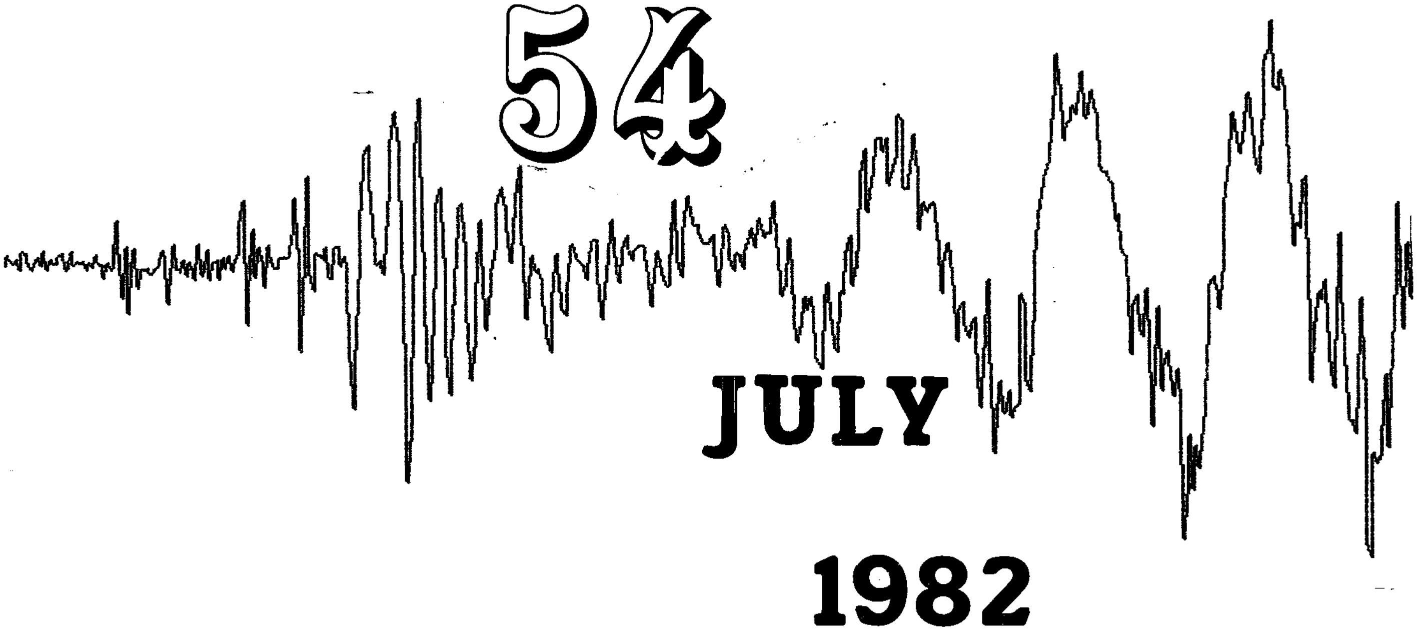
UCLA

WORKING PAPERS
IN PHONETICS
NUMBER
54



JULY
1982

The UCLA Phonetics Laboratory Group

Anthony Davey
Sandra F. Disner
Karen Emmorey
James F. Fordyce
Vicki Fromkin
Shannon Gardner
Manuel Godinez
Hector Javkin
Bruce Hayes
Nonie Holz
Pat Keating
Jenny Ladefoged
Peter Ladefoged
Thomas Lee
Mona Lindau
Wendy Linker
Ian Maddieson
Priscilla McCoy
Jonas Nartey
George Papçun
Ren Hong-Mo
Lloyd Rice
Mika Spencer
Helen Tsai
Diana Van Lancker
Anne Wingate
Andreas Wittenstein

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 for our colleagues in the field of work in progress.

Funds for the UCLA Phonetics Laboratory are provided through;

NSF grant BNS 80-23110
USPHS grant 1 R01 NS18163-01
and the UCLA Department of Linguistics

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

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

UCLA Working Papers in Phonetics 54

July 1982

A Phonetic Answer to a Phonological Question

Anne H. Wingate

Abstract: If acoustic-phonetic facts are used to determine the distinctive feature specification of the post-sibilant stops in English, the parameter most commonly associated with the voicing distinction, VOT, suggests that these stops should be specified [+voice]. Yet generative phonologists have persisted in characterizing these stops as [-voice]. To determine whether some other acoustic-phonetic parameter might account for native-speaker intuitions, production data for matched sentences containing the strings /s#b/, /s#p/, and /#sp/ were analyzed and a perceptual experiment using computer-spliced real speech was conducted. The production data indicated that although the VOT's for /s#b/ and /#sp/ may be identical, the FØ pattern in the following vowel was higher for /#sp/ and more like /s#p/. The results of the experiment indicate that this difference in FØ affects listener judgments, although it ranks low on a hierarchy of perceptual cues. Because any articulatory model posited to account for the higher FØ following the /#sp/ must incorporate increased glottal opening and/or tension, this [p] is less voiced than the [b̥] in /s#b/. Therefore the phonetic facts justify specifying the post-sibilant stops in English as [-voice].

I. Introduction

The twin concepts neutralization and archiphoneme, so important to the Prague phonologists 50 years ago, have been so altered in the theory of generative phonology that Trubetzkoy would find them unrecognizable. Clearly that metamorphosis reflects the redefinition of the phoneme as an entity which incorporates morphological information as well as phonetic contrast. That change--from archiphoneme to archi-segment to fully-specified feature matrices--reflects a new-found certainty of generative phonologists: in principle, there is no longer any doubt about the feature assignments of all segments, whether they occur in a context of neutralization or not.[1]

Given that some kind of "naturalness condition" is one of the cornerstones of generative phonology,[2] any controversy about the phonetic characteristics of a group of segments calls for an investigation of the articulatory and acoustic facts. This is especially true if those segments are not involved in morphological alternations.

Just such a controversy exists over the phonetic characteristics of English stops after word-initial /s/. Generativists have, for a variety of non-phonetic reasons, declared that the stops after word-initial /s/ are phonologically [-voice].[3] This conflicts with some phonetic data which indicate that these stops are phonetically closer to the typical initial allophones of [+voice] stops. That conflict must be taken seriously if one, in principle, looks to phonetic characteristics as the basis for categorizing non-alternating neutralized segments. To complicate matters further, a contemporary phonologist in the Praguian tradition has asserted that these stops are phonetically "ambiguous." (Davidsen-

Nielsen, 1978, p. 208)

This disagreement prompted the perceptual experiment presented here. The results of that experiment are offered as an answer to this question: Is the feature specification [-voice] justified phonetically for the English stops after word-initial /s/?

II. The problem of extracting distinctive feature values from phonetic content

Reliance on surface phonetics in the absence of some contrasting phonetic content presumes that distinctive feature values are discernable from the phonetic (acoustic and/or articulatory) facts alone. For example, there appears to be general agreement that the utterance-final [t] in German *Bund* is phonetically voiceless, and that this represents a neutralization of the voicing contrast between /t/ and /d/. But there are other phonetic realizations of neutralized segments whose feature specifications, some would claim, are *not* so evident. In Trubetzkoy's categorization of archiphoneme representatives, there was a type which he described as a "combinatory variant," with a phonetic realization which combined characterisitics of both oppositions. In fact, Trubetzkoy offered English stops after word-initial /s/ as an example of that category: ". . . in English, where the opposition between voiced lenis *b*, *d*, *g*, and voiceless fortis *p*, *t*, *k*, is neutralized after *s*, a special type of voiceless lenis consonant occurs in that position. . . ." (*Principles*, p. 79-80) More recently, the claim that neutralized segments exist which (a) do not enter into morphological alternations and (b) are phonetically "in between" has been used by Davidsen-Nielsen as an argument for the revival of the notion "archiphoneme." (Davidsen-Nielsen, 1978, pp. 15-19 and p. 158)

Like Trubetzkoy, Davidsen-Nielsen characterizes the English stops after word-initial /s/ as phonetically intermediate:

> These post-initial stops differ from the initial sounds of *pill*, *till*, *kill* by being unaspirated and from the initial sounds of *bill*, *dill*, *gill* by being invariably voiceless, and it is therefore difficult to non-arbitrarily identify them with one rather than the other of the two initial stop series on the basis of phonetic similarity. (1978, p. 17 ff.)

This passage implies that the standard against which neutralized stops should be measured is the initial set. If, instead, one looks at the allophones of voiceless stops in different contexts, it is clear that while the presence of aspiration always implies [-voice], the absence of aspiration does not preclude [-voice]. Therefore, the lack of aspiration of post-sibilant [p,t,k] does not render them "unlike" other voiceless stop allophones. Thus one might argue that aspiration in English is non-distinctive whereas voicing is distinctive and that the phonetic voicelessness of [p,t,k] indicates that they should be classified as [-voice] phonemically.

But such an analysis of the issue raised by Davidsen-Nielsen overlooks one crucial thing: word-initial /bdg/ for most speakers of English are also voiceless unaspirated. (See Table 1.)

Table 1

Voice onset time (VOT) for three different groups of English stops, by place of articulation.

	neutralized stops in the context /#s_/[1]	/bdg/ in word-initial position[2]		/ptk/ in word-initial position[2]	
	mean VOT (msec)	mean VOT (msec)	range (msec)	mean VOT (msec)	range (msec)
labials	12	1	0- 5	58	20-120
alveolars	23	5	0-25	70	30-105
velars	30	21	0-35	80	50-135

[1] Data are for words uttered in the frame "Say___ instead." The mean VOT's are for 5 tokens each from three speakers. (Klatt, 1975)

[2] Data are for words read in isolation from four speakers, excluding tokens (most of which were from one speaker) which contained pre-voiced /bdg/. That exclusion here is based on the fact that Lisker and Abramson, who gathered the data, concluded that whether a speaker pre-voiced /bdg/ was an individual matter. The finding that most English speakers most of the time do not pre-voice word-initial /bdg/ was confirmed by Zlatin (1974), who presented data for 20 adult speakers. In the data presented here, the number of tokens varies for each group, from 51 to 116. (Lisker and Abramson, 1964)

Thus while it is accurate to describe the post-sibilant stops as "voiceless unaspirated," it is not at all obvious how that phonetic characterization translates into the phonological labels [+voice] and [-voice].

