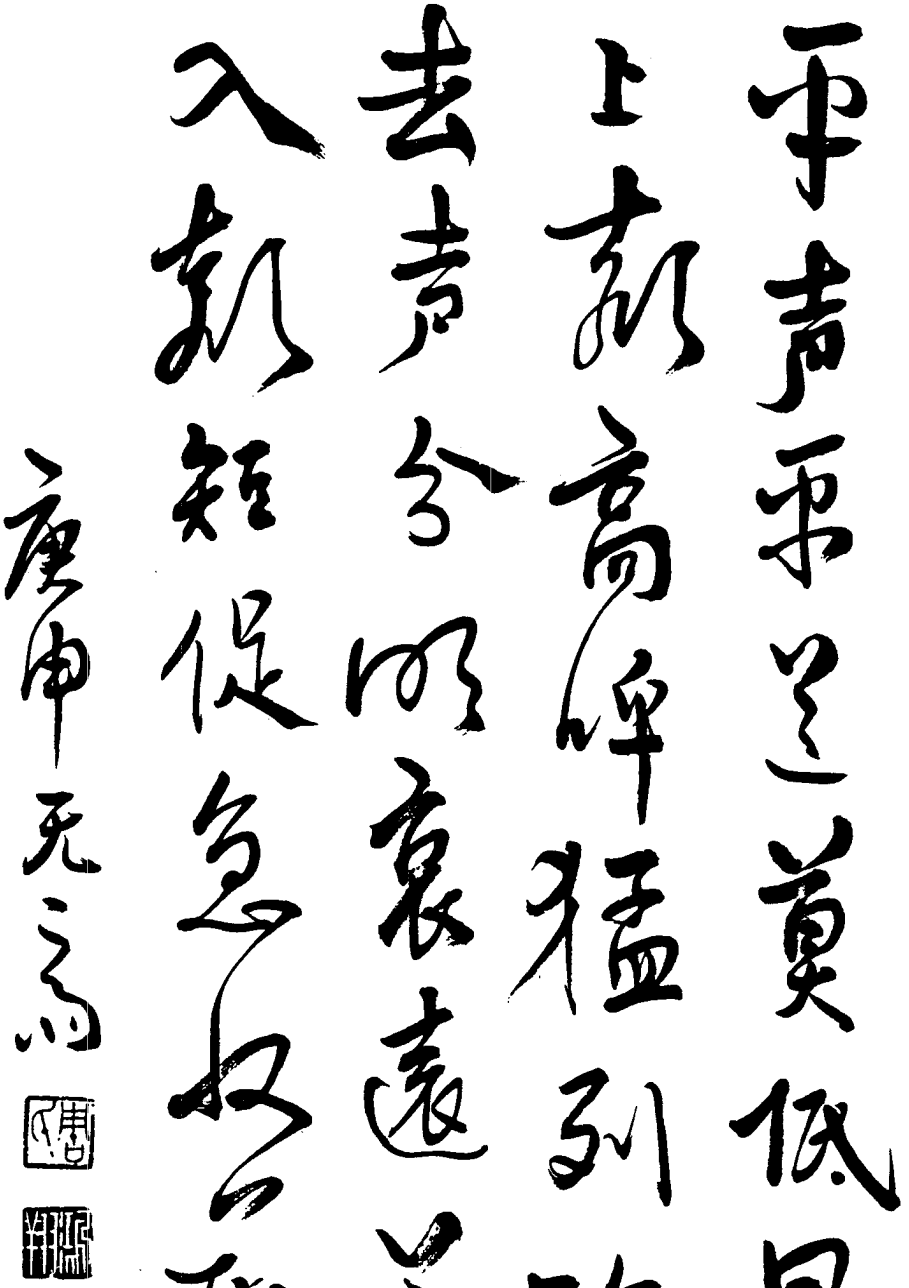


UCLA

wpp 49

APRIL 1980

COVER

Calligraphy by Professor Hung-Hsiang Chou,
Department of Oriental Languages, UCLA

Verses from the Kang-Xi Dictionary of the
Qing Dynasty (1644-1911 A.D.)

Translation from Medhurst's Hok-kien
Dictionary (columns from right to left):

1st: "The Even Tone travels on a level road, neither elevated nor depressed"
2nd: "The High Tone exclaims aloud, being fierce, violent, and strong"
3rd: "The Departing Tone is distinct and clear, gruffly travelling to a distance"
4th: "The Entering Tone is short and contracted, being hastily gathered up"

UCLA Working Papers in Phonetics 49

April 1980

The UCLA Phonetics Laboratory Group

Ron Anderson
Pat Coady
Sandy Ferrari Disner
Jim Fordyce
Vicki Fromkin
Manuel Godinez
Steven Greenberg
Richard Janda
Hector Javkin
Peter Ladefoged
Mona Lindau-Webb
Wendy Linker
Ian Maddieson
Willie Martin
Jonas Nartey
George Papçun
Lloyd Rice
Diane Ridley
Renee Wellin
Anne Wingate
Andreas Wittenstein
Eric Zee

As on previous occasions, the material which is presented here is simply a record for our own use, a report as required by the funding agencies, and a preliminary account of work in progress.

Funds for the UCLA Phonetics Laboratory are provided through:

USPHS grant NS 09780
NSF grant BNS78-07680
and the UCLA Department of Linguistics

Correspondence concerning this series should be addressed to:

Phonetics Laboratory
Department of Linguistics
University of California, Los Angeles
Los Angeles, California 90024

The phonetic inadequacy of phonological specifications of clicks

Peter Ladefoged and Tony Traill

Clicks, like other sounds, may be described in terms of phonological features, as has been shown by Chomsky and Halle (1968), Jakobson (1968) and Ladefoged (1971, 1975). In their discussions of clicks all these authors have been concerned mainly with specifying the phonological oppositions; none of them tried to account for the phonetic facts in any detail. When we examine the ways in which clicks are actually made in different languages, we find that there are a number of difficulties in expressing the phonetic facts in terms of current feature theories.

It might be argued that feature theories are not intended to allow for the description of *all* the phonetic details that characterize a language. But there is at the moment no other way of providing this information. As Chomsky and Halle, 1968, say: "Given the surface structure of a sentence, the phonological rules of the language interact with certain universal phonetic constraints to derive *all grammatically determined facts about the production and perception of this sentence*" (my italics, SPE 293). They make it clear that the physical properties of speech that are *not* taken into account by a phonological description are things "such as the voice pitch and quality of the speaker and also such socially determined aspects of speech as the normal rate of utterance and what has been called by some writers the 'articulation base' .. In addition, phonetic transcriptions [i.e. feature matrices] omit properties of the signal that are supplied by universal rules." (SPE 295). But they note in discussing phonetic differences between languages: "The representation [i. e. feature matrices] must differ [if] the distinction is determined in part by language-specific rules." (SPE 298). Ladefoged has also expressed a similar view: "The systematic phonetic level of description may be said to be that level which specifies all the targets [feature values] necessary for the description of a particular language as opposed to all other languages, but contains no information of the kind that is used simply to specify one speaker of that language as opposed to other speakers." (Ladefoged 1972).

The first problem in accounting for all the phonetic details of a language arises from the fact that all the authors cited above claim that each of their phonological features is a physical scale. They support this position by giving what appear to be definitions of features in physical phonetic terms. But when we try to use these definitions to state measurements of feature values, we find that in most cases it is impossible.

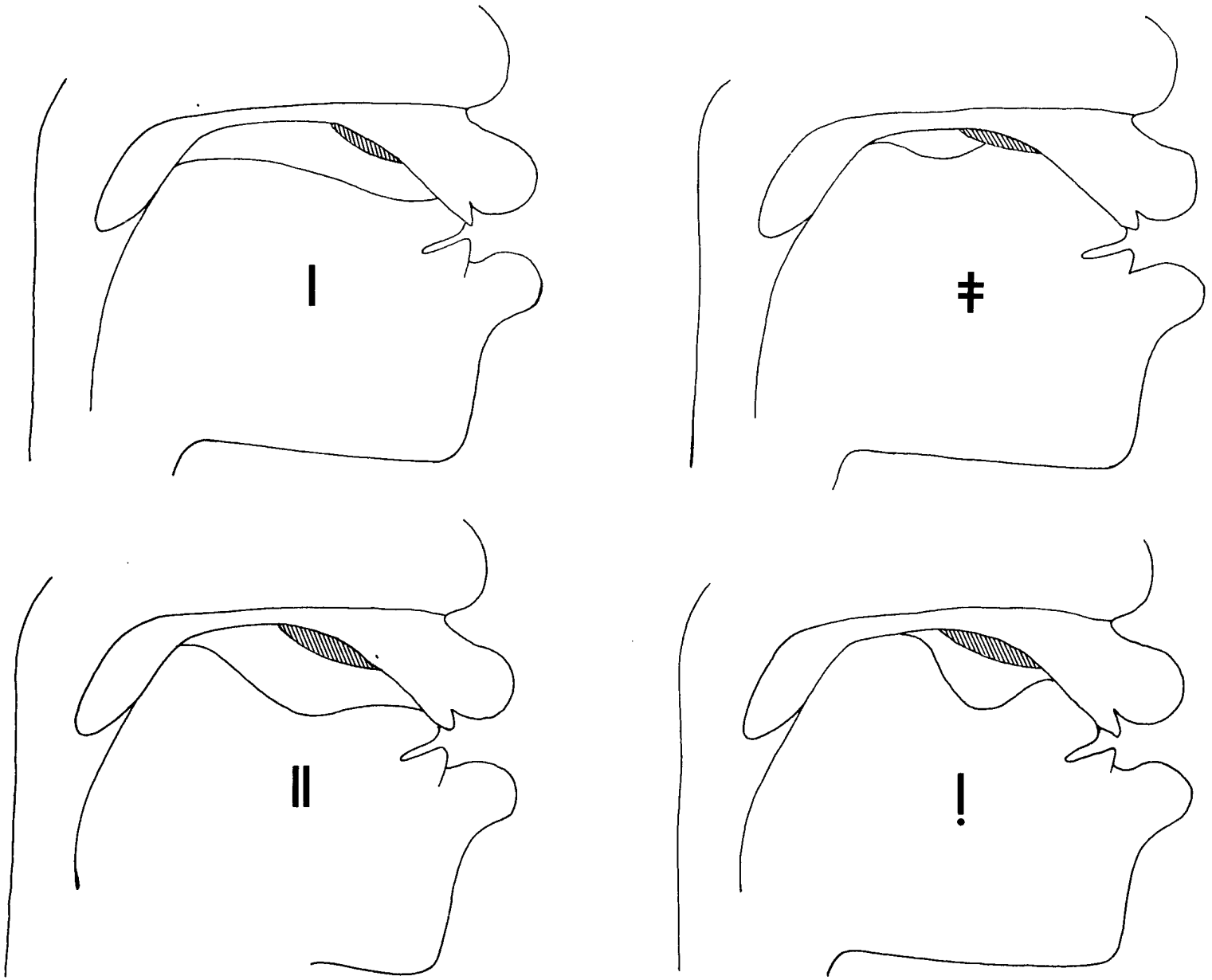


Figure 1. Tracings based on selected frames from cineradiology data, showing the smallest cavity enclosed by the tongue (shaded area) and the largest cavity (lower tongue line) just before the release of the click.

Consider for example, the feature Coronal, which is defined in terms of the raising of the blade of the tongue from its neutral position. If we want to assess whether a given sound is coronal or not we should presumably look at tracings of mid-sagittal x-rays, such as those in figure 1. These four diagrams show the crucial articulatory positions in the production of four clicks which will be discussed in detail later.

Each of the four diagrams shows two positions of the tongue, one in which the tongue is raised so that only in the small shaded area is there no contact between the center of the tongue and the roof of the mouth, and the other, just before the release of the click, in which the body of the tongue has been lowered to produce a suction chamber. Now, if coronal is a scale, we should be able to assess not just whether a sound is coronal or not, but how coronal it is. When faced with data as in figure 1, we find it difficult to define an algorithm that will enable us to do this. We do not know the precise point on the tongue to measure, or when to measure it.

According to Chomsky and Halle 1968: 320, all four of these sounds can be classified phonemically as [+ coronal]. The two sounds in the top row are also classified as [+ anterior], whereas those in the bottom row are [- anterior]. There are no other differences in place of articulation. The sounds on the left are distinguished from those on the right by being [+ delayed primary release] as opposed to [- delayed primary release]. The sound [ǁ] is also [+ lateral].

Chomsky and Halle do not attempt to give for these, *or for any other segments*, a systematic phonetic description of how coronal (or how anterior) they are. But if we are to describe fully the phonetic characteristics of languages we must be able to characterise in physical terms all the consistent differences in place of articulation among sounds. There are obvious differences in the tongue positions that are not specified by the classificatory descriptions. The two [- anterior] sounds in the bottom row have very different tongue positions at the time of release of the click, and the two [+ anterior] sounds in the upper row are different throughout the articulation. If we are operating in terms of a standard generative phonology, the feature specifications must be able to characterise these differences precisely, not just in some vague hand-waiving way indicating that it should be possible to interpret the feature values appropriately. The onus is on proponents of feature theories to show excactly how diagrams such as those in figure 1 can be generated from their specifications.

A similar point can be made with reference to many other features. Thus lateral, a seemingly straightforward feature, is always defined in some way such as "lowering the mid section of the tongue at both sides or at only one side" (Chomsky and Halle 1968: 317). But, again, how does one *measure* this? The part of the tongue involved is very different for a dental lateral as opposed to a palatal lateral. And how about "Rounded sounds are produced with a narrowing of the lip orifice" (Chomsky and Halle 1968: 309)? There are many different ways in which the lip orifice can be narrowed. We agree that we want a single feature, Rounded, to cover them all; but we do not want to regard it as a single physical scale.

The point we are making is not that the definitions of the features are wrong, but that it is inappropriate to expect those features which are necessary for classifying phonological categories to be equally valid for specifying phonetic details in physical terms. There is no reason to presume that phonological features are in a one to one relation with physical scales. Human beings typically classify objects in terms of several physical parameters simultaneously. Thus objects are called heavy or light not just in accordance with their mass, but also in accordance with their size and appearance. It is, in fact, difficult to find perceptual attributes that depend on a single physical scale. The pitch of a sound depends on its intensity and overtone structure, as well as on its fundamental frequency. Color and other aspects of visual perception are all complex phenomena. The linguist's notion that features are elementary properties is a piece of retrograde psychology. Each feature is associated with variations in many phonetic parameters. Furthermore, as the examples cited have shown, the phonetic interpretation of a feature depends on the values of other features that occur at the same time. We always have to map phonological features onto phonetic scales on a many to many rather than a one to one basis.

A second weakness of contemporary linguistic practice is that there are phonetic details which are characteristic of one language in comparison with another, but which are not phonologically relevant in any single language. For example, as has been shown elsewhere (Ladefoged 1980), some languages have a series of laryngealized ('creaky voice') stops [b',d',g'] and others have a series of voiced implosives [ɓ,ɗ,ɠ]; but no language makes a contrast between these two possibilities. From a phonological point of view all these sounds can be classified simply as [+ glottal]. As a result, phonetic differences of this kind are just left out of descriptions that are concerned only with phonological patterns.

A third difficulty for phoneticians who wish to describe languages fully is that standard phonological feature theory does not provide a way of specifying differences in timing. This problem was pointed out many years ago by Fant (1967); but generative phonologists still specify linguistic events simply in terms of sequences of segments. Thus Ladefoged 1975, working along lines proposed by Chomsky and Halle, attempted to describe words in terms of segments composed of features, each of which has a certain value on a physical scale. In these descriptions the values of a feature change from one segment to the next. But this view makes no provision for characteristic differences in timing; it does not provide a way of showing that the value of a phonetic parameter may sometimes change at one moment in a segment, and in another circumstances may change at another. Williamson (1976) has suggested a way of incorporating some variations in timing into phonological specifications of complex items such as affricates. But no one has shown how to specify phonetic timing details such as those we will be exemplifying in our description of clicks.

These three problem areas -- the many to many relation between phonological features and phonetic parameters; the possibility of overlooking phonetic characteristics because they are not contrastive; and the difficulty of specifying differences in timing -- will be exemplified by

reference to the phonology and phonetics of the clicks in two Khoisan languages. We will consider data from one Khoi language, Nama, and one San language, !Xóõ.

Nama

The sounds of Nama have been well described by Beach (1938). It was this description that was used as a basis for interpretations in distinctive feature terms by Chomsky and Halle (1968) and Jakobson (1968). The shortcomings of these interpretations have been discussed at length elsewhere (Traill forthcoming), and will not be considered in detail here; let it suffice to say that they are both procrustean attempts to fit this language into categories that are totally inadequate when one comes to consider a more complex click system such as that of !Xóõ.

Nama has 20 clicks as shown in (1), which gives essentially the same information as may be found in Beach's summary (p. 89), but expressed in terms of what are now more familiar symbols and labels. At this stage in the discussion the labels have been kept as simple as possible. There are four primary types of articulation, for each of which there are five clicks differing in the accompanying actions of the vocal cords and other articulators. Words illustrating these contrasts are given in (2); they are arranged in the same order as in (1), but in the current Nama orthography. (We are very grateful to our Nama consultant, Mr. Johannes Boois, for his skill in finding suitable words, and for his assistance in providing us with instrumental data).

(1) The 20 clicks of Nama.

Accompaniment → / Primary Articulation ↓	voiceless unaspirated k	voiceless aspirated kh	delayed aspiration h	voiced nasal ŋ	glottal closure ʔ
| Dental	k|	k|h	|h	ŋ|	|ʔ
ǂ Palatal	kǂ	kǂh	ǂh	ŋǂ	ǂʔ
! Alveolar	k!	k!h	!h	ŋ!	!ʔ
|| Lateral	k||	k||	||h	ŋ||	||ʔ

(2) Words in the standard Nama orthography illustrating the 20 clicks in the same order as in (1). All these words have a high tone.

|goa	|kho	|ho	|no	|o
(put into)	(play an instrument)	(push into)	(measure)	(sound)
ǂgais	ǂkharis	ǂhais	ǂnais	ǂais
(calling)	(small one)	(baboon's arse)	(turtle dove)	(gold)
!goas	!khoas	!hoas	!noras	!oas
(hollow)	(belt)	(narrating)	(pluck maize seeds)	(meeting)
‖garos	‖khaos	‖haos	‖naes	‖aos
(writing)	(strike)	(special cooking place)	(pointing)	(reject a present)

We investigated the phonetic nature of the phonological oppositions using a number of different instrumental techniques. We made records of the Nama clicks, showing the expiratory nasal air flow, the expiratory oral air flow, the pressure of the air in the pharynx as recorded via a thin tube inserted through the nose,and the waveform as recorded by a microphone in the oral flow mask. Only the pressure record could be adequately calibrated,using instrumentation and techniques described elsewhere (Ladefoged and Traill, forthcoming); the flow records should be considered simply as indications of the relative rate of flow at different moments within an utterance. Furthermore, it should be noted that the frequency response of the air flow measurement system was not flat. (This is an inherent problem with the particular instrument used, an F-J Electronics Electro-Aerometer.) As a result there is an apparent (but unreliable) variation in air flow when vowels are said on different tones. In addition, as we had only a four channel ink-writer available (a Siemens-Elema Mingograf 34) we were unable to record simultaneously the inspiratory and expiratory nasal and oral flows. However, previous observations had shown us that the inspiratory nasal air flowwas irrelevant to our investigation, and there was only a *very* small inspiratory oral air flow associated with the production of the click. As will become apparent, none of the limitations discussed above affect our findings.

Figure 2 shows the aerodynamic behavior recorded during the production of two Nama phrases. The upper part of the figure shows the aerodynamic behavior recorded during the production of the phrase [nes ge a k‖haos] 'This is striking.' The top trace shows that there is expiratory nasal air flow only before the utterance and (with voice vibrations) during the first consonant. The next trace, the oral air flow,shows the voiced flow

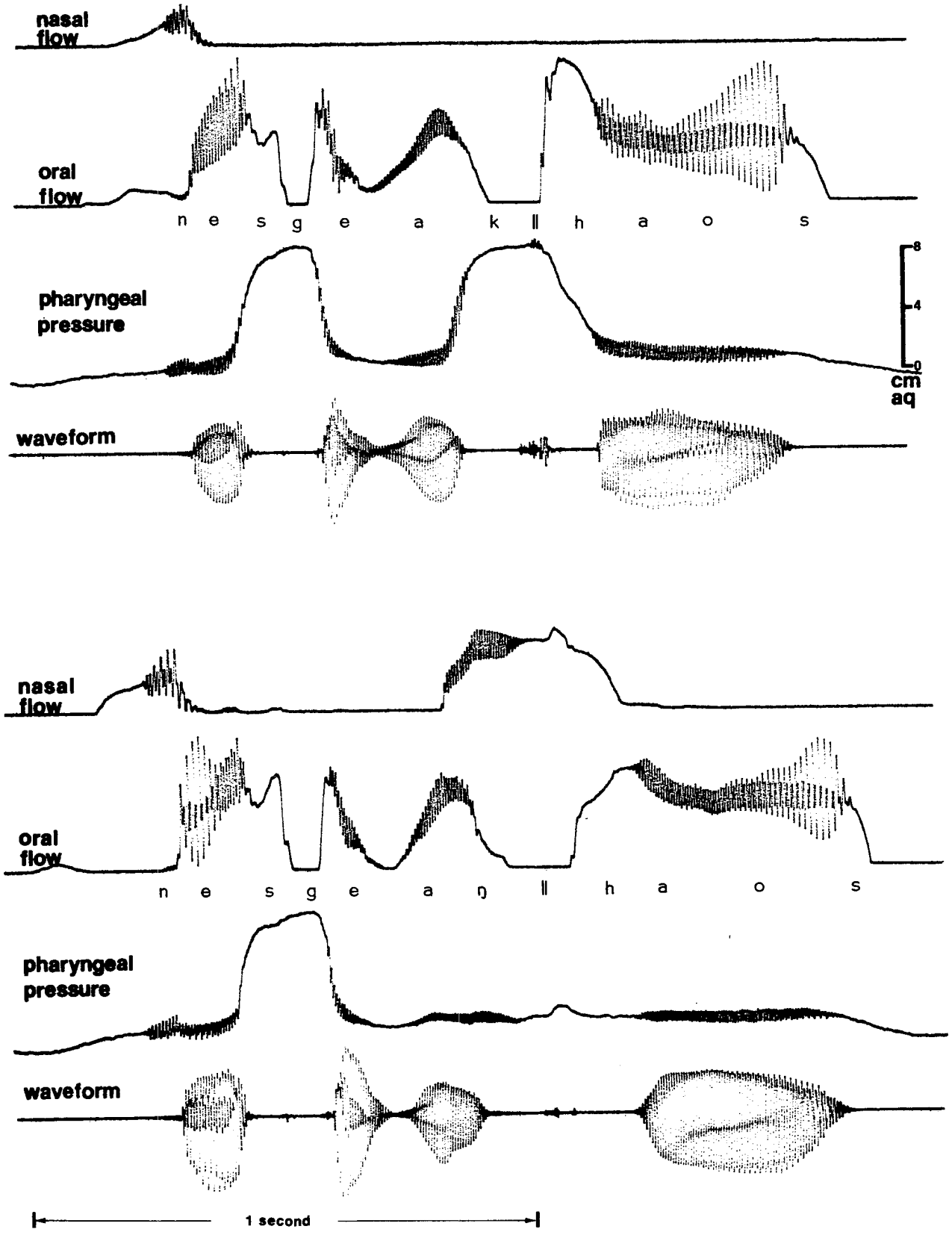


Figure 2. Aerodynamic records of two Nama phrases.

that occurs during the first vowel, the voiceless flow during the fricative [s], no flow during the closure for [g], a sharp increase on the release of this closure, varying amounts during the abutting vowels [e a],no flow during the click closure, a sharp increase in flow after the release of the velar closure flow with voiced vibrations during the vowels,and voiceless flow for the final fricative. The third trace shows that the pressure of the air in the pharynx went up during the fricative [s], remained up for the velar stop [g], showed only voice vibrations during the vowels,increased again during the closure associated with the click (symbolized on the diagram by [k] before the lateral click symbol [ǁ]), fell comparatively slowly at the end of this closure probably due to the slower movement of the back of the tongue associated with the slight affrication present in this release , showed only voice vibrations during the final vowels, and a very small increase for the final [s]. The bottom trace shows the waveform of the sounds, as recorded by a microphone in the oral air flow mask. The burst of noise associated with the release of the click is plainly evident just before the pressure drop that occurs on the release of the velar closure.

The lower part of figure 2 shows the contrasting click [ǁh] which occurs in the phrase [nes ge a ǁhaos] 'This is a traditional cooking place'. Beach (1938) describes clicks such as [ǁh] as being made with delayed aspiration. Figure 2 makes it clear how this delay is achieved. The top line shows that there is voiced nasal air flow not only during the initial nasal consonant, but also as an allophonic nasalization of the vowel before [ǁh]. In fact, after the end of that vowel there is no flow of air through the mouth (as shown by the second line), so there is actually an intrusive voiced velar nasal. It is this nasal escape that stops the pharyngal pressure (shown in the third line) from rising. If it were not for this leak the pharyngal pressure would increase just as it does for [kǁh] in the previous phrase. After the click (which can be seen on the oral waveform record shown in the bottom line), there is an interesting trading relation whereby the voiceless nasal air flow decreases as the oral air flow increases. This is what gives the effect of delayed aspiration.

Figure 3 and 4 show similar records for all the words in (2) pronounced as citation forms. In each case only a part of the vowel after the click consonant is shown. The differences among the four primary articulations are not evident in this type of record, but the differences among all five click accompaniments are readily apparent.

The voiceless unaspirated clicks in the first column have increases in pharyngal pressure duing the click closure comparable to those in the aspirated click [kǁh] which has already been discussed (and which is further exemplified in the second column of figure 4). The clicks in the first column differ from those in the second by having the voice onset occur almost

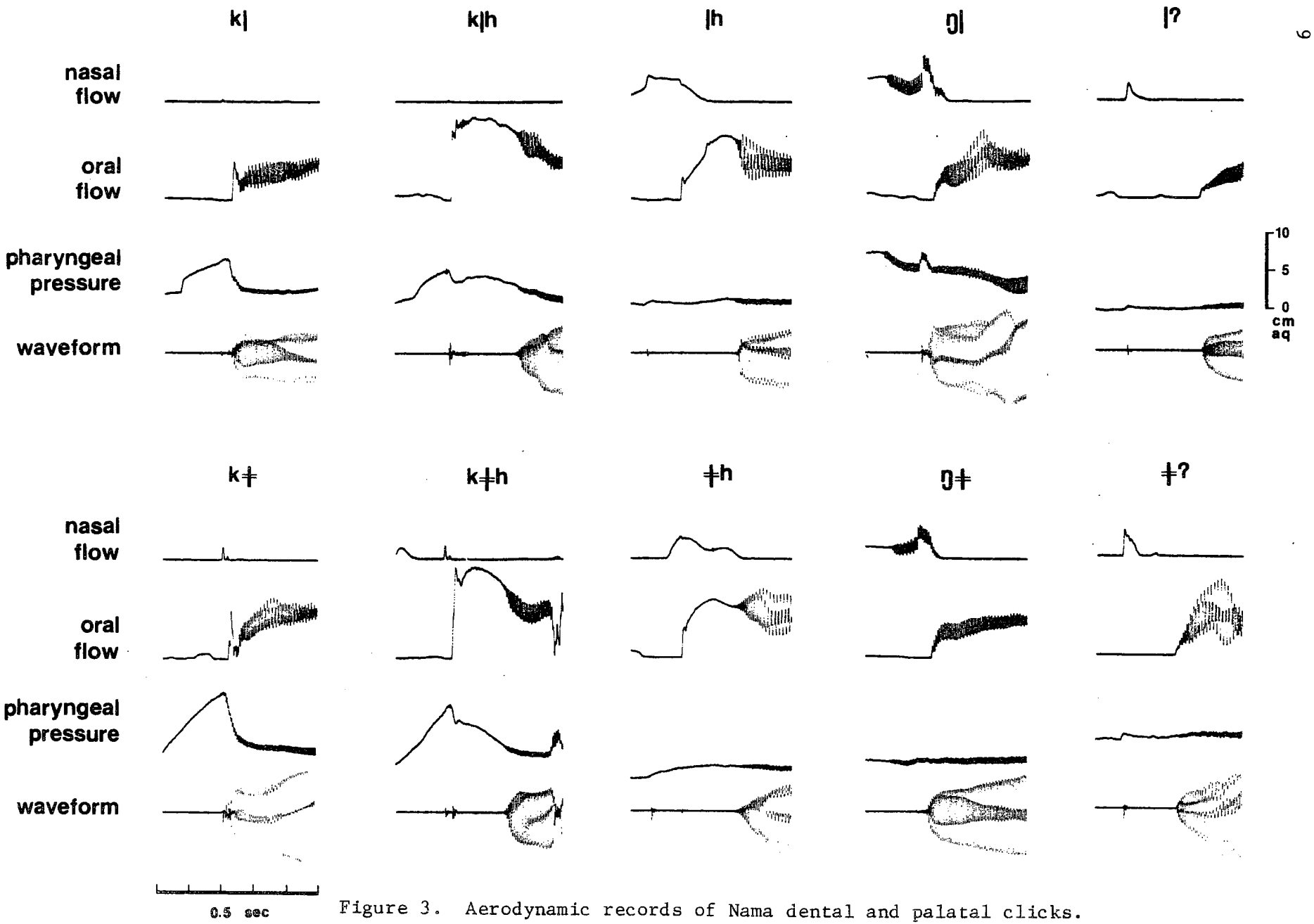


Figure 3. Aerodynamic records of Nama dental and palatal clicks.

immediately after the release of the velar closure. Note also that the pharyngeal pressure falls very rapidly immediately after the release of of the velar closure in the case of the unaspirated clicks in the first column. There is a slower rate of fall in the case of the aspirated clicks in the second column, which is due to the tongue moving away from the roof of the mouth slightly more slowly. The aspirated clicks tend to be affricated after the release of the velar closure.

The third column shows the clicks that have delayed aspiration. The examples in figures 3 and 4 again demonstrate the way in which the nasal air flow prevents the major increase in pharyngal pressure that would otherwise occur. As before, when the velar closure is released there is no sharp increase in oral air flow. Instead the flow of air from the mouth increases slowly as the nasal air flow decreases, so that the oral airflow is sufficiently rapid to cause audible aspiration for only the last part of the period after the click release.

The fourth column shows the nasal click accompaniment, which occurs both before and after the sound of the click. The oral air flow usually starts soon after the click, so that there is what might be regarded as a short nasalized vowel. Before the release of the click the sound is equivalent to a velar nasal consonant.

The clicks that have an accompanying glottal stop are shown in the final column. During the glottal closure there is (naturally) no increase in pharyngal pressure. But there is a very interesting nasal air flow at the release of the click. The voiceless nasal release accompanying the click is a phonetic detail that must be noted in a full description of this language - as, indeed, it was by Beach (1938) in his masterful account. It is possible that it is caused by a raising of the closed larynx while the velum is lowered. The records shown in figures 3 and 4 are typical of the range of nasal emission used by this speaker. The dental, alveolar and palatal clicks all have a considerable amount of nasal air flow at the time of release of the click. The lateral click in the lower part of figure 4 has only very little nasal emission, which is just observable in the original record. We have aerodynamic records of 44 clicks with glottal closure produced by this speaker, and on 90% of them the nasal air flow at the release of the click is of the same order of magnitude as that in the dental, alveolar and palatal clicks in figures 3 and 4. When there is very little nasal air flow it is probably due to a lack of larynx movement rather than the presence of a velic closure. There is no doubt that nasalization is a regular phonetic characteristic of this segment in this language, as Beach correctly noted.

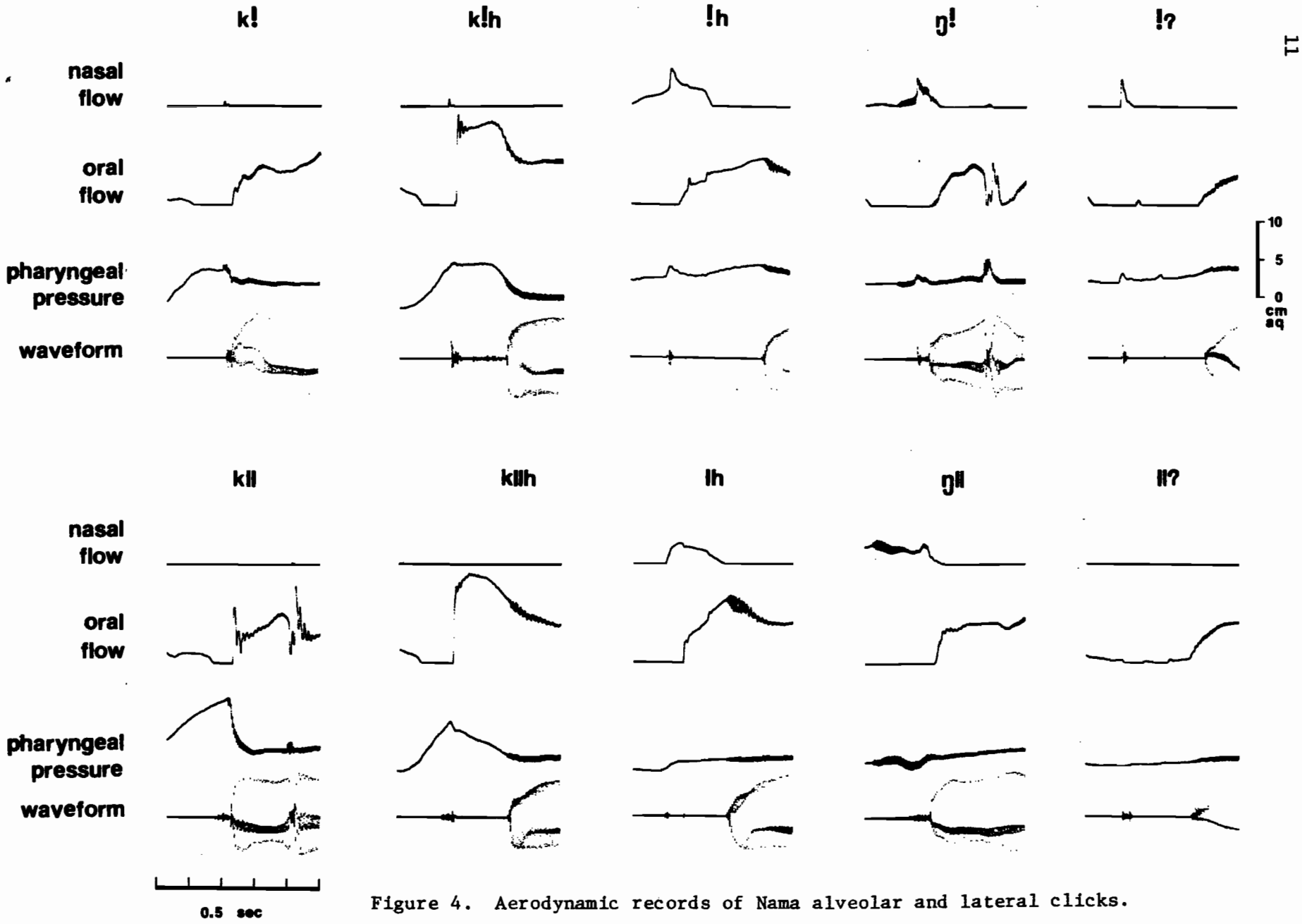


Figure 4. Aerodynamic records of Nama alveolar and lateral clicks.

