster, the perception of the final consonant did cross over from bilabial to alveolar at approximately 102% overlap. The movement toward the percept of assimilation started later (ie. with more gestural overlap), but the end effect was greater than in the two-word condition. In the [db], [alveolar#bilabial] cluster, this crossover, where the perception of the final consonant closure as bilabial became more likely than the perception of it as alveolar, occurred at approximately 97% overlap. Again in the truncated condition the influence of the coarticulated second consonant were perceived earlier, with less overlap, when the bilabial closure was obscuring an alveolar closure than when an alveolar closure overlapped a bilabial closure. These results are discussed below.

4.0 Discussion

The results appear to support the hypothesis that listeners can and do attend to the acoustic consequences of gestural overlap in their perception. The formant transitions of the VC and CV sequences, relative places of articulation, and closure interval were used rather than the presence or absence of a release to determine whether assimilation had occurred. That is, just because there was no release did not indicate to the listener that perceptible coarticulation had occurred. This is of course in line with natural speech perception as consonant clusters without an intervening release are regularly produced, and the accurate perception of both consonants need not be attenuated. (See for example, Wang (1959); Halle, Hughes, and Radley (1957); and Lehiste and Schockey (1972).The fact that the perceptual phenomena described occurred without any acoustic burst information available to the listener confirms that the coarticulatory effects of the two cues of formant

transition and closure interval were sufficient for the perception of assimilation. The nature of both of these cues is the direct consequence of degree of gestural overlap. At low amounts of overlap there was no acoustic effect of C2, and at greater amounts of overlap the acoustic effect of C2 and the total closure interval was prevalent enough to dominate the perception of the place of articulation of the consonant closure.

Not surprisingly, for both consonant clusters a change in responses indicative of assimilation initially appears when the vowel formants preceding the cluster first begin to change due to the co-production of C1 and C2. This can be seen in Figures 2 and 3 where F2 and F3 values for each stimulus are superimposed on the subjects' response functions. The first stimulus in which the "correct", or unassimilated, C1 responses fall below 90 percent for both [bd] and [db] is the first stimulus in which F3 changes by .9 percent from the value for the unequivocal stimuli; the next smaller changes in F3 were .2 percent and .4 percent respectively. There is a direct correspondence between the perception of C1 and the acoustic consequences of the overlap of C2. These acoustic consequences are in turn the direct result of the degree of gestural overlap between the two consonants.

We have also seen two asymmetries in the results. The first is between the full presentation and truncated presentation conditions. The perception of assimilation in the truncated condition requires a greater degree of gestural overlap than in the two-word condition. Second, we find in both conditions an asymmetry between the [bilabial#alveolar] cluster and the [alveolar#bilabial] cluster. The contextual asymmetry and the place of articulation asymmetry will be the focus of the discussion.

With respect to the contextual asymmetry, in the truncated condition, the change to an assimilated response occurs with relatively greater overlap as is shown in Table III.

Table III

Percent overlap at which unassimilated response falls below 90%

	truncated condition	full presentation condition
[bæb dæn]	85% overlap	78% overlap
[bæd bæn]	93% overlap	71% overlap

The acoustic information in the VC1 offset is of course the same in both conditions. We must ask what acoustic information is present in the full condition that is not present in the truncated condition. Of course the larger context, ie. the second word has been removed, but what perceptually relevant acoustic information in determining the first consonant is lost to the listener? Two acoustic cues are lost: the length of the closure interval and the (C2)V onset formant values. It will be argued that the interaction of these two pieces of information are important to the listener in determining place of articulation of C1. First, let us consider the perceptual function of these two pieces of information and then discuss their consequences in the truncated and full presentation conditions.

Ohala (in press-a and -b) has suggested that the common type of consonant assimilation which favors C2, as opposed to a progressive assimilation favoring C1, is due to the perceptual importance placed on the release burst of C2. He argues that the burst is sufficient for determining the place of consonantal closure in the assimilated percept and that the burst is a more reliable and robust cue than the VC1 offset transitions. Of course, the synthesized stimuli which we seek to characterize did not include release bursts. However, other evidence, for example, Whalley and Carrell (1983) and Schatz (1954) show that CV formant transitions are at least as important a cue as the consonantal release bursts. They presented synthesized stimuli with conflicting burst spectra and formant transitions and found that the identification of the consonant corresponded much more reliably to the formant transitions than to the bursts. On the basis of this evidence, we can confidently assert that the type of assimilation favoring C2 in a VC1C2V sequence as described by Ohala and others can also be a function of the salience of the onset formant transitions of the vowel following closure, rather than a sole consequence of place information found in the release burst. This assertion is supported in Repp who concludes:

> While removal of VC transitions from a VCV utterance has a negligible effect, so that their absence can hardly be detected..., removal of the CV transitions dramatically reduced the identifiability of the medial consonant. (Repp, 1978, p. 477)

Because the subjects here did not have this key information in the truncated condition, they would have trouble identifying C2. This thereby created the necessity for a greater amount of overlap before the influence of C2 on V1 was motivating enough for the post-vocalic consonant to be perceived as having the place of articualtion of C2.