The picture gets worse if one considers VOT values for allophones in a wide variety of positions (Umeda and Coker, 1974), or if one looks at VOT data for running speech (Lisker and Abramson, 1964). In both cases, the VOT values for the [-voice] set get smaller and encroach on the regions occupied by /bdg/. Thus the evidene is that the absence of voicing and the short lag in voice onset time which characterizes post-sibilant [ptk] also characterize the unvoiced allophones of /bdg/ and the unaspirated allophones of /ptk/.

Of course, one solution is to treat the apparent phonetic overlapping of post-sibilant [ptk] with the voiceless allophones of /bdg/ as an example of phonemic overlapping--and justify the assignment of [-voice] to the neutralized segments on phonological grounds, e.g., that it promotes simplicity and generality in the grammar.

On the other hand, a phonetic justification for assigning [-voice] to these segments is still possible. If we knew the invariant phonetic correlates of the [±voice] distinction, then the categorization of the neutralized stops would be a simple mapping process. But because those invariant correlates, if they exist, are not known, a phonetic-phonological argument would consist of two stages. First, one would have to demonstrate that there is some significant difference between post-sibilant [ptk] and the word-initial allophones of /bdg/ in spite of their common VOT values. Then one would have to argue that that difference, assuming that it is found, is in general relevant to the distinction between [+voice] and [-voice] sets of allophones, while the more obvious difference between post-sibilant [ptk] and the aspirated allophones of /ptk/ is non-distinctive.

III. Previous production and perceptual studies

A study by Davidsen-Nielsen in 1969 failed to find any differences between post-sibilant [ptk] and word-initial [b̥d̥g̥]. His data were from two American and two British English speakers uttering words beginning with /bdg/, /ptk/, and /sp, st, sk/ embedded in sentences. He measured VOT, closure duration, voicing during closure, and intensity of the noise burst at release. He found no differences in VOT or in burst intensity between [ptk] and the voiceless allophones of /bdg/. He noted but discounted the shorter closure duration of post-sibilant [ptk] (70 msec vs. 110 msec for /#ptk/ and 100 msec for /bdg/). He also dismissed the fact that /#bdg/ were sometimes voiced during closure. He limited his articulatory measurements to peak intra-oral air pressure and concluded that the differences in intra-oral air pressure correlated with the presence or absence of voicing during closure. (The graphs of his pressure tracings, however, appear to show differences in the pattern of rising intra-oral pressure.) His conclusion was that "the interpretation of of /sb, sd, sg/ is preferable if the criterion of 'phonetic similarity' is applied." (1969, p. 17)

However, other evidence in the phonetics literature suggests that the issue is worth pursuing.

Data on Korean, which has three phonemically distinct sets of voiceless stops, show that it is possible to both produce and perceive differences between stops which have overlapping VOT values (Kim, 1965; Han and Weitzman, 1970; Abramson and Lisker, 1972; Hardcastle, 1973; Kagaya, 1974).

Klatt in his 1975 study of production data for /s/ cluster in English states that post-sibilant [t] and [k] frequently have noticeable aspiration in spite of the fact that their VOT values are closer (or almost identical) to those of [d] and [g]. In other words, for [t] and [k] the period between release and voice onset resembles qualitatively that same period in aspirated [t^h] and [k^h] in that it is possible to distinguish, on spectrograms, two different phases; a burst frication (strong excitation of only certain formants) followed by aspiration noise (weaker excitiation of all but the first formant). The aspiration, as distinct from the burst, is not always found in [t] and [k] after word-initial]s], but it is never seen in the voiced stops. (1975, p. 691) In the case of [p], however, the frication burst is too weak and spectrally diffuse for it to be differentiated from aspiration. (p.688)

Testing the acoustic similarity between initial [bdg] and the stops in /s/-clusters was the purpose of three rather similar perceptual experiments, by Lotz, *et al.* (1960), Reeds and Wang (1961), and Davidsen-Nielsen (1969). In all three studies, listeners heard tape recordings of words beginning with /ptk/ and /bdg/ randomly ordered with other words like *spill*, *still*, and *skill* from which the /s/ had been spliced out. Upon the presentation of each token, subjects had to identify the word in a forced choice between /bdg/ and /ptk/. In all three experiments, the subjects heard the truncated words as beginning with /bdg/, e.g., as *bill*, *dill*, or *gill*, demonstrating that the lack of aspiration of these stops, when they are presented as word-initial, makes them tend to be heard as /bdg/. However, in the Reeds and Wang study, it was noted that Wang "observed a relatively consistent cue which distinguishes the set [of] words [beginning with /bdg/] from the truncated set . . . in that the pitch drop appears to be more immediate and abrupt in the . . . words [which were originally /sp, st, sk/]." (p. 80)

McCasland (1977) cited data from a 1965 article by Lisker that suggested phrases like *it's dill* and *it still* may be distinguished from each other by a combination of [s]-duration and stop-closure duration: the duration of word-final [s] is the same as or less than the following stop closure but the duration of word-initial [s] is longer than the stop closure. The mean durations, across all three places of articulation for the stops, as cited by McCasland, would give the following ratios:s/#bdg = .62; s/#ptk = .99; s/non-initial ptk = 1.35.

McCasland devised a listening test in which the duration of [s], the stop closure, and the aspiration were altered on a single taped utterance of *it's still* by adding or removing sections of the audio tape. Although he claimed that the distinction between *it's dill* and *it still* was due to an "interaction of the stop closure and the preceding /s/ segment durations" (p. 227), his results showed in fact that 61% of his unaspirated tokens were ambiguous, i.e., their highest score for any one identification was 64% or less. Moreover, although there seemed to be a correlation between his subjects' responses and the kind of ratios suggested by Lisker's data, there were pockets of ambiguity unexplained by the ratio of the two durations.

Using an experimental paradigm unusual for speech perception but common in psychological experiments, Jaeger (1980) found evidence that listeners unequivocally associate the neutralized stops with the voiceless set. She used both a classical conditioning paradigmn and a category-formation paradigm, in which, respectively galvanic skin response and reaction times (for button pushing) were used as the criteria for how subjects categorized [k] after /#s/. Jaeger attempted to control for orthographic influences by selecting stimuli words in which [k^h] and [sk] were spelled *k*, *c*, *ch*, and *q*. The results showed that in Experiemnt 1, the mean galvanic skin response to the [sk] words was not significantly different from the mean response to non-/k/ words. And in Experiment 2, subjects identified [sk] words 92.6% of the time as belonging to a previously-formed but unnamed /k/ category.

Taken together, these studies indicate the following:

(1) Even when there is control for orthographic influences, English speakers unequivocally assoicate the post-/s/ stop allophone with /k/. If this also holds for labials and alveolars, it cannot be explained by reference to the fact that post-/s/ stops are sometimes aspirated, for [p] in /#s___/ is never aspirated.

(2) Tape-cutting experiments tell us little of how the neutralized stops are perceived in the context /#s___/. They simply demonstrate that the presence or lack of aspiration is the dominant cue for word-initial stops when the word boundary is unambiguous.

(3) The dominant cue for distinguishing the unaspirated strings /#st/ and /s#d/ is probably the relative loudness of the [s], which can be captured to some extent by the ratio of the [s]-closure duration and the stop closure duration. However, this ratio cannot account for all the /#st/ and /s#d/ judgments in McCasland's experiment. Clearly some other factor affected the subjects' responses. The fact that McCasland began with a single utterance may account for the high rate of ambiguous responses. He assumed that the vowel nuclei were interchangeable.

(4) Wang's observation about the pitch difference in vowels following truncated and word-initial stops indicates that the stop-vowel nuclei are not interchangeable. (The fact that subjects did not notice this pitch difference can be attributed to a hierarchy of acoustic cues; in word-initial position, the presence or absence of aspiration is more salient than a difference in pitch pattern in the following vowel.) Wang's comments are particularly interesting because other investigators have shown that the pitch differences in the vowel following obstruents are assoicated acoustically and perceptually with the voicing distinction. The acoustic measurements of House and Fairbanks (1953) and of Lehiste and Peterson (1961) as well as the perceptual experiment of Haggard, Ambler, and Callow (1970) indicate that /ptk/ are associated with a relatively higher pitch in the following vowel, while /bdg/ are assoicated with a lower fundamental frequency. The difference in pitch is found at the vowel onset. The perceptual importance of these pitch differences also holds for the voiced-voiceless distinction in fricatives, especially when VOT values are ambiguous. (Massaro and Cohen, 1976)

These studies suggested that the experiment reported in this paper should be performed. In this experiment, listeners have to distinguish between the strings /s#b/, /s#p/, and /#sp/, embedded in a sentence frame which provides no "top-down" information. Deprived of pauses (an unequivocal cue for placement of word boundaries), subjects must use the acoustic cues in the [s]'s and in the stop-vowel nuclei. Their identifications of different [s]'s paired with different stop-vowel nuclei demonstrate whether listeners distinguish between the unaspirated stops in /s#b/ and /#sp/ solely on the basis of some quality of the [s] or whether the distinction is also affected by something in the stop-vowel nucleus.