We will now consider how these sounds should be represented in phonological terms. We assume that the reasons for wanting to describe the segments of Khoisan languages in terms of features are those given for the use of features in most current phonologies. First we want to distinguish the segments in each language by means of different sets of values of phonological features, so that we can specify all the contrasting lexical items. In more old fashioned terms, we want to be able to distinguish between the phonemes. Second we want to be able to group segments together in terms of their feature values so as to be able to give explanatory descriptions of the phonological patterns that occur.

The data illustrating the phonemic contrasts have been given in (1) and (2). We now have to consider what phonological alternations have to be taken into account. As far as we are aware there is only one that provides any problem for the specification of these sounds in terms of features. Relevant data are given in (3).

(3) tĩi 'my' |ʔuip 'brother-in-law'
[tĩiŋ|ʔuip] 'my brother-in-law'
|ʔoa 'full'
[|ʔoãŋ|ʔoa] 'fill up' (reduplicative causative)
|hop 'friend'
[tĩiŋ|hop] 'my friend'
ǂhaba 'flat'
[ǂhabãŋǂhaba] 'flatten' (reduplicative causative)

It appears from these and other examples that clicks with a glottal closure accompaniment and those with delayed aspiration induce nasalization of a preceding vowel and insertion of a velar nasal. Words beginning with a voiceless or an aspirated click do not induce this nasalization. Further evidence of this phenomenon is provided by the instrumental data in figures 2, where the inserted velar nasal -- a period of nasal air flow and no oral air flow -- can be seen very clearly in the example in the lower part of the figure, but not in the example in the upper part. Accordingly it would be appropriate for the classification to indicate that the clicks with delayed aspiration and those with a glottal closure should be grouped together by means of their feature specification. The shared property among clicks such as [ʔ|] and [|h] is a lack of an audible velar stop release, and a lowering of the velum. As we have seen, they have a comparatively low pharyngal pressure due to the escape of air through the nose. The simplest way of specifying these clicks is to regard them as [+ nasal]. Then we can write a rule inserting a velar nasal as in (4), which correctly indicates the assimilatory nature of the alternation illustrated by the data in (3).

(4) $\emptyset \rightarrow \begin{bmatrix} C \\ -\text{ click} \\ +\text{ nasal} \end{bmatrix} / V \_\_ \begin{bmatrix} +\text{ click} \\ +\text{ nasal} \end{bmatrix}$

In deciding which features to use for classifying Nama clicks we will require only that the specifications be phonetically interpretable in some reasonable way (a vague phrase that will allow us to consider a number of different possibilities), and that they are reasonably (again, vague) parsimonious. As we have noted, the particular values of the features must enable us to distinguish the segments that make up the lexical items, and to account for the phonological patterns that occur.

One way of specifying the Nama clicks within these criteria is given in (5). Only the dental clicks are shown, as the values for the other primary articulations are completely comparable. The classifications in

(5)

	k|	k|h	|h	ŋ|	|ʔ
Voiced	+	-	-	+	-
Glottal	-	-	-	-	+
Nasal	-	-	+	+	+

(5) are phonologically adequate, but some linguists might think that they are not phonetically interpretable in a reasonable way. The unaspirated click has been classified as [+ voice]. This is necessary in order to distinguish it from the corresponding aspirated click. The unaspirated click is, however, completely voiceless, so the phonetic specification rules will have to show that when the value [+ voice] occurs in conjunction with the values [+ click, - nasal] it will have to be interpreted as an abduction (opening) of the vocal cords. The nasal clicks are the only ones that are completely voiced so when the value [+ voice] occurs with [+ nasal] it can be given a more usual phonetic interpretation. Of course the value [- voice] also has to be given two interpretations. In conjunction with [- glottal] it implies an open glottis, whereas in conjunction with [+ glottal] it implies a constricted glottis.

A second classificatory system, given in (6), distinguishes between the aspirated and unaspirated clicks by means of a feature Aspiration. This entails a more redundant phonological specification, using four features to

(6)

	k|	k|h	|h	ŋ|	|ʔ
Voiced	-	-	-	+	-
Aspirated	-	+	+	-	-
Glottal	-	-	-	-	+
Nasal	-	-	+	+	+

distinguish the five possible click accompaniments. But even so the phonetic interpretation is not completely straightforward. The value [- voice] still designates a closed glottis when it is in conjunction with [+ glottal], and an open glottis in other circumstances.

Yet another possible classificatory system is shown in (7). This system has achieved the same descriptive parsimony and phonological adequacy as that in (5), by using the feature Aspiration as a replacement

(7)

	k|	k|h	|h	ŋ|	|ʔ
Aspirated	-	+	+	-	-
Glottal	-	-	-	-	+
Nasal	-	-	+	+	+

for the feature Voice. As a result the phonetic interpretation rules must show that the value [- aspiration] indicates vibration of the vocal cords when it occurs in conjunction with [+ nasal] (or [+ syllabic], etc) and abduction or closure of the glottis in other circumstances. Nama does not have any segments that are phonologically distinguished simply by voicing, even among the non-click segments, and the feature Voice might be considered completely redundant in this language.

It is impossible to choose among the different classifications in (5), (6), and (7), if we maintain an abstract approach in which a language is characterized as a social institution rather than a product of a speaker's competence. But it is important to note that linguists who adopt a mentalist approach cannot do any better. There are no data that would help them to choose which solution is a better reflection of what goes on in a Nama speaker's head. Nor is there any way they can tell what features are required for specifying click languages before investigating these languages. Chomsky and Halle suggest that there is a set of features that reflects the phonetic capabilities of man, irrespective of whether these features are observable in any known language. The feature set thus becomes an *a priori* part of the linguistic theory. But in practice Chomsky and Halle, like all other linguists, determine the feature set by first observing what has to be specified in particular languages and then devising a set that is sufficient for this purpose. They are therefore no better off than us when it comes to choosing among competing solutions. Our main point, however, is that, whichever solution is chosen, the feature specifications will require complex phonetic interpretations.

It is not hard to find further problems in relating any reasonable phonological classification of Nama clicks to the observed phonetic events. Consider, for example, the fact that the back of the tongue moves away from the velum more slowly in the aspirated clicks than in the unaspirated clicks. Any interpretation of the features distinguishing these sounds simply in terms of glottal states would be deficient in that it would not

make this evident. Similarly there is no direct way to relate the phonological classifications to the complex variations in the nasal air flow that we noted in the discussion of the data in figures 2, 3, and 4. All these segments require highly specific accounts of the relative timing of the articulatory movements, the adjustments of the larynx, and the velic valving controlling the nasal airflow. These specifications of timing are plainly not in a one-to-one relation with the phonological features; the notion of features as each relating to a single physical scale seems to be irrelevant to such specifications.

We are not meaning to imply that the phonological descriptions cannot be related to the phonetic facts. But it seems to us that phonological descriptions, which are aiming at classifying segments and accounting for patterns, will always be related to specifications of the phonetic phenomena in a complex way, rather than in terms of simple physical scales. We will now exemplify this point further by reference to another Khoisan language.

The other languge we wish to discuss is !Xóõ, which, as has been shown by Traill (1978), has a considerably more complex set of clicks. There are five primary types of articulation, in that, in addition to dental, alveolar, alveolar lateral, and palatal clicks, as in Nama, there are also bilabial clicks. Each click has one of 16 possible accompaniments, which are exemplified for the dental click by the words in (8). As the other 64 (i.e. 4 x 16) clicks are exactly comparable, words illustrating them will not be listed.

This time, instead of presenting the phonetic data first, as we did for Nama, we will suggest here a possible set of phonological features in (9).

(8) **Words illustrating the 16 dental clicks of !Xóõ.**

1.	k|âa	move off	9.	ŋ|āa	see you
2.	g|áã	work	10.	ɴ̥|u'úi	be careful
3.	q|àa	rub with hand	11.	N̊|Gáa	spread out
4.	q|háa	be smoothe	12.	ʔ|ŋâa	suit
5.	|háa	look for spoor	13.	|q'ɜ̀n	small (pl)
6.	k|xâã	dance	14.	k|q'àa	hand
7.	g|hâa	stale meat	15.	g|q'àã	chase
8.	g|xá'ã	splatter water	16.	|ʔâa	die

(9) **Phonological features for !Xóõ (dental) click accompaniments.**

	k|	g|	q|	q|h	|h	k|x	g|h	g|x	ŋ|	ɴ̥|	ɴ|G	ʔ|ŋ	|q'	k|q'	g|q'	|ʔ
Voiced	-	+	-	-	-	-	+	+	+	-	+	+	-	-	+	-
Aspirated	-	-	-	+	+	-	+	-	-	-	-	-	-	-	-	-
Fricative	-	-	-	-	-	+	-	+	-	-	-	-	-	+	+	-
Glottal	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+
Nasal	-	-	-	-	-	-	-	-	+	+	+	+	-	-	-	-
Uvular	-	-	+	+	-	-	-	-	-	-	+	-	+	+	+	-

We will defer considerations of the place of articulation features, but we will consider the phonological specifications of the click accompaniments as we discuss the phonetic nature of each segment. As before, these specifications are meant to distinguish the phonological oppositions and to classify segments into appropriate natural classes. !Xóõ does not have any phonological alternations that would require these clicks being divided into groups; but the phonological specifications do allow us to state the morpheme structure constraints such as the fact that [+ glottal] vowels cannot occur in stems that contain [+ glottal] consonants. Considerations of the most appropriate way to state simple morpheme structure constraints also influenced our decision to specify all these clicks as units. Many of them could well be called clusters involving sequences of segments; but, following virtually all previous writers on this topic, we decided to avoid a cluster specification. As a result we do not have to consider !Xóõ as breaking one more presumed universal by having initial obstruent clusters with voiced segments followed by voiceless segments. (But we realize that this is a weak reason; the language is so unusual anyway that mere presumed universals should not really be allowed to deter us from any particular analysis).

Our investigations of acoustic and aerodynamic events associated with clicks in !Xóõ were made in the field. Using a stereo tape recorder, we were able to record the regular audio signal and only one other signal such as the pressure of the air in the pharynx, or the nasal or the oral air flow. Our most informative sets of data show the pharyngal air pressure recorded via a tube through the nose by means of a special FM (frequency modulation) system which is described elsewhere (Ladefoged and Traill, forthcoming). Data recorded in this way may be recovered either by demodulating (as explained in the previously cited paper), or by reproducing the FM signal on a sound spectrograph. Figures 5-8 use this latter possibility, superimposing a narrow band expanded scale pressure record above a regular wide band spectrogram. The 16 accompaniments of the dental click are shown as they occur in citation forms.

Figure 5(a) shows the pressure buildup behind the velar closure in the voiceless unaspirated click [k|]. The velar closure is released almost immediately after the release of the anterior click closure, and the vowel also follows immediately. The corresponding voiced click [g|] is shown in 5(b). The increase in pharyngal pressure is somewhat less, as there must be a pressure drop across the vocal cords in order to maintain voicing. The voice vibrations can be seen as small vertical striations near the base line. The unaspirated (or, to be more exact, slightly aspirated) click with a uvular release [q|], in 5(c) has a similar pharyngal pressure increase to that in [k|]. The difference between these two sounds is partly in the length of aspiration (over 6 repetitions of this series of words the voice onset time averaged 20 msec later for the uvular release), and partly in the formant

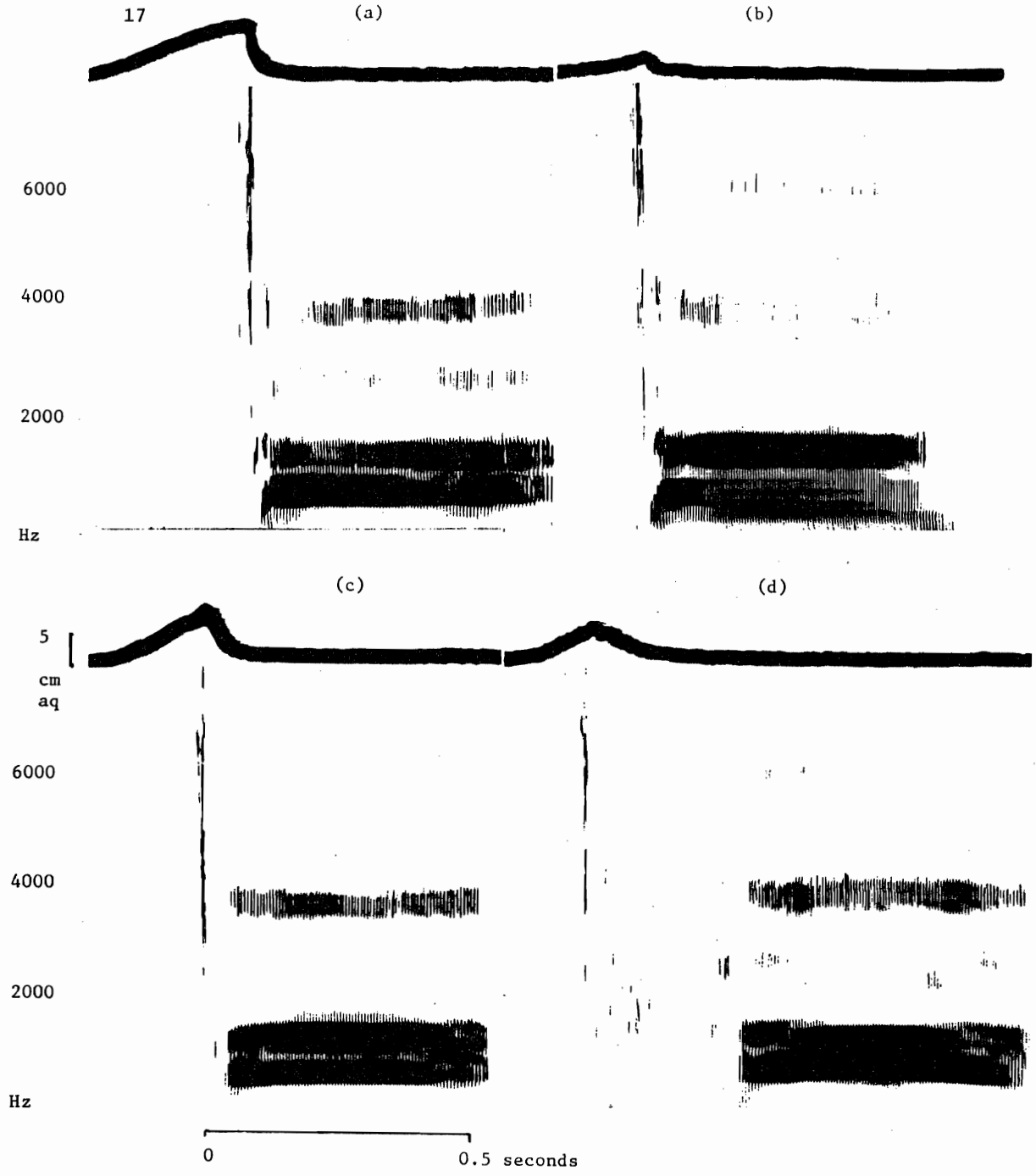


Figure 5. Pharyngal pressure and wide band spectrograms of the !Xóõ words: (a) k|âa (b) g|ãã (c) q|àa (d) q|háa.

transitions. The fully aspirated uvular release may be seen in [q|h] in 5(d). Note the comparatively slow decline in pharyngal pressure associated with the fricative aspect of this release.

From a phonological point of view, the first pair, [k|] and [g|] are appropriately distinguished by the feature Voiced. The second pair [q|h] and [q|] are distinguished by the feature Aspirated; in this language there is no way of distinguishing between all the possible segments without including both these features. The segments [k|] in 5(a) and [q|] in 5(c) differ only in the specification of the feature Uvular release. This change in the value of a single feature has to be interpreted phonetically as corresponding to both a change in the place of release which affects the formant transitions and also a change in the timing of the onset of voice. In addition the slight affrication of [q|h] requires an interpretation of the feature specifications in terms of a slower movement of the back of the tongue than in any of the three other sounds in figure 5. All these phonetic details -- the change in place of release, the change in rate of release, and the change in voice onset time -- can be assigned to a change in the value of the feature Uvular (or the feature High in a Chomsky-Halle framework). But although this is a possible and valid way of describing !Xóõ, these phonetic changes are obviously not *necessary* correlates of the feature changes. They are simply language specific facts which have to be stated as part of the grammar of !Xóõ.

Figure 6 allows us to make further comparisons of the aspirated and affricated click accompaniments. The delayed aspirated release, [|h], is shown in 6(a). This sound is very comparable to the Nama sound which we discussed earlier. There is no increase in pharyngal pressure, and a noticeable delay after the release before the aspiration noise is visible. We do not have nasal air flow records for this sound, so we do not know if the pharyngal pressure was prevented from rising by permitting air to escape through the nose. But we did not hear anything like a voiceless nasal before the release of the click. It may be that in these citation forms the production of pulmonic air pressure is delayed until after the click has been released. We noticed that our Nama speaker for whom we do have nasal air flow records, sometimes pronounced citation forms beginning with a delayed aspirate release without any nasal air flow, although he invariably had some nasal escape when these words occurred in connected speech. It seems likely that the lungs provide a comparatively constant source of pressure, which cannot be turned on and off for particular segments in running speech. The pulmonic pressure build up may be delayed so that it does not start until after the beginning of a word spoken in isolation. But the only way of producing delayed aspiration in a word in the middle of a sentence is by allowing the pharyngal air pressure to escape through the nose.

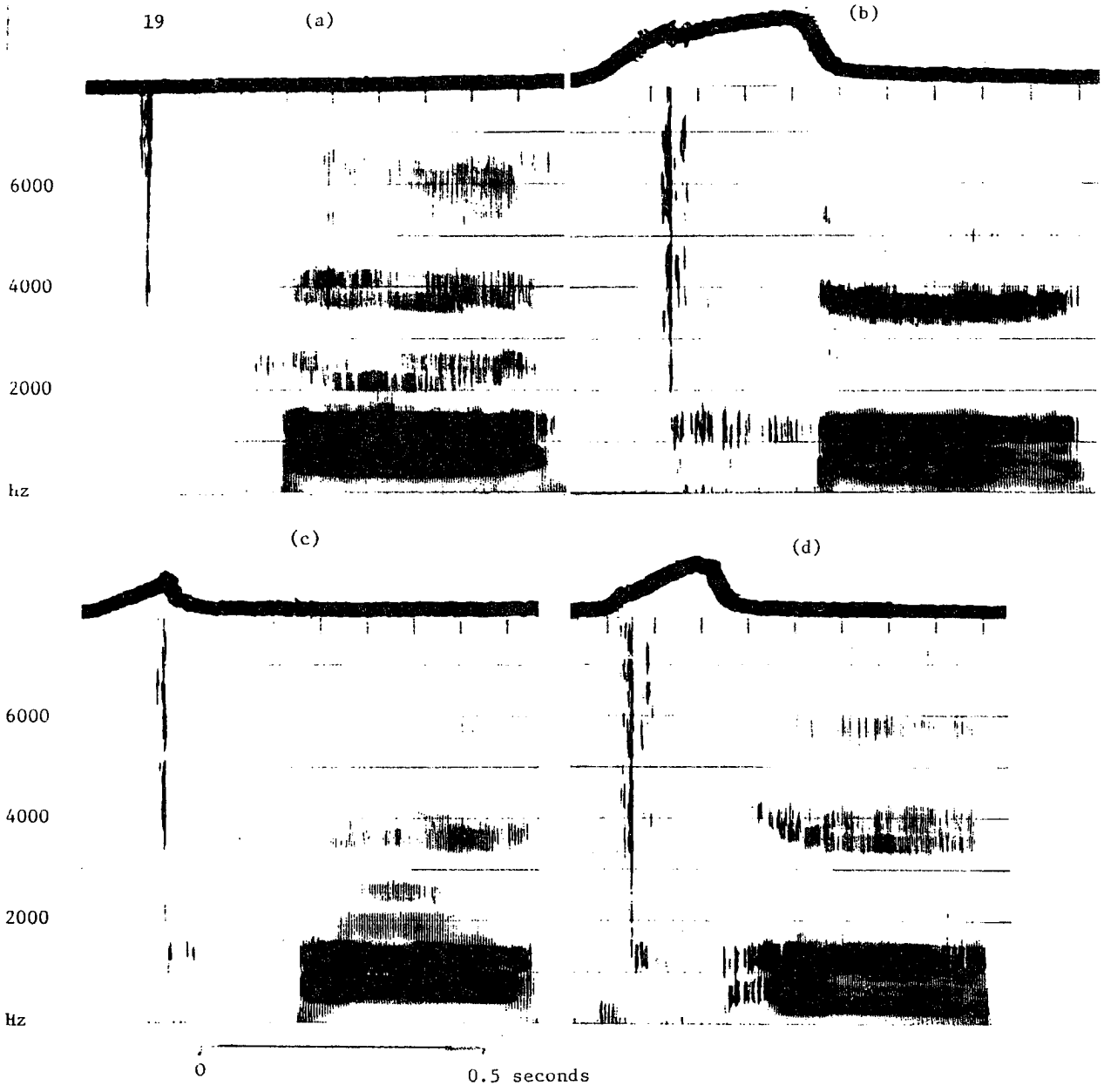


Figure 6. Pharyngal pressure and wide band spectrograms of the !Xóõ words: (a) |háa (b) k|xâã (c) g|hâa (d) g|xá'ã.

In so far as sounds with delayed aspiration are widespread in Khoisan languages, this is an areal rather than a language specific phenomenon. For all these languages the complex phonetic details involved in the pronunciation of clicks with delayed aspiration must be interpreted from the value [+ aspirated] in conjunction with the negative values for all the other features in (9). Once more we may note the many to many relationship between phonetic details and feature values.

The voiceless velar affricate release, [k|x], in 6(b) has a considerable pressure build up during the click closure, which drops slightly on the release of the closure, but then increases again during the fricative. This sound differs from the aspirated uvular release in figure 5(d) by being both more fricative and more velar. These phonetic details may be interpreted in a quite straight forward way from the changes in the two features Aspirated and Fricative.

!Xóõ also has a voiced aspirated click accompaniment, [g|h], as shown in figure 6(c). This sound is nothing like the voiced aspirant [gh] that occurs in Indo-Aryan languages such as Hindi, which have a stop with a breathy voice or murmured release. The !Xóõ sound is a sequence consisting of a regularly voiced velar stop accompanying the click, followed by a period of voiceless aspiration. Sometimes, as in figure 6(c), there are a few voice vibrations immediately after the stop release. We have classified these segments phonologically by means of seemingly contradictory specifications of the glottis. In each case they are said to be [+ voice], which might seem to be at odds with [+ aspirated] in the case of [g|h] and with [+ glottal] in the case of [g|q']. We may regard it as a convention for interpreting the phonetic details of this language that sounds such as [g|h] and [g|q'] which involve contradictory specifications of the state of the glottis realize these specifications by making the voicing come first. An alternative way of dealing with these facts would have been to have used a phonological specification of the kind proposed by Williamson 1976. This would not affect our main point: there is a complex,language specific, many to many relationship between the phonological specification and the phonetic facts of a language.

There is also prevoicing in the click [g|x] shown in figure 6(d). In this case the phonological feature specifications is simply [+ voice]: and, indeed, in some dialectal pronunciations this sound is voiced throughout, so that a more appropriate phonetic transcription would be [g|ɣ]. But in the dialect being described phonetic detail rules must specify that the voicing normally ceases at or shortly after the click release. The pharyngal pressure is maintained (and actually increases) throughout the fricative release. In this particular word, after the click the first vowel is laryngealized, a phonetic(and phonological) aspect of this utterance that is irrelevant to the present discussion, which will be limited to the click segments in these words.

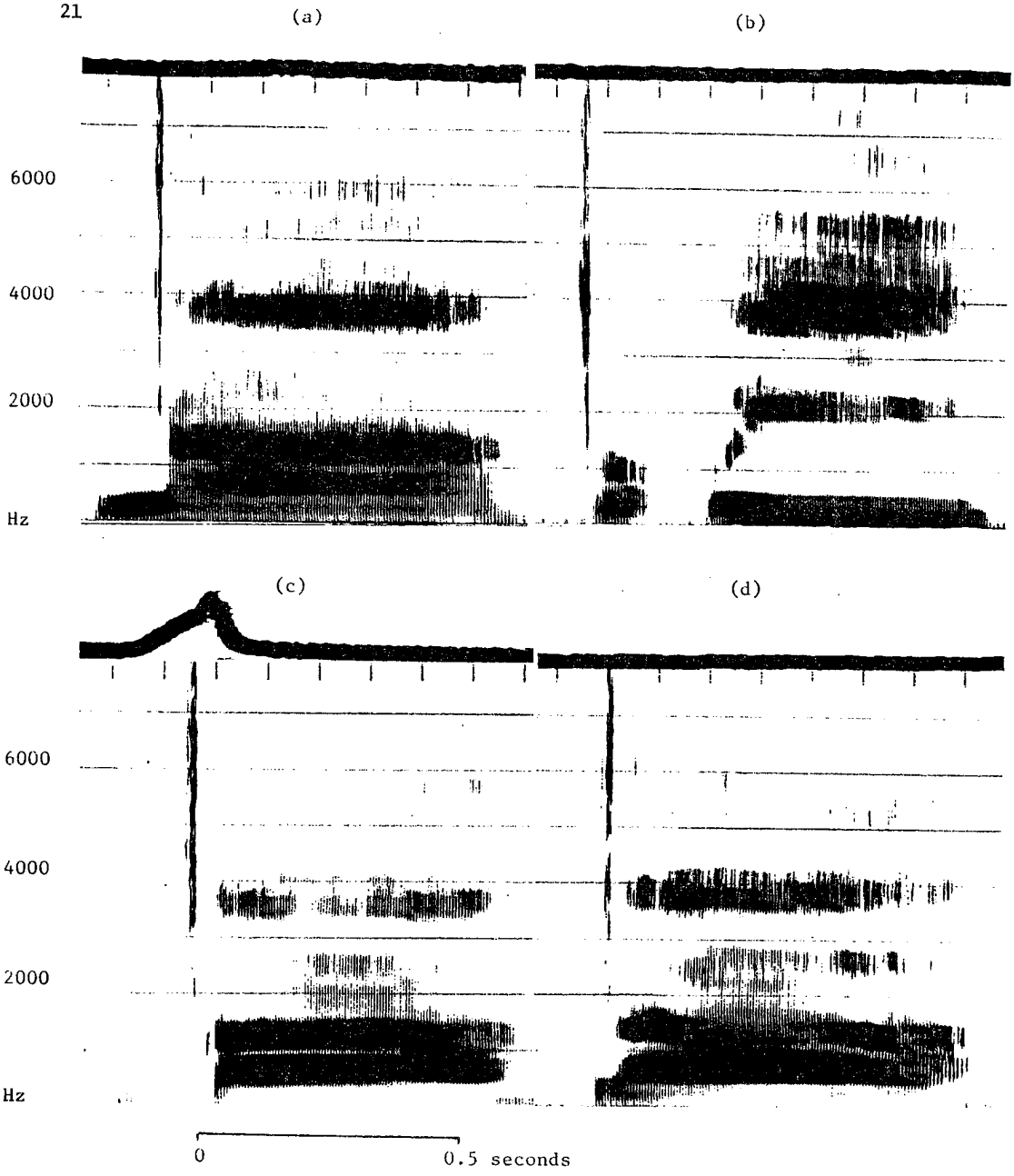


Figure 7. Pharyngal pressures and wide band spectrograms of the !Xóõ words: (a) ŋ|āa (b) ŋ|ú'i (c) ɴ|ǂáa (d) ʔ|ŋâa.

The clicks with nasal accompaniments are shown in figure 7. There is of course no increase in pharyngal pressure in either [ŋ|] or [n̥|], as may be seen in (a) and (b). As in Nama, the click occurs during the last part of the voiced nasal, another detail of the timing which has to be specified in phonetic interpretations of the values of the phonological features. The voiceless nasal also typically precedes the click release. In some senses this sound may be considered to be the unaspirated counterpart of [|h], which, as we have seen, may involve release of air through the nose during the click closure and the first part of the aspiration. The voiceless nasal is not evident on the spectrogram, but it is clearly audible. We recorded the nasal air flow in other examples of this sound and found that in every case air came out through the nose prior to the production of the click.

The prenasalized click with a voiced uvular release is shown in figure 7(c). The nasal part of this click accompaniment may be very short, as in this example. The pharyngal pressure starts rising well before the click release, indicating that there must be a complete stoppage of the air. The pressure continues to rise for a short while after the click. On many occasions the principal auditory cue for this sound is the voiced uvular release, and it may be appropriate to regard it simply as the voiced counterpart of [|q]. It has been classified in (7) as [+ nasal], but the precise phonetic interpretation of this feature value depends on the values of the other features in the segment. As we have been continually emphasizing, the relation between phonological features and phonetic parameters involves a many to many rather than a one to one mapping. The preglottalised nasal click [ʔ|n] shown in figure 7(d) differs from the other sounds with contradictory specifications of the state of the glottis in that in this case the glottal stop occurs before the voicing. There is naturally no increase in pharyngal pressure in this sound, and the click occurs in the middle of the very short nasal.

Other clicks with glottal accompaniments are illustrated in figure 8. The uvular ejective accompaniment [|q'] shown in (a) does not usually have a very large increase in pharyngal pressure, presumably because the glottal closure required for the ejective occurs during the click articulation. The double stop accompaniment [k|q'] shown in (b) consists of a velar stop which is released almost immediately after the production of the click, followed by a uvular ejective during which considerable pharyngal pressure is produced.

The highest pharyngal pressure of all occurs in [g|q'], which is illustrated in figure 8(c). On all 6 of the recordings for which we have valid pharyngal pressure records this prevoiced sound has a higher pharyngal pressure than occurs in the corresponding completely voiceless sound. The instrumental records make it very evident that the second stop is an ejective. The weak velar stop before it is preceded by a small amount of prevoicing which is

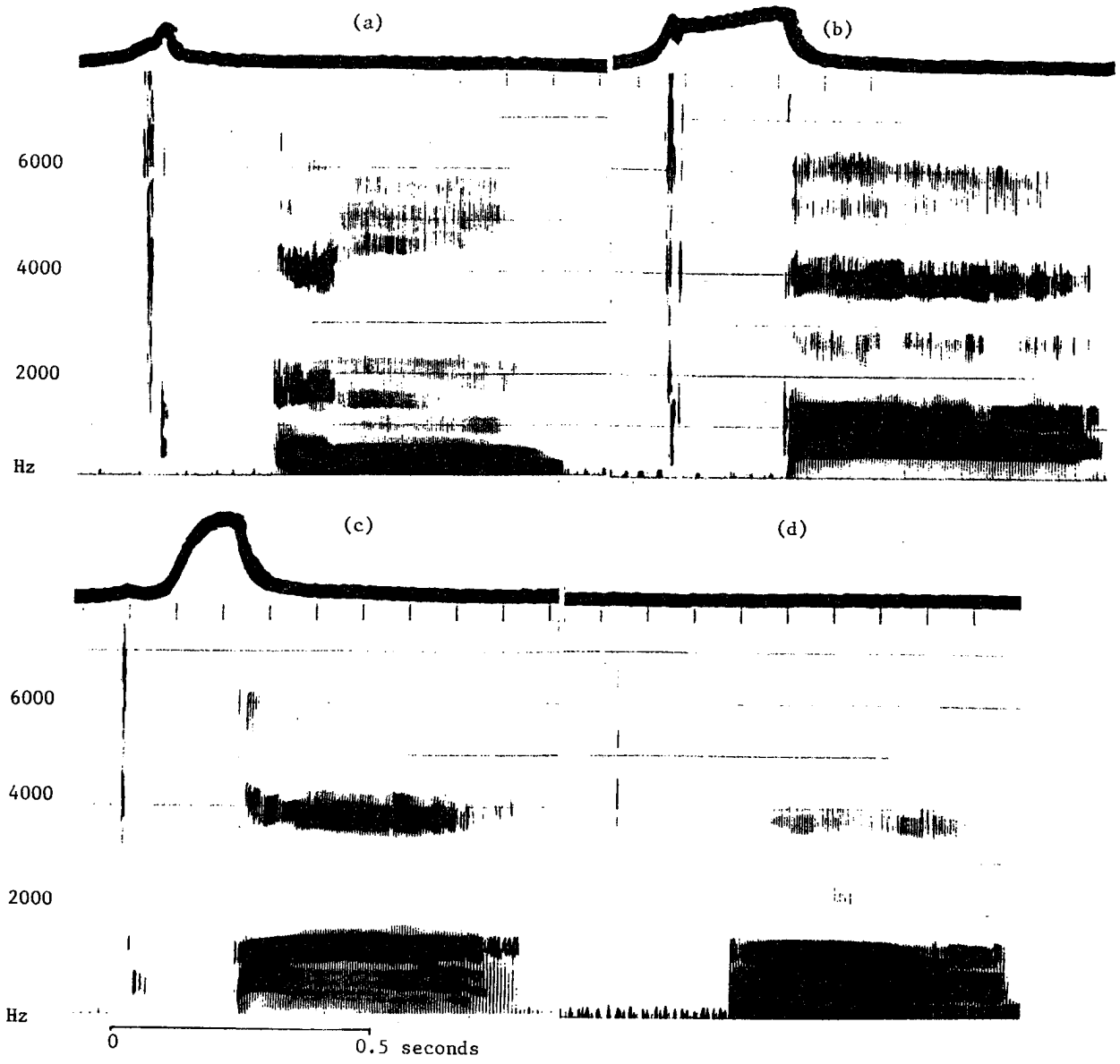


Figure 8. Pharyngal pressures and wide band spectrograms of the !Xóõ words: (a) |q''n (b) k|q'àa (c) g|q'àã (d) |ʔâa.

just visible near the base line. There is a very small increase in pharyngal pressure which is released during the production of the three or four creaky voice vibrations that occur after the click. Then the pressure remains at zero, presumably while the creaky voice has turned into a complete glottal closure, but before the glottis has started rising. The pharyngal air pressure then increases to about 15 cm aq, which is higher than could have been produced by the pulmonic airstream mechanism used in these words. As has been shown elsewhere (Ladefoged, 1967), pulmonic pressures of this magnitude occur only in very loud speech. The pharyngal air pressure goes down when the uvular closure is released. There is only a very short interval before the glottal closure is released and voicing commences for the vowel. This prevoiced stop followed by an ejective uvular release is the most complicated click accompaniment in !Xóõ. The final type of click accompaniment is much simpler, consisting of just a glottal stop as shown in figure 6(d). There is no increase in pharyngal pressure.