There is clearly an interaction between closure interval and these onset transitions. It has been demonstrated by many researchers that if closure interval for a VCCV sequence is reduced significantly, only a single consonant is perceived and that consonant has the place of articulation corresponding to the C2V formant transitions (Repp 1978; Repp 1983b; Lisker 1957; Abbs 1971; Dorman, Raphael, and Liberman 1978; Fujimura, 1975). Repp explains that "the effect may be considered as evidence of perceptual integration over time, with higher weighting of more recent information." (Repp, 1978, p. 473) The listener must be considered to have a perceptual framework in which silence is understood as a linguistically significant piece of information based on their knowledge of the temporal period needed for the articulation of single, geminate, and nonhomorganic consonant closure intervals (Liberman, 1985; Repp 1983b). In the gestural model, *the gestural overlap is the direct cause of changes in temporal length of closure.* Closure length is not independently manipulated as in the experiments of Repp and others.

In the two-word condition, these cues of closure length and CV transition were available to the listener in addition to the VC offset, thereby allowing gestural overlap to create an environment conducive to the perception of assimilation. At large amounts of overlap, the closure interval became short enough that the listener perceived only a consonant with a single place of articulation. In such cases the formant transitions into the vowel of the second word indicated to the listener that this consonant had the place of articulation of C2. In the truncated condition, however, these cues were not available. The listener had only the effects of coarticulation as evidenced in the VC transitions. Due to the paucity of perceptual cues, a relatively greater amount of overlap was necessary before assimilation was perceived, as shown in Table III. That is, without the other cues the

effects of coarticulation on the vowel offset formants needed to be greater than when all three cues were present.

The interaction of vowel offset, closure interval, and vowel onset cues account for the asymmetry found in these data between the truncated and two-word conditions. When all three cues are present, gestural overlap will produce the perception of assimilation with less overlap than if only vowel offset is presented as in the truncated condition. Furthermore, in accordance with the research of others and the observed tendencies of natural language assimilation, this model yields the direction of assimilation of C2 dominating C1 (Repp 1978; Repp 1983; Lisker 1957; Abbs 1971; Dorman, Raphael, and Liberman 1978; Fujimura 1975; Ohala 1989 and in press-a). The results support the likelihood that in natural speech such cases of regressive assimilation can be due to the overlap in consonant cluster articulation.

Now consider the second asymmetry in the results: the constriction location effect. The second asymmetry was between the [alveolar#bilabial] consonant order and the [bilabial#alveolar] consonant order. Recall that we do see a crossover in response frequency for [db] in the truncated condition but not in the two-word condition. In both the truncated and full conditions, the bilabial closure was more effective at obscuring an alveolar closure than the reverse. The results show an earlier crossover point, ie. the point at which an assimilated cluster became more readily perceived than an unassimilated cluster, for the [db] cluster than for the [bd] cluster. In fact, in the two-word condition for [bd], *no* crossover occurred. This result is seen Figure 2 and Table IV.

Table IV

Stimuli at which assimilated response has > 50% frequency
and V(C) formant information

	truncated condition	two-word condition
[bæb dæn]	107 % overlap	no crossover
% change in F3	5.4%	
[bæd bæn]	100 % overlap	100 % overlap
% change in F3	7.6%	7.6%

The explanation for the acoustic differences and their perceptual consequences can be found in the nature of the physical system modeled by the articulatory synthesizer. In this case, subjects are responding to a difference in the acoustic signals as seen in the vowel formant values produced by the articulatory model. Recall that gesture organization is an abstract representation and that the realized movements of individual articulators are influenced by additional, physical considerations.

A bilabial C2 exerts an acoustic influence on the preceding vowel at less overlap than an alveolar C2 (see Figures 2 and 3). When the two consonant gestures are activated simultaneously (100% overlap), the vowel offset formants correspond more closely to bilabial offset formants than alveolar offset formants. These differential effects are demonstrated in Table V.

Table V

Vowel One Offset

	total overlap	single bilabial closure	single alveolar closure	
F2	1483	1397	1601	
F3	2676	2560	2882	Hz

At total gestural overlap, F2 is 6.2% different from the unadulterated bilabial value and 7.4% different from the unadulterated alveolar values, while F3 differs from the bilabial value by 4.5% and from the alveolar value by 7.1%. The bilabial closure is dominating the formant values. The alveolar closure, though, still exerts some influence; the formant values are not exactly the same as those found preceding a single bilabial closure.

The place of articulation of C2 of course also has a differential effect on the V2 onset. As was the case with the V1 offset, the bilabial closure dominates the alveolar in the case of the V2 onset as well. In Table VI, one can see that the vowel onset following the [db] cluster is more greatly affected by the overlapping bilabial gesture than the vowel onset following the [bd] cluster is by the overlapping alveolar gesture. The [db] cluster at 100% gestural overlap precedes an onset in which F2 differs by only 1.8% from the unadulterated bilabial value and F3 by 5%. Following the [bd] cluster at this overlap, the F2 value differs by 5.1 % and the F3 value by 5.2% from the unadulterated alveolar onset values.

Table VI

Vowel Two Onset -- [db]

	total overlap	bilabial onset - no overlap	
F2	1504	1485	
F3	2753	2622	Hz

Vowel Two Onset -- [bd]

	total overlap	alveolar onset - no overlap	
F2	1492	1573	
F3	2707	2854	Hz

When considered in light of the importance of CV onset formants after short closure intervals as described above, this acoustic behavior in the case of the [bd] cluster accounts for the absence of crossover in the two-word condition.