IV. A new experiment

1. Production data.

An initial recording was made of a single male speaker, G.P. saying a set of 15 words. The 15 words consisted of five sets of rhymes beginning with /b/, /p/, or /sp/. Each rhyme consisted of one of the five front vowels followed by a voiceless stop or affricate. The words which began with /b/ and /p/ were spoken in the frame "Say 'that's _____' again," and the words beginning with /sp/ were spoken in the frame "Say 'that ____' again." The words were

beach	bit	bait	beck	bat
peach	pit	pate	peck	pat
speech	spit	spate	speck	spat

Five examples of each sentence were written on 3 x 5 cards and randomly ordered. Additional cards were placed at the front and back of the "deck" to minimize the effects of start-of-list and end-of-list intonation. The 81 tokens were recorded on an Ampex tape recorder with G.P. in a sound booth. Oscillomink tracings of wave-form and amplitude were made of the entire tape, and the duration of the [s]'s and stop closures for 75 of the utterances were measured. (Six had been misread.) These measurements are given in Table 2.

Table 2

Durations based on measurements of oscillomink tracings of waveform and amplitude, to nearest 5 msec.

A. [s] durations	mean	standard dev.	range	no. of tokens
/s#b/	56.3	7.1	40- 70	26
/s#p/	70.7	12.2	56-110	23
/#sp/	100.8	17.1	81-145	26
B. stop cl. durations				
/s#b/	81.5	9.5	70-95	26
/s#p/	74.1	9.0	60-95	23
/#sp/	69.6	5.6	65.80	26

In addition, the ratios of [s]-duration/stop closure duration were calculated for each token (Table 3).

Table 3

Ratios of [s]-duration to stop-closure duration based on measurements of oscillomink tracings.

	mean ratio	standard dev.	range	no. of tokens
[s]/#b	.70	.12	.50- .92	26
[s]/#p	.96	.16	.65-1.40	23
[s]/p	1.37	.35	1.13-2.07	26

From the 75 tokens, 15 were selected. In selecting those 15 tokens, the following criteria were used:

(a) All tokens which sounded ambiguous to the experimenter were excluded.

(b) No /s#b/ which contained a pre-voiced [b] was used. (Out of 27, five had negative voice onset times.)

(c) To minimize temporal dislocation,
(i.) from the /s#b/ tokens, the ones with the longest [s] were selected;
(ii.) from the /#sp/ tokens, the ones with a short [s] but without unusually high intensity were selected.

(d) Of the /s#p/ tokens, ones with the [s] duration which came closest to the [s] duration of the comparable /s#b/ token were selected.

Note that with the /#sp/ tokens, all of the tokens with the shortest [s] were rejected. The intensity tracings on the oscillograms indicated that the shortest [s]'s had extremely high intensity (amplitude) values relative to the adjacent vowel nuclei. Apparently there is a trade-off between the intensity and duration values for word-initial [s].

The 15 tokens which were selected were digitized on an LSI 11/23 computer at the U.C.L.A. Phonetics Laboratory, and a wave-form printout was generated on which more precise measurements could be made. Durations for the following portions of the two words in the phrase "that('s) ____" were measured: first vowel, [t] closure, [s] frication, labial stop closure, voice onset time, and second vowel. The ratio of [s]/stop closure and the sum of $V_1 + V_2$ were calculated. The measurements of the [t] in that('s) and of the two vowels were a check to see whether the rate of speech was similar for each vowel group of three. These measurements revealed that G. P. was remarkably consistent in his rate of speech: the most [t] varied within any one vowel group was 7 msec and the most $V_1 + V_2$ varied for any one set of three was 26 msec in spite of the fact that all tokens were randomly ordered. The values for the other measurements are found in Table 4.

Table 4

Duration measurements (msec) for 15 tokens from computer displayed waveform.

		[s] duration	cl. duration	[s]/cl. ratio	VOT
/s#b/	beach	67	104	.64	14
	bit	53	78	.68	15
	bait	60	86	.70	14
	beck	49	75	.65	15
	bat	65	73	.89	14
/s#p/	peach	73	65	1.12	51
	pit	58	72	.81	42
	pate	57	72	.79	51
	peck	71	73	.97	60
	pat	68	62	1.09	52
/#sp/	speech	94	75	1.25	16
	spit	109	75	1.45	15
	spate	100	66	1.52	14
	speck	93	60	1.55	14
	spat	98	60	1.63	15

From the same wave-form printout which served as the basis for the duration measurements in Table 4, the fundamental frequency pattern for the entire vowel following the /s/-cluster was calculated by measuring the duration of each period pulse. The results of those measurements can be seen in Fig. 1.

From these production data, it is clear that /s#b/ can be distinguished from /#sp/ not only by the marked difference in the [s] duration but also by the difference in stop closure duration and, consequently, the difference in the ratio of the two durations. In addition, it appears that in given pairs matched for following vowel, there is a marked difference in the fundamental frequency (F∅) onset frequencies and patterns for /s#b/ and /#sp/ even though the VOT's are the same (cf. Table 4). For this data, F∅ onset in /#sp/ is always higher than in /s#b/, although it is usually lower than the F∅ onset following /s#p/. (In one of the five it is higher.) Furthermore, if one looks at the over-all pattern of F∅ from the beginning to the end of each vowel, the F∅ slope is least in the /s#p/ tokens (i.e., always negative), and greatest in the /s#b/ tokens (nearly zero or positive). The /#sp/ tokens fall in between but are more like the corresponding /s#p/ tokens. (In the case of the [e^i] set, these patterns are superimposed on a diphthong.)

2. The Hypotheses.

After it was ascertained that the tokens produced conformed to the findings of earlier studies, the perceptual experiment was designed to test the following two hypotheses:

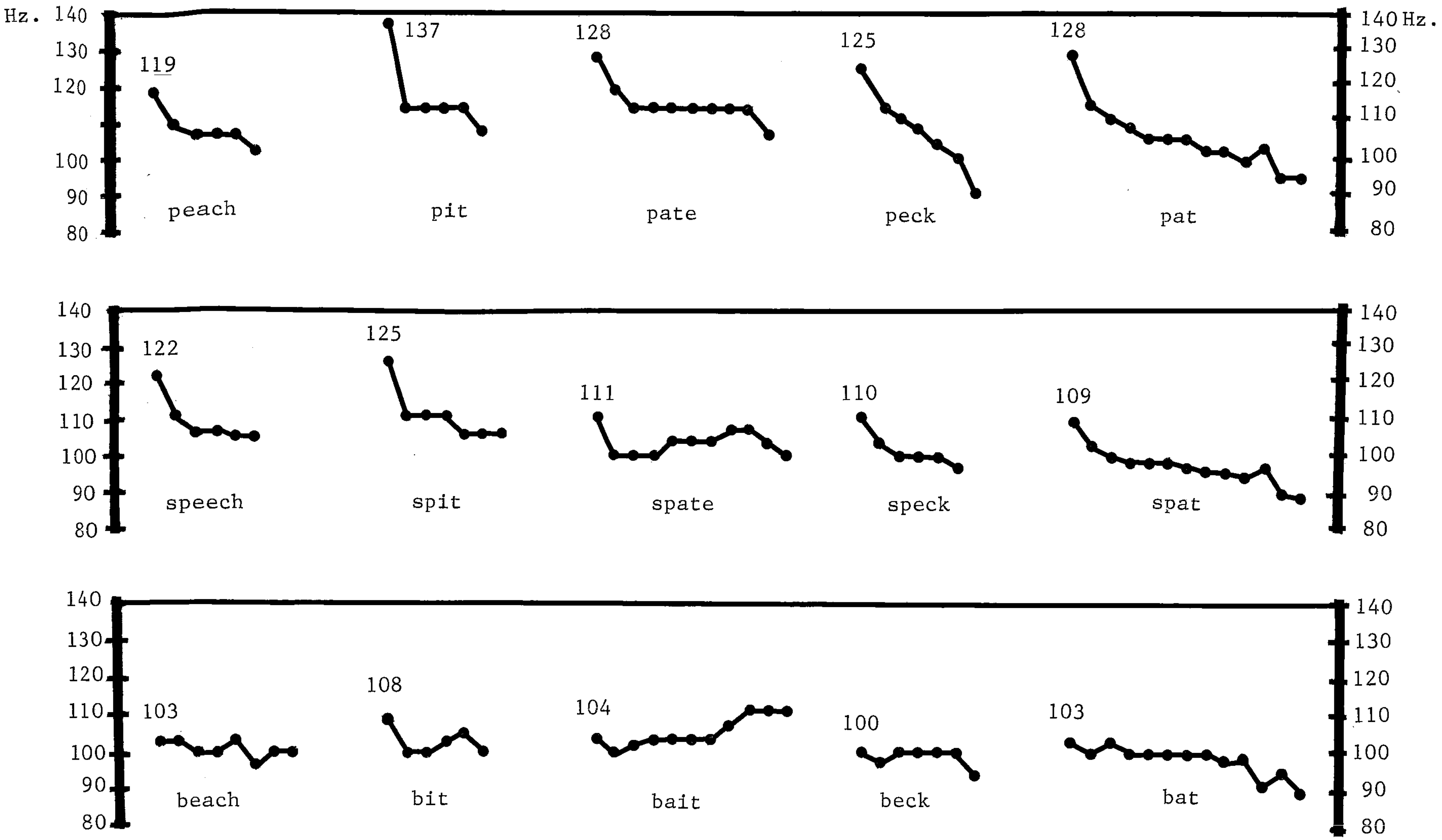


Fig. 1. Fundamental frequency (FØ) in Hz. for the fifteen original utterances. Values were calculated from periodic pulse durations. Each point represents one pulse.

Hypothesis 1. When listeners are presented with stimuli which have the full range of [s] durations, in combination with each of the three types of stops, [b̥, p^{h}, p], they will label stimuli containing aspirated [p^{h}] as /s#p/ close to 100% of the time. Conversely, they will label stimuli containing the unaspirated stops [b̥] and [p] as /s#p/ at levels approaching zero per cent.

Hypothesis 2. Listener responses to unaspirated stimuli will be divided between /s#b/ and /#sp/ judgments. However, the choice between /s#b/ and /#sp/ will be affected not only by which [s] is combined with the unaspirated stop but also by what follows the [s], i.e., the stop + vowel nucleus. In statistical terms there will be two main effects.