The clicks in figure 8 can all be classified as [+ glottal] from a phonological point of view. They have to be grouped together to show that they form a natural class that cannot be followed by a [+ glottal] vowel. But the aerodynamic data make it evident that the feature value [+ glottal] refers to many different laryngeal actions. When in conjunction with some feature values it denotes a closed, non-moving, glottis (as in [|ʔ]), with others a closed, upward moving glottis (as in [|q']), and with others a constricted glottis for laryngealization, followed by a closed, upward moving, glottis (as in [g|q']). Again we see that phonological features are in a many to many relationship with actual phonetic parameters, and that is impossible to specify all the phonetic aspects of languages if we limit each phonological feature to denoting values on a single physical scale.

A similar point may be made with respect to the place of articulation of the actual click sounds themselves (as opposed to their accompaniments). As we mentioned earlier, in this paper we are not concerned with whether segments are better specified in terms of one set of place features or another. Our point is simply to show that the feature values imply one articulatory position when associated with [+ click], and another when associated with [- click]. We have palatographic data on !Xóõ speakers recorded in the field, and cineradiology data on two !Xóõ speakers who were able to visit the University of the Witswatersrand. The cineradiographic procedures will be described more fully in a future publication. As far as we can tell by careful listening and from comparisons of our data and those published by Beach, the clicks of Nama and !Xóõ (and most of the other related languages) do not differ significantly in their place of articulation.

Figure 1, which was discussed briefly earlier in this paper, uses the x-ray data to illustrate the four places of articulation that are common to all these languages. (The !Xóõ bilabial click is not shown.) In each case the shaded area shows the smallest cavity enclosed by the tongue, and the second (lower) tongue line indicates the position which occurred just before the closure was released.

The dental click (in the upper left part of the figure) is produced with a comparatively large suction chamber with its anterior portion in the post alveolar region. If it were not for the confusion it would cause in view of previous descriptions of clicks, we could more properly call this click denti-alveolar, as the contact area definitely includes both the upper teeth and the whole of the alveolar ridge. But this is the term which Beach (1938) and others have applied to the [ǂ] click which we have called palatal. The x-ray data reproduced in the upper right part of the figure, and our palatograms reproduced elsewhere (Ladefoged and Traill, 1980) show that our label for this latter sound is essentially correct. It is true that the tip and blade of the tongue are in contact with the teeth and alveolar ridge, but the forward edge of the click cavity is much further back, and this is the relevant factor. We disagree with Beach in his rejection of the term palatal, as used by earlier writers on these languages. The cineradiology data also demonstrated how the location of the click cavity alters during the production of this click. Figure 1 shows that the tongue contact moved further back while the suction was being developed. At the moment of the release of the click there is no doubt that [ǂ] should be classified as a palatal sound.

The lower part of figure 1 shows the two clicks that have been classified as [+ alveolar]. Note that even when the discussion is limited to clicks, this feature value specifies very different articulatory positions, depending on whether it is in conjunction with [+ lateral] as in the lower left part of the figure, or with [- lateral] as in the lower right.

There are also differences in the rates of the articulatory movements involved in all these clicks, which have to be taken into account in interpretations of the phonological features specifying place of articulation. The two clicks on the left of the figure have comparatively large cavities at the time of release. These dental and (alveolar) lateral clicks are more affricated than the palatal and alveolar clicks on the right. As may be seen from the oscillomink records of Nama in figure 2, in that language (and also in !Xóõ) the lateral click is produced with a considerable fricative noise accompaniment. This noise has a slightly lower mean frequency than the dental click, making it somewhat lower in pitch. The higher frequencies in the dental click do not show up so well in the waveform record. The two clicks on the right of figure 1 are less affricated, but they also differ in the mean frequencies of the bursts of noise associated with their releases. The palatal click has a higher pitch than the alveolar click. The acoustic structure of similar clicks in Naron and Ju|'hõasi have been described by Kagaya (1978) and Snyman (1976), respectively. The point that we wish to make here is just that the acoustic data show that the articulatory movements are slower (thus causing more affrication) in some clicks than in others. Again, these are language specific phonetic details that must be included in a complete grammar.

In the theory of phonology being developed here (and in Ladefoged 1980), there is no God-given set of features, and no need to specify segments in terms of more features than are needed to distinguish the phonemes and to classify sounds into natural classes that may be used in rules explaining observable phonological patterns. The universal characteristics of language depend on (among other things) the 16 or so physiological parameters, which completely specify all possible sounds in all languages (Ladefoged, 1979). Languages use these parameters in various different ways in forming phonemic oppositions and phonological patterns. One might, perhaps, maintain a weakened form of the naturalness condition (Postal 1968) and require that phonological features always reflect *some* of the phonetic facts. This is the approach that has been adopted in the feature specifications suggested here for Khoisan clicks. The major problem with this approach is in validating one phonological description in comparison with another. As we have seen there are alternative ways of classifying the clicks of a language such as Nama, and the preference for one way as opposed to another can be made only in terms of criteria such as the elegance of the one description as opposed to the other.

We doubt there is any salvation in notions such as psychological reality. We suspect that linguists who appeal to criteria of this sort are simply putting off the evil day when they have to admit that their descriptions have no possibility of external validation. We have nothing but admiration and best wishes for any linguist who wishes to go out into the Kalahari desert and try to conduct a test that will discriminate among alternative descriptions that account for the same set of observable phonological patterns. The psychological reality of sound patterns is a valid object of study, but it is not the same thing as phonology. We do not know whether any of the feature specifications proposed above have any testable psychological reality for individual speakers; they simply reflect the most elegant descriptions we can make of the properties of the language, considered as a social institution (including fossilized, non-productive, properties). But we hope that even those who do not share this view will agree that we have shown that a one-to-one mapping between the phonological features proposed here and the observed phonetic facts is not feasible. It is of course possible that some other feature specification and some other set of parameters might be better matched. But there is no reason to believe that this is likely to be the case. Human beings do not usually work that way; observable phonological patterns are nearly always the result of complex, interacting, processes.

Acknowledgments

As always, our thanks to many members of the UCLA Phonetics Lab, and, especially to Ian Maddieson and Mona Lindau-Webb who contributed several useful suggestions. Our thanks are also due to Jan Snyman, UNISA, who provided support for some of the fieldwork and excellent laboratory facilities. A preliminary version of this paper will appear in the proceedings of the 1979 conference on Khoisan languages, UNISA, South Africa.

References

Beach, D. M. (1938) *The Phonetics of the Hottentot Language*. Cambridge: Heffer.

Chomsky, N. and M. Halle (1968) *The Sound Pattern of English*. New York: Harper and Row.

Fant, G. (1973) *Speech Sounds and Features*. Cambridge, MA: MIT Press.

Jakobson, R. (1968) 'Extrapulmonic consonants (ejectives, implosives, clicks)'. MIT Research Laboratory of Electronics, *Quarterly Progress Report* 90. 221-227.

Kagaya, R. (1978) 'Soundspectrographic analysis of Naron clicks: a preliminary report. *Annual Bulletin of the Research Instiutte of Logopedics and Phoniatrics*. 12. 113-125.

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

Ladefoged, P. (1971) *Preliminaries to Linguistic Phonetics*. Chicago: University of Chicago Press.

Ladefoged, P. (1972) 'Phonetic prerequisites for a distinctive feature theory. *Papers in Linguistics and Phonetics to the Memory of Pierre Delattre*, ed. by Albert Valdman, 273-285. The Hague: Mouton.

Ladefoged, P. (1975) *A Course in Phonetics*. New York: Harcourt Brace Jovanovich, Inc.

Ladefoged, P. (1979) 'Articulatory parameters.' *Proc. of the Ninth Intern. Congress of Phonetic Sciences*, ed. by Eli Fischer-Jørgensen, Jørgen Rischel, and Nina Thorsen, 41-7.

Ladefoged, P. (1980) 'What are linguistic sounds made of?' *Language* 56.

Ladefoged, P., and T. Traill (forthcoming) 'Instrumental phonetic fieldwork.' *UCLA Working Papers in Phonetics*.

Postal, P. M. (1968) *Aspects of Phonological Theory*. New York: Harper and Row.

Snyman, J. W. (1976) 'The clicks of ǔ/ʔhõasi and Nama. Paper presented to the African Languages Conference, University of South Africa, Pretoria.

Traill, A. (1978) 'Another click accompaniment in !Xóõ. *Khoisan Linguistic Studies* 5.22-29.

Traill, A. (forthcoming) 'Phonetic and phonological studies of !Xóõ Bushman.

Williamson, K. (1977) 'Multivalued features for consonants.' *Language* 53. 843-871.

Instrumental phonetic fieldwork

Peter Ladefoged and Anthony Traill

When Daniel Jones was about to go off on a field trip someone once asked him what instruments he was going to take with him. "Only these," he said, pointing to his ears. We would agree entirely that by far the most valuable assets a phonetician can have are a trained set of ears. We would add (and we are sure that Daniel Jones would also approve) that the ears should be coupled to highly trained vocal organs that are capable of producing a wide range of sounds. There is no substitute for the ability to hear small distinctions in sounds and to pronounce alternative possibilities. In a fieldwork situation one must be able to ask the language consultant which of two pronunciations sounds better. One of the most efficient procedures for getting results in the field is to test different hypotheses by trying out various vocal gestures of one's own.

But nowadays phoneticians who go out with only their ears and their own vocal apparatus are doing themselves a dis-service. There are three ways in which instrumental aids can be valuable supplements to the field phonetician. Firstly they can sometimes suggest new descriptive possibilities; we have, for example, learned a number of facts about click releases by observing pharyngal pressure records. Secondly instruments allow permanent records to be made so that one can demonstrate the facts to those who do not have access to speakers of the language being described. Thus we are sure that readers would believe our description of the clicks in Bushman languages without any instrumental evidence; but it is nice to have records so that they can see for themselves. Thirdly, instruments enable one to make quantitative descriptions. However good one's ears one cannot, for example, measure and report the duration, in msecs of the aspiration in a set of words; and without measurements one cannot prove that there is a statistically significant difference between one group of words and another, or between the sounds of one language and another.

Recording

Everyone is aware of the value of being able to tape record material so that one can listen to it over and over again and catch subtle nuances of sounds. Furthermore, good recordings are the basis of all subsequent laboratory acoustic analysis. We will not consider here how field recordings can be analyzed by techniques such as sound spectrography. Instead we will simply discuss, in a fairly prescriptive way, how tape recorders can be used in field situations.

Tape recordings can be like audible notebook jottings, or like more formal finished papers that demonstrate the sounds of a language. In either case a recording should *always* begin with a statement of the date, the place, the names of the speakers, and the material. These things will no doubt be written on the box in which the recording is stored; they should also be written on a label which should be stuck on the reel itself. But labels can come off and tapes can easily get put into the wrong box, and become virtually useless unless they have been properly identified.

Tape recordings should also be of the highest quality possible. Interviews sometimes have to take place amidst noises such as background talking and household sounds that the fieldworker can ignore. But on a recording for linguistic phonetic purposes anything other than the voice of the language consultant can be a distraction. We have found that when trying to make a high quality recording it is often a good idea to go off outside. When doing this it may be necessary to protect the microphone from the effect of the wind; and one sometimes has to watch for the noise of the wind in the trees. These problems can be overcome by using a suitable microphone and by finding a large enough open space. Both in everyday parlance and in the technical acoustic terminology, the best recordings are often those made in a free field, rather than in a noisy, reverberant room.

The intensity of the sound being recorded depends on the square of the distance between the source of the sound and the microphone. A baby crying ten feet away many be one quarter of the loudness of a speaker who is five feet from the microphone;but the same baby will be only one hundredth of the loudness of a speaker who is one foot from the microphone. If recordings must be made in circumstances where there is a lot of background noise, make sure that the speaker is a close but constant distance away. Also place the microphone on something soft and at least a few feet from the tape recorder. Check the signal level frequently while making the recording, as speakers often vary their loudness. The signal should always be at the highest permissible level, so as to make sure that it is as far above the background noise as possible. Finally, after making a recording, play it back and verify that you have proper written notes to explain what is on it. As soon as you can, play all your recordings right through, checking them against a complete text. (We try to do this every night, so that we never accumulate data that is not useful).

What sort of machine should be used for making field recordings? We feel that the particular make is less important than whether it meets certain specifications. Firstly it should have a good frequency response: that is to say it should be able to reproduce the complete range of frequencies recorded with the same relative intensities as when they were originally produced. For most linguistic phonetic purposes a flat response ($\pm$ 2 db) from 60 Hz to 14,000 Hz is more than satisfactory. Secondly the

tape recorder should have a good S/N (signal to noise) ratio. The difference between the maximum signal that can be recorded without overloading and the noise that is present in the absence of any signal should be at least 45 db. Thirdly speed variations, whether of the short term kind known as flutter, or the day to day kind due to variations in the batteries and motors, should be less than 0.1%. In addition, the tape recorder, like any other piece of apparatus taken into the field, should be tough and reliable. It should be possible to drop it on the floor and kick it around for a few minutes without too much damage. We have heard stories of professional tape recorders that have been fished out of rivers and found to be still working. The stories are probably apocryphal, but they provide a good standard to aim for.

Before going into the field, the tape recorder should be thoroughly checked out. This means making sure that the batteries are not only fully charged, but also are capable of maintaining their charge for as long as is required. In addition the heads should be cleaned and demagnetized (simple procedures that may increase the frequency response by as much as 5,000 Hz), and a frequency response check run. Record frequencies from 50 Hz to 15,000 Hz at about half octave intervals, checking with a meter that they are all being produced by the signal generator at the same level. Then observe on a meter the relative intensities of these signals when they are played back, so as to make sure that the tape recorder really has a range of 60-14,000 Hz, $\pm$ 2 db. In addition, check the S/N ratio, and, if the necessary equipment is available, the speed constancy. If a stereo tape recorder is being used (and, for reasons that will become apparent in the next section, this is often desirable), check that both channels work equally well, and that the signals recorded on each of them are completely separated.

Aerodynamic data

Acoustic analyses made from good quality tape recordings can provide large amounts of data. But they often do not indicate in an unambiguous way important articulatory facts such as the direction of the airstream or the timing of movements of the vocal organs, particulary those during voiceless closures. The best way of gaining information on these phonetic parameters is by recording four aerodynamic parameters: (1) the pressure of the air in the mouth behind any bilabial closure; (2) the pressure of the air in the pharynx, behind any alveolar or velar closure; (3) the flow of air in and out of the nose; (4) the flow of air in and out of the mouth.

Examples of the use of such data for elucidating articulatory descriptions of West African languages have been given by Ladefoged (1968). In these examples the data were recorded live by subjects in the laboratory, using a four channel inkwriter. Unfortunately, we have as yet been unable to find an adequate, light weight, battery operated, ink writer for use in the field. Accordingly the required data must somehow be recorded on a tape recorder, and brought back to the laboratory for later reproduction. This involves making an FM recording.

original as for MPP

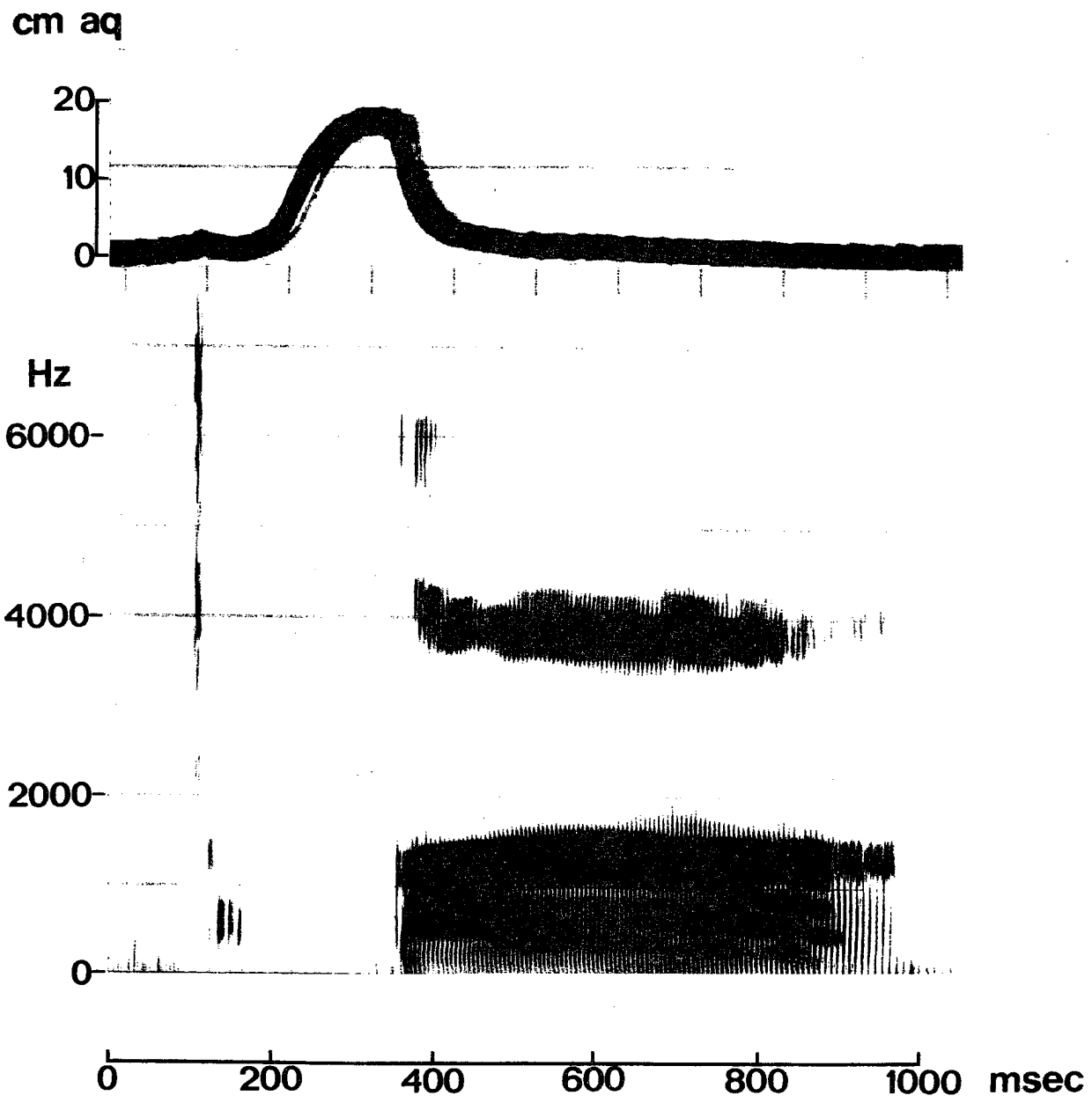


Figure 1. Pharyngal air pressure record and wide band spectrogram for the !Xóõ word g|q'àã. Time intervals are expressed in msec from an arbitrary starting point.

Frequency modulation (FM) recording is a process whereby a relatively slowly varying signal, is turned into a variation in frequency. Thus a zero pharyngal pressure may be made to correspond to a base frequency of 1,500 Hz. As this pressure increases to, say, 10 cm aq (i.e. the pressure required to hold up a column of water 10 cm high) this frequency may be made to increase to 2500 Hz. In this way variations in pressure may be changed into variations in frequency which sound like a fluctuating whistle. This sound can be recorded on the second channel of a stereo tape recorder while the regular audio signal is recorded on the first channel. The frequency variations may later be changed back (demodulated) to signals that correspond to the original pressure variations. These signals can be written out on an inkwriter in the laboratory.

As an alternative to demodulating the frequency variations, both channels of the stereo recording may be analysed on a sound spectrograph. This will result in a visible record of the speech with a line corresponding to the FM signal superimposed on top of it. Figure 1 illustrates data on one of the clicks of !Xóõ, a Bushman language described by Traill (forthcoming) recorded in this way. In this case the outputs of the two channels of the stereo tape recording were analyzed separately and then combined by placing one spectrogram above the other, aligning them by reference to a third spectrogram which showed the two channels analyzed simultaneously. The lower part of the figure shows a wide band spectrogram, with the regular full scale of 0-8,000 Hz. The upper part of the figure is a narrow band expanded scale analysis of a 1,500 Hz signal which has been frequency modulated by a signal corresponding to the pressure of the air in the pharynx.

A great deal of information can be obtained from analyses of this kind. Thus figure 1 shows that there are voicing vibrations visible near the base line in the period before the release of the click, which occurs at 110 msec after the arbitrary starting point of this spectrogram. During this period there is also a small increase in pharyngal pressure due to the stoppage of the air behind the velar closure that occurs in the course of the production of the click. This part of the sound is therefore equivalent to a voiced velar stop [g̬]. The pharynx pressure goes down immediately after the click, so this velar stop must have been released. There follow three or four creaky voice vibrations, lasting until time 170 msec. During the period 150-170 msec the pharyngal pressure is almost at zero, so there was probably no oral closure. At about time 170 there must be both an oral and a velic closure, otherwise the pharyngal pressure could not have started increasing. Considering the magnitude of this pressure increase (almost 20 cm aq), there is probably a glottalic airstream mechanism. The lungs do not produce pressures this high, except in fairly loud speech (Ladefoged, 1967), which this was not. Presumably, after the creaky voice vibrations, the glottis closed completely, and then came sharply upward. (This we know is likely to be true from our visual observations of this sound.) The closure (a uvular one as we know from listening -- we could not be sure from the spectrogram) was released at time 350, and the pharyngal pressure fell rapidly. Less than 5 msec later, the spectrogram indicates the release of a glottal stop into the vowel. An appropriate IPA transcription of this sequence might therefore be [g̬ʔ']. In symbols more familiar to scholars of click languages it would be g|q'. (We have independent auditory, visual, and kinesthetic evidence for the fact that the click is dental.)

With the equipment we have been using in the field to date, we have been able to record only the regular audio signal on one channel and one other signal. Accordingly we have had to choose which additional parameter we wished to record. As an alternative to recording air pressure behind the lips or in the pharynx, we can record the rate of flow of air out of the mouth through a mask, or out of the nose through tubes connected to the nostrils. In either case the airflow is recorded by detecting the very small increase in air pressure that occurs when the air flows through a wire mesh. This pressure is converted into a varying frequency in exactly the same way as the oral pressure variations discussed above.

We may note here some requirements of a pressure/flow system for making recordings in the field. It should obviously be light weight, compact, and battery operated. As with the tape recorder, it should be rugged enough to withstand the pounding it is likely to get while being transported into places where there are no roads. The air pressure system should be capable of measuring $\pm$ 25 cm aq. The air flow system should be constructed so that it distinguishes on a single channel between ingressive and egressive air flow rates up to 2 litres/second. It should have a flat frequency response, so that it does not give different readings for sounds with identical flow rates said on different pitches.

There are a number of practical points to take into consideration when making pressure and flow recordings. The tubes used should be short and as thick as the speaker can conveniently tolerate, so that they have a high enough frequency response to show some voicing vibrations. Long thin tubes act as acoustic filters and cut out the small pressure variations associated with voicing. Generally, tubes should be sealed at the end, with small holes at the sides near the tip to let the pressure in. Tubes that are open at the end often become full of mucous. We have found infant feeding tubes, size 12 French, to be suitable.

When recording the pressure of the air in the mouth one end of the tube should be connected to the pressure transducer, and the other end should be held (by the speaker) so that it is just behind the lips. The speaker should be discouraged from sucking on the end of the oral tube, as this leads to the tube becoming full of saliva, which will lower its frequency response.

The pharyngal pressure tube should be inserted through the nose so that its open end rests on the back wall of the pharynx about 1 cm below the uvula. The speaker should first be given a practical demonstration by the field-worker of how easy it is to pass a tube through the nose into the pharynx. Take a clean, sterile, tube (tubes should be cleaned with disinfectant and boiled for 30 minutes if they have to be reused). Hold it about 12 cm from the tip, and moisten it with saliva. If it has a slight natural curve, make sure that this curve is pointing downwards. Then, still holding the tube about 12 cm from the end, push it straight (i. e. horizontally, not upward) back into the nose until some obstruction is reached. It is advisable to be fairly rapid about this part of the proceedings. Many people find the

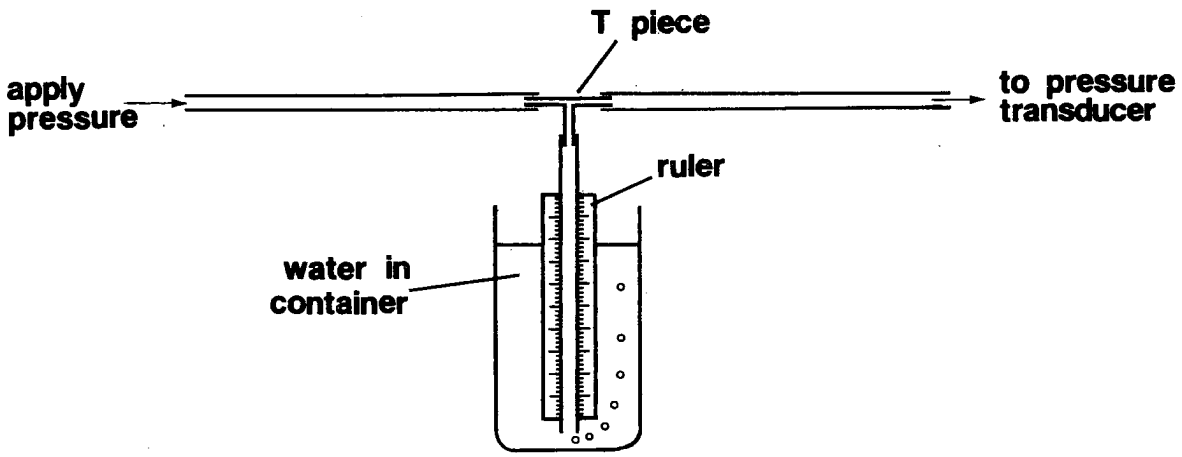


Figure 2. System for calibrating air pressure in the field.

most difficult part of the whole thing to be the irritation that may occur while they are swithering just inside the nostril. Once an obstruction has been reached, take a mouthful of water, and, while pushing the tube gently further in, swallow a small quantity. Keep pushing gently, swallowing, and breathing in through the nose until the tube has passed over the top of the velum into the pharynx. If there is any difficulty in getting the tube to pass round the velum, twist it slightly and go on swallowing and pushing straight back. Remember to hold your hand high above the level of the nose, so that the tube is never pushed upward, but always goes straight back (and, hopefully,down). Nearly everybody finds it easier to pass the tube through one nostril rather than the other, so if at first you don't succeed, try again with the other one.

For an average male adult one will want to pass about 15 cm of the tube into the nose in order to locate the open end properly. It is a good idea to make a mark by putting a thread around the tube at slightly more than this distance before beginning the experiment. As soon as the tube is located properly in the pharynx the speaker should relax quietly for a moment in order to avoid a gag reflex. Check the location of the tube by using a flashlight (a small, focused pen light) while looking into the speaker's mouth. It may be necessary to use some object such as a spoon or a spatula to depress the speaker's tongue in order to see the end of the tube. If it is correctly located about 1 cm below the uvula,fasten the tube to the cheek just outside the nose with a small sticking plaster.

When recording pharyngal pressure through a nasal catheter, it is necessary to keep the tube free from mucus. Before attaching the end to the pressure transducer, connect it to a rubber bulb (the one from the palatography spray to be discussed below is very suitable) and blow some air through it. Repeat this procedure at frequent intervals throughout the recording.

Pressure recordings may be calibrated using a water manometer, or more simply (and only slightly less accurately) by the following procedure. Connect one arm of a T piece to the pressure transducer, and the center of the T to a tube attached to a ruler in a container of water as shown in Figure 2. Blow gently into the other arm of the T until bubbles come out of the tube. At this time the pressure exerted on the transducer will be equivalent to the depth of the water (or, to be more exact, the distance between the end of the tube and the surface of the water). Repeat this procedure with varying water levels, so that signals are produced over the entire range of interest (probably up to 15 cm aq). When these signals are reproduced a graph can be drawn, showing the relation between the variations in the signal and the set of known pressures.

Variations in airflow are slightly more difficult to record than variation in air pressure. Nearly all systems for recording air flow involve measuring the amount of flow in terms of the very small amount of pressure that it produces in specified circumstances. (The only major exception to this generalization is the hot wire anenometer which has not been produced in a portable, field, version.) We have found that the most suitable way of measuring air flows in speech is by measuring the slight pressure build up that occurs when the air flows through a fine wire mesh. This system has a substantially flat frequency response, usually with a slight peak around 5-10 Hz, associated with the volume of air contained in the mask. It will measure both egressive and ingressive flows by converting them to small positive and negative pressures which can be transduced. It can be calibrated by connecting it to a commercially available flow meter. With a little practice one can learn to blow steadily at a number of different rates, so that one can observe the flow meter readings that correspond to different signals. More accurate measures can be made in the laboratory by using a vacuum cleaner controlled by a variable voltage, reversing the drive to the fan, so that the device blows instead of sucking.

We have found it best not to try to use a face mask that is divided into compartments for the mouth and the nose. The flow of air from the nose is usually very small indeed, and it is simplest to gather it by two tubes ending in bulbs (nasal olives) lightly inserted into the nostril. These tubes are then joined and led out from the face mask into a tube containing a wire gauze. In practice it is difficult to collect air from both nostrils if one nostril is also being used for a catheter in the pharynx. This nostril is best sealed off completely, and the nasal flow recorded through the other one.

Finally in this section we must note briefly a luxury that it is very pleasant to have when recording aerodynamic data in the field. This is a battery operated, portable, storage oscilloscope. This device makes it possible to see instantly the form that the pressure and flow signals will have when they are written out more permanently on an inkwriter. As a result it is possible to check in the field that one event occurs before another -- and this may lead to a modification of one's hypothesis, or of the set of data to be recorded in future experiments.

Articulatory data

There are two simple techniques for obtaining articulatory data showing positions of the vocal organs. Firstly, photographs can be taken of the lips. Secondly, palatographic records can be made, showing both the part of the roof of the mouth that has been touched by the tongue in a particular word, and the part of the tongue that made the contact. Articulatory data of these kinds is shown in Figure 3 (from Ladefoged 1968),

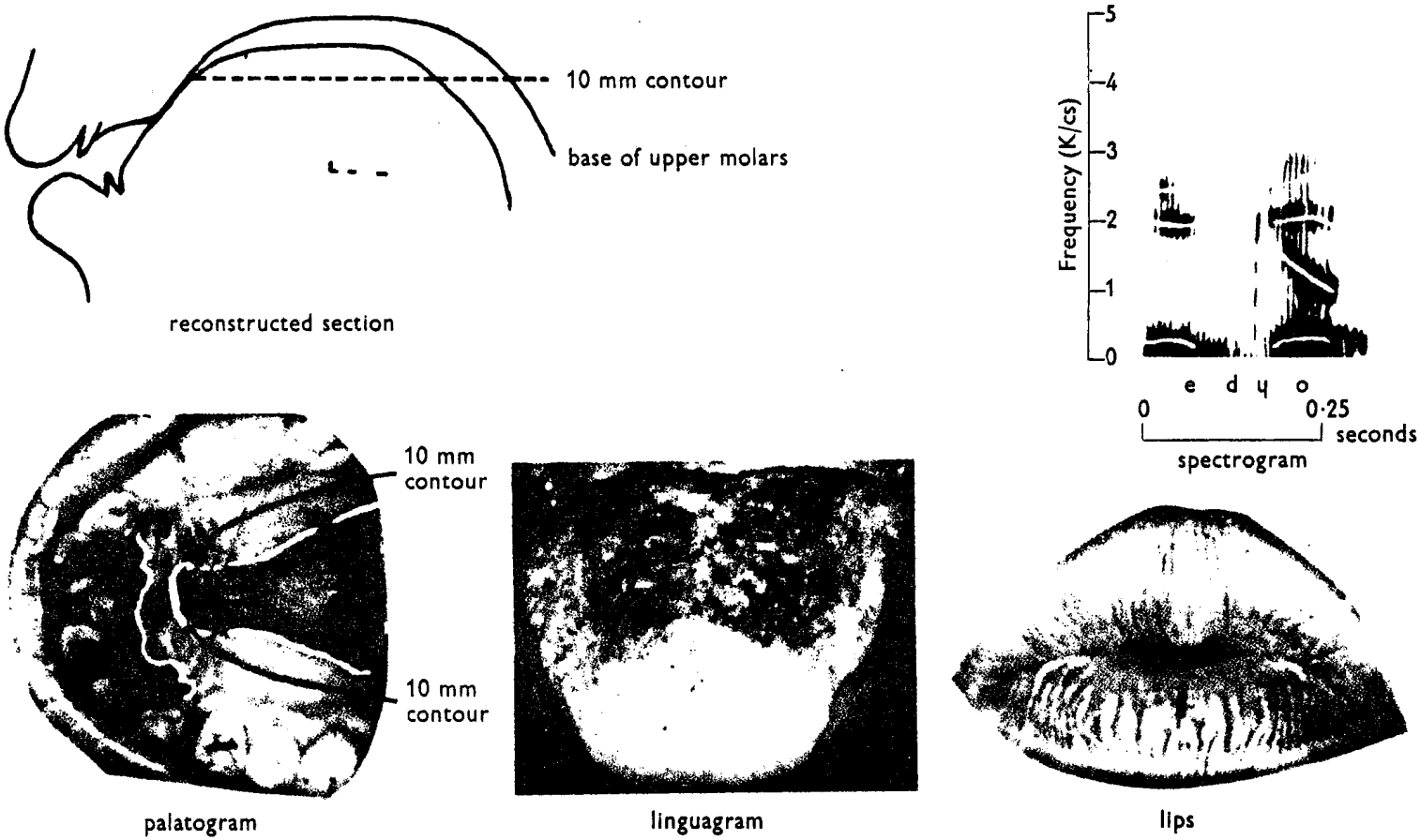
which illustrates the pronunciation of a labialized and palatalized prepalatal affricate in the word *edwo* [édʮò] 'a yam' in Late, a Guang language with close affinities to Akan, spoken in West Africa.