One must consider why the labial gesture dominates the signal when the gestures are activated simultaneously, ie. co-produced. It might be hypothesized that this is because a bilabial gesture is created anteriorly to an alveolar gesture. However, the fact that the bilabial closure is more frontal is not precisely the significant aspect creating its dominant effect in a coarticulated cluster. In examining subjects' production of sentences containing consonant clusters in which a second, or incipient, constriction co-occurs with the C1 constriction, Zsiga and Byrd (1990) found that the vowel formants preceding a coarticulated [d#k] were a statistically more accurate predictor of C2 than those preceding a [d#b] cluster. Clearly the relative anteriority of C2 was not the determining characteristic. Instead, the velar constriction was shown to have a strong effect on V1 because even a small, secondary constriction in the velar region creates large changes in the formant offset values of a preceding vowel in a [Vd#k] sequence. In the clusters tested for perception of assimilation here, a secondary constriction in the labial region creates a greater change in alveolar offset formants than a secondary constriction in the alveolar region creates in a primarily bilabial offset. The further distinction of stifffness between the labial and tongue tip articulators and the effect of this distinction will be discussed below.

4.1 The experimental results and natural speech

If we wish rely on our results as being reflective of perception of assimilation in natural speech, we must establish a degree of accord between the gestural model which generated the experimental stimuli and the characteristics of articulation in natural speech. Only then can we confidently compare our experimental results with certain phenomena in natural speech. Nolan rejects the use of an articulatory synthesizer as an experimental tool for producing stimuli with adequate realism to test the perception of assimilation (Nolan,

1989). Admittedly, Nolan's ideal of a "fully-perfected, comprehensive, and tractable articulatory synthesizer" (Nolan, 1989, p. 5) is not available. However, the discussion below will attempt to briefly show that a high enough level of correspondence exists between natural articulation and articulation as characterized by the gestural model for conclusions drawn from these experimental results to be extended to a discussion of assimilation in natural speech.

We must confirm that the gestural model used in this experiment implements gestures in a way conforming to implementation by human speakers. One key element in this implementation relevant to these results is the stiffness or velocity of the articulators. In the gestural model, a labial gesture is released more slowly than is a tongue tip gesture because the tongue tip has a higher stiffness or frequency of oscillation causing it to return more quickly to its equilibrium position. This difference in articulator velocity is supported by experimental results with real speakers. Kuehn and Moll (1976) found in cineradiographic studies of articulatory velocity that "the tongue tip generally moved faster than either the lower lip or tongue dorsum in both coordinate [maxillary and mandibular] systems." (Kuehn and Moll, 1976, p. 309) In their comparison of articulatory velocity in relation to consonant phone types and transition types (VC_1VC_2 yielding VC_1, VC_2, and C_1V transitions), alveolar closure was consistently faster than bilabial closure (Kuehn and Moll, 1976, p. 312). They additionally note that "the rank ordering of velocities for different phone-to-phone gestures is maintained regardless of a change in speaking rate." (Kuehn and Moll, 1976, p. 313) Two older studies of diadochokinesis also report results showing the tongue tip to have the fastest articulator velocity. Hudgins and Stetson found that "[t]he tip of the tongue is obviously the fastest member in all cases [out of the set of five tested articulators]." (Hudgins and Stetson, 1937, pp. 92-93) Lundeen's diadochokinetic results show a higher mean number of syllables per second for both [də] and [tə] than for either [pə] or [bə] (as well as for all of the ten other tested syllables), although the difference was not found to be statistically significant (Lundeen, 1950, p. 56).

The second crucial element in the construction of the experimental stimuli is the use of co-production to model assimilation. The other more common model of total assimilation is that characterized by many feature geometry analyses. The gestural overlap model makes a prediction which is not made by theories relying on the delinking or deletion of a feature or set of features. Browman and Goldstein's gestural model characterizes the perceptual deletion such as that occurring in [pɚfəkt ˊmɛmɚɹi] ---> [pɚfəkˊmɛmɚɹi] as a case of increased gestural overlap which produces perceptual hiding or complete assimilation of the alveolar gesture (Browman and Goldstein, 1989b). The gestural analysis of overlap predicts that the gesture is not deleted but rather that it remains physically and is only acoustically unrealized. A feature geometry type analysis considers the set of features representing the alveolar gesture to be delinked. Browman and Goldstein note: "To interpret a delinked gesture as one that is articulatorily produced, but auditorily hidden, would require a major change in the assumptions of the [feature geometry] framework."(Browman and Goldstein, 1989b, p. 220) Ohala and Nolan offer similar objections. Ohala emphasizes that "[c]laims that features can be unlinked have not been made with any evident awareness of the full phonetic complexity of speech, including not only the anatomical but also the aerodynamic and the acoustic-auditory principles governing it." (Ohala, 1989, p. 26) Nolan voices the concern that "a phonological notation of this kind is still too bound to the notions of discreteness and segmentality to be appropriate for modelling the detail of assimilation." (Nolan, 1989, p. 11) The prediction that gestures do not necessarily disappear articulatorily when they become perceptually

absent has in fact been evaluated using X-ray microbeam data obtained from a speaker producing the utterance above [pɚfək'mɛmɚɹi], noted above (Browman and Goldstein, 1987). Browman and Goldstein conclude:

> ...the alveolar closure gesture at the end of 'perfect' was produced in the fluent context, with much the same magnitude as when the two words were produced in isolation, but it was completely overlapped by the constrictions of the preceding velar closure and the following labial closure. Thus, the alveolar closure gesture was acoustically hidden. (Browman and Goldstein, 1989b, p. 215)

In response to Nolan's concerns, Hayes has in fact suggested a phonological rule of assimilation which retains feature spreading but does not include any delinking, thereby yielding a "complex" segment (Hayes, 1989) which still retains some specification of the place of the assimilated segment. However, he firmly maintains the existence of a separate phonetic component of the grammar which "translates the autosegments of the phonology into quantitative physical targets" (Hayes, 1989, p. 5). As this solution seems to be merely shifting the "problem" of the gradient nature of the assimilation phenomena to another component of the grammar, we will conclude that "a [single] dynamical description *simplifies* the relation between categorial and continuous characterizations of articulation, which is desirable from both a practical and theoretical perspective" (Browman and Goldstein, 1990, in press-b; emphasis added).

Again we see that the implementation of gestures used in the model for this experiment is in accordance with articulator movement in natural speech. The gestural model offers the additional theoretical advantage of representational simplicity because assimilation is modeled as a direct consequence of the input representation rather than any rule, action, or condition altering the input representation. If the claims made by the gestural analysis can in fact be supported by the relevant data as seems to be the case, this is evidence for the correctness of analyzing assimilation as the result of a unitary articulatory process of gestural overlap (Browman and Goldstein, 1989b).

Nolan (1989) raises two additional objections to a model of assimilation as co-production such as that offered by Fowler (1985, p. 254). First, he says, "it is not clear how the continuum [ie. gradual nature] of articulation-types [observed in assimilation sites] could be explained mechanically."(Nolan, 1989, p. 12) Clearly the gestural model of coarticulation is aimed at achieving exactly this mechanical gradation of co-production. The manipulation of gestural overlap in this experiment is used to determine the perceptual effect of such gradation of coarticulation. Secondly, Nolan remarks that cross-linguistic variation in place assimilation must be taken into account by a model relying on the actions of the vocal mechanism. The gestural model allows this because canonical phasing rules are clearly language-specific. The gestural analysis actually makes some predictions with respect to cross-language differences in casual speech processes. It claims that, with the same increase in overlap during casual speech, a language having a very small degree of canonical overlap will show fewer assimilations and deletions than a language which has a greater canonical overlap (Browman and Goldstein, 1989b). This claim could be evaluated by a comparative examination of casual speech processes in languages with systematically released stops, such as Georgian (Browman and Goldstein, 1989b; Anderson, 1974) and a language with comparatively more articulatory overlap such as English.

Having briefly addressed the correspondence between the gestural model and real speaker articulation as well as the theoretical advantages offered by the model to the linguist, a comparison of the experimental results with observed phenomena in natural speech is in order.

Different types of place assimilation do not occur with equal regularity in language. In many languages, partial and complete assimilation favoring velars and labials over alveolars occurs frequently. English, for example, shows patterns such as those in Table VII:

Table VII

"good morning" ---> [gəmoɹnɨn]
"fixed cars" ---> [fɪkskɑɹs]
"dust broom" ---> [dəsbɹum]
"last girl" ---> [læsgɚl]

However, the reverse is more uncommon. Similarly, "One might ask why labial-velar or velar-labial assimilations [as opposed to assimilations involving alveolars] do not occur (at least not frequently), given the possibility of their overlapping...A possible explanation lies in the fact (Kuehn and Moll, 1976) that tongue tip movements show higher velocities than do either tongue dorsum or lip movements (which are about equivalent to each other). A slower movement might prove more difficult to hide" (Browman and Goldstein, 1987, p. 18). In the experimental results described above, the bilabial release succeeds more effectively in producing an assimilated consonant cluster, while obtaining this effect with an alveolar consonant in second position requires a more substantial change in the formant transitions--ie. greater amount of gestural overlap. If co-produced, the faster gesture is indeed more readily hidden due to overlap with the slower gesture, than is the slower gesture equally overlapped by the faster one. In the gestural model of assimilation, the acoustic information specific to the overlapping gesture has perceptually replaced the acoustic information specific to the overlapped gesture. The asymmetrical perceptual assimilation of alveolar consonants found in the experimental results may be due to both the acoustical consequences of the dynamic characteristics of the gestures as well as the differing magnitude of formant transitions found for alveolar and bilabial closures; alveolars as a rule showing larger formant transitions than bilabials. The experimental data are reflective of assimilation processes seen in natural speech.