3. Stimulus set.

The 15 tokens originally selected to serve as the core set were redigitized, copied, and edited by computer to create a total of 45 stimulus utterances: one complete set was left intact, unedited, and the other two sets were altered by interchanging the [s]'s. In order to indicate the original source of the [s] segments in the edited stimuli, a system of superscripts will be used. Thus an [s] with a superscript [$^{\#b}$] will indicate that this [s] was one originally recorded before a word-initial /b/. So, for example, the three phrases in the [e^{i}] vowel group would be retranscribed as

that's bait	[ɣæts$^{\#b}$ b̥e^{i}t]
that's pate	[ɣæts$^{\#p}$ p^{h}e^{i}t]
that spate	[ɣæt s^{p}pe^{i}t]

In the editing process, nothing on either side of the [s]'s was altered, e.g., no changes were made in the stop closure durations. Thus the portion of the syllable following the [s], the stop + vowel nucleus and final consonant, was left intact. To designate the three types of stop and the following vowel, the phonetic notation for the three stops will be used: [b̥], [p^{h}], and [p]. Within each vowel group, six new utterances were constructed, giving the following 3 x 3 set of types (asterisks mark the unedited types):

[s$^{\#b}$ b̥]*	[s$^{\#p}$ b̥]	[s^{p} b̥]
[s$^{\#b}$ p^{h}]	[s$^{\#p}$ p^{h}]*	[s^{p} p^{h}]
[s$^{\#b}$ p]	[s$^{\#p}$ p]	[s^{p} p]*

The computer editing of these tokens was done on the PDP-11/34 at the U.C. Berkeley Phonology Lab. The audio tape, originally recorded at 7½" per second, was played at half speed and was digitized at a sampling rate of 20,000 Hz. The low-pass filter used had a cut-off of 9.5 kHz. By playing the tape at half speed, frequencies up to 18 kHz were preserved. The purpose was to maintain the integrity of the [s]'s, which often have frequencies as high as 10-12 kHz.

Copies of the new audio tape were made on the U.C.L.A. Phonetics Laboratory's Tandberg recorders. These copies were then spliced by hand so that there was an interstimulus interval of 4 seconds. The stimulus tape consisted of one randomized set of the original 15 unedited types followed by three complete sets of 45, each set with a different quasi-random order. (The randomized orders were adjusted so that no two stimuli of the same vowel type or same consonant cluster type followed each other.) The 150 tokens ((15 + (3 x 45)) were then divided into seven "sections" corresponding to the 7-paged, multiple-choice answer sheet used by the subjects to record their judgments. Additional intervals were spliced into the tape for page-turning instructions.

4. Procedure.

The test was administered to three groups of U.C.L.A. undergraduate linguistics students (a total of 21), who were paid for the one-hour session. Subjects filled out questionnaires on their language background and received instructions before they heard the stimulus tape. The responses of three subjects who were from bilingual families and had learned a second language at the same time they had learned English were excluded from the analysis. All the remaining 18 subjects were born and grew up in the United States; none had learned a second language until after the age of 10.

After completing the questionnaire, the subjects received the following instructions in written and oral form:

> "You are about to participate in a speech perception experiment. Although the format can be described as a listening 'test,' it is a test of the material, not of you, the listener. There are no right or wrong answers. The responses you give will add to our knowledge of how English speakers hear certain sounds.
>
> You will hear a total of 150 sentences in the first section and 135 in the second section. Each of the sentences consists of the frame 'Say ___ again,' with one of the following 15 phrases inserted in the frame:
>
> | | | | | |
> |---|---|---|---|---|
> | that's beach | that's bit | that's bait | that's beck | that's bat |
> | that's peach | that's pit | that's pate | that's peck | that's pat |
> | that speech | that spit | that spate | that speck | that spat |
>
> Note that each phrase contains an s. In some cases, the s is at the end of 'that's' and in other it begins the second word. Your task is to listen for the second word in the phrase 'that('s)___' and to check it off on your reponse sheet. For example,
>
> beach _____ peach ✓ speech _____ .
>
> Because you will hear each sentence only once before the next one is presented, be sure to mark your response as soon as you have heard each sentence. Check the first word that comes to mind. Please try to make a response to each of the sentences you hear.

> Any 'pattern' you find in the answers is accidental. Moreover, it is unlikely that you will hear an equal number of each of the three types (words beginning with *b*, *p*, and *sp*). Your best strategy--and the one most helpful to us--is to concentrate on each sentence as it is presented on the tape.
>
> The test is divided into sections. Each section corresponds to one of your answer pages. To help you keep track of where you are, you will be told when to turn the page. *The first section is for practice.* It is shorter than the other sections. Its purpose is to familiarize you with the voice on the tape. If you must leave some answers blank on the practice page, do not worry about it. The answers on that page will not count."

The purpose of the short set was to familiarize the subjects with the voice on the tape and the procedure of checking off responses after hearing each token only once (forced choice). Subjects heard the remaining three randomized sets of 45 tokens twice. Between the two halves of the listening test, each of which lasted 15 minutes, the subjects were given a 10-minute break during which the answer sheets from the first half were collected. Playing the stimulus tape twice provided six responses from each subject to each of the 45 types for a total of 108 responses to each type.

The subjects took the test in a small classroom in the U.C.L.A. Phonetics Laboratory. The tape was played on one of the Tandberg tape recorders over two ceiling-mounted loud speakers. The listening stituation was not ideal. The classroom is vulnerable to environmental noises from the building's heating and ventilation equipment in an adjacent room as well as to noises from the adjacent offices.

5. Results.

The raw identification scores across all subjects and across all vowels are given in Figure 2. In Figure 2, there is a separate graph of responses for each of the nine stimulus types (data collapsed across vowel sets).

The total number of responses to each cluster category was 540 (18 subjects x 6 presentations of each cluster category x 5 vowel groups). Recall that at each presentation, subjects selected from three possible responses: words indicating they had heard /s#b/, /s#b/, and /#sp/. Within each box in Fig. 2 are three points corresponding to the number of responses in each of the three response categories (/s#b/, /s#p/, /#sp/).

Note first that aspirated [p^h] stimuli and two response groups, /s#b/ and /#sp/, are almost mutually exclusive. That is, when the stimulus contained unaspirated stops, [b̥] or [p], the /s#p/ responses were very low. This fact is made explicit in Figures 3 and 4.

Fig. 3 shows that the unaspirated stimuli elicited responses from each subject which were /s#b/ 53.7% of the time (mean across the 18 subjects) and

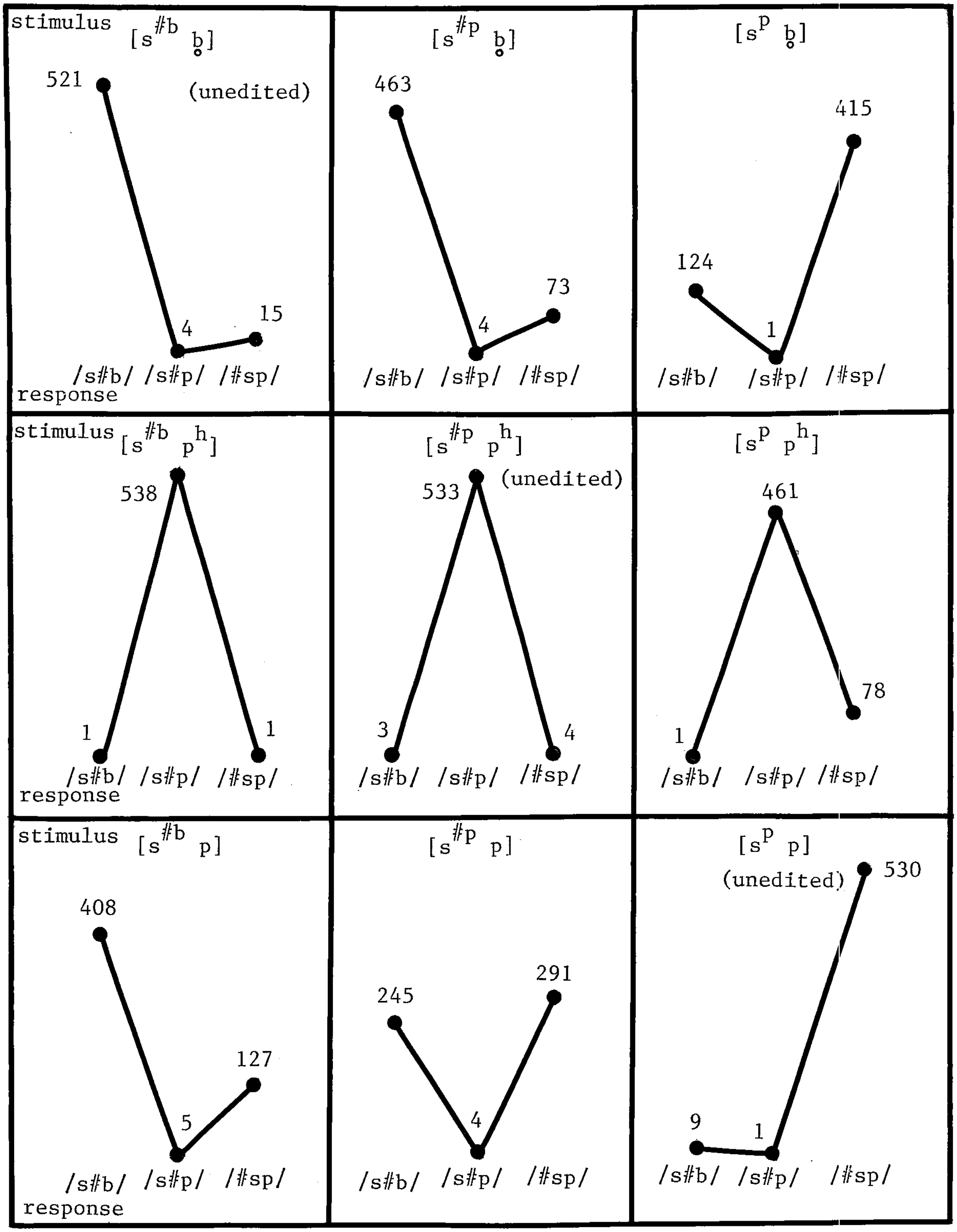


Fig. 2. The stimulus types (at top of each square) and their responses by category (bottom of each square), across all vowels. Total number of responses to each stimulus type was 540, divided among /s#b/, /s#p/, and /#sp/. The unedited stimuli are labeled. The other six had [s]"s replaced:

$[s^{\#b}]$ < s#b $[s^{\#p}]$ < s#p $[s^{p}]$ < #sp

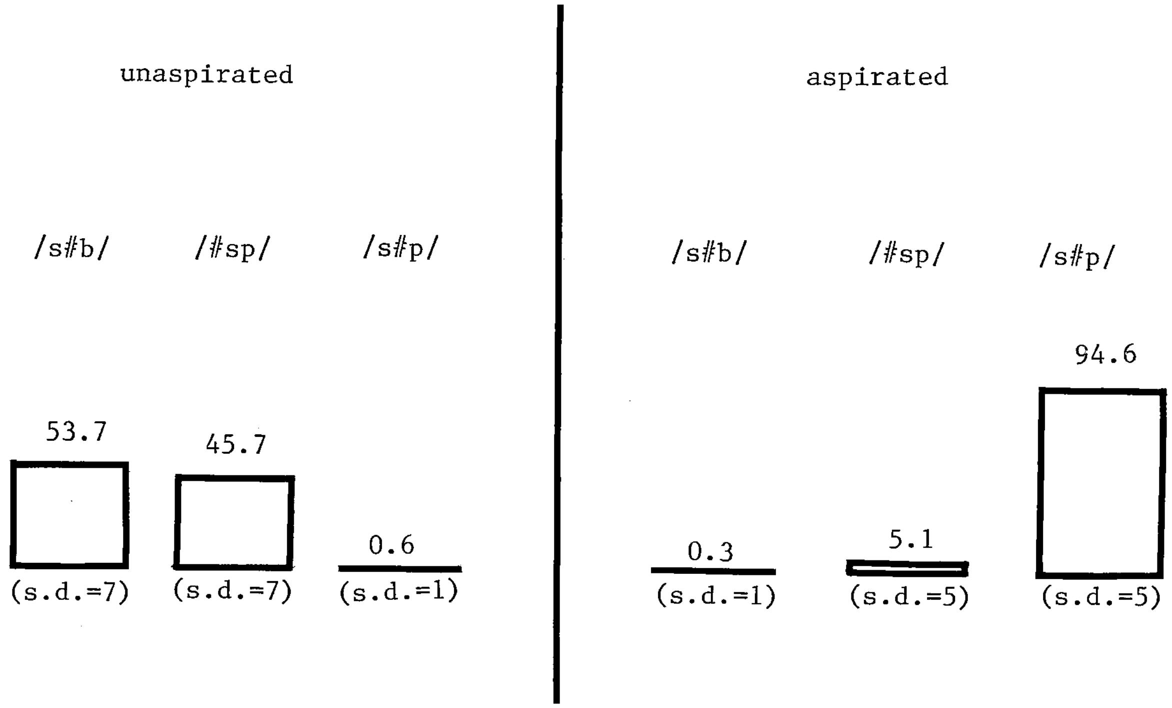


Fig. 3. Per cent of unaspirated and aspirated stimuli identified as /s#b/, /#sp/, and /s#p/. Per cent values represent means across eighteen subjects.

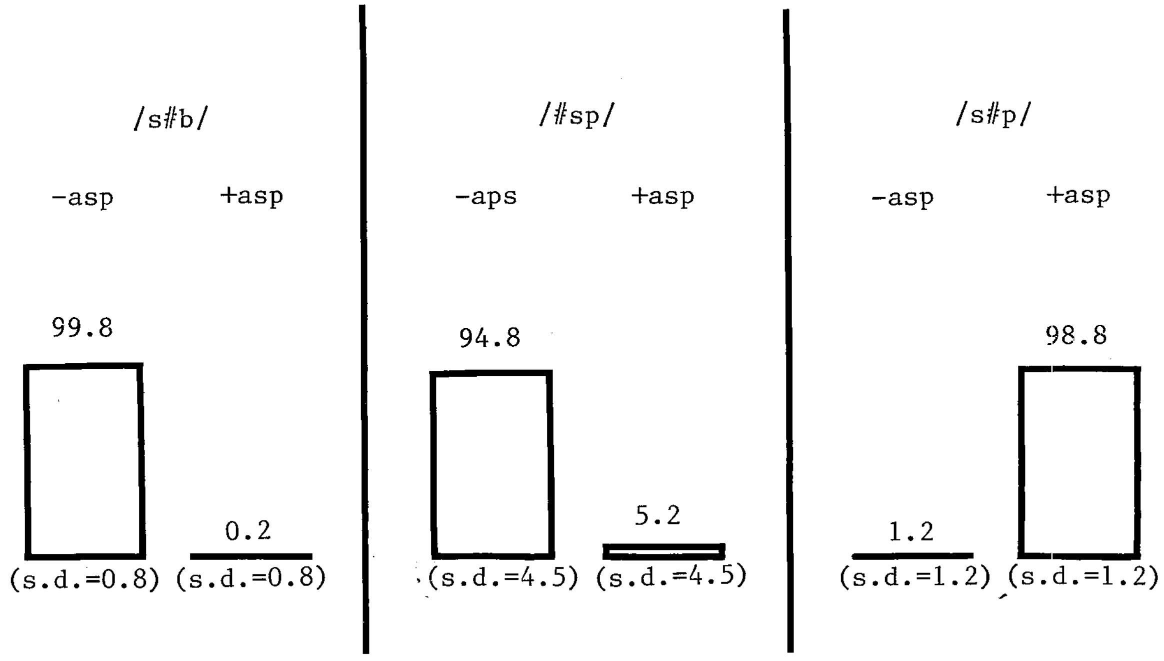


Fig. 4. Per cent of each of the three identifications, /s#b/, /#sp/, and /s#p/, elicited by unaspirated and aspirated stimuli. Per cent values represent means across eighteen subjects.

45.7% of the time /#sp/. In contrast, the aspirated stimuli were identified by each subject as /s#p/ 94.6% of the time.

Fig. 4 shows for each response category across all vowels the proportion which was elicited by aspirated and unaspirated stimuli. Of the total number of /#sp/ responses from each subject, 94.8% were elicited by unaspirated types. The /s#b/'s had a similar proportion (99.8%) prompted by unaspirated stimuli, while 98.8% of the /s#p/ responses from each subject were prompted by aspirated stimuli.

Fig. 5 presents the scores in a different form. Here the data are organized by response category first and then by stimuli so that one can more easily see how different [s]'s and different stop-vowel nuclei contributed to the pattern of responses. Within each response box, each symbol represents the percentage of responses to one of the the 9 different stimuli in a vowel group which were given the label in that box by the subjects. The combination of symbol (square, empty circle, filled circle) and line (straight, dashed, and dotted) encodes the stimulus type. By following the connected lines, one may see how different [s]'s (represented by the three different kinds of symbols) affected the distribution of responses.

In Fig. 5 one may see that the percentage of responses in the categories with word-final /s/'s (/s#p/ and /s#b/) decreases as the length of the [s] increases. (Compare the [s] durations and corresponding distribution of responses in Table 5.) This is to be expected because longer [s]'s are usually associated with word-initial /s/. Conversely, the one category with word-initial /s/ (/#sp/) increases as the length of the [s] increases.

One should also note that the /s#b/ and the /#sp/ responses to unaspirated stimuli are almost mirror images of each other. The combined scores for these stimuli (straight lines or dotted lines) often come close to 100 % (within the same-vowel group). In other words, as /s#b/ scores go down in response to unaspirated stimuli, /#sp/ scores go up.

From Fig. 5 and Table 5 it is also apparent that increasing the duration of the [s], that is, changing the /s/ from [s$^{\#b}$] to [s$^{\#P}$] to [s^{P}], does not affect stimuli with [b̥] and [p] equally. This pattern--that there are differences in responses associated with different stop-vowel nuclei--appears to hold for most of the 15 response groups. Thus the percentages in Fig. 5 and in Table 5 suggest that Hypothesis 2 is supported by the data, that not only the kind of [s] but also the particular stop-vowel nucleus affects listeners' identifications of unaspirated stimuli.