The photograph of the lips in the lower right of figure 3 shows that the lips are closely rounded during the closure for this sound. With a little practice it is quite possible to take photographs at appropriate moments in the words being investigated. Ask the speaker to repeat the sound in a phrase, over and over again, while you get ready to take the photograph. Have a tape recorder running, and make sure that the camera is close to the microphone, so that there is a clear recording of the click of the shutter when you eventually take the photograph. You can find out in the field roughly when the photograph was taken by playing the recording at half speed. Later, spectrograms can be used to locate the precise moment.

The palatographic data in figure 3 is even more informative. The procedures involved have been well established for many years (Abercrombie 1956; Ladefoged 1957). In essence it usually nowadays consists of covering the upper surface of the mouth with a dark powder, saying a word, inserting a mirror into the mouth, photographing the area wiped clean by the tongue, and then photographing the tongue so as to show which parts of it have become covered with powder. There are Polaroid cameras available (see Appendix) that have dental mirrors attached, making it very easy to photograph the roof of the mouth. Photographs of the roof of the mouth are usually called palatograms, and those of the tongue are called linguagrams.

In the word recorded in figure 3, the linguagram shows powder on the center and front of the tongue, but not on the tip or the blade. We may conclude, therefore, that the center and front of the tongue touched the roof of the mouth, but that the tip and blade must have been held down so that they played no part in the articulation of this consonant.

The palatogram in the lower left of figure 3 has a white line added to show the limits of the area contacted by the tongue. The extent of this area was clearly visible on the original photograph, but, as is often the case, we considered it advisable to draw a line on the photograph to counterbalance the losses due to reproduction. The palatogram shows that in the midline the powder has been removed from only a small part of the roof of the mouth just behind the alveolar ridge. In addition, a great deal of powder has been wiped away from the sides of the hard palate, leading us to conclude that the sides of the tongue must have been raised in that region. It is this raising of the sides and body of the tongue that gives the sound its palatalized quality.



W. O. Gyampoh (Late): édɥó 'yam' 26 Jan 62

Figure 3. Data illustrating the pronunciation of a labialized pre-palatal stop in Late.

It must always be remembered that the contact areas reflect the sum of the articulatory contacts that occurred in the pronunciation of each of these words; they do not show the position at any one particular moment. In addition the photographs of the tongue show it when it has been slightly stuck out of the mouth,and is therefore not in the same shape as it was when producing any of the sounds.

The illustrations of the sagittal sections of the vocal organs in figure 3 were made possible through the use of dental impressions of the oral cavity made in the field. There is no reason for these impressions to be made using a tray of the kind that dentists use. As the outer surfaces of the teeth play no role in the production of speech, we can simply neglect them. All we need is an impression of the inner surfaces of the roof of the mouth. For our purposes it is necessary to use an alginate impression material. Other substances which set harder and cannot be cut are of no use. The easiest way to make an impression is to mix a sufficient quantity of the material, and place it on the back of the palatography mirror. Get the speaker to lean slightly forward, and then insert the mirror with the material on it into the mouth. Press the mirror firmly against the upper teeth, allowing some of the material to flow out of the mouth around the upper lip. A good impression for phonetic purposes should be made with sufficient material to indicate (at least roughly) the shape of the upper lip and the curvature of the soft palate. The palate will, of course, be in a lowered position, as the speaker will have been breathing through the nose while the impression material is setting. The impression material around the lips sets slightly more slowly than that inside the mouth, where it is slightly warmer. When you can see that the material around the lips is firm, it is quite safe to remove the mirror from the mouth, first rocking it, raising and lowering it slightly, so as to break the seal.

If the impression is to be kept for any length of time it must be put under water to prevent it drying and shrinking. Otherwise, take it off the mirror and trim the base flat so that it is parallel to the plane of the teeth. If the mirror really was pressed firmly against the upper teeth while the impression was being made, this should involve no more than the removal of excess material from around the sides. The impression may then be cut in half in the mid-sagittal plane, and a tracing of the upper surface made as in the upper left of figure 3. The exact positions of movable structures such as the lips and the soft palate have to be estimated, but if care has been taken to have sufficient impression material around the lips and as far back in the mouth as possible, the sagittal diagram will be reasonably accurate. Palatograms should always be accompanied by diagrams of this kind. It has long been established that sagittal sections provide the most useful representations of speech sounds.

The impression of the roof of the mouth may also be used to construct contour lines at fixed distances from the plane of the teeth. A line showing all points 10 mm below the highest point of the palate has been superimposed on the palatogram in figure 3. In order to draw such a line put the two halves of the impression material together again, and cut them horizontally (i.e. with the blade of the knife parallel to the surface of the impression material corresponding to the plane of the teeth). It is then possible to measure the distance of this plane from the roof of the mouth, and to draw a line round it. In order to superimpose this contour line accurately on the photograph it is first necessary to be certain that the photographs that are being used are life size, a fact that may be checked by photographing a ruler put in the place of the mouth. Fortunately, cameras such as the Polaroid dental camera (see appendix) automatically produce photographs that are always life size, in focus, and correctly lit.

We may conclude this section by noting some practical points in connection with palatography. Firstly, care should be taken in selecting appropriate words. We are often interested in comparing the places of articulation of comparable sounds. Accordingly words must be chosen that contain these articulations, and do not contain any other similar articulations that might overlap with them. Thus when investigating the difference between s and ʃ in English one should use words such as "sop - shop" rather than "sot - shot." Similarly one should use either a range of vowels ("seep-sheep, sip-ship, same-shame, Sam-sham, sop-shop, etc") or, if this is not possible, just open vowels which might be expected to have less effect on the consonant articulation. As with all instrumental phonetic investigations, time spent selecting suitable words is a good investment.

When doing palatography, one should allow the speaker to practice the task extensively. It is important to get the speakers to relax after the upper surface of the mouth has been sprayed, so that when they say the word being investigated they do so naturally. It also requires practice to stick the tongue out of the mouth the same way every time. It is obviously important to date and label the photographs as soon as they are taken. In addition, again as with all instrumental data, it is preferable to make records of several different speakers saying a few utterances rather than one or two speakers repeating a large number of different utterances. Ideally one would like to get a dozen speakers of the same dialect each repeating a dozen times all the contrasts to be investigated. But in a world in which time and effort are limited it is most important to find out the properties of the language that speakers have in common, rather than the details of an individual's pronunciation. The best kinds of instrumental investigations are those to which one can apply statistical techniques such as analysis of variance, which allows one to compare differences among individuals speaking different languages. When the differences between groups are statistically large in comparison with differences within a group then one can say that the languages represented by the groups really are phonetically different.

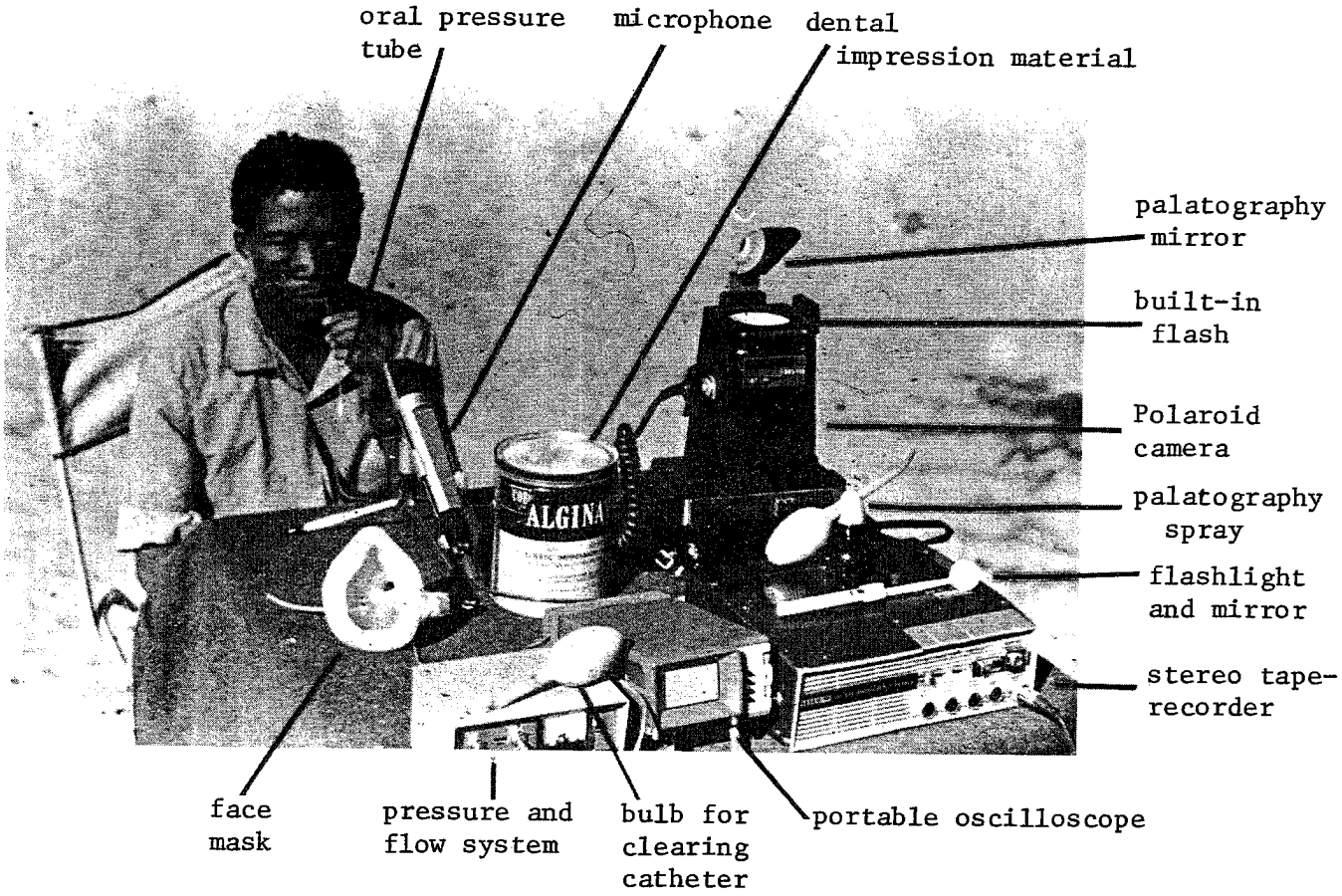


Figure 4. A portable phonetics laboratory.

We may review much of the quipment we have been discussing here by reference to a photograph (figure 4) taken on a recent fieldwork trip in the Kalahari desert. The speaker, Guhmsa, is inserting a tube into the nose, so that we might make the pharyngal pressure records exemplified in figure 1. The stereo tape recorder, FM pressure/flow system, and portable storage oscilloscope are on the left side of the table; the palatography camera and other equipment is on the right. The appendix gives a complete list of all the apparatus, including, where appropriate, manufacturers, model numbers, and 1979 prices.

References

Abercrombie, D. (1956) "Palatography" *Zeitschrift fur Phonetik.*

Ladefoged, P. (1957) "Use of palatography" *J. Speech and Hearing Disorders.*

Ladefoged, P. (1968) *A Phonetic Study of West African Languages.* Cambridge Univ. Press.

Ladefoged, P. (in press) "What are linguistic sounds made of." *Language.*

Ladefoged, P. and Traill, A. (1980) "Phonological features and phonetic details of Khoisan languages" in J. Snyman (ed.) *Proceedings of the Khoisan languages seminar.* Pretoria: University of South Africa.

Traill, A. (forthcoming) Phonetics and phonology of !Xhóõ Bushman.

Appendix

Check list of equipment for instrumental phonetic fieldwork.

1. Uher 4200 stereo recorder ($650.00).
   Uher of America, New Jersey

2. Pressure transducers
   Pitran R-pressure transistor, model PT-L2 and PT-H2 ($75.00 each)
   Stow Laboratories, Inc. Kane Industrial Drive, Hudson, Mass. 01749

3. Infant feeding tubes (pressure catheters)

4. Flashlight and dental mirror (for checking catheter position)

5. Plaster (for affixing catheters)

6. Bulb for clearing catheter

7. Disinfectant
   1. Webcol alcohol prep
   2. Unitek cleansing towelette
   3. Menda no fog dental mirror polish

8. T Tube and ruler (for calibrating pressure)

9. Air flow transducers
   Fleisch pneumotachograph # 1-7319 ($300.00)
   Dynasciences/Whittaker Corp. Towenship Line Rd, Blue Bell, Pa. 19422

10. Airflow meter

11. FM recording system

12. Tektronix portable storage oscilloscope
    Oscilloscope model 214
    Tektronix, P.O. Box 44408, San Francisco, Ca. 94144

13. Polaroid dental camera
    Polaroid Camera, CU-5 ($700.00)
    Polaroid Corp.

14. Palatography spray

15. Alginate impression material

16. Rubber mixing bowl

17. Spatula

18. Long sharp knife (for cutting impressions)

The ability of listeners to identify voices

Peter and Jenny Ladefoged

In many court cases people are identified simply by listeners who heard their voices. We have been concerned with various kinds of cases: a woman claimed to be able to identify a man who raped her by hearing his voice in a police line up; someone overheard a member of a street gang she knew outside her window; an employee surreptisiously taped a conversation showing her boss was biassed against her; relatives recorded evidence concerning an inheritance; and people have often made identifications from tapped telephone conversations. How accurate are the judgments of listeners in all these circumstances? The short answer is that we do not know, but there is often reason to doubt their reliability.

Bricker and Pruzansky (1976) have presented a comprehensive review of the literature on speaker recognition. Their 72 references include numerous unpublished lab reports and orally presented papers. About two thirds of their references are concerned primarily with machine techniques for speaker recognition or verifications, and are irrelevant to the present discussion. The remainder are concerned with recognition by human listeners, but even so some of them have no bearing on legal situations. We will discuss here published studies that are relevant in forensic proceedings, together with some previously unreported experiments of our own. The principal studies reported in the literature are summarised in Table. 1.

One of the earliest relevant studies from a legal point of view is that of McGhee (1937). Her research was prompted by the fact that Colonel Lindbergh identified the defendant Hauptmann by his voice almost three years after the offense was committed. McGhee summarises many early legal cases concerning the admissibility of voice identity as evidence. She notes that telephone conversations have been admitted into evidence for many years, "the uncertainty of recognition going only to the weight of evidence." McGhee then reports her own experimental investigations of voice recognition. The general procedure was to have unknown voices talking behind a screen. In some experiemnts there was a single voice on one day and five voices (including that one) either one day, two days or up to five months later. In these experiments 83% of the listeners identified the correct speaker one or two days later, but only 68% made correct identifications after a two week interval. In another experiment, when there were five unknown voices on the first day, only one of which occurred in a series of five voices two days

Table 1. Summary of some of the principal studies

Author(s)	Number of speakers	Conditions	(Mean) % recognized
McGhee	5	One day later	83
		two weeks later	68
		three weeks later	51
		five months later	13
Stevens et al	4	tape loops	94
	4 + others	may be unknown	90
Clarke & Becker	4	naive listeners	58
		speech students	63
		practised ditto	68
Bricker & Pruzansky	10	familiar voices	98
Compton	9	familiar, single vowel	54
Garvin & Ladefoged	12 (10 familiar)	might be familiar single vowel	42
Ladefoged	11 (out of 25)	assumed to be familiar	98

later, only 46% of the listeners made a correct identification. (As listeners were always choosing one out of five speakers, in all these experiments one would expect 20% of the listeners to be correct simply by chance). McGhee's summary notes "the reliability of court procedure in accepting testimony of positive identification of a defendant by his voice, in consideration of the length of time interval and the common fallibility of memory, would seem to be relatively low in the light of the present experiments designed to test the validity of such a procedure."

Many of the other studies are more concerned with laboratory situations that are not so relevant to most legal cases. Stevens, Williams, Carbonell, and Woods (1968), used recordings of words or very short phrases. In some of their experiments listeners could hear, over and over again, a loop of tape playing a recording of an unknown voice. They could also hear, as often as they liked, loops of four voices. In experiments in which they were told that the unknown voice was one of the four voices the error scores were about 6%. In experiments in which they did not know whether the unknown voice was

one of the four, they failed to identify the voice about 10% of the time. More importantly, even under these nearly ideal listening circumstances, they wrongly identified some other speaker as the unknown voice 6-8% of the time. We would presume that the error rate would be very much higher given poorly made tapes with a high background noise. Stevens et al also point out that there are large differences in the ability of their subjects to do this task.

In another fairly similar experiment conducted by Clarke and Becker (1969) a recording of an unknown voice was followed by recordings of four voices, one of which was the same as the unknown voice. This study provides evidence that both professional background and additional practice affect the number of correct responses. Naive listeners achieved only a 58% correct recognition, but five graduate students of speech made 63% correct identifications on the first day of the experiment. After several weeks of practice in which they became familiar with the voices of the 20 speakers involved, they made 67% correct identifications; they were still wrong about one time in three.

The experiments reviewed so far have been mainly concerned with the recognition of speakers not previously known. However, many legal cases are concerned with the identification of familiar voices. Experience suggests that there may be no difficulty in identifying the voice of an intimate friend; but even with such people one may be misled; and the identification is often not beyond a reasonable doubt. Casual acquaintances can easily be confused, and people with different accents may be very hard for an outsider to distinguish. Before discussing experiments illustrating these points it is useful to consider some general points concerning accents. It is a common observation that people of other races all look alike. The same kind of thing is true about speakers with accents very different from our own. They have an overwhelming similarity to one another, from our point of view, which masks what they may regard as very noticeable differences.

Some years ago Spencer (1957) remarked this point with reference to the perception of differences in accent. As he put it whether you have an accent or not depends on who is listening to you: "accent is in the hearer, not the speaker." Spencer notes that to an outsider most Yorkshire accents are indistinguishable; but to someone who lives in that region, there are numerous locally distinct forms of speech. Similarly many English people think that all Americans sound alike; and we find that many Americans mistake our British accents for Australian ones. Clearly the closer one is to a given group, the easier it is to distinguish particular members of that group. For all of us, an accent is something that somebody else has.

A similar point has been made in the speaker recognition literature. Bricker and Pruzansky (1976) quote Williams (1964, not seen) as saying "speaker recognition depends not only on the individual characteristics of

each of the speakers but also on the characteristics of the other speakers with whom he is being compared." Hecker (1971) notes that several studies have shown that some pairs of voices are more confusable than others. The similarities between different voices may be due to chance or to things such as family background. Rosenberg (1973) reports a study in which a speaker was confused with his twin brother 96% of the time. We have been involved in a legal case (People vs. Alcantara) in which a similar mistake was made. The prosecution originally alleged that a recorded voice was that of one person when subsequent evidence showed clearly that it was her sister.

Bricker and Pruzansky (1966) found that listeners were very good at identifying people they knew quite well. In their experiments 16 subjects listened to high quality recordings of 10 male speakers who were members of their lab group, each saying the same 5 sentences. When listening to these sentences in a random order, six subjects correctly identified all 50 sentences; the worst score was 46 correct; and the mean over all speakers and listeners was 98% correct. It is important to note that in this experiment the listeners were told who the 10 speakers would be, and had pictures of them in front of them during the listening task. They had also practised the task using similar sentences in which the speakers identified themselves. Thus this was a closed set discrimination task, rather than a test of a person's ability to identify a voice without knowing the range of familiar voices available, or even knowing for sure whether the voice to be identified was known or unknown.

Bricker and Pruzansky also played shorter, edited, utterances to their listeners, and suggest that these results support Pollack et al (1954) who say "we believe that the duration of the speech sample *per se* is relatively unimportant, except insofar as it admits a larger or smaller statistical sampling of the speaker's speech repertoire." It appears that even very short utterances convey some information about the speaker. Compton (1963) showed that listeners could identify familiar speakers from recordings of a single vowel. In a forced choiced task, 15 listeners, after a training period, could identify 9 speakers correctly 54% of the time. Again, this was a closed set discrimination task, in that the listeners knew that the speaker was one of a group of nine. Garvin and Ladefoged (1963) reported a similar experiment, but in this case the four listeners were given no prior training, and did not know in advance who the speakers were. They were told that they knew most of the speakers reasonably well, but there would also be some (in fact two out of 12) whom they did not know. After listening to high quality recordings of the vowel [ɑ], each lasting about two seconds, three of the four listeners could name four out of 10 speakers, and the fourth listener could name 5 speakers.

Our own experiments (Ladefoged 1978) also showed that listeners were very good at identifying the voices of speakers they know well. In these experiments listeners were not told in advance who the speakers were, but were allowed to presume that they were all associated with the UCLA Phonetics Lab group. Nine of the 10 listeners correctly identified all the 11 speakers who were associated with the lab group, usually within a few seconds of the start of the recording. The tenth listener confused two of the speakers who were not very well known to him. This experiment was not a closed set discrimination task in that there were at least 14 other people in the lab group who might have been among the speakers.

This experiment was also designed to show how expectations can affect the response. One is far more likely to identify a voice as a given person's if one is expecting to hear that person's voice. This may lead to some errors in identification. In the experiment being described a twelfth speaker was recorded who was not part of the lab group. Five out of the 10 listeners wrongly identified this voice as one of the two Blacks in the lab group, neither of whom had been recorded. Because the voice was that of a Black speaker who had some superficial resemblance to the two Blacks in the lab group,and because they were expecting to hear members of the group, half the listeners made a wrong identification.

We know of at least one legal case in which expectation was part of the cause of a misidentification. In People vs. Kalkin, narcotics agents telephoned a particular room at a hotel and made arrangements for a narcotics deal. They subsequently charged Mr. Kalkin, the person to whom the room was rented, alleging that they had spoken to him on the phone. But he had happened to be out of the room at the time of the call, and the agents had actually spoken to an associate of his, who later admitted to the conversation. We obtained a recording of this second person and are in no doubt whatsoever that the recording which had been used as evidence was much more similar to this recording than to recordings we also made of Mr. Kalkin. The prosecution later agreed to a stipulation that the voice on the recording in evidence was not that of Mr. Kalkin. They had apparently misidentified the voice because they had expected Mr. Kalkin to be answering the phone in that room.

In all the experiments discussed so far the set of possible speakers has been comparatively small. But Pollack et al (1954) have shown that the error rate goes up appreciably when the number of response alternatives is increased. In real life situations an identification often has to be made by considering a very large number of possible speakers. With this problem in mind we conducted a preliminary experiment on the ability of a listener to identify the voices of a wide range of friends and family interspersed with a number of unfamiliar voices. The subject was P. L., the first author, and the experimenter was J. L., the second author.

We estimated that at the time of the experiment we could identify a group of about 100 people who were reasonably well known to P. L., and whose voices he might be expected to be able to recognize. Recordings of 29 of these people were made by J. L. The speakers in this sample were selected in part by considerations of who was available, and who could be recorded without P. L. being aware that the recording was being made. But care was also taken to record only a proportion of those who were easily accessible, so as to make sure that P. L. could not presume that any given person would be in the sample. Thus recordings were made of P. L.'s mother, but not of his father. The voices considered as unfamiliar consisted of a group of 11 people who were totally unknown to P. L. and a further 13 people with whom he was not completely unacquainted, but whose voices he had not heard more than once or twice. All the recordings were made on a portable tape recorder with a frequency response of 50-12,000 Hz $\pm$ 3 db, and a signal to noise ratio better than 40 db. Three types of material were recorded: (1) The single word "hello". (2) A one sentence description of a picture (3) A further 30 seconds of continuous speech describing the same picture. Two additional types of material were derived from these recordings. Firstly, in order to check whether prosodic cues alone were sufficient for identification, the material was re-recorded after having been played back through part of a pitch extraction circuit which produced a saw-toothed wave that varied in frequency and amplitude. This buzzing type of sound conveyed the rhythm, intonation, and durational characteristics of the original speech without conveying any segmental information. Secondly, in order to check whether any of the speakers could be identified simply by their vocabulary or syntax, J. L. re-recorded all the material in her own voice.

P. L. listened to all the recordings using high quality equipment in a sound treated room. There were five parts to the listening test. He first heard all the examples of the buzzing sound that conveyed only prosodic information; next he heard all the material re-recorded in J. L.'s voice; then he heard all 52 examples of *hello;* then the 52 single sentences; and finally all the 30 second passages. The listening was done on a number of occasions over a period of several days. P. L. tried to identify each voice by name, noting for each identification whether he was certain, or fairly certain or whether it was only possibly the named person. Each recording was played through only once. He was not told whether he had made a correct identification or not until after the whole experiment had been completed, so as to make sure that he would not be biased by knowing which familiar voices had been recorded.

Listening to the buzz conveying only the prosodic cues did not enable him to identify anybody. Listening to the material re-recorded in J. L.'s voice he could sometimes again gain information such as whether it was a man or a woman speaking, and he could occasionally infer something about

their degree of familiarity with J. L. But he felt that usually he could gain very little specific information. This is a point which is hard to quantify. Undoubtedly vocabulary and style do convey some personal information. But, as is also the case for the prosodic data, by themselves they rarely convey enough information to permit identification. In the present experiment P. L. did identify one re-recording in J. L.'s voice as being possibly his mother-in-law. This identification was correct.

The main part of the experiment was listening to the 52 examples of *hello*, the single sentences, and the longer passages. The results are given in table 2, which indicates that 31% of the 29 familiar voices were correctly identified from the single word *hello*; 66% from a single sentence; and 83% from a longer sample of their speech. The proportion recognized with a higher degree of certainty increased steadily as the amount of material heard became longer.

Table 2. Number of speakers correctly identified with different degrees of certainty. There were 29 familiar voices interspersed with 24 other speakers.

Type of material	possible	fairly certain	certain	total	percentage
"Hello"	4	4	1	9	31
single sentence	8	3	8	19	66
30 seconds of speech	3	2	19	24	83

Table 2 does not show some information that is relevant to the present discussion. What is very important from a legal point of view is that on a number of occasions P. L. thought that he had heard one person when he had actually heard another. When listening to the 30 second passages, 24 people were recognised correctly, but in addition three wrong identifications were made. Two of these were cases in which an unfamiliar voice was mistaken for a familiar voice, and the third was a case in which a familiar voice was incorrectly identified. Thus about 11% of the identifications were wrong.

Almost equally interesting is the fact that the familiar voices who were not recognised included some who were very well known to P. L. Thus he did not recognize his own mother when she said *hello*, merely noting that the voice was that of a "familiar, low pitched, woman." He still did not recognize her when she said a single sentence; and it took the whole of the 30 second passage before he identified her, and then only in the "possibly" category. One reason for this may be that the sample contained 14 women who spoke with the same accent as hers (British English, RP), all of whom he found confusingly similar, despite being a speaker with that accent himself.

We have left till the end the question of *how* people recognize voices. Answers have been suggested by Voiers (1964) and Holmgren (1967) both of whom factor analysed rating scales such as intense-mild, hard-soft, sharp-dull. But as Bricker and Pruzansky (1976) comment in summarizing this literature: "We are left with the conclusion that the physical and perceptual correlates of speaker recognition by listening remain obscure."

From a linguistic point of view it seems preferable to go back over half a century to the work of Sapir (1927) who suggested that there are five ways in which individual voices differ from one another: (1) voice quality; (2) voice dynamics, by which Sapir meant that we would now call prosodic features such as intonation and rhythm, as well as relative continuity and speed; (3) pronunciation, which we would refer to as phonetic and phonological variations among segments; (4) vocabulary; and (5) style, including syntactic variations. In recognition situations we make use of various combinations of these factors. In our experiment described above, for example, on one occasion P.L.'s notes indicate that he heard a given voice as having North Country vowels, tried to remember who came from that area, and then realized that the voice was Trevor's. But on another occasion, when hearing a Gaelic influenced form of English, he did not go through that sort of process at all,but noted, instantly, Freddie. It seems foolish to presume that there is *a* way of recognizing voices. We can all do this task to some extent, no doubt using one technique on one occasion,and another on another. No doubt, also, we are all likely to make mistakes perhaps more than 10% of the time. We can only hope that our mistakes do not have serious, legal, consequences.

References

Bricker, P. and Pruzansky, S. (1976) 'Speaker recognition' in *Contemporary Issues in Experimental Phonetics*. (ed. N. Lass) Academic Press: New York.

Bricker, P. and Pruzansky, S. (1966) 'Effect of stimulus content and duration on talker identification'. *J. Acoust. Soc. Amer.* 40.6, 1441-1449.

Clarke, F. R. and Becker, R. W. (1969) 'Comparison of techniques for discriminating among talkers.' *J. Speech and Hearing Research.* 12, 747-761.

Compton, A. J. (1963) 'Effects of filtering and vocal duration upon the identification of speakers, aurally.' *J. Acoust. Soc. Amer.* 35.11, 1748-1752.

Garvin, P. and Ladefoged, P. (1963) 'Speaker identification and message identification in speech recognition.' *Phonetica.* 9. 193-199.

Hecker, M. (1971) *Speaker Recognition: An Interpretive Survey of the Literature.* ASHA Monograph Number 16, American Speech and Hearing Association, Washington, D.C.

Holmgren, G. L. (1967) 'Physical and psychological correlates of voice recognition.' *J. Speech and Hearing Research.* 10, 57-66.

Ladefoged, P. (1978) 'Expectation affects identification by listening' *Lang. and Speech.* 21.4, 373-4.

McGhee, F. (1937) 'The reliability of the identification of the human voice.' *J. Gen. Psychol.* 17, 249-271.

McGhee, F. (1944) 'An experimental study in voice recognition'. *J. Gen. Psychol.* 31, 53-65.

Pollack, I., Pickett, J.M. and Sumby, W.H. (1954) 'On the identification of speakers by voice.' *J. Acoust. Soc. Amer.* 26, 403-406.

Rosenberg, A. E. (1973) 'Listener performance in speaker verification tasks.' *IEEE Transactions, Speech Acoustics and Signal Processing* ASSP-23, 176-182.

Sapir, E. (1927) 'Speech as a personality trait.' *Amer. J. Sociol.* 32, 892-905.

Spencer, J. (1957) 'Received pronunciation: Some problems of interpretation.' *Lingua,* 7.

Stevens, K. N., Williams, C. E., Carbonell, J.R. and Woods, B. (1968) 'Speaker authentication and identification: a comparison of spectrographic and auditory presentations of speech material.' *J. Acoust. Soc. Amer.* 44, 1596-1607.

Voiers, W. D. (1964) 'Perceptual bases of speaker identity.' *J. Acoust. Soc. Amer.* 36, 1065-1073.

Williams, C. E. 'The effect of selected factors on the aural identification of speakers.' Section III of Report ESD-TDR-65-153 Electronics Systems Division, Air Force Systems Command, Hanscom Field (1964). [Not seen; cited in Bricker and Pruzansky, 1976].

Bibliography of X-ray Studies of Speech

Jonas N. A. Nartey

1. Introduction

This bibliography lists the publications dealing with or containing some x-ray data on speech that we have copies of in our files at the UCLA Phonetics Laboratory. We are aware that this list by no means exhausts the material in the field. We are publishing this solely as a working bibliography for our own use and as a convenience to any other interested users.

Preceding the bibliography itself is a partial annotation of the studies grouped according to topic. In this section, care has been taken to avoid duplication of the surveys completed by Macmillan and Kelemen (1952) and Simon (1961). This has been achieved by concentrating only on the more recent work. As the main reason for the present compilation was to locate as many possible of the published tracings or photographs, the annotations tend to concentrate on this aspect and may not reflect the main point of the paper as the author or another reader might perceive it.

2. Annotations

2.1 Works relating to experimental and analytical techniques

Bjork (1961) contains technical details of X-ray method and dosages.

Hardcastle's 1974 survey paper describes the application of various instrumental techniques used to obtain information on lingual activity during speech production. He suggests that even though cinefluorography, electromyography, and electropalatography are by far the most reliable techniques available, there is still room for improvement in any one of them. The paper includes one simplified tracing of a still x-ray photograph (6 x 5 cm) of Hardcastle's own production of [s]. The tracing represents the upper two-thirds of the oral cavity.

Houde's (1967) Ph.D. dissertation studied the movements of the tongue body and their relationships to the phonemic representation of speech. He used nonsense trisyllabic words throughout the study as spoken by an American English speaker. He suggests point-parameterization as a reliable means of dealing with the physical characteristics of the articulators. Curiously enough he includes only one 9 x 9 cm x-ray frame and a tracing of the same in the entire dissertation.

Kent (1972) includes three tracings (approximately 4 x 4 cm) showing superimposed tongues. Only one tracing includes the complete oro-nasal cavity. Motion films of varying speed (100 or 150 frames per second) were used in the study. At least two speakers of American English served as subjects. The sounds were taken from actual words, e.g. [k] from "coffee", and [t] from "toying". Kent investigated a number of techniques in the analysis of cinefluorographic data in the study of speech. He concluded that a point-parametized analysis is particularly advantageous in that it allows for the recording of spatio-temporal detail.

Kent, Netsell, and Bauer (1975) investigated x-ray techniques and their application to articulatory mobility in dysarthric subjects. They recommended the use of radio-opaque markers as they allow time/motion studies of discrete articulatory points.

Moll (1960 and 1965) deals with x-ray techniques and procedures in speech research. He notes (1965) that it is impossible to select measures with which to describe completely the position of the tongue.

Perkell (1969) describes various methods of measuring the resonance cavities, with reference to fixed points. From his data he suggests a "simplified model" for vowel production. He further suggests that variation in muscle activity corresponding to tenseness is more isotonic in vowels and more isometric for consonants.