The gestural overlap model may provide a useful tool with which to consider the types of multiply-articulated segments in the world's languages. Earlier, the function of co-production for the listener was debated. It was suggested that at certain degrees of gestural overlap, coarticulation allowed parallel transmission whereby information about an upcoming event whose primary acoustic and articulatory correlates occur later in time is available to the listener. We have seen that at large amounts of gestural overlap assimilation may occur, causing a perceptual loss, ie. the perceptual information specific to a phonological segment ceases to be effectively transmitted. This range of variability of the

consequences of co-production is a result of the gradient nature of gestural overlap possible within the framework of the linguistic gestural model. This question of recoverability must be addressed when considering multiply-articulated segments. Such segments are linguistically simultaneous and not sequential articulations, although such articulations may involve an asynchrony in timing to ensure that information about both places of articulation are available to the listener in order to distinguish them from singly articulated stops (Maddieson and Ladefoged, 1989). They are distinguished from stop sequences by "the principle that single complex segments have duration comparable to that of simple segments of the same phonetic class in the same environment" (Maddieson and Ladefoged, 1989, p. 130) and by the absence of release bursts for the individual constrictions (Maddieson and Ladefoged, 1989). These characteristics suggest that such segments as [k͡p] are produced with some type of gestural overlap (although I will not try to formulate precisely how such segments would be represented within the gestural framework in terms of stiffness, phasing, etc.). Crucially, Maddieson and Ladefoged (1989) note that such segments are limited to two constriction locations and that "no segments with more than two primary linguistically relevant articulations" are possible (Maddieson and Ladefoged, 1989, p. 117). They note that the "phonetic cues involved [in multiply-articulated segments] cannot be used effectively to signal the presence of three or four simultaneous articulations" (Maddieson and Ladefoged, 1989, p. 117).

We have seen, in the case of *perfect memory* for example, that a sequence of three gesturally overlapped stops may result in the perceptual loss of one of the segments. However, because each constriction location corresponds with a phonologically independent segment and because the segments are associated with a word in the listener's lexicon, the listener will be able to "reconstruct" the utterance based on his or her linguistic competence. However, in the case of a multiply-articulated segments, a constituent constriction location consistently unassociated with any overt acoustic representation in the signal will not be recoverable by the listener. This situation would arise because, in gestural terms, the gestures composing such segments are highly overlapped, and the acoustic realization of all three constriction locations would be impossible. Decreasing overlap would result in the loss of the distinction between such multiply-articulated segments and the similar stop sequences by yielding releases and/or increasing duration beyond that typical of a similar single segment in a like environment. Clearly phonologically contrastive segments must involve perceptual contrasts. Such contrasts are impossible with triple-articulations, and this fact is reflected by the absence of such segments in the world's languages. The gestural overlap mechanism producing assimilation in a sequence like *perfect memory* can be extended to help explain the impossibility of a gestural organization yielding perceptually salient information associated with all three constriction locations within a multiply-articulated segment whose component gestures are simultaneous or highly overlapped.

Interestingly, Maddieson and Ladefoged (1989) comment that Yeletnye is the only language of which they are aware that has phonetic segments with simultaneous labial and coronal closures--it has labial-alveolar and labial-post-alveolar stops and nasals. The rarity of such multiply-articulated segments as compared to labial-dorsal and nasal double articulations could perhaps be accounted for by the acoustic results shown in Tables VI and VII and the consequent perceptual dominance of the labial element observed in this experiment's simultaneously articulated [bd]/[db] clusters. Maddieson and Ladefoged go on to note that two languages, Dagbani and Nzema, have [t͡p] as an allophone of another phoneme. This occurs in Dagbani and Nzema before front vowels for /k͡p/ (Wilson &

Bendor-Samuel 1969, Ladefoged 1968, Berry 1955, Chinebuah 1963), and as a variant of /t^{w}/ for at least some speakers of Abkhaz (Catford 1972, 1977, and Khaidakov 1955). However, in Dagbani and Nzema the allophone is predictable. Photographs taken by Catford of an Abkhaz speaker show that the labial contact is "light" and involves "considerable forward protrusion of the lips and the contact is between the surfaces"--quite distinct from the normal /p/ articulation. (Maddieson and Ladefoged, 1989, p. 133) Maddieson and Ladefoged comment that the only other languages frequently cited as having labial-alveolars, the two Chadic languages of Margi and Bura (Maddieson and Ladefoged, 1989; Hoffman 1963, Ladefoged 1968, Halle 1983),

> ...on closer examination prove to have labial+alveolar sequences rather than double articulations (Maddieson, 1983, 1987). For example, /bd/ is considerably longer than /b/ or /d/ and *a labial release can be detected well before the alveolar one*...(emphasis added)(Maddieson and Ladefoged, 1989, p. 133)

Abkhaz is the only case of a labial-coronal segment other than Dagbani, Nzema, and Yeletnye. This additional case of Abkhaz has the characteristic of rounding which would provide distinguishing cues in the acoustic signal to differentiate these segments from a "normal" bilabial stop or a [bd] sequence such as occurs in Bura and Margi. The rarity of labial-coronal multiply-articulated segments might be explained in part by the experimental results obtained in this study in that simultaneously activated bilabial and alveolar gestures showed an acoustic dominance of the bilabial gesture in both the preceding and following vowels (see Tables V and VI). This might make such multiply-articulated segments less distinct from a purely bilabial constriction than is desirable for the phonetic inventory of a language.