To test Hyposthesis 2, a two-way analysis of variance was done on the number of /s#b/ identifications only. (Recall that /#sp/ identifications were mirror images of /s#b/ responses.) Because aspirated stimuli elicited negligible /s#b/ responses, the scores analyzed were for the 30 unaspirated stimuli. Thus the two factors analyzed were the type of [s] (that is, [s$^{\#b}$], [s$^{\#P}$], or [s^{P}]) and the type of unaspirated stop ([b̥] or [p]). As predicted, there were two statistically significant main effects ($p < 0.01$). There was also significant interaction for four of the five vowel groups.[5] A Newman-Keuls post-hoc test was done on the three sub-goups of [s] types to pin-point

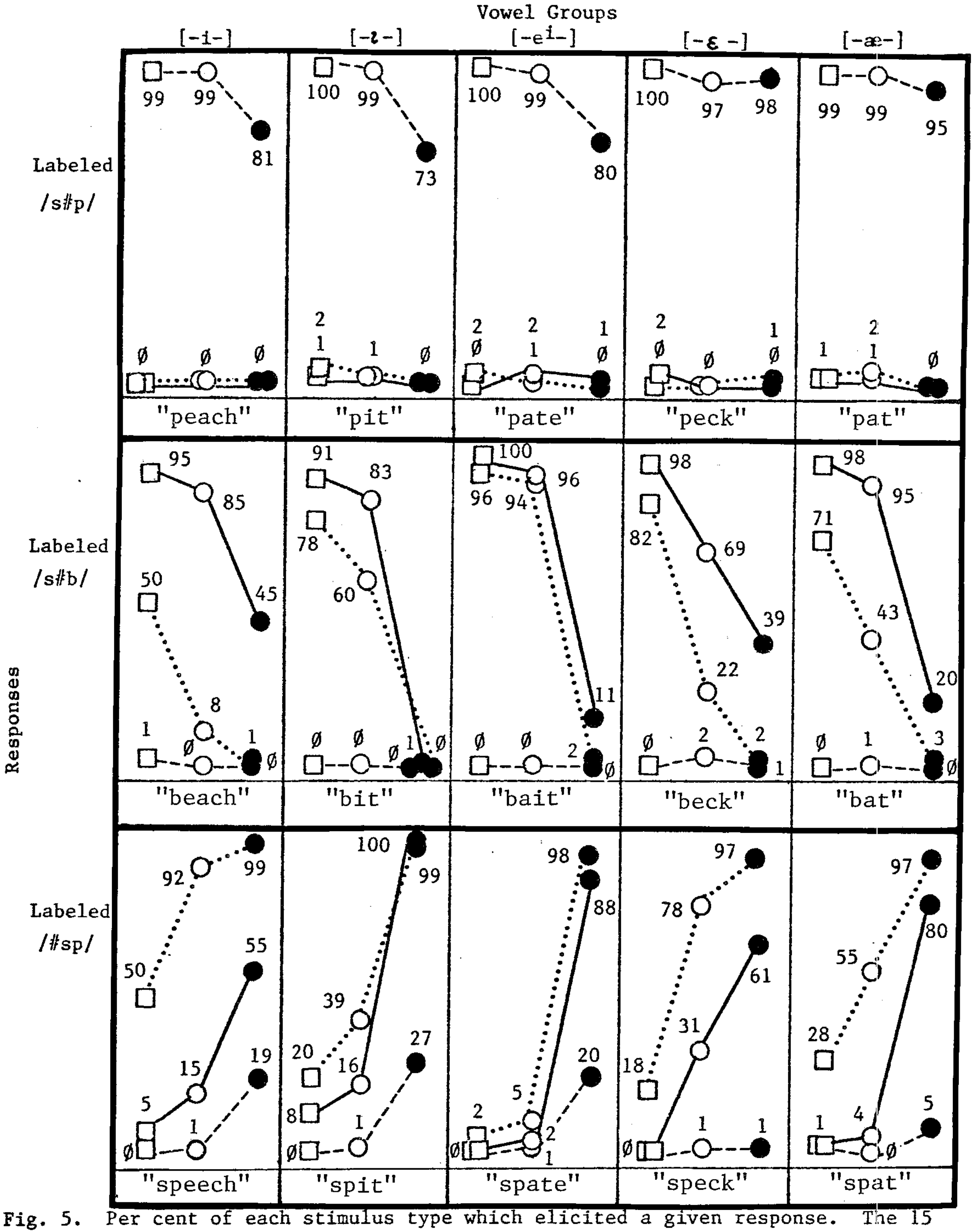


Fig. 5. Per cent of each stimulus type which elicited a given response. The 15 responses (words) are organized by cluster type (rows) and vowel groups (columns). Within each response box are 9 points, one for each of the 9 stimuli in that vowel group. Each point represents the per cent of presentations of that stimulus which elicited the response in that box. (Compare Table 5.) The stimuli are encoded by the combination of symbol and line:

□ = s#b ○ = s#p ● = sP ——— = b̥ ------ = pʰ .... = p

Table 5

Duration measurements and ratios for original and edited stimuli with corresponding /s#p/, /s#b/, and /#sp/ identifications. (Holding constant the stop-vowel nucleus type while changing the [s]'s.)

Stimulus type		[s]duration	cl.duration	[s]/cl.	identifications		
		(msec)	(msec)	ratio	/s#p/	/s#b/	/#sp/
1. peach	[s$^{\#b}$ p^{h}]	67	65	1.03	.99	.01	.00
	[s$^{\#p}$ p^{h}]	73	65	1.12	.99	.00	.01
	[s^{p} p^{h}]	94	65	1.45	.81	.00	.19
2. pit	[s$^{\#b}$ p^{h}]	53	72	.74	1.00	.00	.00
	[s$^{\#p}$ p^{h}]	58	72	.81	.99	.00	.01
	[s^{p} p^{h}]	109	72	1.51	.73	.00	.27
3. pate	[s$^{\#b}$ p^{h}]	60	72	.83	1.00	.00	.00
	[s$^{\#p}$ p^{h}]	57	72	.79	.99	.00	.01
	[s^{p} p^{h}]	100	72	1.39	.80	.00	.20
4. peck	[s$^{\#b}$ p^{h}]	49	73	.67	1.00	.00	.00
	[s$^{\#p}$ p^{h}]	71	73	.97	.97	.02	.01
	[s^{p} p^{h}]	93	73	1.27	.98	.01	.01
5. pat	[s$^{\#b}$ p^{h}]	65	62	1.05	.99	.00	.01
	[s$^{\#b}$ p^{h}]	68	62	1.10	.99	.01	.00
	[s^{p} p^{h}]	98	62	1.58	.95	.00	.05
6. beach	[s$^{\#b}$ b̥]	67	104	.64	.00	.95	.05
	[s$^{\#p}$ b̥]	73	104	.70	.00	.85	.15
	[s^{p} b̥]	94	104	.90	.00	.45	.55
7. bit	[s$^{\#b}$ b̥]	53	78	.68	.01	.91	.08
	[s$^{\#p}$ b̥]	58	78	.74	.01	.83	.16
	[s^{p} b̥]	109	78	1.40	.00	.00	1.00
8. bait	[s$^{\#b}$ b̥]	60	86	.70	.00	1.00	.00
	[s$^{\#p}$ b̥]	57	86	.66	.02	.96	.02
	[s^{p} b̥]	100	86	1.16	.01	.11	.88
9. beck	[s$^{\#b}$ b̥]	49	75	.65	.02	.98	.00
	[s$^{\#p}$ b̥]	71	75	.95	.00	.69	.31
	[s^{p} b̥]	93	75	1.24	.00	.39	.61
10. bat	[s$^{\#b}$ b̥]	65	73	.89	.01	.98	.01
	[s$^{\#p}$ b̥]	68	73	.93	.01	.95	.04
	[s^{p} b̥]	98	73	1.34	.00	.20	.80

(continuation of Table 5)

Stimulus type		[s]duration	cl.duration	[s]/cl.	identifications		
		(msec)	(msec)	ratio	/s#p/	/s#b/	/#sp/
11. speech	[s$^{\#b}$ p]	67	75	.89	.00	.50	.50
	[s$^{\#p}$ p]	73	75	.97	.00	.08	.92
	[s^{p} p]	94	75	1.25	.00	.01	.99
12. spit	[s$^{\#b}$ p]	53	75	.71	.02	.78	.20
	[s$^{\#p}$ p]	58	75	.77	.01	.60	.39
	[s^{p} p]	109	75	1.45	.00	.01	.99
13. spate	[s$^{\#b}$ p]	60	66	.90	.02	.96	.02
	[s$^{\#p}$ p]	57	66	.86	.01	.94	.05
	[s^{p} p]	100	66	1.51	.00	.02	.98
14. speck	[s$^{\#b}$ p]	49	60	.82	.00	.82	.18
	[s$^{\#p}$ p]	71	60	1.18	.00	.22	.78
	[s^{p} p]	93	60	1.55	.01	.02	.97
15. spat	[s$^{\#b}$ p]	65	60	1.08	.01	.71	.28
	[s$^{\#p}$ p]	68	60	1.13	.02	.43	.55
	[s^{p} p]	98	60	1.63	.00	.03	.97

the location of the significant difference. As could be predicted from looking at the graphs in Fig. 5 and the values in Table 5, the difference in /s#b/ scores between [s$^{\#b}$] and [s$^{\#p}$], between [s$^{\#p}$] and [s^{p}], and between [s$^{\#b}$] and [s^{p}] were all significant ($p < 0.01$) except for [s$^{\#b}$] and [s$^{\#p}$] in the case of "bait." A one-way analysis of variance was also done for the three sets of /s#p/ responses ("peach," "pit," and "pate") which seemed affected by the change from [s$^{\#p}$] to [s^{p}]. As expected, the effect on /s#p/ responses of inserting word-initial [s] into those three utterances was significant ($p < 0.01$).

6. Interpreting the analysis of variance.

Ascribing specific acoustic causes to the two main effects and their interaction is not possible given the design of this experiment. The independent variable labeled "[s]-type" in the statistical analysis could be the absolute duration of the [s] (or the integrated amplitude over that duration) or it could reflect the ratio of the [s]-duration to the following stop closure (among other things). Those two separate factors, the [s]-duration and the [s]/closure ratio, were not varied orthogonally in this experiment. The effect of interchanging the [s]'s without changing anything in the stop was that as the [s]'s were changed, so were the [s]/closure ratios. In other words, increasing the duration of the [s] within a given utterance resulted in an increase in the [s]/closure ratio as well. One may compare these changes in absolute duration, and the corresponding changes in the ratio, with the changes in listener identifications in Table 5. For example, in all of the 15 stop-vowel groups, the number of /#sp/ responses increases as the duration of the [s] increases, but there is no way to tell whether this increase in /#sp/ responses is due to the increase in absolute duration or to the increase in the [s]/closure ratio.

Interpreting the acoustic cues which are assoicated with the stop-vowel nucleus variable is also difficult. Again, varying the stop-vowel nucleus also affected the [s]/closure ratio because the same [s]'s were paired with different stop closures.