Strenger (1968) includes eleven 10 x 12 cm, two 7 x 11 cm, and three 4 x 5 cm x-ray photographs. These are all stills of the consonants [f,m,n,ŋ] and one instance of the vowel [u] spoken by Swedish subjects. All parts of the speech organs are represented. The study extends Strenger's 1963 consonant studies to cover vowel articulation. He stresses the necessity of describing the different speech sounds with reference to the fixed physical rest point, which he defines as "that position in which all the parts of the organs of speech are in a state of rest, and when respiration can continue without the passage of air being in any way obstructed or altered.".

Subtelny, Pruzansky, and Subtelny (1957) discuss some of the problems of existing x-ray techniques and suggest modifications that might help capture the fact that speech is dynamic and not "static".

Truby (1962) based his studies on Swedish, and American English vowels preceded by CL clusters. He pointed out that the best means of collecting physical data on speech is the simultaneous use of radiographical techniques and sound spectrographs.

2.2. Studies principally concerned with vowels

Barth and Grassman (1907) contains x-rays of all the German vowels.

Benard's 1970 study was on selected vowel sounds of Australian English. He published only three tracings (approximately 5 x 4 cm) taken from motion films at a speed of 24 frames per second. Included in the tracings are the upper pharynx, upper incissors, the alveo-palatal ridge, the velum, and the upper two-thirds of the tongue. The tracings were from two of the twelve adult male speakers used in the study. The vowels shown from the study are [u], [i], [ɒ], and [ɛ] in the frame [hVd]. Each tracing consists of two superimposed vowels. He found among other things, that the tongue crests for [ɛ] and [u] are very close together, and that the perceptual difference is attributable to the effect of lip rounding.

Brichler-Labaeye (1970) is an extended work on the articulations of French vowels. It contains three hundred and ninety-nine 3 x 7 cm and one 8 x 11 cm tracings, all of which include two superimposed vowels. The main problem with this book is that it has tracings of only the anterior two-thirds of the vocal tract. This obviously excludes data on the area of the epiglottis and position of the larynx.

Curry (1938) includes eight 5 x 7 cm x-rays with superimposed tracings of the American English vowels [i:], [u:], and [a:], spoken by a female subject. The tracings were from a movie film with a speed of 60 frames a second.

As indicated by the title, Fant's (1964) paper explores the relationship between formants and cavities. There are a number of interesting remarks, one being that the highest point of the tongue is not an accurate articulatory representation of the vowel [ɑ], since the major difference between [ɑ] and all front vowels is the relative narrow pharynx. Most of his data was based on his monumental (1960) book.

Fónagy (1976) studied Hungarian vowels said with varying emotions, namely: anger, hate, sadness, and irony. (The study is published in French).

Gendron (1962) compared both oral and nasal vowels of Canadian and Parisian French. His main conclusion is that Canadian vowels generally tend to be more lax than their Parisian counterparts.

Holdbrook and Carmody (1937) included 184 single tracings from still photographs. The approximate size of the tracings, which include all of the oral cavity, is 3 x 2.5 cm. The number of speakers per language are French - 2 females and 1 male, Spanish - 1 male and 1 female, American English - 2 males and 1 female and 1 speaker each of British English, Italian, German, Polish,and Russian. The study was based on steady state vowels.

Lindblom and Sundberg (1969-70) applied a quantitative model of cardinal vowel production to the distinctive features of Swedish vowels, suggesting how such "cardinal" vowels may be used to teach language in the classroom. The study included six 6 x 6 cm superimposed tracings and one 9 x 10 cm single tracing.

MacNeilage and Scholes (1964) studied movements of the tongue during the production of American English vowels. Using both EMG and x-ray data, the authors were usually able to show the muscles controlling the tongue during the production of vowels.

Moll (1962, 1967, 1971) studied velopharyngeal closure and the systematic movement of the velum during speech. Seemingly contrary to other authors (e.g. Perkell, 1969), Moll suggests (1962) a greater velopharyngeal closure for high vowels than low ones.

Skaličková (1967) did a comparative study of English and Czech vowels. She indicated among other things that the acoustic differences between [e] and [i] are not necessarily achieved through differences in tongue position, but that mandibular variation is also involved. Her data indicate that the so-called front vowels of English are further back than the Czech ones. Included are fifteen 10 x 12 cm tracings taken from connected speech of 2 Czech and one English subject.

Ladefoged (1964) included radiographic data on 13 vowels in Ngwe and 10 vowels in Igbo, each from one speaker. A principal conclusion is that Igbo vowel harmony can be characterized in terms of advancement of the tongue root - since all the so called "tense" vowels have a wider pharynx than all the so called "lax" vowels. This book also includes tracings of the implosive [ɓ] in Igbo and the labiodentalized fricative [z^{v}] in Kutep.

Painter (1973) provided x-ray data on the feature 'covered' in Twi (Akan) vowel harmony. The article includes 14 tracings, one of which superimposes 10 vowels on one frame, of the approximate size 5 x 5 cm. The tracings are from a motion film at a speed of 24 frames per second. The alveo-palatal ridge, the velum, the pharyngeal wall, the epiglottis, and most of the tongue, are visible in the tracings. One adult male speaker of Twi served as subject. The sounds were extracted from short Twi sentence pairs of the type "mì dì"and "mì dì". The data points to the fact that in Twi, /e/ is more like /i/ than /ɪ/, and /ɛ/ is more like /ɪ/ than /e/ both in tongue height and tongue root position.

Lindau's 1975 dissertation was a study of vowels in the African languages Akan, Ateso, Dho-Luo, Igbo, and Ijo. There are a number of original superimposed 6 x 8 cm tracings obtained from 4 speakers of Akan. She suggested that in Akan, at least, vowel harmony can be described in terms of [± advanced tongue root], or [± expanded]. (cf. Painter, 1973).

Strenger (1969) showed that oral vowels in French actually have a higher tongue position than nasal vowels.

Takeuchi's 1961 studies were based on Japanese monosyllables. Thirteen 5 x 6 cm x-ray photographs are included here. The main body of the text is written in Japanese, but the tables are in English.

Wängler (1961) published both the original x-rays and superimposed tracings of steady state articulations of all the German sounds. One problem with his tracings is that he has two epiglottises for a number of the sounds, thus indicating he was not very sure of the outlines. (cf. Wängler under 2.3 below).

Kent and Moll (1972) investigated tongue body articulation during vowel and diphthong articulation in American English speakers. Their findings point to the importance of tongue body displacement in determining the rate of tongue body movement.

Pétursson's 1974 study dealt mainly with the vowels of modern Icelandic. This study includes 26 tracings (approximately 4 x 3 cm) from motion films at a speed of 19 frames per second. Each tracing is a composite of at least three stages of a given vowel. With the exception of the tongue root and the lower pharynx, all of the oro-nasal cavity is shown in the tracings. The subject was a native Icelander. The sounds were extracted from real words, e.g. /ɪ̯/ from "bytta", and [ɛ̄] from "Bettu". In this study Pétursson showed that using lateral cineradiography alone, one could arrive at a vowel system that is totally different from traditional ones.

Hegedüs (1937) included fourteen 2 x 3 cm tracings of vowels from German, Hungarian, Spanish, and American English. Due to the size one cannot do much with the tracings. He showed the similarities as well as differences between the vowels of the language in question.

Subtelny, Pruzansky and Subtelny (1957) included one (5 x 7 cm) still x-ray of sustained phonation of American English [u]. Even though all parts of the oral and nasal cavity are shown, the x-ray is not very readable. An x-ray taken from Norris (1934) is also included.

Sovijärvi (1962) was mainly on Finnish diphthongs.

2.3. Studies principally concerned with consonants

Kent and Moll (1975) were interested in the articulatory timing of /spr/, /spl/, and /skw/ sequences in American English speakers. Their study endorsed a preprogramming model over a feedback model of motor timing.

Monnot and Freeman (1972) compared the Spanish single-tap /r/ with American /t/ and /d/ in post-stress intervocalic position. The study has six 4 x 4 cm tracings of an x-ray movie of the speed 24 frames per second. The tracings show all aspects of the oro-nasal cavity from just above the glottis to the lips. These are from one out of three speakers each of Spanish and American English. The sounds were all extracted from continuous speech, e.g. [a,ɾ,ə] from "water", and [e,ɾ,i] from "*Ibérica*". They concluded that the two sounds are so alike that language teachers may appeal to students' knowledge of one in producing the other.

Pétursson (1971) studied [θ], [ð], and [s] in Icelandic. He published 6 tracings (about 5 x 4 cm) from high speed x-ray films. The tracings show every articulator in the upper two-thirds of the oro-nasal cavity. The lower pharynx and tongue root were not included. All tokens were from the same speaker, a native of south Iceland. The sounds were from real words of the language, e.g. "hjálpaðu." He showed that all three consonants were produced in the alveolar region, thus refuting earlier suggestions that Icelandic [θ] and [ð] were apico-dental articulations.

Subtelny, Oya, and Subtelny (1972) continued their earlier studies of sibilants. They included three 7 x 7 cm and two 4 x 4 cm tracings of [s] and [z] from real American English words such as "Suzy", "sister", "saw", and "sustain". Their findings support earlier observations that vowel reduction is a characteristic feature of English. Also, they suggested that open vowels had a greater effect on lip opening for sibilants than did closed vowels.

Wood (1975) investigated the differences between the English and Swedish dental stops. He published 15 tracings (approximately 7 x 8 cm). These are from x-ray motion films at a speed of 75 frames per second. The tracings include the lower jaw, the tongue, and sometimes, the alveo-palatal ridge. They also include a slight approximation of the epiglottis. He used two subjects (one Swedish and one English). The sounds were extracted from continuous speech, e.g. [n] from *nitton*, [t] from *nitton*, and [t] from *sjuttio*. All tracings have preceding vowels superimposed. The [s]- like quality of the English [t] burst was attributed to a laminal occlusion.

Feldman (1972) studied Brazilian Portuguese utterance-final [ɸ] and [ɰ], noting that the two sounds are very much alike both acoustically and articulatorily. The study includes four 4 x 3 cm tracings from still x-rays, showing every feature of the oro-nasl cavity above the larynx. Due to the size of the print one could could conclude that most of the outlines were only approximated. There was no information on the specific number of speakers used, though they were identified as both male and female adults of four major Brazilian Portuguese dialects. Both nonsense and real words were used.

Giles and Moll (1975) examined x-ray pictures of selected allophones of English [l]. The study includes 22 single and 6 heavily superimposed tracings from a high speed motion film (150 frames per second). The tracings (approximately 3 x 3 cm) are of the upper two-thirds of the tongue, and the alveo-palatal ridge only; the velum, for instance is not included. Three adult speakers of Midwestern American English served as subjects. All tokens were from continuous speech samples. Among their findings, the authors showed that there are similar constraints on the shape of the back portion of the tongue for vowels and the /l/ allophones studied. Also, prevocalic /l/ allophones showed a more anterior tongue position than postvocalic ones.

Wängler's 1961 study includes thirty 7 x 9 cm original x-rays and superimposed tracings of steady state articulations of all the German sounds. Regular photographs of the mouth, and also palatograms accompanying the x-rays.

Delattre (1971) studied pharyngeal features in Arabic, German, Spanish, French and American English. He included a total of 157 tracings of approximately 2 x 2 cm from a motion film of the speed 24 frames per second. Even though the entire cavity (from the lower pharynx to the lips) is shown, users may find that the tracings are too small to make much of. The number of tokens range from one to four per sound. Both sentences and minimal pairs were employed as frames for the sounds studied. Examples of words used are German "stache", Spanish "jiba", and American English "later." He concluded that many speech sounds that had been classified as velars were more pharyngeal than velar. He pointed out that German /x,ɾ/, Spanish /x/, and French /R/ are all marked by pharyngeal constrictions.

Hetzron (1969) based his work on the articulation of semitic laryngeals. He includes 6 tracings of Arabic [ʕ], of the very small size of 1 x 1 cm. He points out that due to the very low nature of pharyngeal constriction accompanying Arabic laryngeals, the velum is pulled away from the pharyngeal wall, thus creating a velic opening almost identical to that found in nasal vowels.

2.4. Studies concerned with coarticulation between segments

Carney and Moll (1971) investigated fricative consonant vowel co-articulation in American English. The publication includes ten tracings (5 x 3 cm), each being a composite of 3 or 4 superimposed utterances, taken from a motion film of 100 frames per second. Only the upper two-thirds of the oral cavity is shown here. In addition to the confusion created by superimposing, these tracings only seem to approximate the velum. The subjects were two young adults (one male, one female). Each token of the form [hVCV] utilized one or two of the vowels /i,a,u/ and one of the consonants /s,z,f,v/, resulting in nonsense words of the kind [hisi] and [havu]. Their findings confirm earlier suggestions of vowel-to-vowel tongue movement with a superimposition of the fricative consonant. Of particular interest is their finding that fricative production is more affected by post-consonantal vowels than by pre-consonantal vowels.

Daniloff and Moll (1968) used high speed cineradiography for their study, involving the syllables [bu], [stu], [stru], and [nstru] spoken by three English speaking subjects. The most interesting result is that the lip rounding for [u] begins as early as possible in the sequence of preceding consonants.

Sovijärvi (1959) showed how the central portions of Finnish short [ɑ] and [i] are influenced by labials, palatals, velars, and dentals. The publication includes seven tracings (approximately 4 x 4 cm). Two of these have extra vowels superimposed on them. All articulators in the oro-nasal cavity (including the lower pharynx) are shown. He used 2 subjects (one adult and one child), saying real Finnish words. The speed of the film used was 24 frames per second.

2.5. Studies involving altered, unusual or defective speech

Croatto (1966) studied French subjects with various malformations in the larynx and pharyngeal cavity in general.

Laczkowska (1961) studied the function of the velum in the speech of boys between the ages of 7 and 12. He used both normal and pathological subjects (including stuttering, dyslalia, and hypernasality). He found marked differences in the velar function in normal and pathological cases.

Novák (1972) based his study on a group of Czech children with Down's syndrome and another group of mentally retarded children without Down's syndrome. He explained the fusion of the first two formants in certain vowels produced by children with Down's syndrome as due to "loose connection of resonating cavities."

Huizinga (1931) included ten 11 x 13 cm tracings of the vowels [i,u, e,o,a], said once with ventriloquist voice, and once with normal voice. The x-rays are not particularly readable. The tracings are not consistent with respect to what is included. His 1932 article included four 6 x 9 cm x-rays of [i] and [u] said once normally and once with a bite-block, and ten 3 x 5 cm tracings of [a,i,œ,o,e] also spoken with and without the bite block.

Ondráckova's 1973 book, with its very rich appendix of x-ray photographs, studied the articulation of sung, spoken, and whispered Czech long vowels. Included are four 10 x 15 cm, twenty-four 3 x 3 cm, and seventy-four 7 x 5 cm x-rays. One of her findings indicates the progressive enlargement of the oral cavity from whispered through spoken to sung vowels. She maintains that while there is no regular specification for the oral cavity, the overall relation between the capacity of the oral cavity and that of the throat is important for discriminatory vowel perception. Her 1961 results (based on tongue and velum movements during singing) revealed air to be the most satisfactory natural substance for various x-ray pictures.

Sundberg (1969-70) mainly deals with articulatory differences between sung and spoken Swedish vowels. His study shows that spoken vowels have higher first and second formant frequencies than sung vowels. The publication includes one 15 x 15 cm tracing of superimposed spoken and sung vowels.

Putnam and Ringel (1976) used American English subjects with temporarily induced oral sensory deprivation to investigate mandible position. Their findings show among other things that more than half the time, inferior/superior mandible position relative to the maxilla is either closer or less close than traditionally claimed.

Weinberg's 1968 study compared normal and defective [s] articulation. The selection of his subjects was based on variations in incisor dentition. He found excessive fronting of the tongue tip to be the main cause of defective [s] articulation.

3.0 Bibliography

Ali, L. H., and R. G. Daniloff (1972) 'A cinefluographic-phonologic investigation of emphatic sound assimilation in "Arabic"'. In A.Rigault and R. Charbonneau (eds.) *Proc. of the 7th Intern. Congress of Phonetic Sciences*. Montreal 1971. The Hague: Mouton.

Amerman, J. D., R. Daniloff, and K. Moll (1970) 'Lip and jaw coarticulation for the phoneme /æ/.' *J. Speech and Hear. Res.* 13: 146-161.

Barth, E. and E. Grassman (1907) 'Röntgenographische Beiträge Zur Stimmphysiologie.' *Archiv für Laryngologie und Rhinologie.* 19: 396-407.

Benard, J. R. L-B (1970) 'A Cine-x-ray study of some sounds of Australian English.' *Phonetica.* 21: 138-150.

Benson, D. (1972) 'Roentgenographic cephalometric study of palato pharyngeal closure of normal adults during vowel phonation.' *Cleft Palate Journ.* 9.9: 43-50.

Björk, L. (1961) 'Velopharyngeal function in connected speech, studies using tomography and cineradiography synchronized with speech spectrography.' *Acta Radiologica Suppl.* 202.

Bothorel, A. (1971) 'A propos du Breton parlé à Drjol: quelques observations sur les consonnes geminées.' *Travaux du L'institut de phonétique de Strasbourg.* 3: 195-233.

Brichler-Labaeye, C. (1970) 'Les voyelles françaises: mouvements et positions articulatoires à la lumière de la radiocinématographie.' *Bibl. fr. et romane* A, 18. Paris: Klincksiek.

Butt, A. H. (1973) Articulation with reduced oral sensory control: a cineradiographic study. Ph.D. dissertation. Purdue University.

Carmody, F. J. (1936) 'Radiographs of thirteen German vowels.' *Archives Néerlandaises de Phonetique Experimentale.*

Carmody, F. J. (1941) 'An x-ray study of pharyngeal articulation.' *University of California Publications of Modern Philology.* 21.5: 377-384.

Carney, P. J. and K. L. Moll (1971) 'A cinefluorographic investigation of fricative consonant-vowel coarticulation.' *Phonetica.* 23: 193-202.

Chiba, T., M. Kajiyama (1958) 'The vowel: its nature and structure.' *Tokyo-Kaiseikan Pub. Co.* Tokyo. Second edition published by Phonetic Society of Japan.

Chlumsky, J. (1938) *Radiographies des voyelles et des semivoyelles françaises.* Prague.

Croatto, L. (1966) 'Les malformations congénitales du tractus vocal.' In *Phonétique et Phonation.* (Ed. by A. Moles and B. Vallancien, Masson et Cie: Paris. 91-112.

Curry, R. (1938) 'The physiology of the contralto voice.' *Archives Néerlandaises de Phonetique Experimentale.* 14: 73-79.

Daneš, F., B. Hála, A. Jedlička, and M. Romportl (1954) *O Mluveném Slové.* Praha.

Daniloff, R. and K. Moll (1968) 'Coarticulation of lip rounding.' *J. of Speech and Hear. Res.* 11.4: 707-721.

DeClerk, J. L., L. S. Landa, D. L. Phyje, and S. I. Silverman (1965) 'Cinefluorography of the vocal tract.' In D. E. Cummins (ed.) *Proc. of the 5th Int. Congress on Acoustics.*

Delattre, P. (1971) 'Pharyngeal features in the consonants of Arabic, German, Spanish, French, and American English.' *Phonetica.* 23: 129-155.

Fant, G. (1961) *Acoustic Theory of Speech Production, with Calculations Based on X-ray Studies of Russian Articulations.* The Hague: Mouton.

Fant, G. (1965) 'Formants and cavities.' *Proc. of the 5th Int. Congress of Phonetic Sciences.* (Ed. by E. Zwirner, and W. Bethge.) Basel. 120-141.

Fant, G. (1968) 'Analysis and synthesis of speech processes.' *Manual of Phonetics.* (Ed. by B. Malmberg.) Amsterdam. 173-277.

Fant, G. (1969-70) 'Distinctive feature and phonetic dimensions.' STL-QPSR 2-3/1969, 1-18.

Feldman, D. (1972) 'On utterance-final [ɤ] and [u] in Portuguese.' *Papers in Linguistics and Phonetics to the Memory of Pierre Delattre.* (Ed. by Albert Valdman.) The Hague: Mouton. 129-142.

Fónagy, I. (1976) 'La mimique bucale. Aspect RAdiologique de la vive voix.' *Phonetica.* 33: 31-44.

Gay, T. (1974) 'A cinefluographic study of vowel production.' *J. Phonetics* 2: 255-266.

Gendron, J-D. (1962) 'La méthode radiographique appliquée à la comparaison des articulations vocaliques en Français Canadien et en Français Parisien.' *Proc. of the 4th Intern. Congress of Phonetic Sciences.* Helsinki, 1961. The Hague: Mouton. 155-166.

Giles, S. B., and K. L. Moll (1975) 'Cinefluographic study of selected allophones of English /l/.' *Phonetica.* 31: 206-227.

Hála, B., and M. Sovak (1955) *Hlas-řeč-sluch.* (základuvěciz fonetiky a logopedie). Praha.

Hála, B. (1956) *Nature Acoustique des Voyelles.* Nákladem Karlovy University. V Praze.

Hála, B. (1959) *English Vowels in Phonetic Pictures.* Prague.

Hála, B. (1962) *Uvedení do Fonetiky Češtiny: Na Obecne Fonetickém Základě.* Praha.

Hardcastle, W. J. (1974) 'Instrumental investigations of lingual activity during speech: A survey.' *Phonetica.* 29: 129-157.

Hegediis, L. (1937) 'Roentgenaufnahmen von ungarischen Vokalen.' ANPE 13: 72-77.

Hetzron, R. (1969) 'Two notes on semitic laryngeals in East Gruage.' *Phonetica.* 19: 69-81.

Holbrook, R. T., and F. J. Carmody (1937-41) 'X-ray studies of speech articulation.' *University of Calif. Press Publications in Modern Philology.* 20.

Horn, W. (1936) 'Experimental phonetic und Sprachgeschichte.' *Proc. of the 2nd Intern. Congress of Phonetic Sciences.* ed. by D. Jones and D. B. Fry. 12-18. Cambridge.

Houde, R. A. (1967) 'A study of tongue body motion during selected speech sounds.' Ph.D. thesis. University of Michigan.

Huizinga, E. (1931) 'Recherches sur un ventriloque Néerlandais.' *Archives Néerlandaise de Phonetique Experimentale.* 6: 66-87.

Huizinga, E. (1932) 'Über die stello, wo der Charakter des Selbstlautes Gebildet Wird.' *Archives Néerlandaises de Phonetqiue Experimentale.* 17: 104-117.

Huizinga, E. and A. Moolenae-Bijl (1941) 'Kompensierung im asatzstuck bei der Bildung von Selbstlauten.' *Archives Néerlandaises de Phonétique Expérimentale.* 17: 1-8.

Husson, R. (1962) *Physiologie de la Phonation.* Masson et Cie: Paris.

Jones, D. (1922) *An Outline of English Phonetics.* New York.

Jones. E. (1956a) *The Pronunciation of English.* 4th edition. Cambridge: The University Press. 1st edition, 1909.

Jones, S. (1929) 'Radiography and Pronunciation.' *Proc. of the Society of Radiographers. The British Journ. of Radiology.* N.S. 2: 149-150.

Kent, R. D. (1972) 'Some considerations in the cinefluorographic analysis of tongue movements during speech.' *Phonetica.* 26: 16-32.

Kent, R. D. and K. L. Moll (1969) 'Vocal-tract characteristics of the stop cognates.' *J. Acoust. Soc. Amer.* 46.6: 1549-1555.

Kent, R. D. and K. L. Moll (1972) 'Cinefluographic analysis of selected lingual consonants.' *J. Speech Hearing Res.* 15: 453-473.

Kent, R. D. and K. L. Moll (1972) 'Tongue body articulation during vowel and diphthong gestures.' *Folia Phoniat.* 24: 278-300.

Kent, R. D. and K. L. Moll (1975) 'Articulatory timing in selected consonant sequences.' *Brain and Language* 2: 304-323.

Kent, R. and R. Netsell (1975) 'A case study of an altaxic dysarthric: cineradiographic and spectrographic observations.' *J. Speech and Hearing Dis.* 40: 119-134.

Kent, R. D., R. Netsell, and L. L. Bauer (1975) 'Cineradiographic assessment of articulatory mobility in the dysarthrias.' *J. Speech and Hearing Dis.* 40: 467-480.

Kirkpatrick, J. A. and W. R. Olmsted (1959) 'Cinefluorographic study of pharyngeal function related to speech.' *Radiology.* 73: 557-559.

Kojima, G. (1971) 'Problemies de coarticulation en Japonais.' Travaux de L'institute de Phonetique de Strasburg. 3: 158-172.

Laczkowska, M. (1961) 'Concerning the function of the velum.' *Folia Phoniat.* 13: 107-111.

Ladefoged, P. (1964) *A Phonetic Study of West African Languages.* Cambridge: University Press.

Ladefoged, P., J. DeClerk, M. Lindau, and G. Papçun (1972) 'An auditory-motor theory of speech production.' *UCLA Working Papers in Phonetics.* 22: 48-75.

Lindau, M. (1975) 'Features for vowels.' *UCLA Working Papers in Phonetics.* 30.

Lindau, M. (1979) 'The feature expanded.' *Journ. of Phonetics.* 7, (Forthcoming).

Lindblom, B. and J. Sundberg (1969-70) 'A quantitative theory of cardinal vowels and the teaching of pronunciation.' *STL-QPSR* 2/3/1969. 19-26.

Lindqvist, J., M. Sawashima, and H. Hirose (1973) 'An investigation of the vertical movement of the larynx in a Swedish speaker.' *Annual Bulletin* 7. Research Inst. of Logopedics and Phoniatrics. U. of Tokyo.

Macmillan, A. S., and G. Keleman. (1952) 'Radiography of the supraglottic speech organs, a survey.' *A. M. A. Archives of Otolaryngology* 55: 671-688.

MacNeilage, P. F. and G. N. Scholes (1964) 'An electromyographic study of the tongue during vowel production.' *J. Speech and Hearing Res.* 7: 209-232.

Mazlvoâ, V. (1967) *Vijslovnost na Zábřežsku.* [La pronunciation aux environs de Zábřek]. Prague.

Miller, E. R. (1959) 'Cinefluorography in practice.' *Radiology.* 73: 560-564.

Miletić, B. (1933) *Izgovor srpskochravataskich glasova* (Pronunciation des son serbocroates). Belgrade.

Moll, K. L. (1960) 'Cinefluorographic techniques in speech research.' *Journ. Speech and Hearing Res.* 3.3: 227-241.

Moll, K. L. (1962) 'Velopharyngeal closure in vowels.' *Journ. Speech and Hearing Research.* 5.1: 30-37.

Moll, K. L. (1965) 'Photographic and radiographic procedures in speech research.' *ASHA Reports.* 1: 129-139.

Moll, K. L. (1971) 'Investigation of the timing of velar movements during speech.' *Journ. Acoust. Soc. Amer.* 50.2: 678-684.

Moll, K. L. and T. H. Shriner (1967) 'Preliminary investigation of a new concept of velar activity during speech.' *The Cleft Palate Journ.* 4.1: 58-69.

Monnot, M. and M. Freeman (1972) 'A comparison of Spanish single-tap /r/ with American /t/ and /d/ in post-stress intervocalic position.' *Papers in Linguistics to the Memory of Pierre Delattre.* ed. by Albert Valdman, The Hague: Mouton. 409-416.

Navarro Tomás, T. (1916) 'Siete Vocales Españolas.' *Revista de Filologia Española.* 3: 51-62.

Novák, A. (1972) 'The voice of children with Down's Syndrome.' *Folia Phoniatrica.* 24: 182-194.

Öhman, S. E. G. (1965) 'X-ray studies of articulatory dynamics.' *Proc. of the 5th Intern. Congress of Phonetic Sciences.* Munster 1964 (Bassel-New York), 16-23.

Öhman, S. E. G. (1966) 'Coarticulation in VCV ytterances: Spectrographic measurements.' *Journ. Acoust. Soc. Amer.* 39. 151-168.

Ondráčková, J. (1960b) *Artikulace českyéh zpivaných samohlásek.* Dissertation.

Ondráčková, J. (1961) 'The movement of the tongue and the soft palate in the singing of vowel.' *Folia Phoniat.* 13. 99-106.

Ondráčková, J. (1964) *Rentgenologický Výzkum Articulace Českých Vokálu.* Praha.

Ondráčková, J. (1973) *The Physiological Activity of the Speech Organs: An Analysis of the Speech Organs During the Phonation of Sung, Spoken and Whispered Czech Vowels on the Basis of X-ray Method.* The Hague: Mouton.

Painter, C. (1973) 'Cineradiographic data on the feature 'covered' in Twi vowel harmony.' *Phonetica.* 28. 97-120.

Parmenter, C. E. and S. N. Trevino (1932) 'Vowel positions as shown by x-ray.' *The Quarterly Journ. of Speech.* 18. 351-369.

Parmenter, C. E., and C. A. Bevans (1933) 'Analysis of speech radiographs.' *American Speech.* 8.3: 44-56.

Perkell, J. S. (1969) *Physiology of Speech Production: Results and Implications of a Quantitative Cineradiographic Study.* Res. Monograph No. 53. The M.I.T. Press, Cambridge, Mass.

Peterson, G. E. (1957) 'Laryngeal vibrations' *Manual of Phonetics.* Ed. by L. Kaiser, Amsterdam.

Pétursson, M. (1971) 'Etude de al réalisation des consonnes islandaises þ,ð,s, dans la prononciation d'un sujet islandais à partir de al radiocinématographie.' *Phonetica.* 23. 203-216.

Pétursson, M. (1974) 'Peut-on intepréter les donées de la radiocinématographie en fonction du tube acoustique à section uniforme?' *Phonetica.* 29. 22-79.

Putnam, A. H. B., and R. L. Ringel (1976) 'A cineradiographic study of articulation of two talkers with temporarily induced oral sensory deprivation.' *Journ. Speech and Hearing Res.* 19. 247-266.

Retard, G. L. A. (1972) 'L'Agni-variété dialectale sanvii phonologie, analyses tomographiques documents.' *Annales de L'université d' Abidjan* - série H-v-Fascicules 1.

Russell, G. O. (1928) *The Vowel.* The Ohio State U. Press: Columbus.

Russell, G. O. (1929-30) 'The mechanism of speech.' *Journ. Acoust. Soc. Amer.* 1: 83-109.

Russell, G. O. (1931) *Speech and Voice.* Macmillan Co. New York.

Russell, G. O. (1933) 'First preliminary x-ray consonant study.' *Journ. Acoust. Soc. Amer.* 5. 247-251.

Russell, G. O. (1936) 'Synchronized x-ray, oscillograph, sound and movie experiments, showing the fallacy of vowel triangle and open-closed theories.' *Proc. of the 2nd Intern. Congress of Phonetic Sciences.* London, 1935. Cambridge. 198-204.

Santerre, L. (1971) 'La de limitation d'un continuu phonétique.' *Travaux de L'institute de Phonetique de Strasbourg.* 3. 185-195.

Simon, P. (1961) 'Films radiologiques des articulations et les aspects génétique des son du langage.' *Orbis.* 10. 47-68.

Skaličkova, A. (1955) 'The Korean vowels.' *Archivu Orientální.* 23. 29-51.

Skaličkova, A. (1967) 'A radiographic study of English and Czech vowels.' *Phonetica Pragnesia.* 1. 29-44.

Smith, T. (1971) 'A phonetic study of the function of the extrinsic tongue muscles.' *UCLA Working Papers in Phonetics* 18.

Sonninen, A. (1962) 'Paratasis-gram of the vocal folds and the dimensions of the voice.' *Proc. of the 4th Intern. Congress of Phonetic Sciences.* Helsinki, 1961. The Hague: Mouton.

Sovijärvi, A. (1959) 'Über die Veranderlichkeit der Zungenstellung und der entsprachenden akustischen Schwankungsgebiete der Vokale auf Grund eines Röntgentonfilms gesprochener finnischer Sätze', Ph 4, Supplement, 74-84.

Sovijärvi, A. (1962) Röntgenkinematografisch-akustische Untersuchungen über die Artikulation der Diphthonge. *Proc. of the 4th Intern. Congress of Phonetic Sciences.* Helsinki, 1961. The Hague: Mouton. 111-128.

Strenger, F. (1963) 'Untersuchung Schwedischer Konsonanten nach der indirekten Palatogramm-Methode.' *Zeitschrift für Phonetik.* 16: 211-216.

Strenger, F. (1968) 'Radiographic, palatographic, and labiographic methods in phonetics.' *Manual of Phonetics*, ed. by B. Malmberg. Amsterdam 334-364.

Strenger, F. (1969) *Les Voyelles Nasales Francaises.* Gleerup, Lund.

Subtelny, J. D., S. Pruzansky, and J. Subtelny (1957) 'The application of roentgenography in the study of speech.' *Manual of Phonetics.* ed. by L. Kaiser. Amsterdam: No. Holland Publishing. 166-179.

Subtelny, J. D. and J. Subtelny (1962) 'Roentgenographic techniques and phonetic research. *Proc. of the 4th Intern. Congress of Phonetic Sciences.* Helsinki, 1961. The Hague: Mouton. 129-146.

Subtelny, J. D., J. C. Mestre, and J. D. Subtelny (1964) 'Comparative study of normal and defective articulation of /s/ related to mal-occlusion and deglutition.' *J. Speech Disorders.* 29. 269-285.

Subtelny, J. D., N. Oya, and J. D. Subtelny (1972) 'Cineradiographic study of sibilants.' *Folia Phoniatrica.* S. Karger, Basel, 24. 30-50.

Sundberg, J. (1969) 'Articulatory differences between spoken and sung vowels in singers.' Stockholm RIT STLQPSR 1. 33-46.

Takenuchi, Y. (1961) 'Perceptual segmented Japanese monosyllables.' *Studia Phonologica*, ed. by Hisanosuke Izui, Univ. of Kyoto, Japan. 70-85.

Truby, H. M. (1962) 'Synchronized cineradiography and visual-acoustic analysis.' *Proc. of the 4th Intern. Congress of Phonetic Sciences.* Helsinki, 1961. The Hague: Mouton. 265-279.

Wada, T., M. Yasumoto, N. Ikeolea, Y. Fiyuki, and R. Yoshinaga (1970) 'An approach for the cinefluorographic study of articulatory movements.' *Cleft Palate Journal* 7. 502-522.