Similarly, the experimental results obtained here may help explain the distribution of secondary vowel articulations in the world's languages. We have seen that vowel formant transitions are not only smaller for alveolar closures than for labial closures, but that they are also more affected by an overlapping labial constriction than an overlapping alveolar constriction, although both gestures may be occurring simultaneously, see Tables V and VI. A minor change in position of the tongue tip does not change the perceived vowel quality as significantly as a secondary constriction in another region. The asymmetrical sensitivity of vowels to secondary constrictions in the labial and velar regions versus in the alveolar region is evidenced by the fact that vowels are often velarized or rounded/labialized but they are not "alveolarized," ie. they do not occur with a secondary articulation at the alveolar ridge; Ladefoged (1982, pp. 210-212) lists only palatalization, velarization, pharyngealization, and labialization as secondary articulations. In fact, secondary constrictions as vocalic modifications never seem to involve any exclusively tonge-tip secondary articulation.

The implementation of the gestures in the experimental stimuli is reflective of articulation in natural speech, and the perceptual results seen in this experiment are in accordance with types of assimilation seen in natural speech, as well as facts of multiply-articulated segments and secondary vowel articulations. These facts allow the confident assertion that the gestural model of coarticulation indeed has explanatory value in describing the relationship between representation, articulation, and perception in assimilation processes.

4. Concluding Remarks

In summary, these results demonstrate a correlation between the perception of assimilation and the acoustic consequences of gestural overlap within consonant clusters. In addition to the influence of coarticulation on the preceding vowel offset formants, it was also found that gestural overlap has a perceptually significant effect on closure interval which also contributed to the perception of assimilation. It was suggested that the dominance of C2 in the perception of an assimilated consonant cluster sequence can be attributed to the perceptual weight accorded the CV onset formants. The perceptual significance of closure interval together with vowel onset formants were demonstrated by the attenuated perception of assimilation in the truncated condition in which these cues were not present. We have seen that the presentation of the stimuli in the full, two-word context facilitates the perception of assimilation rather than increasing the likelihood that the relevant acoustic information will be filtered out or perceptually neutral. Lastly, we have attempted to explain an asymmetry in the results which pointed to the perceptual dominance of a bilabial closure over an alveolar closure in this data. This asymmetry is seen in the diminished perception of assimilation in the [b#d] cluster as compared to the [d#b] cluster. We suggested that the acoustic behavior of bilabial and alveolar overlapped constrictions differ in the response of the vocal tract to these places of constriction during vowel production due in part to the greater intrinsic stiffness or velocity of the tongue tip gesture. The asymmetry obtained in this perceptual experiment was found to conform to assimilation processes in natural connected speech.

I have proposed viewing assimilation in connected speech as a process which can be modeled in terms of degree of coarticulation and have further suggested the reasons why this is a correct model of assimilation phenomena in natural speech. The percept of a consonant cluster will depend on the amount of gestural overlap in its production. This overlap will interact with the physical characteristics of the vocal tract to determine the acoustic make-up of the utterance. This acoustic information in conjunction with the listener's linguistic knowledge of speech is what the listener will use in forming his or her perception. With small amounts of gestural overlap such as found in all normally spoken speech, the acoustic consequences of gestural overlap will not be substantial enough to create a perception of an assimilation. If gestural overlap increases, as occurs sometimes in fast or casual speech (Zsiga and Byrd, 1990; see also Barry, 1985 and Kerswill 1985), the listener may then deem the set of acoustic cues powerful enough to indicate that assimilation has taken place. The gestural overlap model is able to characterize this gradient nature of assimilation as a casual speech process and yield perceptual results in agreement with previously observed phenomena in natural speech. Crucially, the gestural model adds to our understanding of the complex interactions of articulator movements and temporal organization of articulatory gestures.

References

Abbs, M. H. *A study of cues for the identification of voiced stop consonants in intervocalic contexts.* Unpublished Ph.D. dissertation, University of Wisconsin, 1971.

Anderson, S. R. *The Organization of Phonology.* New York:Academic Press, 1974.

Barry, Martin. A palatographic study of connected speech process. *Cambridge Papers in Phonetics and Experimental Linguistics*, 4:1-16, 1985.

Berry, Jack. Some notes on the phonology of the Nzema and Ahanta dialects. *Bulletin of the School of Oriental and African Studies*, 17:160-165, 1955.

Browman, Catherine and Louis Goldstein. Tiers in articulatory phonology, with some

implications for casual speech. *Haskins Laboratories Status Report on Speech Research*, SR-92, 1987. to appear in J. Kingston and M.E. Beckman (eds.), *Papers in Laboratory Phonology I: Between the Grammar and the Physics of Speech.* Cambridge: Cambridge University Press.

_____. Some notes on ayllable atructure in articulatory phonology. *Phonetica*, 45:140-155, 1988.

_____. "Targetless" schwa: An articulatory analysis. *Proceedings of the Second Conference of Laboratory Phonology*, Edinburgh, Scotland, 28 June 1989 - 3 July 1989, (in press-a).

_____. Articulatory gestures as phonological units. *Phonology* 6,2:201-252, 1989b.

_____. Gestural specification using dynamically-defined articulatory structures. to be published in *Journal of Phonetics*, special issue on Phonetic Representation, 1990, (in press-b).

Catford, John C. Labialization in Caucasian languages, with special reference to Abkhaz. *Proceedings of the Seventh International Congress of Phonetic Sciences*, ed. A. Rigualt and R. Charbonneau. Mouton, The Hague:679-682, 1972.

_____. Mountain of tongues: The languages of the Caucasus. *Annual Review of Anthropology*, 6:293-314, 1977.

Chinebuah, Isaac K. The category of number in Nzema. *Journal of African Languages*, 2:244-259, 1963.