The general correlation between [s]/closure ratio and /#sp/ responses can be seen in Fig. 6 as well as some deviations from the trend. In two of the groups, the ones labeled "speech" and "spat," each increase in the [s]/closure ratio has a corresponding increase in /#sp/ responses. What is noteworthy, however, is that both groups show a marked increase at [s$^{\#p}$ p]. In the case of "spat," the sharp increase between [s$^{\#b}$ p] and [s$^{\#p}$ p] can't be attributed to either a difference in the stop-vowel nucleus (for both are [p]) or to a difference in the absolute duration of the [s]'s. But the amplitude of [s$^{\#p}$] was greater than that for [s$^{\#b}$]. This was visible on both the original oscillomink tracings and on the digitized wave-form. This suggests that all [s]-durations should be interpreted as an estimate of their "loudness" (amplitude integrated over duration). (A similar difference in amplitude was visible for [s$^{\#b}$] and [s$^{\#p}$] in the "spit" group, which would account for the sharp increase in /#sp/ responses between [s$^{\#b}$ p] and [s$^{\#p}$ p].) However, in the case of the "speech" group one cannot attribute the sharp increase between [s^{p} b̥] and [s$^{\#p}$ p] to such a difference in the [s]'s (duration or amplitdue) for those values for these [s]'s *decrease*. Thus that particular increase in /#sp/ responses may have something to do with the difference between [b̥] and [p]. In two other vowel groups there is additional support for the interpretation that changing the stop-vowel nucleus from [b̥] to [p] affects the number of /#sp/ responses positively. In the case of "spit" the positive effect of increasing the [s]/closure ratio is *reversed* between [s$^{\#b}$ p] and [s$^{\#p}$ b̥]--when the stop is changed from [p] to [b̥]--in spite of the fact that the [s$^{\#p}$] in [s$^{\#p}$ b̥] is louder than the [s$^{\#b}$] in [s$^{\#b}$ p]. A similar situation occurs in "speck": [s^{p} b̥] reverses the increase in /#sp/ responses, in spite of the fact that both the absolute length of the [s] and the [s]/closure ratio have increased.

It is possible that the difference in /#sp/ responses here attributed to some difference between [b̥] and [p] could be explained by some factor outside the clusters, e.g., amplitude patterns over the two syllables on either side or variations in the rate of speech in the carrier phrase. But even if the amplitude of the [s]'s are considered in relation to the overall amplitude envelope, that still could not explain the fact that in one instance a much *shorter* [s] was heard as word-initial when it was paired with [p] (in "speech") and in another instance a much longer [s] was heard as word-final when it was paired with [b̥] (in "speck"). Thus from this data it appears that there is some acoustic cue associated with the stop + vowel nucleus such that [p] elicits more /#sp/ responses than does [b̥].

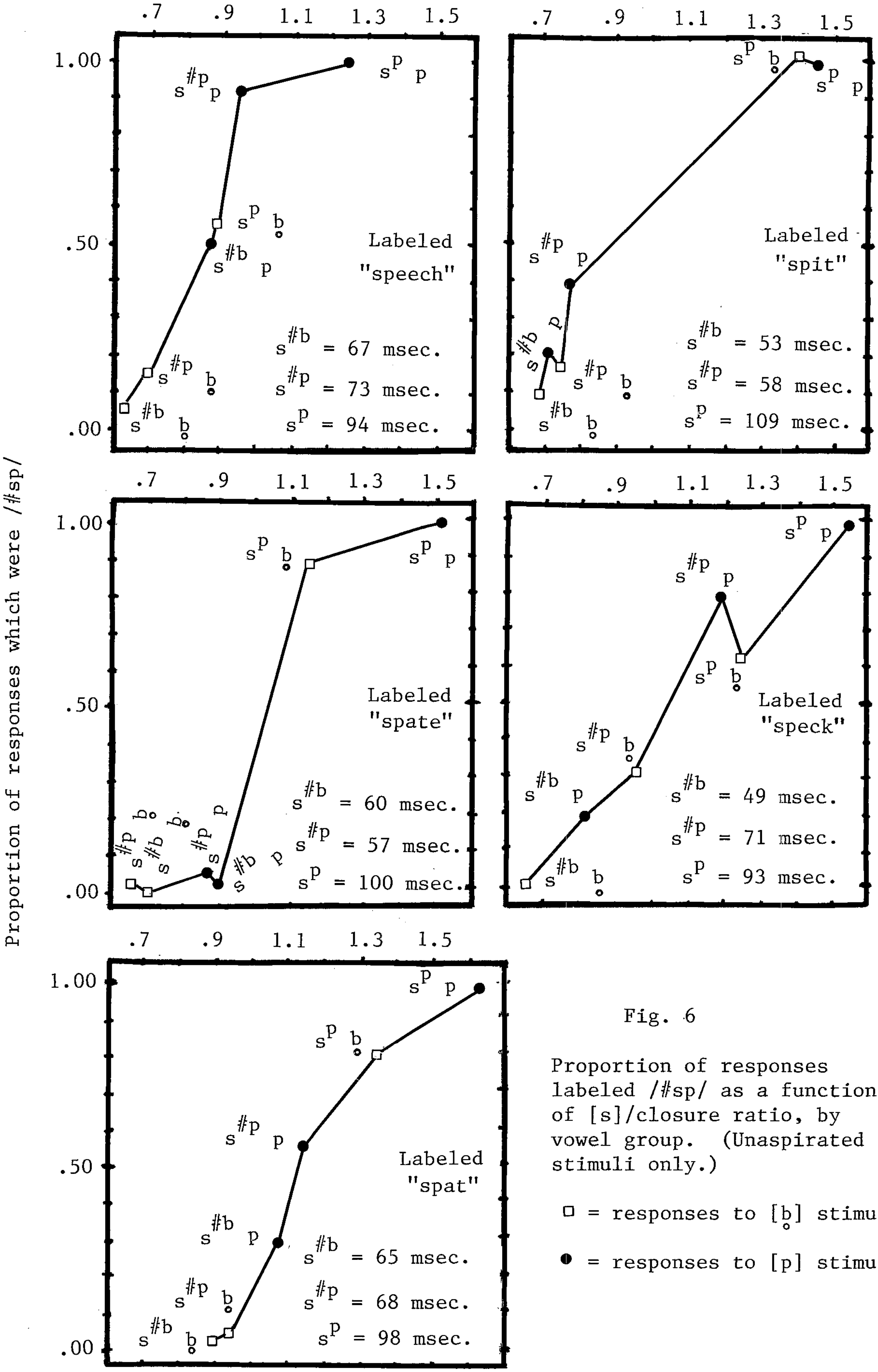


Fig. 6

Proportion of responses labeled /#sp/ as a function of [s]/closure ratio, by vowel group. (Unaspirated stimuli only.)

□ = responses to [b̥] stimuli

● = responses to [p] stimuli

Values on the abcissa indicate ratio of [s]/closure duration for stimuli.

V. Discussion and Conclusion

The results of this experiment indicate that English-speaking listeners use a hierarchy of acoustic cues to distinguish the strings /s#p/, /s#b/, and /#sp/. The presence of aspiration, which signals a word-initial voiceless stop, takes precedence. However, if the [s] preceding the aspirated stop is long and loud enough, it may counteract the marked aspiration of the [p^{h}] and switch the percept to /#sp/. In the absence of aspiration, the most important cue in distinguishing the two remaining possibilities, /s#b/ and /#sp/, is the "loudness" of the [s]. That "loudness" as a cue can be estimated by the duration of the [s], but that is only an estimate, for [s]'s of the same duration will elicit different responses if their amplitudes are different. Thus the notions "word-final [s]" and "word-initial [s]" are probably best described by using an integrated amplitude over the duration of the [s]. It is misleading, however, to assume that it is the [s] alone which signals the placement of the word boundary in the two clusters /s#b/ and /#sp/. For there is a positive correlation between the ratio of the [s]-duration and the proportion of tokens identified as /#sp/; there is a corresponding negative correlation between that ratio and the number of /s#b/ identifications. Thus at least one characteristic of the stop--its closure duration--may affect the placement of the word boundary.

There is, however, in this data a residue effect which cannot be explained by attributing it either to the [s] itself or to the [s]/closure ratio. Although the acoustic factor involved is low on the hierarchy of acoustic cues used to distinguish the three strings, it is there nonetheless: voiceless unaspirated [b̥] is not interchangeable perceptually with voiceless unaspirated [p] given similar values for the preceding [s] and the [s]/closure ratio.

This effect might be dismissed as an artifact of the tokens selected or the method employed to construct the stimuli if it were not for the fact that another aspect of the production data, the variation in fundamental frequency, supports the claim that [b̥] is different from [p]. That is, the production data indicate an acoustic difference between the two segments which corresponds to the perceptual difference. The acoustic difference is in the FØ onset frequencies and the FØ pattern in the following vowel.

The possibility that the FØ trajectory is a contributing factor in the different responses elicited by [p] and [b̥] is supported by Lehiste and Peterson (1961). They found a "different distribution of the fundamental frequency movement" in vowels following voiced and voiceless consonants. (p. 74)

If further research establishes that the fundamental frequencey differences observed in this limited set of data hold in general for the contrasting stings /s#b/, /s#p/, and /#sp/, then any articulatory model for the distinction between the word-initial [+voice] and [-voice] stops will have to take into account two facts about strings containing /#sp/:
(1) they have the same VOT as strings with /s#b/; (2) they have fundamental frequency patterns which are different from /s#b/ and more like /s#p/.

One could hypothesize that the difference in FØ is due to the size of the glottal opening either during the [s] or during the stop closure or both. Kim (1970) speculates that the glottal widening for English [p] in /#sp/ is of the same magnitude and the same duration as for [p^h] in /#p/ but that it begins earlier, during the [s], because a wide-open glottis is not incompatible with the voicelessness of the [s]. Further, if one assumes that the widening gesture has a constant duration across various contexts, then the short VOT for [p] in /#sp/ can be accounted for: the glottis, having reached its maximum opening earlier in /#sp/ than in /#p/, begins to close earlier and is almost completely adducted by the time of release.