Wängler, H-H. (1958) *Atlas deutscher Sprachlaute.* Berlin: Akademie-Verlag.

Weinberg, B. (1968) 'A cephalometric study of normal and defective /s/ articulation and variations in incisor dentition.' *J. Speech Research.* 11. 288-300.

Wood, S. (1975) 'What is the difference between English and Swedish dental stops.' *Working Papers, Phonetics lab. Lund University* 10. 173-193.

Wood, S. (1975) 'The weakness of the tongue-arching model of vowel articulation.' *Working Papers, Phonetics Lab. Lund University* 11. 55-108.

Zwirner, E. (1936) 'Speech and speaking' *Proc. of the 2nd Intern. Congress of Phonetic Sciences.* ed. by Daniel Jones and D. B. Fry. Cambridge. 239-245.

The phonetic function of rise and decay time in speech sounds: A preliminary investigation

Vincent J. van Heuven

[Paper presented at the 9th International Congress of Phonetic Sciences, 6-11 August 1979]

Speech sounds may differ characteristically in a number of ways, among which are normally listed spectral composition, periodicity, intensity and duration. Though a certain average intensity may be associated with a particular (type of) speech sound (e.g. vowels have greater intensity than consonants; vowel intensity increases with degree of openness), no speech sound has a level intensity throughout its duration. It has been suggested that the time interval during which the amplitude envelope increases more or less steadily at the beginning of a speech sound (rise time) or decreases at the end of a sound (decay time), may additionally differentiate between (classes of) speech sounds.

A stylised oscillogram of a sound, e.g. a vowel, produced in isolation, may look as in figure 1.

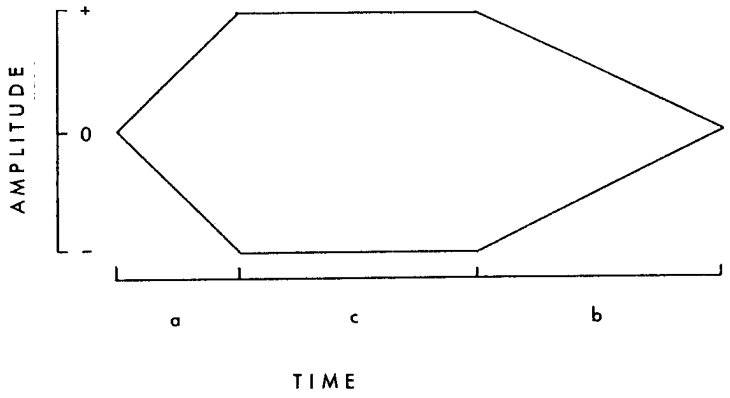


Figure 1. Definition of (a) rise time, (b) decay time, and (c) steady time in an isolated vowel with stylised amplitude envelope(all scales have arbitrary units).

There are, of course, numerous complications in deriving such a stylisation from an actual oscillogram, due to assymmetry of positive and negative halves of the wave form, the gradual nature of the amplitude contour, and disturbances in the monotonicity of intensity build-up and decay. Though high inter-individual agreement can be found when investigators are asked to visually stylise such oscillograms, a satisfactory algorithm achieving the same is still wanting.

The function of rise and decay times in speech has been investigated in much less detail than any of the other acoustic parameters mentioned. Yet, a review of the (scanty) literature will reveal their potential relevance.

Survey of the literature

Cohen, Slis and 't Hart (1963) found in a perceptual experiment, in which speakers of Dutch, American English, and French were instructed to manipulate a blind knob controlling the rise time of a selection of synthetic vowels, that the preferred rise time differed systematically for each of the three languages. Within the one language they studied in greater detail, Dutch, rise time variations were randomly distributed over the 12 vowels used.

In a separate experiment, in which two parameters could be simultaneously controlled by the subjects, decay time rather than steady state duration was found to divide the vowels into three categories, which coincide with the traditional classes of 'short' (mean decay time 40 msec), 'half-long' (80 msec), and 'long' (140 msec) vowels, which is a contrast with phonological implications in Dutch.

In the Cohen *et al* (1963) experiments F0 varied as a function of the amplitude envelope, thus giving the subjects a double cue, which would warrant replication of the experiment with adequate control for F0. Also, the vowels were presented in isolation, which is an unnatural condition, especially since phonemically short vowels are prohibited in word final position in Dutch (and, *a fortiori*, in isolation) by phonotactic constraint. As a point of interest, however, preliminary acoustic measurements of spectrally similar pre-pausal short-long Hausa vowel cognates in actual words turned out to be adequately distinguishable by 35 *versus* 115 msec decay time (Newman and van Heuven, 1978). No perceptual follow-up has been given to this study, yet.

Malécot (1975) claims that a difference in vowel amplitude rise time is the primary cue by which an utterance initial French glottal stop can be differentiated from a non-glottalized vowel onset (30-50 *versus* 70-75 msec). Vowel shortening and faster decay time were found as acoustic correlates of this distinction in utterance final position with typical values

of 370/90 msec for gradual onsets and 165/40 msec for a glottal stop ending. It is unclear from his report how many measurements underlie these values nor what their variability is. Moreover, the description of the perceptual check on these findings is not specific enough to enable us to establish the relative contributions of the various parameters involved in the contrast.

In an unpublished dissertation Gerstman (1957) found that rise time was the principal cue in the American English affricate-fricative distinction in perception of isolated synthetic *Ca*-syllables. In a recent re-analysis of this study (van Heuven, 1979) I showed that when rise and steady times of friction noise bursts are identically sampled for both parameters, while observing the same ranges, rise time explains a larger percentage of the response variance than steady time (44 *versus* 36%).

Using the maximum rate of short-term power increase (which is related to rise time) as a measure of the abruptness of noise burst onset, Kunisaki, Higuchi and Fujisaki (1978) could categorise the manner of articulation of Japanese voiceless affricates and fricatives in isolated *CV*-syllables uttered by one speaker, without a single error.

Cutting and Rosner (1974) found that stimuli differing by a 20 msec increment of friction noise rise time in *Cad*-syllables with a constant 410 msec vowel duration are poorly discriminated (generally below 60% correct), except at the affricate-fricative boundary, which was located at 45 msec rise time (over 75% correct). Thus, rise time appears to be perceived categorically. In non-speech control conditions involving discrimination and identification of 'plucked' *versus* 'bowed' excitation of a string, the same 20 msec increment in the rise time of sine and sawtooth waves was virtually inaudible over the whole stimulus range (below 60% correct discrimination), except around a musical category boundary at 40 msec for sines or 35 msec for sawtooth waves.

It should be pointed out that in the Cutting and Rosner (1974) experiments rise time added to the overall duration of the stimulus, so that two cues are confounded. This will be of marginal consequence for the 'musical' stimuli, which were all longer than 1 sec, but may well have influenced the perception of the speech stimuli, whose longest rise time (80 msec) added about 50% to the overall friction noise duration.

Just noticeable difference of rise and decay time

It is generally acknowledged that speech perception is often better understood against a background of purely psycho-acoustic facts. In particular, information on just noticeable differences (JND's) of an acoustic

parameter may help to decide what precision should be observed in the analysis of acoustic data, and may provide a partial explanation for the inventories of (phonemic) contrasts in languages (see e.g. Lehiste, 1970: 5,11; Gandour, 1978; 56-57).

Such information may shed light on the question whether it is realistic to claim that a 50/70 msec difference in the rise time of an utterance initial vowel would be sufficient to cue a contrast, as Málecot (1975) did, if a ternary opposition along a vowel decay time continuum from 0 to 150 msec (Cohen *et al.*, 1963) would be at all audible, or even if rise time differences could possibly outweigh steady time differences (Gerstman, 1957; van Heuven, 1979).

For these reasons we felt that a careful estimation of JND for rise and decay time, in a purely psycho-physical experiment, would be of interest to the study of speech perception.

Method

8 phonetically non-naive Dutch subjects were asked to adjust a blind knob controlling the rise or decay time (depending on the condition) of a matching signal, which was presented 1000 msec after a reference stimulus, until they considered the two signals to be identical. Stimulus pairs were repeated every 4200 msec until the subject indicated to be satisfied with his match. Only the final adjusted value was noted down. In half of the trials the subject was requested to start from a maximally short rise/decay time, in the other trials he would start from a maximally long rise/decay time. In 4 different signal conditions either rise or decay time of either gated 1000 Hz sine waves or white noise bursts were adjusted. The variable slope of the reference stimulus could take one of the following values: 0, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, or 100 msec, while the invariant slope was held constant at 50 msec. The steady steady portion of the stimulus lasted 400 msec. Overall stimulus duration was affected by decay time only, as indicated in figures 2a and 2b:

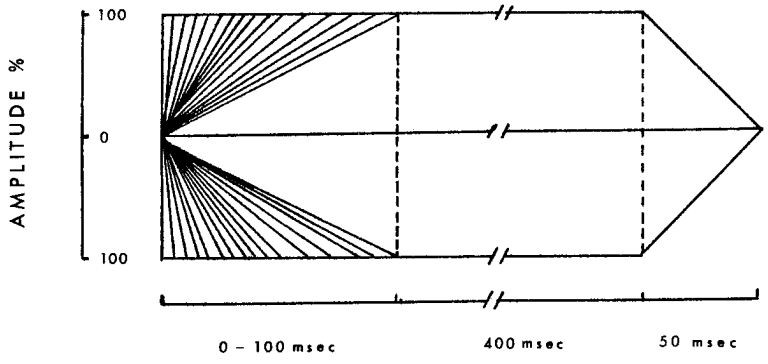


Figure 2a. Stimuli for (a) rise times and (b) decay times as used in adjustment experiment.

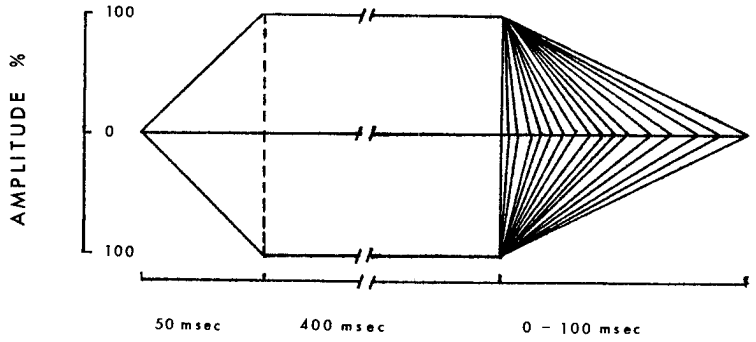


Figure 2b. Stimuli for (a) rise times and (b) decay times as used in adjustment experiment.

The stimuli were presented through headphones at 60 dB above threshold. The experiment was run in 4 sessions per subject, a different signal condition for each session.

Results

1024 measurement points were obtained (8 subjects * 4 signal conditions * 2 movements * 16 stimulus values). Figure 3 presents the mean rise/decay times of the adjusted signals and their standard deviations, as functions of the reference rise/decay time.

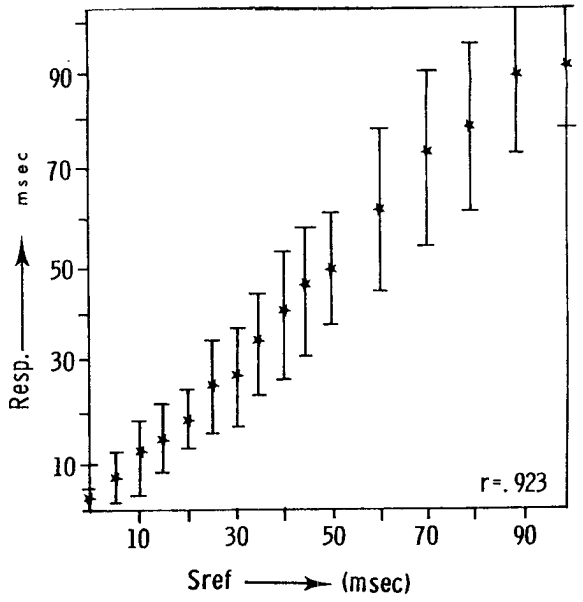


Figure 3. Means and standard deviations of adjusted rise/decay time as functions of the reference stimulus. Each dot represents 64 measurements.

Figure 3 reveals no general tendency toward over or underestimation of the reference values. Towards the 100 msec extreme, however, some underestimation seems to point a ceiling effect there, which has probably been caused by the fact that, due to instrumental limitations, no rise/decay times could be produced by the subject of values exceeding 103 msec.

Standard deviation of adjustment is normally taken as a measure for JND. As is apparent from figure 4, where SD of adjustment is plotted as a function of the reference rise/decay time. JND increases more or less linearly with the reference value, and thus turns out to be a constant fraction of the reference value.

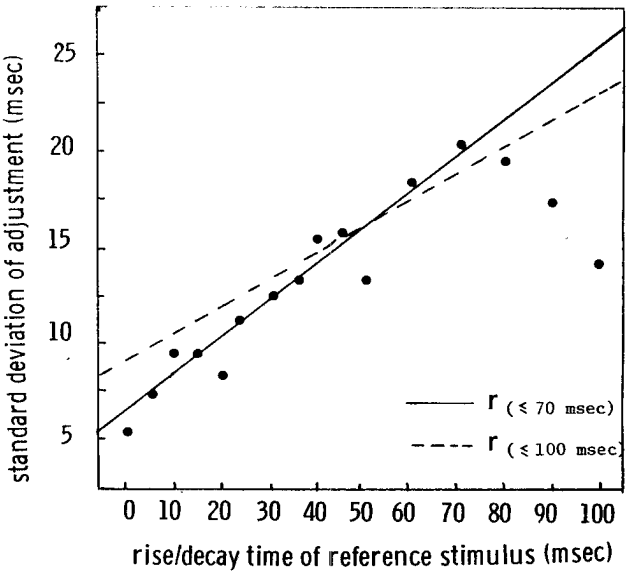


Figure 4. Standard deviation of adjustment as a function of the reference rise/decay time, with and without inclusion of stimulus values greater than 70 msec.

Plotting this Weber fraction, as in figure 5, shows that the JND is roughly 25% of the reference value, with a marked increase at the 0-extreme, which is normally found in the literature on JND's for temporal phenomena (cf. Henry, 1948; Small and Campbell, 1962).

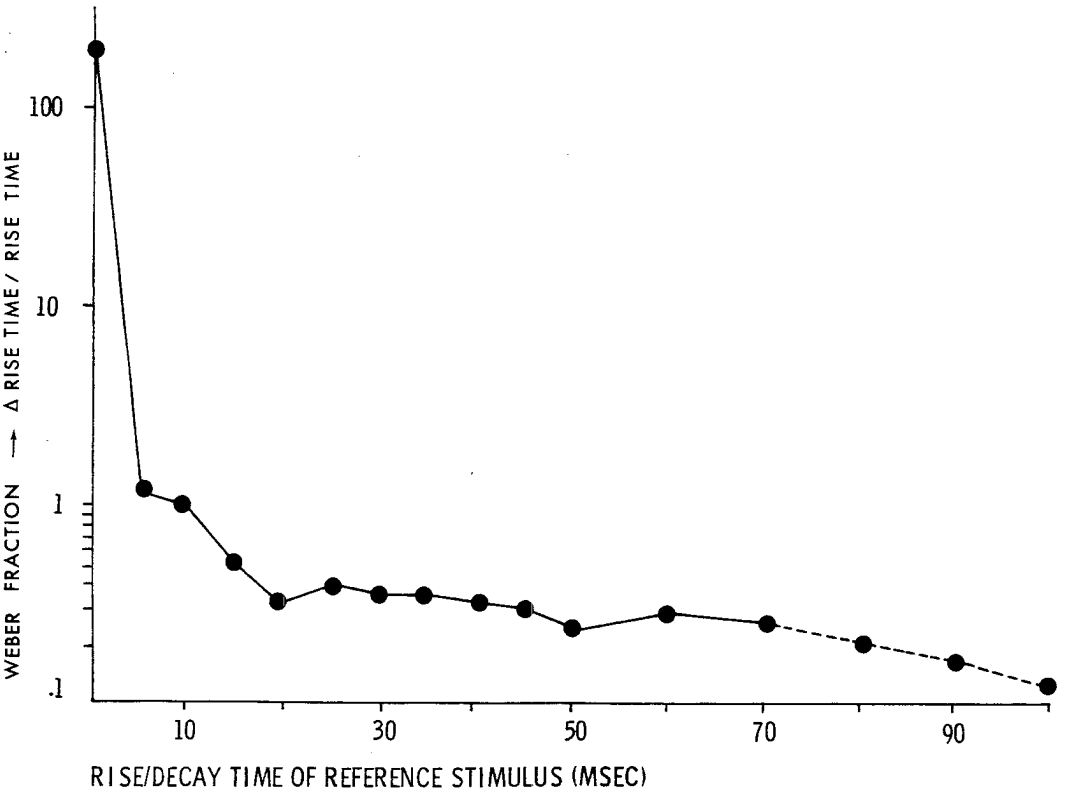


Figure 5. Weber fractions as a function of rise/decay time of the reference stimulus.

Figure 6, which presents the information in figure 4 separated out for the four signal conditions, shows that the curves for the sine-decay and noise-rise conditions generally overlap, that the JND's for the sine-rise condition are somewhat smaller, and that a remarkable increase in discriminability occurs with noise-decay stimuli at the upper half of the continuum.

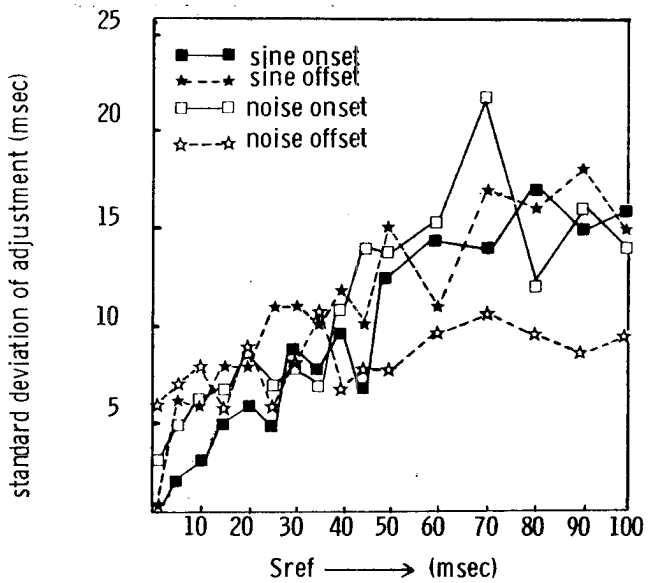


Figure 6. Standard deviation of adjustment as a function of reference rise/decay time, separated out for the 4 signal conditions.

Discussion

The results of this experiment indicate that the phonetic contrasts mentioned earlier are acoustically characterized by rise or decay time differences that are all well above threshold, and which may therefore function adequately in language. It should be emphasised that the JND for rise and decay phenomena, as established in this experiment (25%) is larger than JND's reported in the literature on other temporal phenomena (for a survey cf. Lehiste, 1970). On the basis of this observation it is not immediately clear how rise time differences can be the most effective cue in the affricate-fricative distinction in English, as claimed by Gerstman (1957). This matter is discussed at greater length in van Heuven (1979).

Finally, no traces of categorical perception were found in our experiment, whose results in general strongly contradict the Cutting and Rosner (1974) findings. For an elaborate discussion of the discrepancies in the results of these two studies the reader is referred to van Heuven and van den Broecke (1978, 1979).

Acknowledgement

This research was funded in part by a grant from the Netherlands Organization for the Advancement of Pure Research (Z.W.O.).

References

Cohen, A., Slis, I.H. and Hart, J. 't. (1963). "Perceptual Tolerances of Isolated Dutch Vowels," *Phonetica* 9, 65-78.

Cutting, J. E. and Rosner, B. S. (1974). "Categories and Boundaries in Speech and Music," *Perception and Psychophysics* 16, 564-570.

Gandour, J. (1978). "The Perception of Tone," in *Tone*, edited by V.A. Fromkin. Academic Press, New York), pp. 46-71.

Gerstman, L. J. (1957). "Perceptual Dimensions of the Friction Portions of Certain Speech Sounds," Ph.D. dissertation, New York University, (Unpublished).

Henry, F. M. (1948). "Discrimination of the Duration of a Sound," *Journal of Experimental Psychology* 38, 734-742.

Heuven, V. J. van. (1979). "The Relative Contribution of Rise, Time, Steady Time,and Overall Duration of Noise Bursts to the Affricate-Fricative Distinction in English: a Re-analysis of Old Data," in *Speech Communication Papers Presented at the 97th Meeting of the Acoustical Society of America*, edited by J.J. Wolf and D. H. Klatt (The Acoustical Society of America, New York), Pp 307-311 (Abstracted in *Journal of the Acoustical Society of America*, 65, S1, 79).

Heuven, V. J. van and Broecke, M.P.R. van den. (1978). "Auditory Discrimination of Rise and Decay Time in Tone and Noise Bursts," Progress Report of the Institute of Phonetics Utrecht 3(2), 49-69.

Heuven, V.J. van and Broecke, M.P.R. van den. (1979). "Auditory Discrimination of Rise and Decay Time in Tone and Noise Bursts: Two Experiments." (submitted to *Journal of the Acoustical Society of America*).

Kunisaki, O., Higuchi, N. and Fujisaki, H. (1978). "Extraction of Acoustic Features and the Classification of the Voiceless Affricates /ts/ and /ch/ in Japanese," (Faculty of Engineering, University of Tokyo) (abstracted in *Journal of the Acoustical Society of America* 64, S1, 179-180).

Lehiste, I. (1970). *Suprasegmentals*. M.I.T., Cambridge MA.

Malécot, A. (1975). "The Glottal Stop in French," *Phonetica* 31, 65-78.

Newman, R. A. and Heuven, V. J. van. (1978). "Temporal Aspects of Final Vowels in Hausa: Some Phonetic Observations," (Dept. of African Linguistics, Leyden University) (Unpublished).

Small, A. M. and Campbell, R. A. (1962). "Temporal Differential Sensitivity for Auditory Stimuli," *American Journal of Psychology* 75, 401-410.

van Heuven, V.J. (see Heuven, V.J. van).

Peak intraoral air pressure in [p] *as a function of* F_0

Eric Zee

1. INTRODUCTION

There have been a number of studies investigating the differences in intraoral air pressure between voiced and voiceless stop consonants. The general agreement among investigators is that voiceless stop consonants are produced with higher peak intraoral air pressure than their voiced cognates (Black, 1950, Malecot, 1955, 1966, Warren, 1964, Subtelny, Worth and Sakuda, 1966, Arkebauer, et al, 1967, Ringel, House and Montgomery, 1967, Löfqvist, 1971, Brown, McGlone and Profitt, 1973, Bernthal, 1978). The differences in pressures have usually been attributed to the effect of glottal resistance (Malecot, 1955, Warren, 1964, Arkebauer, et al, 1967). During the production of voiced stop consonants, the adduction of the vocal cords increases the resistance to pulmonic air entering the vocal tract and thus reduces the pressures in the oral cavity. With regard to the consonantal position in the syllable, it has been shown that for voiceless stop consonants the peak intraoral air pressure is higher in medial position , lower in initial position and lowest in final position (Malecot, 1955, 1968, Arkebauer, et al, 1967, Lisker, 1970, Löfqvist, 1971). The relation between tone and peak intraoral air pressure, however, has not been explored. It is the purpose of this study to investigate whether the peak intraoral air pressure in a voiceless unaspirated stop [p] is affected by the tone on the following vowel. As variations in fundamental frequency may be accompanied by changes in subglottal pressure (Ladefoged, 1963, Öhman & Lindqvist, 1966) and as the subglottal pressure and the intraoral air pressure are more or less the same during the production of a voiceless stop consonant (Ladefoged, 1967, 1968, Netsell, 1969, Scully, 1969), it is reasonable to expect that the peak intraoral air pressure in [p] may vary according to changes in subglottal pressure for the production of different tones on the vowel that follows [p].

2. PROCEDURE

In our investigation, four Chinese dialects, Standard Mandarin, Chungking (Western Mandarin), Cantonese and Shanghai, were used. In each dialect, different tones may occur on a CV syllable, where C is a voiceless unaspirated bilabial stop [p] to form words of different meaning, as shown in the following:

Mandarin			Chungking		
44	[pi]	'to oppress'	45	[pi]	'sex organ'
35	[pi]	'nose'	21	[pi]	'to oppress'
214	[pi]	'to complete'	41	[pi]	'to compare'
51	[pi]	'closed'	13	[pi]	'to avoid'

Cantonese			Shanghai		
55	[pei]	'sorrow'	53	[pi]	'edge'
33	[pei]	'secret'	34	[pi]	'to change'
22	[pei]	'nose'			

The numerals to left of each word denote levels of pitch, '5' being the highest and '1' being the lowest. Thus a combination such as '51' means a high-low falling tone, whereas '13' is a low-mid rising tone. A reading list for each dialect was prepared containing 10 repetitions of each word in the sentence frames below:

Mandarin:	[uo	cian-tsai	tu	___	kei	ni	$t^hiŋ$]
	I	now	read		for	you	listen
Chungking:	[ŋo	cian-tsai	tu	___	tcɪ	ni	t^hin]
	I	now	read		for	you	listen
Cantonese:	[ŋo	yiu	tvk	___	pei	nei	t^hian]
	I	want	read		for	you	listen
Shanghai:	[ŋu	yɪo	do?	___	pə?	ɪ	$t^hiŋ$]
	I	want	read		for	him	listen

Four male native speakers, one for each dialect, all in their late thirties, participated in the experiment. Each speaker was asked to read the sentences at a normal rate of speech. He was also asked not to emphasize the test word.

During the readings intraoral air pressure was sensed with a polyethylene tube (120 mm in length and 2.5 mm internal diameter). The tube was inserted into the oral cavity between the upper and lower lips and was maintained near the center line in the oral cavity and parallel to oral airflow. The length of the tube inside the month was approximately 40 mm. Outside the mouth, the tube was attached to a pressure transducer. Transduced intraoral air pressure signals were amplified and displayed on an oscilloscope. To obtain measureable records the oscilloscope was in turn connected to one channel of an oscillomink and a microphone was connected to another. The intraoral air pressure recording system was calibrated in cm H_2O against a U-tube water manometer. The accuracy of the recording was calculated to be within the range of ± 0.24 cm H_2O. The recording of the intraoral air pressure and the utterances was performed in a sound treated booth. The audio signals were analyzed using a PDP-12 computer. Fundamental frequency measurements for each test word were obtained every 10 msec by the Cepstrum method, using a window of 51.2 msec.

3. RESULTS

Table Ia presents the values (in cm H_2O) of the peak intraoral air pressure for [p] associated with different tones on the following vowel or diphthong for each of the four speakers. Also shown are the means and the standard deviations of these values for [p] associated with each tonal category. The mean of the F_0 onset values for the vowel or the diphthong associated with each tonal category is given at the bottom of the table. The F_0 onset refers to the average of the first three data points of the output of the Cepstrum analysis. We can see in general a higher mean peak intraoral air pressure is associated with a higher mean F_0 onset, with

SPEAKER	Mandarin				Chungking				Cantonese			Shanghai	
TONE TYPE	44	35	214	51	45	21	51	13	55	33	22	53	34
PEAK INTRA-ORAL AIR PRESSURES ($cm\ H_2O$)	4.36	3.76	3.80	4.52	3.76	4.00	4.76	4.24	4.60	4.56	4.24	3.56	3.12
	4.24	3.88	3.88	5.36	4.76	3.60	4.52	3.36	4.64	4.36	4.12	3.56	2.52
	4.44	3.24	3.76	5.24	4.52	3.76	4.00	3.60	4.56	4.20	4.12	4.12	3.24
	3.76	3.52	3.12	4.60	3.49	4.52	4.12	3.60	4.64	4.44	4.28	3.80	3.00
	4.60	3.64	3.36	4.24	3.24	3.76	4.24	3.60	4.80	4.32	4.12	3.64	3.00
	4.00	3.68	3.52	5.32	4.64	3.00	5.00	3.24	4.80	4.36	4.00	3.92	3.32
	4.56	3.52	4.00	4.44	4.12	3.12	3.88	3.40	4.52	4.24	3.76	4.92	3.44
	4.12	3.24	4.00	4.48	3.64	4.76	4.12	3.60	4.52	4.52	4.36	3.20	3.32
	4.10	3.52	2.64	4.64	4.00	4.24	4.24	--	4.52	4.00	3.88	4.12	3.52
	4.60	--*	--	--	4.24	--	4.88	--	4.36	4.64	3.88	4.12	3.36
$\bar{X}$	4.28	3.56	3.56	4.76	4.04	3.86	4.38	3.58	4.60	4.36	4.08	3.82	3.18
S.D.	0.28	0.22	0.46	0.43	0.51	0.60	0.39	0.29	0.13	0.19	0.19	0.32	0.29
F_0 ONSET (Hz) $\bar{X}$	216	140	99	248	150	112	171	99	188	150	137	166	126
F_0 ONSET (Hz) S.D.	8.2	3.0	5.4	10.6	4.7	2.6	5.6	2.7	6.3	4.5	5.0	15.4	5.7

* '--' token mispronounced.

Table Ia. Values ($cm\ H_2O$) of the peak intraoral air pressure for [p] associated with different tones on the following vowel or diphthong, and mean F_0 onsets (Hz) of the vowel or diphthong for each of the four speakers.

SPEAKER	TONE TYPES & MEAN F0 ONSET	MEAN F0 ONSET DIFFERENCE	PEAK INTRAORAL AIR PRESSURES MEAN	t-SCORE	SIGNIFICANCE
Mandarin	44 (216 Hz) 35 (140 Hz)	76 Hz	4.28 3.56	6.27	p <0.005
	44 (216 Hz) 214(99 Hz)	117 Hz	4.28 3.56	4.14	p <0.005
	44 (216 Hz) 51 (248 Hz)	32 Hz	4.28 4.76	2.93	p <0.005
	35 (140 Hz) 214(99 Hz)	41 Hz	3.56 3.56	0.53	Non-sig.
	35 (140 Hz) 51 (248 Hz)	108 Hz	3.56 4.76	7.56	p <0.005
	214(99 Hz) 51 (248 Hz)	149 Hz	3.56 4.76	5.74	p <0.005
Chungking	45 (150 Hz) 21 (112 Hz)	38 Hz	4.07 3.86	0.85	Non-sig.
	45 (150 Hz) 51 (171 Hz)	21 Hz	4.07 4.38	1.64	Non-sig.
	45 (150 Hz) 13 (99 Hz)	51 Hz	4.07 3.58	2.49	p <0.025
	21 (112 Hz) 51 (171 Hz)	59 Hz	3.86 4.38	2.38	p <0.025
	21 (112 Hz) 13 (99 Hz)	13 Hz	3.86 3.58	1.17	Non-sig.
	51 (171 Hz) 13 (99 Hz)	72 Hz	4.38 3.58	5.09	p <0.005
Cantonese	55 (188 Hz) 33 (150 Hz)	38 Hz	4.60 4.36	3.17	p <0.005
	55 (188 Hz) 22 (137 Hz)	51 Hz	4.60 4.08	6.98	p <0.005
	33 (150 Hz) 22 (137 Hz)	13 Hz	4.36 4.08	3.37	p <0.005
Shanghai	53 (166 Hz) 34 (126 Hz)	40 Hz	3.82 3.18	4.62	p <0.005

Table Ib. Results of the grouped data t-test (one tailed) for the peak intraoral air pressure for [p] for any given tone compared pairwise with each other tone for each of the four speakers.

an exception, however, that in the speech of the Mandarin speaker the mean peak intraoral air pressure in [p] associated with both 35 and 214 is the same, which is 3.56 cm H_2O, although the values of the mean F_0 onset for these two tonal categories are 140 Hz and 99 Hz respectively.

For each speaker, the peak intraoral air pressure for [p] for any given tone was compared pairwise with each other tone using grouped data t-tests (one tailed). The results, shown in Table Ib, indicate that the difference between any such two sets of values is significant at the 0.005 level for the Cantonese and the Shanghai speakers. This is also true for the Mandarin speaker, except for the pair which involves tones 35 and 214 referred to above. In the speech of the Chungking speaker, the difference is significant at the 0.025 level between the tone pairs 45 and 13, 21 and 51, and 51 and 13, however, non-significant between the pairs 45 and 21, 45 and 51, and 21 and 13. In other words, only those pairs of tones whose onsets differ by 3 or more in the traditional 5-point notation show a significant difference.

There seem to be some idiosyncracies in the relationship between the peak intraoral air pressure and the F_0 onset for each speaker. For instance, in the speech of the Mandarin speaker, the difference in the mean peak intraoral air pressure is highly significant at the 0.005 level for [p]'s associated with tones 44 and 51, but non-significant for [p]'s associated with tones 35 and 214. However, the difference in the mean F_0 onset is 32 Hz between tones 44 and 51 and 41 Hz between tones 35 and 214. This seems to suggest that for this particular speaker a greater difference in F_0 onset is not necessarily the result of a significant difference in peak intr oral air pressure. In this case, the level of the F_0 onset seems to have a role in determining when a difference in F_0 onset is associated with a significant difference in the peak intraoral air pressure. The mean F_0 onsets of tones 44 and 51 are at a higher pitch level (248 Hz and 216 Hz respectively) than those of 35 and 214 (140 Hz and 99 Hz, or 138 Hz and 104 Hz if only the first data point is taken to be the F_0 onset). Thus, a smaller difference in F_0 onset between two tones may mean a significant difference in peak intraoral air pressure if the F_0 onset of the tones is at a higher pitch level. On the other hand, for the speech of the Chungking speaker a greater difference in F_0 onset between two tones, irrespective of their pitch level, is the factor which determines a significant difference in peak intraoral air pressure. In the speech of both the Mandarin and the Chungking speakers, the difference between peak intraoral air pressures associated with the F_0 onsets of a lower pitch level (35 and 214 for the Mandarin speaker, and 21 and 13 for the Chungking speaker) is not significant. However, this is not so in the speech of the Cantonese speaker. For this speaker, despite that both 33 and 22 have a lower F_0 onset, the difference in peak intraoral air pressure between [p]'s associated with these tones is significant.