Dorman, Michael F., Lawrence J. Raphael, and Alvin M. Liberman. Some experiments on the sound of silence in phonetic perception. *Journal of the Acoustical Society of America.* 65 (6):1518-1532, June 1979.

Fant, G. Some remarks from the viewpoint of speech research. *Speech Motor Control: Proceedings of an International Symposium on Speech Motor Control*, held at the Wenner-Gren Center, Stockholm, May 11 and 12, 273-277, 1981.

Fowler, Carol A. Production and Perception of Coarticulation among Stressed and Unstressed Vowels. *Journal of Speech and Hearing Research*:127-139, 1981.

_____. Current perspectives on language and speech production: a critical overview. In R. Daniloff (ed.), *Speech Science: Recent Advances* San Diego, Ca.:College-Hill, 193-278, 1985.

Fujimura, O. *A look into the effects of context--some articulatory and perceptual findings.* Paper presented at the 8th International Congress of Phonetic Science, Leeds, England, 1975.

Halle, Morris. On distinctive features and their articulatory implementation. *Natural Language and Linguistic Theory*, 1:91-105, 1983.

Halle, M., G. W. Hughes, and J. P. A. Radley. Acoustic properties of stop consonants. *Journal of the Acoustical Society of America*, 29:107-116, 1957.

Hardcastle, W. and R. Morgan. An instrumental analysis of articulation disorders in children. *British Journal of Disorders of Communication*, 6:47-65, 1982.

Hayes, B. Assimilation as spreading in Toba Batak. *Linguistic Inquiry*, 17:467-499, 1986.

_____. Comments on the paper by Nolan. *Proceedings of the Second Conference of Laboratory Phonology*, Edinburgh, Scotland, 28 June 1989 - 3 July 1989 (in press).

Hoffman, Carl. *A Grammar of Margi.* Oxford University Press, Oxford, 1963.

Hudgins, C. V. and R. H. Stetson. Relative speed of articulatory movements. *Archives néerlandiaises de Phonétique experimentale*, 13:85-94, 1937.

Kaisse, Ellen M. and Patricia A. Shaw. On the theory of lexical phonology. *Phonology*

Yearbook, 2: 1-30, 1985.
Kelso, J. A. S. and B. Tuller. Intrinsic time in speech production: Theory, methodology, and preliminary observations. In E. Keller and M. Gopnik, eds. *Sensory and Motor Processes in Language*. Hillsdale, NJ:Erlbaum.
Kerswill, P.E. A sociophonetic study of connected speech processes in Cambridge English: an outline and some results. *Cambridge Papers in Phonetics and Experimental Linguistics*, 4:1-39, 1985
Khaidakov, Said M. Balkharskij Dialekt Lakskogo hazyka. *Trudy Instituta Jazykoznanija*, 3, Akademija Nauk S.S.S.R., Moscow, 1955.
Kuehn, D.P. and K. Moll. A cineradiographic study of VC and CV articulatory velocities. *Journal of Phonetics*, 4:303-320, 1976.
Ladefoged, Peter. *A Phonetic Study of West African Languages (second edition)*. Cambridge University Press, Cambridge, 1968.
_____. *A Course in Phonetics (Second Edition)*. Harcourt Brace Jovanovich, Inc., 1982.
Lehiste, I. and L. Schockey. On the perception of coarticulation effects in English VCV syllables. *Journal of Speech and Hearing Reasearch*, 14:500-506, 1972.
Liberman, A. How Abstract Must a Motor Theory of Speech Perception Be? *Haskins Laboratories Status Report on Speech Research*. SR-44:1-15, 1975.
_____. Reading Is Hard Just Because Listening Is Easy. *Haskins Laboratories Status Report on Speech*. SR-95/96:145-150, 1988.
Liberman, A., F. Cooper, D. Shankweiler and M. Studdert-Kennedy. Perception of the speech code. *Psychological Review* 74:231-436, 1967.
Liberman, A. and I. G. Mattingly. The motor theory of speech perception revised. *Cognition*, 21:1-36, 1985.
Lindblom, B. The interdisciplinary challenge of speech motor control. *Speech Motor Control: Proceedings of an International Symposium on Speech Motor Control*, held at the Wenner-Gren Center, Stockholm, May 11 and 12, 1981.
Lisker, L. Closure duration and the intervocalic voiced-voiceless distinction in English. *Language*. 33:42-49, 1957.
Lundeen, C. W. The relationship of diadochokinesis to various speech sounds. *Journal of Speech and Hearing Disorders*, 15:54-59, 1950.
Maddieson, Ian. The analysis of complex phonetic elements in Bura and the syllable. *Studies in African Linguistics*, 14:285-310, 1983.
_____. The Margi vowel system and labiocoronals. *Studies in African Linguistics*, 18:327-355, 1987.
Maddieson, Ian and Peter Ladefoged. Multiply-articulated segments and the feature hierarchy. *UCLA Working Papers in Phonetics*, 72:116-138, March, 1989. expanded version of a paper presented at the 1988 Annual Meeting of the Linguistic Society of America, New Orleans, December 27-30, 1988.
Mattingly, I.G. and A. M. Liberman. Specialized perceiving systems for speech and other biologically significant sounds. In G. M. Edelman, W. E. Gall, and W. M. Cowan (eds.), *Functions of the auditory system*. New York:Wiley, (in press-a).
Mattingly, I.G. and A. M. Liberman. Speech and auditory modules. In G. M. Edelman, W. E. Gall, and W. M. Cowan (eds.), *Signal and sense: Local and global order in perceptual maps*. New York:Wiley, (in press-b).
Nolan, Francis. The Descriptive role of segments: Evidence from assimilation. *Proceedings of the Second Conference of Laboratory Phonology*, Edinburgh, Scotland, 28 June 1989 - 3 July 1989 (in press).
Ohala, John. The phonetics and phonology of aspects of assimilation. to appear in J.