In order to determine the degree of association between the values of the peak intraoral air pressure for [p] and the values of the F_0 onset for

SPEAKER	df	CORRELATION COEFFICIENT (r)	SIGNIFICANCE
Mandarin	35	0.800	0.01
Chungking	35	0.536	0.01
Cantonese	28	0.678	0.01
Shanghai	18	0.803	0.01

Table II. Results of the correlation analysis for the values of the peak intraoral air pressure for [p] and the values of the F_0 onset for the following vowel or diphthong.

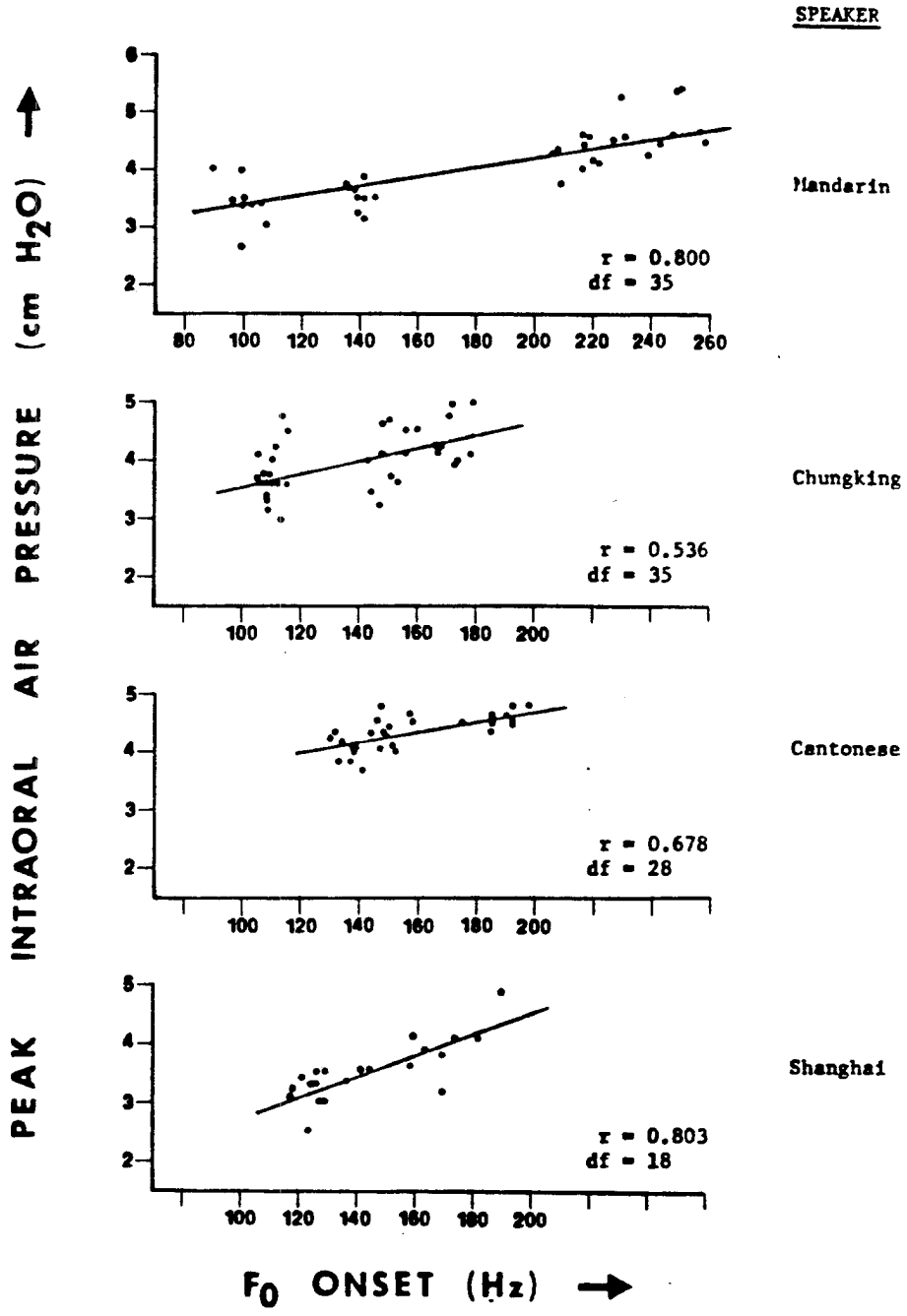


Figure 1. Scatter diagram for the peak intraoral air pressures plotted as a function of F_0 onset variation for each of the four speakers.

the following vowel or diphthong, a Pearson correlation analysis was performed for the speech of each of the four speakers. The results, shown in Table II, indicate that there is an overall positive correlation between the two sets of values (significant at the 0.01 level) for all four speakers. Also shown in the table are the values of correlation coefficient (r). Notice that the correlation coefficient for the Chungking speaker is only 0.536 which is the smallest compared to the r values for the other speakers. This is basically due to the high variance in the Chungking data. As shown in Table Ia, for the Chungking speaker, the standard deviations of the values of the peak intraoral air pressures in [p]'s associated with tones 21 and 45 are 0.60 and 0.51 respectively which are quite large compared to the values of the other standard deviations. Thus, in the speech of the Chungking speaker the peak intraoral air pressures in [p]'s do not correlate well with the F_0 onsets of the two tones, despite a good overall correlation.

Our findings are summarized by four scatter diagrams, as shown in Figure 1, for the speech of each of the four speakers, in which the peak intraoral air pressures are plotted as a function of F_0 onset variation. Also presented in the figure is a linear regression line in each of the four diagrams. Our findings indicate that in the speech of the four speakers the peak intraoral air pressures in [p]'s differ according to the F_0 onset of the tone on the following vowel or diphthong.

4. DISCUSSION

Despite the fact that a higher F_0 is usually accompanied by a higher subglottal pressure (Ladefoged, 1963, Öhman and Lindqvist, 1966), it has been shown that changes in subglottal pressure cannot be the only contributing factor to F_0 variation (Ladefoged, 1963, Öhman and Lindqvist, 1966, Ohala, 1973, 1978). Our findings support the hypothesis that pitch variations are attributed to factors other than subglottal pressure, that is, to laryngeal factors. We, of course, have presupposed that the peak intraoral air pressure and the subglottal pressure are equivalent during the production of a voiceless stop consonant (Ladefoged, 1967, 1968, Netsell, 1969, Scully, 1969). We have shown that in our study the peak intraoral air pressures have a good correlation with the F_0 onsets of the tone on the following vowel or diphthong in most cases. This suggests that the onset subglottal pressure is higher for a tone with a higher F_0 than for a tone with a lower F_0 onset. However, the effect of subglottal pressure on F_0 has been found to be approximately from 2.5 Hz/cm H_2O (Ohala and Ladefoged, 1969, Hixon, Mead and Klatt, 1971) to 2.9 Hz/cm H_2O (Ohala 1978, based on Ohman and Lindqvist, 1966). In our results the mean F_0 onset variations (Table Ib) are far too great to be entirely attributed to the changes in subglottal pressure. The difference between the mean peak intraoral air pressures for any pair of tones never exceeds 1.5 cm H_2O, e.g., in the speech of the Mandarin speaker, the mean F_0 onset differences between tones 44/214, 35/51, and 214/51 are 117 Hz, 108 Hz and 148 Hz respectively, whereas the corresponding differences in the mean peak

intraoral air pressure are merely 0.72 cm H_2O, 1.20 cm H_2O, and 1.20 cm H_2O. Furthermore, in the speech of the same speaker, the difference in the mean F_0 onset between the tones 35 and 214 is 41 Hz, however, there is no difference in the mean peak intraoral air pressure in [p]'s associated with these two tones. These cases indicate that the difference in subglottal pressure cannot account for the large differences in F_0. Similar cases are also found in the speech of the other three speakers (see Table Ib), although the differences in the mean F_0 onset are in a smaller magnitude. Furthermore, as we have described there are high variances in the peak intraoral air pressure in [p]'s associated with tones 21 and 45. This shows that for some tones this speaker does not use pressure to control F_0. Thus, our results provide evidence in support of the hypothesis that variation of F_0 cannot be principally attributed to change in subglottal pressure. Although the changes in subglottal pressure cannot possibly account for the large variations of F_0 onset, the good correlation between the F_0 onsets and the peak intraoral air pressures in [p] nevertheless indicate that the changes in subglottal pressure that accompany the F_0 variations may be a facilitating, but not a necessary, physiological effort for manipulating pitch change in natural speech.

The issue of tone effects on consonants has been controversial. Hyman (1973b) and Hyman and Schuh (1974) denied that such an effect is possible. However, Maddieson (1978) has shown that consonants are affected by tone in terms of diachronic changes of consonants, synchronic difference in consonants and in terms of small but measurable phonetic differences. Our results are in agreement with Maddieson (1978) as far as the phonetic effect of tone on consonant is concerned, as we have shown how the peak intraoral air pressure in a voiceless unaspirated stop is affected by the F_0 onset of the vowel or diphthong that follows the consonant.

5. CONCLUSION

We have shown that (1) there is a good correlation between the peak intraoral air pressure in [p] and the F_0 onset of the tone on the following vowel or diphthong; (2) the amount of change in subglottal pressure cannot account for the large variation of F_0; (3) in terms of peak intraoral air pressure, a consonant is affected by tone.

References

Arkebauer, H.J., Hixon, T.J. and Hardy, J.C. (1967) 'Peak intraoral air pressures during speech.' *J. of Speech and Hearing Research.* 10, 196-208.

Bernthal, J.E. (1978) 'Intraoral air pressure during the production of /p/ and /b/ by children, youths and adults.' *J. of Speech and Hearing Research* 21, 361-371.

Black, J.W. (1950) 'The pressure component in the production of stop consonants. *J. of Speech and Hearing Disorders* 15, 207-210.

Brown, W.S. and McGlone, R.E. and Profitt, W.R. (1973) 'Relationship of lingual and intraoral air pressure during syllable production.' *J. of Speech and Hearing Research* 16, 141-151.

Hixon, T.J., Mead, J. and Klatt, D.H. (1971) 'Influence of forced transglottal pressure changes on vocal fundamental frequency.' *Journ. Acoust. Soc. Amer.* 49, 105.

Hyman, L.M. (1973b) 'Consonant types and tone'. *Southern California Occasional Papers in Linguistics* 1. University of Southern Calif. Los Angeles.

Hyman, L.M. and Schuh, R. (1974) 'Universals of tone rules: evidence from West Africa. *Linguistic Inquiry* 5, 81-116.

Ladefoged, P. (1963) 'Some physiological parameters of speech. *Language and Speech.* 6, 109-119.

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

Ladefoged, P. (1968) 'Linguistic aspects of respiratory phenomena. *Annals of the New York Academy of Sciences* 155, 141-151.

Lisker, L. (1970) 'Subglottal air pressure in the production of English stops.' *Language and Speech.* 13, 215-231.

Löfqvist, A. (1971) 'Some observations on supraglottal air pressure. Phonetics Laboratory, Lund University, *Working Papers* 5, 27-44.

Maddieson, I. (1978) 'Tone effects on consonants.' *J. of Phonetics* 6, 327-344.

Netsell, R. (1969) 'Subglottal and intraoral air pressure during the intervocalic contrast of /t/ and /d/. *Phonetica* 20, 68-73.

Ohala, J.J. (1973) 'The physiology of tone.' In L. Hyman (Ed.) *Consonant Types and Tone. Southern California Occasional Papers in Linguistics* 1, 1-14.

Ohala, J.J. (1978) 'Production of tone.' In V. Fromkin (Ed.) *Tone: a Linguistics Survey.* N.Y. Academic Press, 5-32.

Ohala, J.J. and Ladefoged, P. (1969) 'Further investigation of pitch regulations in speech.' *UCLA Working Papers in Phonetics* 14, 12-24.

Öhman, S. and Lindqvist, J. (1966) 'Analysis-by-synthesis of prosodic pitch contours.' Speech Transmission Laboratory, Stockholm, *QPSR* 4/1965, 1-6.

Ringel, R., House, A. and Montgomery, A. (1967) 'Scaling articulatory behavior: intraoral air pressure.' *Journ. Acoust. Soc. Amer.* 42, 1209A.

Scully, C. (1969) 'Problems in the interpretation of pressure and air flow data in speech.' University of Leeds, Phonetics Dept. Report 2. 53-92.

Subtelny, J.D., Woth, J.H. and Sakuda, M. (1966) 'Intraoral pressure and rate of flow during speech.' *J. of Speech and Hearing Research* 9, 498-518.

Warren, D.W. (1964) 'Velopharyngeal orifice size and upper pharyngeal pressure-flow patterns in normal speech.' *Plast. Reconst. Surg.* 33, 148-162.

The Effect of Aspiration on the F_0 of the Following Vowel in Cantonese

Eric Zee

1. INTRODUCTION

Conflicting results have been reported with regard to the effect of voiceless aspirated stop consonants on the F_0 onset of the following vowel. Han and Weitzman (1967, 1970) demonstrated that for their Korean subjects the F_0 onsets of vowels following the aspirated stops [p^h, t^h, k^h] are much higher than the F_0 onsets of vowels following the weak aspirated stops [p, t, k], although the difference between the F_0 onsets of vowels after the aspirated stops and the strong unaspirated stops [P, T, K] is much smaller. A separate study of the Korean stops (Kagaya, 1974) produced inconclusive results with respect to the effect of the aspirated stops on the F_0 onset of the following vowel as the two subjects who participated in the experiment produced conflicting results. Jeel (1975) reported that in the speech of six Danish speakers the F_0 onset of a vowel after the asiprated stops [p^h, t^h, k^h] is consistently higher than that of a vowel following the unaspirated stops [p, t, k]. In Erickson (1975), eight of the eleven Thai subjects had a higher F_0 onset for a vowel following the aspirated stop [p^h] than for a vowel following the unaspirated stop, whereas the other three produced opposite results. Ewan (1976) reported that in the speech of a Japanese and a Thai speaker the F_0 onset of a vowel after a voiceless aspirated stop [p^h] is also higher. However, in the analysis of the speech of a Thai speaker, Gandour (1974) presented data showing that the F_0 onset of a vowel after the voiceless aspirated stops [p^h, t^h] is slightly lower than the F_0 onset of a vowel after the unaspirated counterparts. Kagaya and Hirose (1975) showed that in the speech of a Hindi speaker the F_0 onset of a vowel following an aspirated stop is also slightly lower. Hombert & Ladefoged (1977) concluded that voiceless aspirated and voiceless unaspirated stops have similar effects on the F_0 of the following vowel. It seems no agreement can be reached with respect to the effect of the voiceless aspirated stops on the F_0 onset of the following vowel. The disagreement certainly requires further research in this aspect. In the present study we investigate the difference between the effect of [p^h] and [p] on the F_0 onset of the following diphthong [ei] in Cantonese. In this Chinese dialect, the high tone, historically the Ying-Ping tone, may occur on the vowel in both syllable types [p^hV] and [pV]. Etymologically, neither [p^h] nor [p] were derived from [b^h], so a difference in F_0 onset of the following diphthong between the syllable types cannot be attributed to an earlier voicing difference.

2. PROCEDURE

A reading list was prepared containing 10 repetitions of two Cantonese words, [p^hei] 'to spread' and [pei] 'sorrow' in the sentence frame below:

[ŋo	yiu	tvk ___	pei	nei	t^hian]
I	want	read	for	you	listen

Other dummy meaningful words were added to the reading list in order to avoid monotony which may be caused by the limited number of the test words. The tokens in the reading list were arranged in a random order. Three male

native Cantonese speakers participated in the investigation. They were undergraduate students in their early twenties. Each speaker was asked to read the word list at a normal rate of speech. The recording was performed in a single session for each speaker in a sound treated booth. The recorded tapes were analyzed using a PDP-12 computer. A fundamental frequency measurement for each test word was obtained every 10 msec by the Cepstrum method with a window size of 51.2 msec. Also obtained every 10 msec were the intensity (rms) values of the test words, using a square window of 51.2 msec.

3. RESULTS

The F_0 contours of the vocalic portion of all the tokens are shown in Figure 1 for the speech of the three speakers. Each dot represents one data point of the output of the Cepstrum analysis and the time interval between any two successive dots is 10 msec. On the left of the figure are the F_0 contours for the 10 tokens associated with the aspirated stop [p^h] and on the right are the F_0 contours for the 10 tokens associated with the unaspirated stop [p]. It can be easily seen that for all three speakers the values of the initial data points are greater for the F_0 contours associated with [p^h] than those associated with [p].

Table I shows the values of the F_0 onsets and the intensity (rms) onsets for the tokens associated with [p^h] or [p] for all three speakers. The F_0 onset is defined as the mean value of the first three data points of the output of the Cepstrum analysis. Similarly, the intensity onset is the mean value of the first three data points of the output of the intensity measurement. These numbers thus reflect values in the first 71.2 msec following the consonant. Also shown in the table are the means ($\bar{X}$) and the standard deviations (S.D.) of the F_0 and intensity onsets for each set of 10 tokens associated with [p^h] or [p]. In the bottom of the table are the results (t-scores) of the grouped data t-tests (one tailed) between the F_0 onsets associated with [p^h] and [p], and between the intensity onsets associated with [p^h] and [p]. These results show that the difference between the F_0 onsets associated with [p^h] and those associated with [p] are highly significant at the 0.005 level for all three speakers. This is also true for the difference between the intensity onsets associated with the two types of stop consonants.

The values of the F_0 and the intensity onsets for all the tokens by all three speakers shown in Table I are plotted in Figure 2. The circles and the filled circles represent the F_0 onsets associated with [p^h] and [p] respectively, and the empty squares and the filled squares represent the intensity onsets associated with [p^h] and [p] respectively. The circle and the empty square, or the filled circle and the filled square in the same column refer to the F_0 and the intensity onsets of a single token. We can see that for all the tokens the F_0 onsets associated with [p^h] are higher than those associated with [p] for all three speakers. However, the corres-

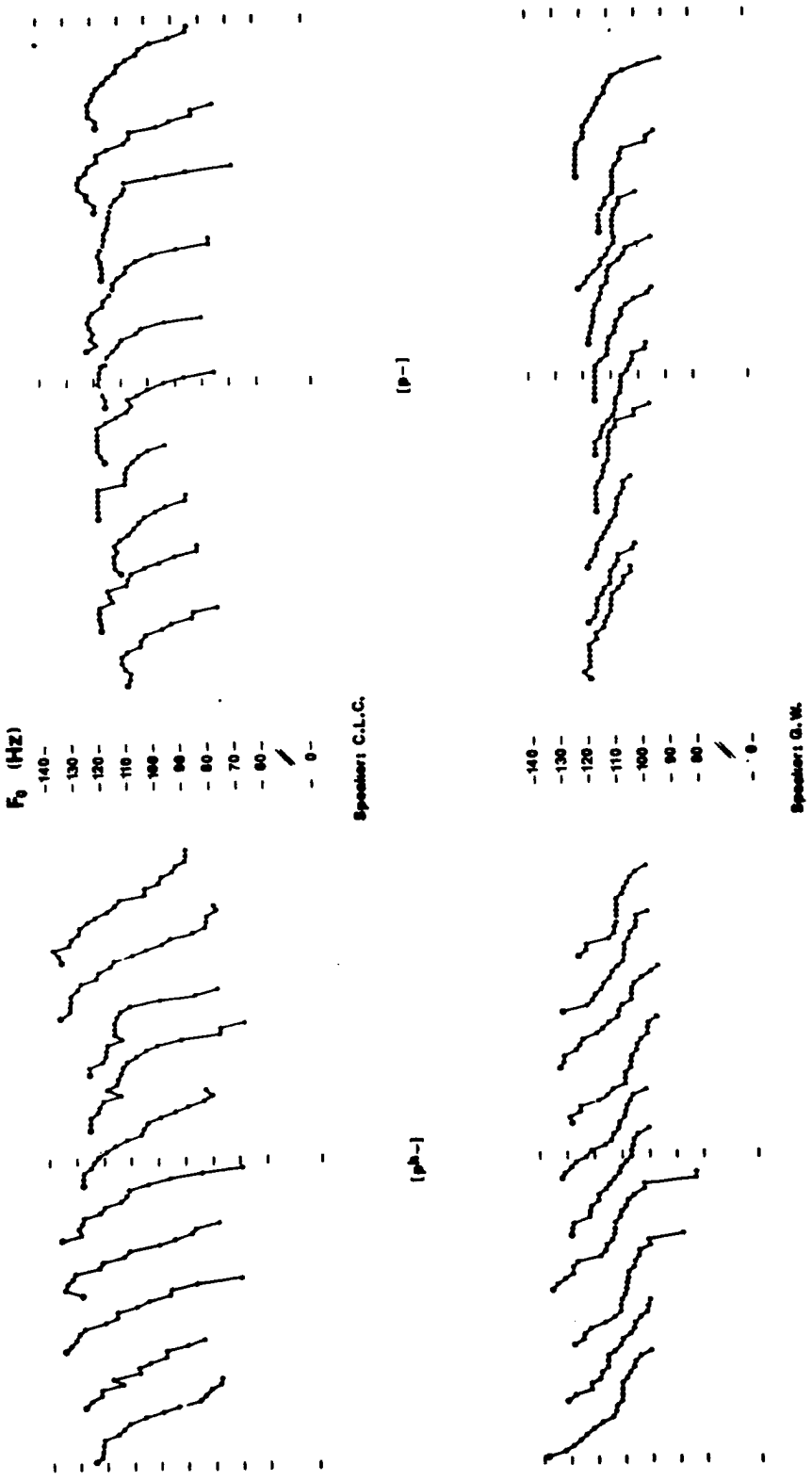


Figure 1. Fo contours of the diphthong [ei] following [p^h] (on the left) and [p] (on the right) by speakers C.L.C. and G.W.

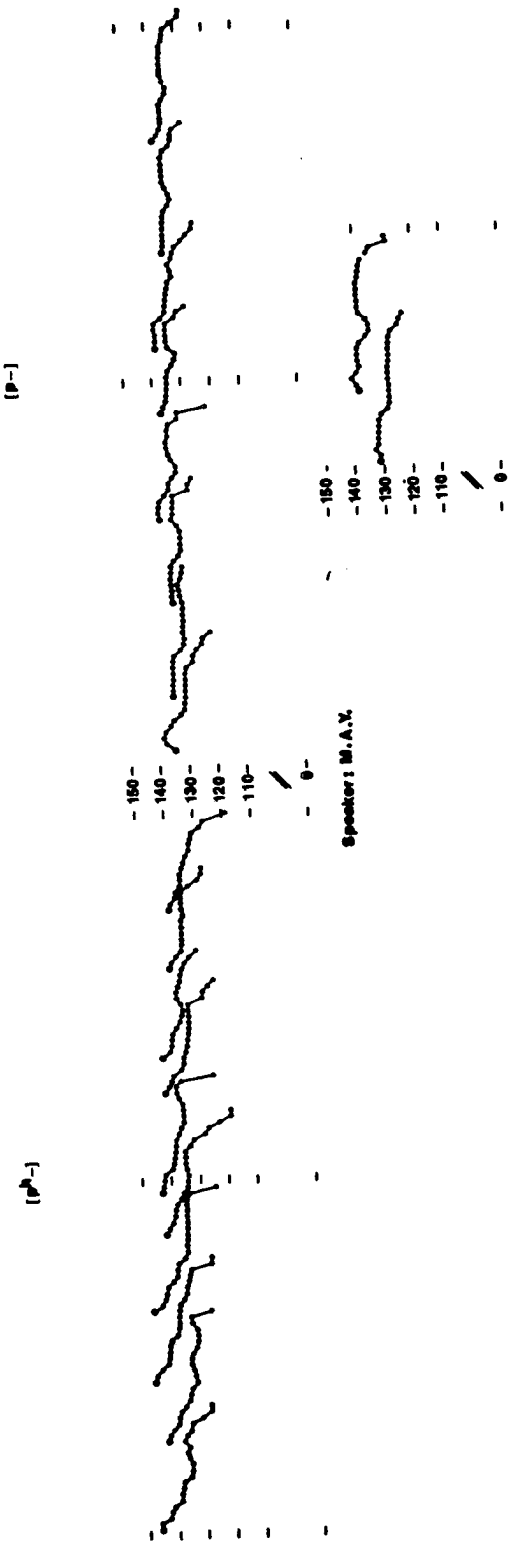


Figure 1. F0 contours of the diphthong [el] following [pʰ] (on the left) and [p] (on the right) by speaker M.A.Y.

	[p^h]		[p]		[p^h]		[p]		[p^h]		[p]	
	rms	F0	rms	F0	rms	F0	rms	F0	rms	F0	rms	F0
	1284	127.4	1969	116.2	881	143.2	1268	139.9	1266	122.8	1436	119.4
	1245	131.9	1619	116.2	992	142.3	1011	136.6	1331	125.5	1588	108.8
	1227	124.4	2106	122.2	955	146.3	1183	135.2	1395	133.1	1483	112.7
	1228	127.2	1754	116.4	1020	145.1	1123	138.7	1304	132.6	1460	119.1
	1021	126.0	1731	118.0	994	144.7	1036	131.9	1364	131.1	1394	116.4
	1182	134.1	1428	113.6	1073	141.1	1101	135.1	1330	128.0	1419	121.1
	1125	128.4	1690	119.5	1061	138.2	1158	135.0	1347	125.0	1557	116.2
	1021	130.2	1672	120.0	964	140.1	1142	138.8	1429	123.6	1598	118.0
	910	123.0	1697	117.9	1042	140.0	1120	136.3	1378	132.8	1421	118.8
	1229	131.8	1698	117.2	950	138.8	1121	136.7	1325	136.4	1433	120.1
$\bar{X}$:	1147	128.4	1738	117.7	993	142.0	1126	136.4	1347	129.1	1497	117.1
S.D.:	124	3.5	185	2.4	59	2.6	72	2.3	47	4.8	77	3.8
Speaker:	G.W.				M.Y.				C.L.C.			

	F0		rms		F0		rms		F0		rms	
	[p^h]	[p]	[p^h]	[p]	[p^h]	[p]	[p^h]	[p]	[p^h]	[p]	[p^h]	[p]
t-score:	7.9		8.4		4.8		4.5		6.2		4.7	
p:	<.005		<.005		<.005		<.005		<.005		<.005	

Table I. Values of the F_0 onsets and the intensity onsets for the tokens associated with [p^h] or [p] for all three speakers, and results (t-scores) of the grouped data t-tests (one tailed) between the F_0 onsets associated with [p^h] and [p] and between the intensity onsets assoicated with [p^h] and [p].

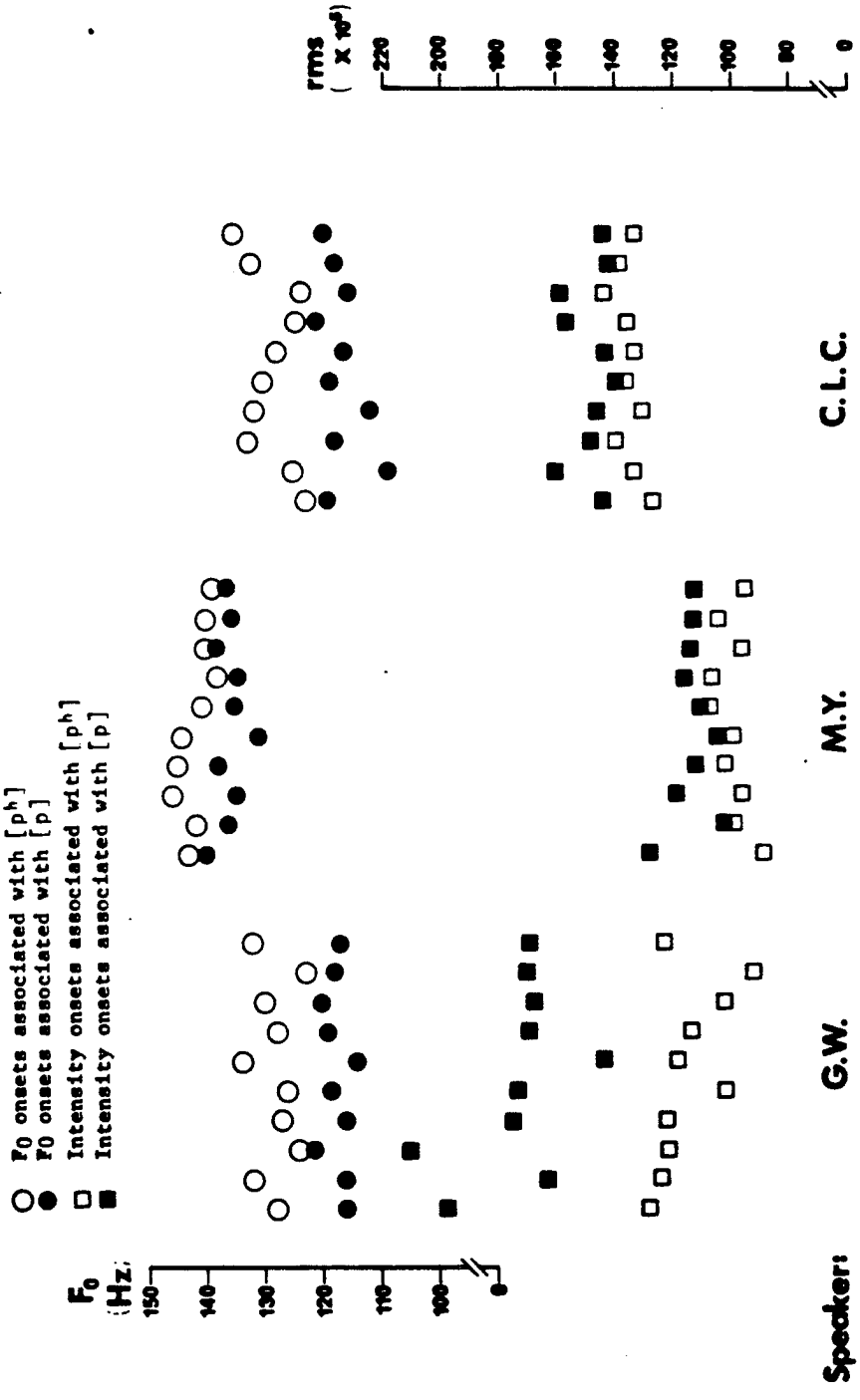


Figure 2. Plotted values of the F0 and the intensity onsets for all tokens by all three speakers.

ponding intensity onsets associated with [p^h] are lower than those associated with [p], and this is true for all three speakers. Thus, we have shown that in the speech of three Cantonese speakers the high tone on the diphthong [ei] after [p^h] has a higher F_0 onset but a lower intensity onset than the same tone on the same diphthong following [p].

4. DISCUSSION.

The effect of the voiceless aspirated stops on the F_0 onset of the following vowel not only varies according to different languages, but also differs according to individual speaker within a language, for instance, both in Kagaya (1974) on Korean and Erickson (1975) on Thai, speakers of the same language have produced different results. Our findings in Cantonese have not contributed to resolve the issue, nevertheless, the Cantonese data has increased the number of the languages in which voiceless aspirated stops raise the F_0 onset of the following vowel. That the intensity onset of the diphthong [ei] following [p^h] is always lower in Cantonese seems to imply that the subglottal pressure at the onset of the diphthong is also lower. The fact that the F_0 onset of the diphthong following [p^h] is always higher indicates that a higher F_0 may be produced even with a decreased subglottal pressure. However, it is not clear at this point why opposite results are produced by speakers of different languages, or by speakers of the same language. In order to have a better understanding of the causes of such differences, future investigations should obtain data on airflow, subglottal pressure, larynx height, glottal aperture, and vocal cord lengths at the onset of a vowel following [p^h] or [p].

REFERENCES

Erickson, D. (1975). Phonetic implications for a historical account of tonogenesis in Thai. In J.G. Harris & J.R. Chamberlain (eds.), Studies in Tai Linguistics in Honor of W.J. Gedney. Bangkok, Central Institute of English Language, Office of State Universities, 100-111.

Ewan, W.G. (1976). Laryngeal behavior in speech. University of California, Berkeley Ph.D. Dissertation.

Gandour, J.T. (1974). Consonant types and tone in Siamese. Journal of Phonetics 2, 337-350.

Han, M.S. & Weitzman, R.S. (1967). Acoustic features in the manner-differentiation of Korean stop consonants. Studies in the Phonology of Asian Languages, V. Los Angeles, Acoustic Phonetics Research Lab, University of Southern California.

Han, M.S. & Weitzman, R.S. (1970). Acoustic features of Korean /P,T,K/, /p,t,k/ and /p^h,t^h,k^h/. Phonetica 22, 112-128.

Hombert, J.-M. & Ladefoged, P. (1977). The effect of aspiration on the fundamental frequency of the following vowel. UCLA Working Papers in Phonetics 36, 33-40.

Jeel, V. (1975). An investigation of the fundamental frequency of vowels after various Danish consonants, in particular stop consonants. Annual Report of the Institute of Phonetics, University of Copenhagen 9, 191-211.

Kagaya, R. (1974). A fiberoptic and acoustic study of the Korean stops, affricates and fricatives. Journal of PHonetics 2, 161-180.

Kagaya, R. & Hirose, H. (1975). Fiberoptic electromyographic and acoustic analyses of Hindi stop consonants. Annual Bulletin of Research Institute of Logopedics and PHoniatrics 9, 27-46.

A spectrographic investigation of Mandarin tone Sandhi

Eric Zee

INTRODUCTION

"In Mandarin there are four tones for stressed syllables. If the average range of the speaker's voice is divided into four equal intervals separated by five points: 1 low, 2 half-low, 3 middle, 4 half-high, and 5 high, any tone can be fiarly well presented by giving its starting and ending pitch, and in the case of circumflex tones, the turning point. Moreover, if we use a short vertical line as a reference line for ordinates and plot a simplified graph to its left, with time as abscissa and pitch as ordinate, we get a letter-like symbol to represent the tone, as in the fast column of the following table:

Tone	Chinese name	Description	Pitch	Graph
1st	Inpyng	high-level	55:	˥
2nd	Yangpyng	high-rising	35:	˧˥
3rd	Shaang	low-dipping	214:	˨˩˧
4th	Chiuh	high-falling	51:	˥˩

... Tone sandhi is the change in the actual value of tones when syllables are spoken in succession.

... When a syllable is completely unstressed, its tone disappears and is said to be atonic or in the neutral tone." (Chao, 1948, p. 24-25, 27)

This paper is a spectrographic investigation of the tone sandhi rules in Mandarin Chinese. It consists of three parts, each dealing with one of the following claims with regard to the tonal phenomena in the language:

CLAIM I:

In a bisyllabic compound, tone 3 /214/ changes to tone 2 [35] in normal speech when it is followed by another tone 3 /214/ (Chao, 1948), that is,

/214 + 214 / → [35 + 214] (Rule 1)

CLAIM II:

In a trisyllabic compound, tone 2 /35/ on the second syllable changes to tone 1 [55] for speech at conversational speed, when it is preceded by another tone 2 /35/ or tone 1 /55/ and followed by

any except the neutral tone (Chao, 1948, 1968, Cheng, 1968), i.e.,

$$/ \begin{Bmatrix} 55 \\ 35 \end{Bmatrix} + 35 \ X/ \rightarrow [\begin{Bmatrix} 55 \\ 35 \end{Bmatrix} + 55 + X] \quad \text{(Rule 2)}$$

$$X \neq \text{neutral tone}$$

CLAIM III:

In rapid speech, if the tones on all syllables are tone 3 /214/ in a sentence such as,

/lau ll mai çiau pl/ (214 214 214 214 214)
'Old Li buys small pen'

then,

(a) tone 3 on the first syllable changes to tone 2 [35],
(b) tone 3 on the second, third and fourth syllables change to tone 1 [55], and
(c) tone 3 on the last syllable remains unchanged (Cheng, 1968),

that is,

/214 + 214 + 214 + 214 + 214/ → [35 + 55 + 55 + 55 + 214] (Rule 3)

In the following sections, we will investigate whether these claims are valid. As the claims were based on the result of impressionistic analysis, they may not accurately describe the actual changes in the fundamental frequency contour when the syllables are juxtaposed. Spectrographic analysis demonstrates physically the actual changes in the fundamental frequency contours for the words and sentence concerned. Thus, the purpose of our study is to provide acoustical data relevant for the analysis of tonal phenomena we have just described.

Two female native Peking speakers who recently came to study in this country participated in our investigation. They, both in the late twenties, were born and grew up in Peking. Their speech is representative of the style of the younger generation in Peking as judged by other native Peking speakers who also arrived in this country recently. Spectrograms were made from their recorded speech. Fundamental frequency measurements were made on the spectrograms.

SECTION (I)

In order to determine whether tone 3 /214/ actually becomes tone 2 [35] when it is followed by another tone 3 /214/, we may compare the fundamental frequency contours of the following two sequences of tones:

(a) /214 + 214/
(b) / 35 + 214/

As tone 2 /35/ is not reported to change in /35 + 214/, the F_0 contours of these two sequences of tones should be the same. Five pairs of bisyllabic compounds were chosen on the basis that they were commonly used. One of the paired bisyllabic compounds had the tone sequences /214 + 214/, and the other /35 + 214/. The five pairs of bisyllabic compounds are shown below:

/tʰu kaɪ/	(214 214)	'land re-distribution'
/tʰu kaɪ/	(35 214)	'to retouch'
/tʂau xuo/	(214 214)	'to look for fire'
/tʂau xuo/	(35 214)	'on fire'
/tɕʰi ma/	(214 214)	'at least'
/tɕʰi ma/	(35 214)	'to ride a horse'
/mai ma/	(214 214)	'to buy a horse'
/mai ma/	(35 214)	'to bury a horse'
/fən tʂʰaŋ/	(214 214)	'a flour factory'
/fən tʂʰaŋ/	(35 214)	'graveyard'

In order to avoid speakers' conscious effort to contrast the paired bisyllabic compounds, two separate reading lists were made for two separate recording sessions. In one reading list only the bisyllabic compounds with the tone sequence /214 + 214/ were included, and in the other only those with the tone sequence /35 + 214/. In each reading list other dummy meaningful bisyllabic compounds were added to avoid monotony. The test compounds were arranged in a random order. Each test compound was repeated four times in the reading list, and it was placed in the carrier frame as below:

[uo çian-tʂaɪ tu ___ gei ni tʰin]
I now read for you listen

Speakers were instructed to read the word list slightly faster than the normal rate of speech. The recordings were made either in a sound treated booth or in a quiet room.

Forty spectrograms of the bisyllabic compounds (1 pair x 5 bisyllabic compounds x 4 repetitions) were made for the speech of each speaker. Samples are shown in Figures 1a-1b for Speaker Q.M. and Figures 2a-2b for Speaker Y.H.J. (spectrogram size 78% reduced from the original). Each figure shows a pair of bisyllabic compounds, with the upper member having the tone sequence /35 + 214/, and the lower member the tone sequence /214 + 214/. The beginning and end points of the harmonics that were measured to obtain F0 values are marked with an arrow.

Tables Ia, Ib and Ic show the F0 values for the beginning, dip (when there is a dip) and end points of the pitch contour of the first syllable in these paired bisyllabic compounds.

The shapes of the F0 contours of the first syllables in the bisyllabic compounds with the /35 + 214/ tone sequence and those with the tone sequence /214 + 214/ are similar in each pair. However there are two distinct patterns. The first pattern applies to the shapes of the F0 contours (see Figure 1a for Speaker Q.M. and Figure 1b for speaker Y.H.J.) of the first syllables in /tʰu kaɪ/ (35 + 214) 'to retouch' and /214 + 214/ 'land re-distribution'. These are quite similar and they are both rising. The F0 measurements for this pair of compounds for each speaker are given in Part I of Table Ia. These show that the tone on the first syllable in /tʰu kaɪ/ (214 + 214) 'land re-distribution' has indeed changed from a dipping to a rising contour for both speakers, as claim I predicts. This is also true for the compound /tɕʰi ma/ (214 + 214) 'at least' for both speakers. The measurements of this pair are given as part III in Table 1b. Note that the end point of the rise appears markedly higher if the underlying tone was originally rising (35).

However the /214/ tone on the first syllable of /214 + 214/ compounds does not always change to a rising contour. It sometimes retains its lexical shape as a dipping contour. This may be seen in the spectrograms (1c, 1d) of the phrase /tʂau xuo/ 214 + 214 'to look for fire'. The F0 measurements for this phrase and its pair /35 + 214/ 'on fire' are given as Part II of Table 1a. It can be seen that not only does /tʂau/ 214 retain its dipping contour but /tʂau/ 35 has also assumed a dipping contour. In this case the members of the pair have still become similar but they have done so because a rising tone has changed to a dipping tone.

Despite the similarity of contour this second pattern also retains some trace of the underlying distinction between /35/ and /214/. The F0 values for the dip and end points of the pitch contour are higher for the first syllable in the compound with underlying /35 + 214/ than for the compound with underlying /214 + 214/. For example, the average F0 values for the dip and end points are 213.8 Hz and 316.6 Hz respectively for [tʂau] in /tʂau xuo/ (35 + 214) and 199.9 Hz and 244.4 Hz respectively for [tʂau] in /tʂau xuo/ (214 + 214) for Speaker Q.M. and correspondingly 202.0 Hz/293.5 Hz and 182.1/217.4 Hz for Speaker Y.H.J.

Table Id presents the durations (in seconds) for each of the bisyllabic compounds for both speakers. We can see that the duration of the bisyllabic compounds never exceeds 0.57 sec for Speaker Q.M. and 0.71 sec for Speaker Y.H.J., which shows that these bisyllabic compounds were produced at a fairly fast speech rate by both speakers.

CLAIM I above states:

/214 + 214/ → [35 + 214] (Rule 1)

But, according to our acoustical data the only time the tone /214/ on the first syllable changes to a rising contour is when the vowel in the first syllable is preceded by aspiration. Furthermore, as we have shown, when it does change to a rising contour, the end point is not as high as it is when the first syllable is underlying tone 2 /35/. In addition the tone /214/ on the first syllable does not always change to a rising contour, that is, it sometimes retains its lexical tone shape. Thus, CLAIM I (or Rule 1) does not apply to the data provided by our speakers. On the other hand, tone 2 /35/ on the first syllable changes to a dipping contour when the vowel or diphtong in the first syllable is preceded by aspiration. Based on the observation, we formulate the following rules:

$$/35/ \rightarrow [215] \; / \; C \underset{[-Asp]}{\_\_\_} \; /214/ \qquad \text{(Rule 4)}$$

$$/214/ \rightarrow [34] \; / \; C \underset{[+Asp]}{\_\_\_} \; /214/ \qquad \text{(Rule 5)}$$

Rule (4) and Rule (5) correctly describe the data provided by the two native speakers of Peking. Tone change in the bisyllabic compounds with the tone sequences /35 + 214/ and /214 + 214/ in Mandarin Chinese is less straightforward than what was previously claimed to be.

SECTION II.

In this section we will investigate whether the F0 contour of the 2nd-tone /35/ on the second syllable changes to [55] for speech at conversational speed when it is preceded by another 2nd-tone /35/ or a 1st-tone /55/ and followed by any except the neutral tone (CLAIM II), or,

$$/ \begin{Bmatrix}55\\35\end{Bmatrix} + 35 + X/ \rightarrow [\begin{Bmatrix}55\\35\end{Bmatrix} + 55 + X] \qquad \text{(RULE 2)}$$

$$X \neq \text{neutral tone}$$

The following trisyllabic compounds were used for investigation:

/tuŋ nan fən/	(55 35 55)	'Southeast wind'	from Cheng 1968
/çian ʐən tçaŋ/	(55 35 214)	'cactus'	
/mei ian faŋ/	(35 35 55)	'(a name)'	from Chao, 1968
/çi iaŋ şən/	(55 35 55)	'American ginseng'	
/san nian tçi/	(55 35 35)	'third grade'	
/tsʰuŋ iou piŋ/	(55 35 214)	'onion oil cake'	
/tuŋ xɤ̆ ian/	(55 35 51)	'East Riverside'	
/fən şuei liŋ	(55 35 214)	'watershed'	
/şuei nəŋ fei/	(35 35 55)	'Who can fly?'	
/xai mei uan/	(35 35 35)	'not yet finished'	
/xan şu piau/	(35 35 214)	'thermometer'	

The two female native speakers of Peking described in Section (I) read the above words at a fast speech rate. The recording was made in a sound-treated booth or a quiet room. Spectrograms of each of the above test trisyllabic compounds were made for both speakers and FO values were calculated for the beginning and end and, where relevant, for the central portion of a dipping contour. Table II shows the FO values obtained for Speaker Q.M. The trisyllabic compounds listed in Table II are arranged according to the shape of the FO contour on the second syllable as indicated in the last column of the table. We can see that tone 2 /35/ on the second syllable does not always change to [55], that is, HIGH LEVEL. In fact, in most cases (9 out of 11) the tone does not. It may change to a MID LEVEL [33] (approximately) as illustrated by compounds #3 and #4, to a HIGH DIPPING [535] (approx.) as illustrated by compound #5, or it may remain unchanged, as illustrated by compounds #6-11.

Table III shows the FO measurements of the trisyllabic compounds provided by Speaker Y.H.J. Tone 2 /35/ on the second syllable in this case changes to even more different shapes. It changes to HIGH LEVEL [55] (approx.), as illustrated by compound #1, to MID LEVEL [33] (approx.), as illustrated by compound #2 and #3, to FALLING [31] (approx.), as illustrated by compounds #4 and #5, and to DIPPING [313] (approx.), as illustrated by compounds #8-#11, or it may remain unchanged, that is, RISING [35] (approx.), as illustrated by compounds #6 and #7. Notice that for Speaker Y.H.J. there is only one out of eleven cases in which tone 2 /35/ changes to HIGH LEVEL [55].

Our findings based on the data provided by the two native Peking speakers apparently do not conform to CLAIM II (thus Rule 2). The discrepency between our acoustical data and the earlier impressionistic studies may be due to the fact that the FO contour on the second syllable is difficult to perceive becuase of its short duration. Table IV shows the durations (in seconds) for the first and the second syllables in each of the trisyllabic compounds. We can see that for Speaker Q.M. the total duration for the two syllables in any trisyllabic compounds does not exceed 0.38 sec and for Speaker Y.H.J. 0.43 sec. As the duration is so short, we can understand why the changing FO on the second syllable was not perceived by the authors of the earlier studies. To conclude, we have demonstrated that CLAIM II is not a valid generalization at the productive and acoustic levels. It may remain valid as an observation of perceived tone changes.

SECTION (3)

CLAIM III (Cheng, 1968) is concerned with tone change at the sentence level in Mandarin Chinese. In rapid speech (faster than fast speech, according to Cheng, 1968), if the tones on all the syllables in a sentence are 3rd-tones /214/, then, (a) the 3rd-tone on the first syllable changes to a 2nd-tone [35], (b) the 3rd-tone on the second, third and fourth syllables change to a 1st-tone [55], and (c) the 3rd-tone on the last syllable remains unchanged, or,

/214 + 214 + 214 + 214 + 214/ → [35 + 55 + 55 + 55 + 214] (RULE 3)

In this section, we will test whether CLAIM III is phonetically valid. The same sentence given in Cheng (1968) was used in our investigation and it is repeated in the following:

/lau li mai çiau pi/ (214 + 214 + 214 + 214 + 214)
'Old Li buys small pen'

The two female native speakers of Peking were instructed to read the sentence first in fast speech and then in the fastest speed they possibly could. Spectrograms of the sentence produced at the fastest speed are shown as Figures 2 (Speaker Q.M.) and 3-4 (Speaker Y.H.J.).

Table V shows the F0 values for the beginning and end points of the F0 contour on each one of the first four syllables [lau] and [li] [mai] [çiau] in the sentence produced at these two different speeds by Speakers Q.M. and Y.H.J. These values were obtained from the spectrograms shown in Figures 5a, 5b, 6a and 6b. Also shown are the difference between these F0 values in each syllable and the total duration of the four syllables produced with two speeds for both speakers. Table VI, which is based on the values shown in Table V, shows the shape of the F0 contour on each one of the first four syllables for Speakers Q.M. and Y.H.J. As far as the tone on the first syllable [lau] is concerned, it has indeed changed from Dipping /214/ to a RISING contour [35] in both speeds and for both speakers. However, the shapes of the F0 contours on the other syllables [li], [mai], [çiau] are far from being level, except for the cases of [çiau] in the fastest speech for Speaker Q.M. and [li] in fast speech for Speaker Y.H.J. They are either RISING or FALLING. Furthermore, Speakers Q.M. and Y.H.J. have produced opposite results for the syllable [mai]. The shapes of the F0 contours on the syllable are RISING in both speeds for Speaker Q.M., but FALLING in both speeds for Speaker Y.H.J.

We have demonstrated that in most cases the tones /214/ on the second third and fourth syllables /li/, /mai/, /ciau/ do not change to HIGH LEVEL [55] as Cheng (1968) has claimed that they should, although the tone on the first syllable in the sentence did change to RISING in all cases. Thus, CLAIM III is erroneous insofar as the tone change on the second, third and fourth syllables is concerned. The difference between our results and CLAIM III may be again attributed to the short durations of the syllables. As shown in Table V, the total time of the four syllables [lau li mai çiau] is only 0.83 sec (fast speed) or 9.55 sec (fastest speed) for Speaker Q.M. and 0.79 (fast speed) or 9.53 sec (fastest speed) for Speaker Y.H.J.

CONCLUSION

Results of the spectrographic analysis of the data provided by two female native speakers of Peking have shown that none of the three claims proposed by Chao, 1948, 1968 and Cheng, 1968 applies to tone sandhi at the productive or acoustic level in the speech of today's young generation of Peking Mandarin speakers.

		I		II	
		[tʰu] in /tʰu+kaɪ/ (214+214) 'land redistribution'	[tʰu] in /tʰu+kaɪ/ (35+214) 'to re-touch'	[tʂau] in /tʂau+xuo/ (214+214) 'to look for fire'	[tʂau] in /tʂau+xuo/ (35+214) 'on fire'
		Beg. End	Beg. End	Beg. Dip End	Beg. Dip End
Speaker Q.M.		244.0/306.2	238.3/360.6	229.9/201.4/248.3	236.7/216.7/310.0
		245.7/313.3	238.3/340.1	218.9/190.2/241.8	253.3/226.7/363.3
		231.5/306.4	243.6/333.0	234.7/204.7/241.5	236.7/203.3/286.7
		241.7/320.1	241.7/326.7	223.3/203.3/245.3	240.0/208.3/306.4
	$(\bar{X})$	240.9/311.5	240.5/340.1	226.7/199.9/244.4	241.7/213.8/316.6
Speaker Y.H.J.		235.3/289.2	232.3/297.9	235.3/180.4/270.6	247.1/207.8/290.2
		238.8/293.5	230.4/299.0	240.8/180.3/274.6	236.0/200.0/284.0
		228.9/283.6	222.2/297.9	239.2/187.7/269.4	244.0/200.0/302.0
		227.3/282.8	230.8/292.3	232.0/180.0/270.9	244.8/200.0/297.9
	$(\bar{X})$	232.6/287.3	229.0/296.8	236.8/182.1/271.4	243.0/202.0/293.5

Table Ia. F0 values for the beginning, (dip) and end points of the pitch contour on the 1st syllable in the bisyllabic words: /tʰu kaɪ/ (214+214) 'land re-distribution', /tʰu kaɪ/ (35+214) 'to retouch', /tʂau xuo/ (214+214) 'to look for fire', and /tʂau xuo/ (35+214) 'on fire' for both Speakers Q.M. and Y.H.J.

		III			IV				
		[tɕʰi˥] in /tɕʰi˥+ma/ (214+214) 'at least'		[tɕʰi˥˧] in /tɕʰi˥˧+ma/ (35+214) 'to ride a horse'		[mai˨˩˦] in /mai˨˩˦+ma/ (214+214) 'to buy a horse'			[mai˥˩] in /mai˥˩+ma/ (35+214) 'to bury a horse'		
		Beg.	End	Beg.	End	Beg.	Dip	End	Beg.	Dip	End
Speaker Q.M.		231.3/264.6		239.4/315.0		216.8/211.7/283.2			234.7/227.1/316.3		
		222.9/266.7		237.5/298.3		229.6/204.1/275.2			227.1/221.9/331.6		
		218.8/264.6		245.8/318.8		213.2/203.4/279.4			250.0/237.2/290.8		
		230.8/248.0		237.2/279.9		218.7/213.9/262.0			228.4/221.1/302.9		
	$(\bar{X})$	226.0/261.0		240.0/303.0		219.5/208.3/275.0			235.1/226.8/310.4		
Speaker Y.H.J.		226.8/268.5		240.1/297.8		220.0/198.0/268.0			220.4/200.0/277.5		
		243.3/277.3		252.3/301.8		212.2/200.0/269.4			220.0/206.0/288.0		
		234.2/279.3		240.8/283.8		216.0/200.0/265.5			216.3/208.2/292.2		
		242.0/280.1		232.9/292.3		208.0/176.0/260.0			220.0/200.0/280.0		
	$(\bar{X})$	236.6/276.3		241.5/293.9		214.1/193.5/265.7			219.2/203.6/284.0		

Table Ib. F_0 values for the beginning, (dip) and end points of the pitch contour on the 1st syllable in the bisyllabic words: /tɕʰi ma/ (214+214) 'at least', /tɕʰi ma/ (35+214) 'to ride a horse', /mai ma/ (214+214) 'to buy a horse', and /mai ma/ (35+214) 'to bury a horse' for both Speakers O.M. and J.H.J.

V

		[fən] in /fən+tʂʰaŋ/ (214+214) 'flour factory'			[fən] in /fən+tʂʰaŋ/ (35+214) 'graveyard'		
		Beg.	Dip	End	Beg.	Dip	End
Speaker O.M.		244.8	221.9	291.7	260.2	239.8	301.0
		250.0	219.4	306.1	255.1	239.8	321.4
		244.9	219.4	295.9	240.0	229.6	316.3
		245.0	220.0	305.0	240.3	230.8	317.3
	$(\bar{X})$	246.2	220.2	299.7	248.9	235.0	314.0
Speaker Y.H.J.		239.9	216.7	268.7	238.7	200.0	283.9
		235.5	201.3	258.1	241.9	212.9	277.4
		244.8	214.2	269.7	258.1	214.8	295.5
		232.2	209.7	260.0	251.6	214.8	279.4
	$(\bar{X})$	238.1	210.5	264.1	247.6	210.6	284.5

Table Ic. F_0 values for the beginning, dip and end points of the pitch contour on the 1st syllable in the bisyllabic words: /fən tʂʰaŋ/ (214+214) 'flour factory' and /fən tʂʰaŋ/ (35+214) 'graveyard' for both Speakers O.M. and Y.H.J.

	Speaker Y.H.J.					Speaker Q.M.				
					($\bar{X}$)					($\bar{X}$)
[t^hu kaɪ]/214+214/	0.60	0.58	0.55	0.58	0.57	0.58	0.55	0.54	0.49	0.54
[t^hu kaɪ] /35+214/	0.66	0.68	0.61	0.65	0.65	0.53	0.50	0.55	0.55	0.53
[tʂau xuo]/214+214/	0.58	0.53	0.62	0.60	0.58	0.61	0.53	0.62	0.53	0.57
[tʂau xuo] /25+214/	0.66	0.65	0.62	0.68	0.65	0.48	0.55	0.52	0.53	0.52
[tɕ^{h}i ma]/214+214]	0.52	0.52	0.54	0.68	0.57	0.53	0.49	0.52	0.53	0.52
[tc^{h}i ma] /35+214/	0.60	0.60	0.61	0.60	0.60	0.50	0.49	0.52	0.50	0.50
[maɪ ma]/214+214/	0.62	0.61	0.61	0.53	0.59	0.50	0.49	0.48	0.53	0.50
[maɪ ma] /35+214/	0.60	0.55	0.56	0.55	0.57	0.55	0.44	0.48	0.47	0.49
[fən tʂ^{h}aŋ]/214+214/	0.64	0.65	0.65	0.66	0.65	0.53	0.58	0.56	0.53	0.55
[fən tʂ^{h}aŋ] /35+214/	0.67	0.67	0.72	0.76	0.71	0.59	0.62	0.55	0.53	0.57

Table Id. The durations (in second) of all the bisyllabic word by Spkrs. Q.M. and Y.H.J.

	Trisyllabic word	Lexical tones	1st Syllable Beg.	1st Syllable End	1st Syllable Diff.	2nd Syllable Beg.	2nd Syllable End	2nd Syllable Diff.	Pitch contour on the 2nd syllable
# 1.	[fən şuei ɻɻŋ] 'watershed'	(˥ ˧ ˨˩)	322.5/329.4		(+ 5.9)	310.4/294.1		(-16.3)	HIGH LEVEL
# 2.	[san nian tçi] '3rd grade'	(˥ ˧ ˧)	328.4/338.2		(+ 9.8)	328.4/318.6		(- 9.8)	
# 3.	[tuŋ nan fəŋ] 'Southeast wind'	(˥ ˧ ˥)	330.9/316.2		(-14.7)	315.0/310.0		(- 5.0)	MID LEVEL
# 4.	[tsʰuŋ lou piŋ] 'onion oil cake'	(˥ ˧ ˨˩)	354.9/326.9		(-28.0)	320.7/312.8		(- 7.9)	
# 5.	[çian ʐən tçaŋ] 'cactus'	(˥ ˧ ˨˩)	319.0/312.8		(- 6.2)	317.3/298.1/322.1		(-19.2)(+24.0)	DIPPING
# 6.	[tuŋ xɤ lan] 'East Riverside'	(˥ ˧ ˩˨)	326.9/346.2		(+19.3)	284.0/308.0		(+24.0)	
# 7.	[çi laŋ şən] 'American ginseng'	(˥ ˧ ˥)	331.9/310.9		(-21.0)	297.8/316.2		(+18.4)	
# 8.	[şuei nəŋ fei] 'Who can fly?'	(˧ ˧ ˥)	283.9/296.3		(+12.4)	319.4/328.7		(+ 9.3)	
# 9.	[mei lan faŋ] '(a personal name)'	(˧ ˧ ˥)	271.6/289.8		(+18.2)	296.3/310.0		(+13.7)	RISING
# 10.	[xai mei uan] 'not yet finished'	(˧ ˧ ˧)	274.0/307.7		(+33.7)	328.0/333.0		(+ 5.0)	
# 11.	[xan şu piau] 'thermometer'	(˧ ˧ ˨˩)	288.5/314.1		(+25.6)	317.5/336.5		(+19.0)	

Table II. F_0 values for the beginning and end points of the pitch of the 2nd (and the 1st) syllable in the trisyllabic words with the lexical tone on the 1st syllable being either HIGH LEVEL or MID-RISING for Speaker Q.M.

Trisyllabic words		Speaker Q.M.	Speaker Y.H.J.
[tuŋ nan fən]	'Southeast wind'	0.29	0.38
[çian ʐen tçaŋ]	'cactus'	0.38	0.43
[mei lan faŋ]	'(a personal name)'	0.31	0.38
[çi laŋ ʂən]	'American ginseng'	0.34	0.38
[san nian tçi]	'third grade'	0.36	0.38
[tsʰuŋ iou piŋ]	'onion oil cake'	0.29	0.38
[tuŋ xɤ lan]	'East Riverside'	0.32	0.29
[fən ʂuei liŋ]	'watershed'	0.26	0.41
[ʂuei neŋ fei]	'Who can fly?'	0.29	0.43
[xai mei uan]	'not yet finished'	0.26	0.38
[xan ʂu piau]	'thermometer'	0.38	0.34
	$(\bar{X})$	0.32	0.38

Table IV. The total duration (in sec) for the first and second syllables in each of the trisyllabic words.

Spkr.	Speed	[lau] Beg.	End	Diff.	[lɪ] Beg.	End	Diff.	[mai] Beg.	End	Diff.	[çlau] Beg.	End	Diff.	Total duration
Q.M.	Fast	288.9/325.9		(+37.0)	308.0/209.9		(-98.0)	211.6/256.6		(+45.0)	236.1/259.3		(+23.2)	0.83 sec
	Fastest	259.3/327.2		(+67.9)	332.9/287.0		(-45.9)	259.3/281.5		(+22.2)	281.5/277.7		(- 3.8)	0.55 sec
Y.H.J.	Fast	203.7/270.0		(+66.3)	250.0/242.1		(- 7.9)	229.6/170.4		(-59.2)	192.6/237.1		(+44.5)	0.79 sec
	Fastest	204.6/222.2		(+17.6)	236.1/209.9		(-26.2)	200.0/181.9		(-18.1)	187.8/197.8		(+10.0)	0.53 sec

Table V. The F_0 values for the beginning and end points of the F_0 contour on each one of the first four syllables in the sentence [lau lɪ mai çlau pɪ] 'Old Li buys a small pen.' produced with two different speed by Speakers Q.M. and Y.H.J.

Spkr.	Speed	[lau]	[li]	[mai]	[çlau]
O.M.	Fast	RISING	FALLING	RISING	RISING
	Fastest	RISING	FALLING	RISING	LEVEL
Y.H.J.	Fast	RISING	LEVEL	FALLING	RISING
	Fastest	RISING	FALLING	FALLING	RISING

Table VI. The shape of the F_0 contour on each one of the first four syllables in the sentence [lau li mai çlau pi] 'Old Li buys a small pen.' produced with two different speeds by Spkrs. O.M. & Y.H.J.

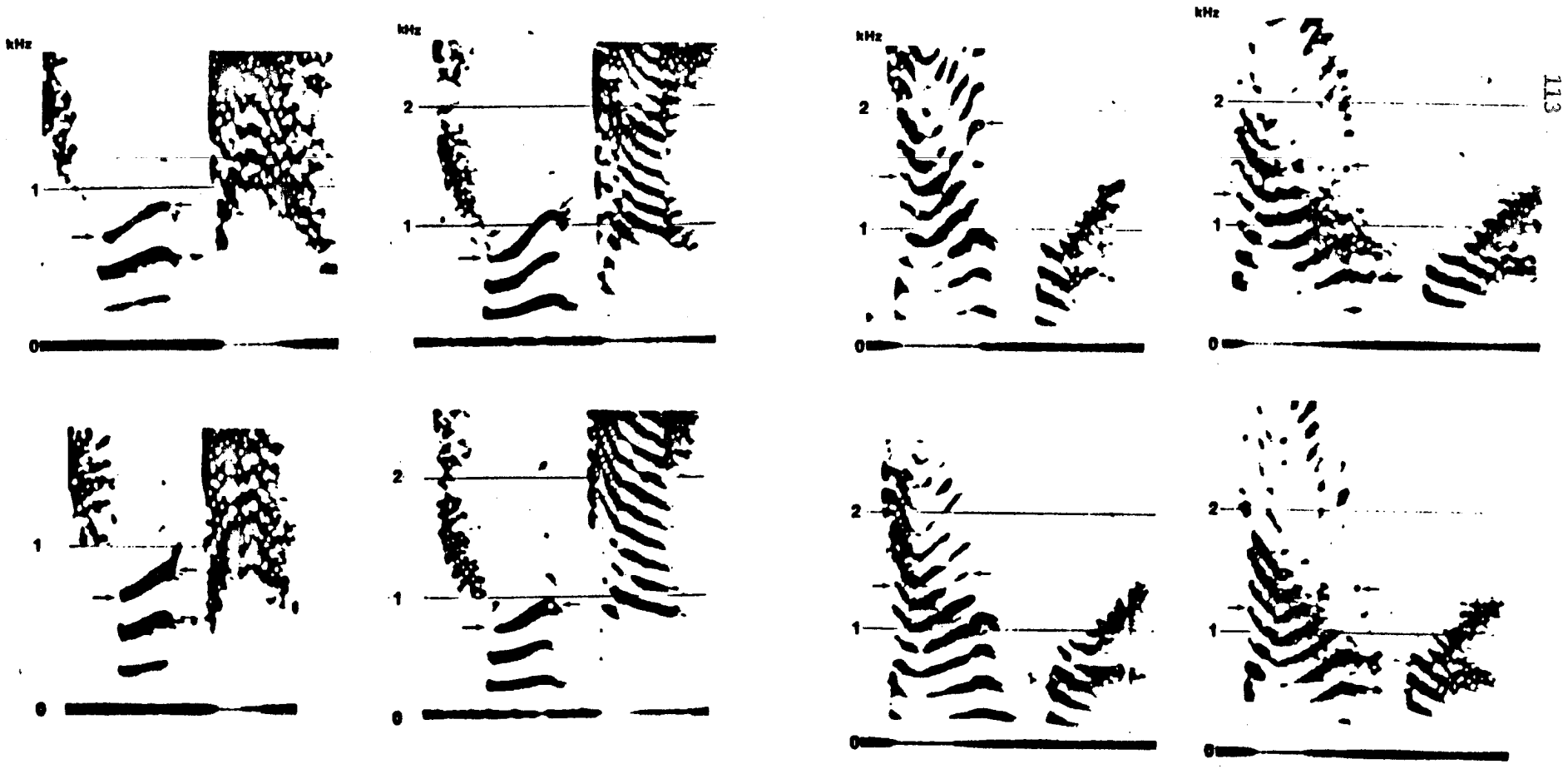


Figure 1. Sample spectrograms of bisyllabic pairs spoken by Q.M. (a,c) and Y.H.J. (b,d). On the left /tʰu kai/ (35 + 214) 'to retouch' (top, a and b) and (214 + 214) 'land redistribution' (bottom, a and b). On the right /tʂau xuo/ (35 + 214) 'on fire' (top, c and b) and (214 + 214) 'to look for fire' (bottom, c e d).

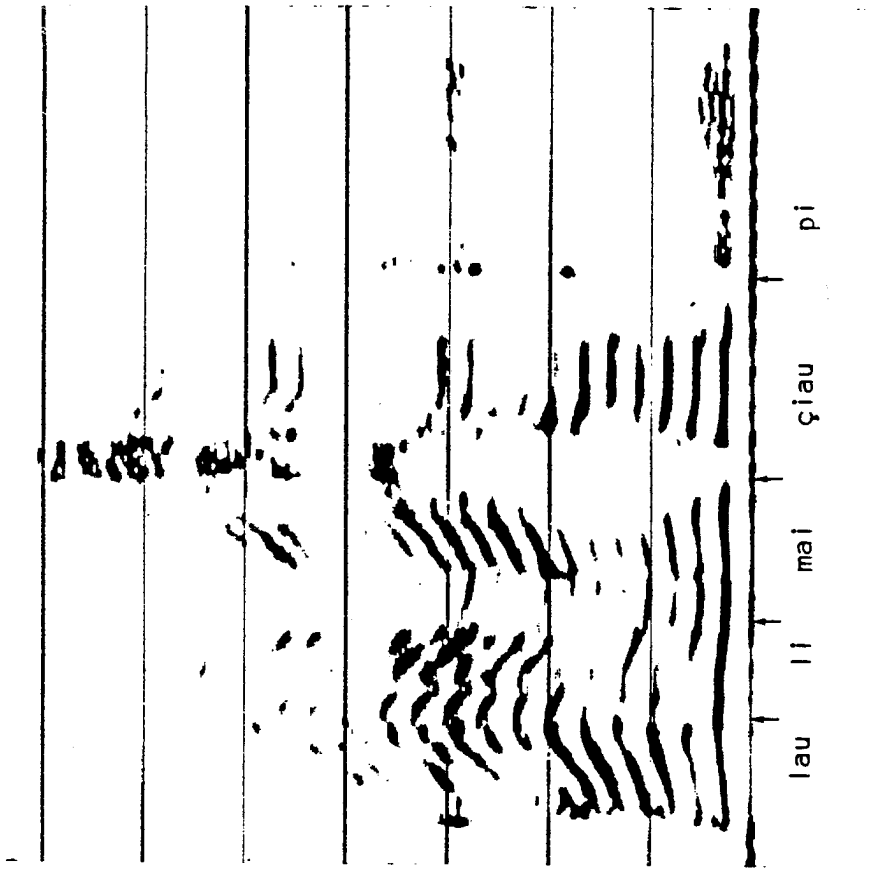


Figure 2. Spectrogram of /lau li mai çiau pi/ (214+214+214+214+214) 'Old Li buys a small pen' in the fastest speech by Speaker Q.M.

Figure 3. Spectrogram of /lau li mai çiau pi/ (214+214+214+214+214)
'Old Li buys a small pen' in the fastest speech by Speaker Y.H.J.

References

Chao, Y.R. (1948) *Mandarin Primer.* Cambridge, Harvard University Press.

Chao, Y.R. (1968) *A Grammar of Spoken Chinese.* Berkeley, University of California Press.

Cheng, C.C. (1968) *Mandarin Phonology.* Ph.D. Dissertation, University of Illinois.