Kingston and M.E. Beckman (eds.), *Papers in Laboratory Phonology I: Between the Grammar and the Physics of Speech.* Cambridge: Cambridge University Press, (in press-a).

_____. The segment: primitive or derived? *Proceedings of the Second Conference of Laboratory Phonology*, Edinburgh, Scotland, 28 June 1989 - 3 July 1989, (in press-b).

Repp, Bruno. Perceptual integration and differentiation of spectral cues for intervocalic stop consonants. *Perception and Psychophysics*. 24 (5):471-485, 1978.

_____. Perception and Production of Two-Stop-Consonant Sequences. *Haskins Laboratories Status Report on Speech Research.* SR-63/64:177-194, 1980.

_____. Coarticulation in sequences of two nonhomorganic stop consonants: Perceptual and acoustic evidence. *Journal of the Acoustical Society of America*, 74 (2):420-427, August, 1983a.

_____. Bidirectional contrast effects in the perception of VC-CV sequences. *Perception and Psychophysics*, 33 (2):147-155, 1983b.

Rubin, P., T. Baer and P. Mermelstein. An articulatory synthesizer for perceptual research. *Journal of the Acoustical Society of America.* 70:321-328, 1981.

Saltzman, E. Task dynamic coordination of the speech articulators: A preliminary model. In H. Heuer and C. Fromm eds. *Generation and modulation of action patterns (Experimental Brain Research Series* 15). New York:Springer-Verlag, 1986.

Saltzman, E. and J. A. S. Kelso. Skilled actions: A task dynamic approach. *Psychological Review*, 94:84-106, 1987.

Saltzman, E., L. Goldstein, C. P. Browman and P.E. Rubin. Dynamics of gestural blending during speech production. Presented at First Annual International Neural Network Society (INNS), Boston, September 6-10, 1988.

Saltzman, E. and K. G. Munhall. A dynamical approach to gestural patterning in speech production. *Ecological Psychology*, 1:333-382, 1989.

Scharf, Donald J. and Ralph N. Ohde. Physiologic, acoustic, and perceptual aspects of coarticulation : Implications for the remediation of articulatory disorders. *Speech and Language*:153-247, 1981.

Schatz, C. D. The role of context in the perception of stops. *Language*, 30:47-56, 1954.

Wang, W. S. Transition and release as perceptual cues for final plosives. *Journal of Speech and Hearing Research*, 2:66-73, 1959.

Walley, Amanda and Thomas Carrell. Onset spectra and formant transitions in the adult's and child's perception of place of articulation in stop consonants. *Journal of the Acoustical Society of America*, 73 (3):1011-1022, March, 1983.

Wilson, W. A. A. and John T. Bendor-Samuel. The phonology of the nominal in Dagbani. *Linguistics*, 52:56-82, 1969.

Zsiga, E. and D. Byrd. Acoustic evidence of overlap in consonant sequences. presented at the 118th meeting of the ASA, San Diego, 1990.

Announcements

Bibliography

We have started using the Macintosh program EndNote to compile the bibliographies used in our papers. This is a very useful program that organizes bibliographic data so that references can be printed out in almost any style. It also forms a valuable database, which can be searched in different ways. (For example one can ask for a list of all entries containing the words 'formant' and 'vowel' in the title, or containing'stop' and not 'nasal'.) So far we have about 1,500 entries in the database, but there are many omissions which we are attempting to remedy. We would like to share these with others who are also using EndNote. We will send anyone who sends us an EndNote bibliography a merged version of our bibliography and theirs.

A Course in Phonetics

A third edition of *A Course in Phonetics* is in preparation. Comments, criticisms and suggestions based on the present edition will be welcome. Anyone who sends in a copy of the present edition so that I can see any marginalia can expect not only a grateful acknowledgement but also their copy returned, a new copy when the third edition appears, and anything I can supply from our UCLA software. Changes now scheduled include (1) deleting most of the last chapter and replacing it by a review of much of the material in the book; this would be done by reference to recent publications of the International Phonetic Association; (2) changing to current IPA symbols throughout (mainly using [ɪ, ʊ] for [ɪ, ʊ] in the transcriptions of English, substituting the current click symbols [ǀ, ǁ, ǃ] for [ɪ, ɔ, ʗ], and the IPA retroflex symbols [ʈ, ɖ, ɳ, ʂ, ʐ] for [ṭ, ḍ, ṇ, ṣ, ẓ]; and (3) altering the illustrative material in the tables to bring it into line with that in the UCLA Hypercard database, The Sounds of the World's Languages.

But what else should be done? P.L.