Kim does not specify how the glottal gesture for /#sp/ might differ from the one for /s#b/. It is possible that for /s#b/ the glottis either does not open so wide for the (word-final) [s], or that it begins to close earlier for the [b̥] than for the [p]. In any case, one possible explanation for the FØ difference between /#sp/ and /s#b/ is that, at comparable points during stop closure, the glottis may be more open for [p] than for [b̥].

If in fact that is what happens, then the degree of glottal opening at release may also be slightly larger for /#sp/ than for /s#b/--but only slightly, because the time lag between release and voice onset is so short. With other places of articulation, however, the time lag is greater, increasing the likelihood that the glottis is wider ar release for /#st/ and /#sk/ than for /s#d/ and /s#g/. This would explain Klatt's 1975 findings, that the [t] in /#st/ and the [k] in /#sk/ are sometimes aspirated.

An alternative explanation for the difference in FØ between /#sp/ and /s#b/ is that the articulatory gesture for /#sp/ involves a stiffer vocalis and/or a raised larynx. Thus the state of the glottis during the stop closure in /s#b/ would be one of lax adducted cords, in /#sp/ tense adducted cords, and in /s#p/ tense abducted cords. In this model, the differences in FØ onset values between /s#b/ and /#sp/ would be due to the difference in glottal tension during closure, and the difference in FØ onset between /#sp/ and /s#p/, when it occurs, would be due to a difference in glottal opening at release. Such a model would be supported if it could be shown that the degree of glottal opening during stop closure is the same or almost the same for /s#b/ and /#sp/, for then it would be necessary to posit laryngeal tensing (e.g., in the cricothyroid, lateral cricothyroid, or vocalis) if pitch changes within an utterance containing voiceless stops could not be fully accounted for by changes in transglottal pressure. (See Hirano, Ohala, and Vennard (1969) and Ladefoged (1967).)

It is also possible that there is a trading relationship between glottal widening and glottal tensing such that there is a set of motor commands which control both to either facilitate voicing or prevent voicing. It is known that both may be factors in the initiation of vocal cord vibration (Halle and Stevens, 1971) and in the regulation of fundamental frequency (Hirano, Ohala, and Vennard, 1969). If such were the case, then one would expect to find that either articulatory gesture could vary but not independently of the other. In other words, there would be two sets of constraints on how much glottal width and tensing could vary: one set for [+voice] segments and another set for [-voice] segments.

Whatever articulatory model emerges, the fact remains that the difference in FØ for the two strings /#sp/ and /s#b/ can only be explained by a difference in glottal states during the stop closure. Because of what is known about the relationship between glottal opening, tenseness of the laryngeal muscles, pitch differences, and voicing, whether the articulatory difference between [b̥] and [p] is due to glottal width or tenseness or both does not matter here. Both increasing glottal width and increasing glottal stiffness reduce the probability that periodic vibrations of the vocal cords will begin (Halle and Stevens, 1971). In that sense, the [p] in /#sp/ is more "voiceless" or "devoiced" than the unvoiced [b̥] in /s#b/. That alone should be sufficient justification for categorizing [p] as phonemically [-voice].

It would support this categorization if one could demonstrate that the glottal state for [p] has something in common with the glottal states for other /p/ allophones, something not shared with voiceless [b̥]. Such a finding would represent an invariant articulatory correlate for the phonetic feature [± voice]. Two possible candidates for such an articulatory correlate would be

> (1) For a given closure duration, glottal widening or glottal tenseness (or a combination of the two) is constant for all the allophones of /ptk/. (This constant relationship would be independent of the timing relationship between stop closure release and voice onset.)
>
> (2) The devoicing of the allophones of /ptk/ (including those following word-initial /s/) is an active process, assisted by motor command to widen the glottis and/or tense the laryngeals and/or tense the supra-glottal vocal tract. This active devoicing would be in contrast to the passive devoicing of the voiceless allophones of /bdg/, which would be accounted for by the cessation of airflow due to supraglottal closure.

In conclusion, the production data and experimental results presented here demonstrate that the English stop [p] after word-initial /s/ is both acoustically and perceptually different from the voiceless allophone of word-initial [b̥]. There is evidence in the phonetic literature (Klatt, 1975) that this difference is even more marked for the alveolar and velar stops. This difference can only be accounted for by positing different glottal states for [ptk] and [b̥d̥g̊]. In spite of the fact that the acoustic difference may be attended to by listeners only when they are deprived of "top-down" cues and when other acoustic cues have ambiguous values, the fact remains that the difference is there. Thus the phonetic facts justify characterizing these stops as phonemically [- voice].

Footnotes

[1] Although generative grammars permit unspecified features in the lexicon, those features are, in fact, fully-specified by redundancy rules. Thus the analyst is forced to commit himself at the systematic phonemic level, even in the case of contextually-determined loss of contrast. See Stanley (1967) and Chomsky and Halle (1968), Chapter 8.

[2] Both Postal (1968) and Anderson (1974) make general statements to the effect that "phonemic forms will be as much as possible like phonetic forms"

(Anderson, p. 43). Kiparsky (1968) and (1976), in discussing neutralized, non-alternating segments in German and Sanskrit, argued that "on general grounds, one would expect the underlying form to be that which is closer to the surface." (Kiparsky (1976), p. 164.)

[3] For instances in which generativists have assumed that these post-sibilant stops are phonologically [-voice], see Chomsky (1957), p. 236, note 5; Halle (1962), p. 340 ff.; Chomsky and Halle (1968), p. 178 and p. 418; Schane (1973), p. 43. For an explicit argument based on Jakobson's implicational universals, see Houlihan and Iverson (1979), p. 55.

[4] Klatt (1974) stated that "word boundaries are likely to occur before long [s] but not after long [s]." (p. 61)

[5] The interaction was not significant for the /e^i/ vowel group. For the /i, ɪ, æ/ groups, interaction was significant ($p < 0.010$). For the /ɛ/ group, significance may have been due to the fact that the analysis of variance design used is generally prescribed for random model experiments while this experiment is of the mixed model type with repeated measures. The analysis of variance used was more conservative than necessary.

[6] There is some indirect support for Kim's hypothesis that the abduction-adduction gesture is constant across contexts but independent of release in Summerfield and Haggard (1977). They found that total periods of devoicing for voiceless aspirated stops in a set of bisyllabic nonsense words "tend to be more invariant than either the period of devoicing preceding oral release or the VOT itself. Possibly . . . it is the moment of oral release that is varied within a fixed time frame of abduction-adduction." (p. 446, note 7)

Bibliography

Abramson, A. S., and L. Lisker. (1972) "Voice timing in Korean stops," *Proceedings of the 7th International Congress of Phonetic Sciences*, 139-146.

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

Chomsky, N. (1957) "Review of Jakobson and Halle, Fundamentals of Language," *International Journal of American Linguistics*

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

Davidsen-Nielsen, N. (1969) "English stops after initial /s/," *English Studies* L, 4, Aug. 1969.

Davidsen-Nielsen, N. (1978) *Neutralization and Archiphoneme: Two Phonological Concepts and Their History*. (Publications of the Dept. of English, University of Copenhagen, Vol. 7)

Haggard, M., S. Ambler, and M. Callow. (1970 "Pitch as a voicing cue," *Journal of the Acoustical Society of America* 47, 2 (Part 2), 613-617.

Halle, M. (1962) "Phonology in generative grammar," *Word* 18, 54-72. Reprinted in *The Structure of Language* (1964), ed. J. Fodor and J. Katz (New Jersey: Prentice Hall), 334-352.

Halle, M., and K. N. Stevens. (1971) "A note on laryngeal features," *Quarterly Progress Report of the Research Laboratory of Electronics*, MIT, 101, 198-213.

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

Hardcastle, W. J. (1973) "Some observations of the *tense-lax* distinction in initial stops in Korean," *Journal of Phonetics* 1, 263-272.

Hirano, M. J. Ohala, and W. Vennard. (1969) "The function of laryngeal muscles in regulating fundamental fequency and intensity in phonation," *Journal of Speech and Hearing Research* 12, 3, 616-628.

Houlihan, K. and G. K. Iverson. (1979) "Functionally-constrained phonology," in *Phonological Theory*, ed. D. A. Dinnsen (Bloomington: Indiana University Press), 50-73.

House, A. S., and G. Fairbanks. (1953) "The influence of consonantal environment upon secondary acoustical characteristics of vowels," *Journal of the Acoustical Society of America* 25, 105-113.

Jaeger, J. (1980) "The psychological reality of phonemes revisited," *Report of the Phonology Lab* (Berkeley), No. 5, 6-50.

Kagaya, R. (1974) "A fiberscopic and acoustic study of the Korean stops, affricates and fricatives, " *Journal of Phonetics* 2, 161-180.

Kim, C.-W. (1965) "On the autonomy of the tensity feature in stop classification (with special reference to Korean stops)," *Word* 21, 3, 339-359.

Kim, C.-W. (1970) "A theory of aspiration," *Phonetica* 21, 107-116.

Kiparsky, P. (1968) "Linguistic universals and linguistic change," *Universals in Linguistic Theory*, ed. E. Bach and R. Harms, 170-202.

Kiparsky, P. (1976) "Abstractness, opacity, and global rules," *The Application and Ordering of Grammatical Rules*, ed. A. Koutsoudas, 160-186.

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

Klatt, D. (1975) "Voice-onset time, frication, and aspiration in word-initial consonant clusters," *Journal of Speech and Hearing Research* 18, 686-706.

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

Lehiste, I., and G. Peterson. (1961) "Some basic considerations in the analysis of intonation," *Journal of the Acoustical Society of America* 33, 4, 419-125.

Lisker, L. and A. S. Abramson, (1964) "A cross-language study of voicing in initial stops: aocustic measurements," *Word* 20, 384-422.

Lotz, J., A. Abramson, L. Gerstman, F. Ingemann, and W. Nemser. (1960) "The perception of English stops by speakers of English, Spanish, Hungarian, and Thai: a tape-cutting experiment," *Language and Speech* 3, 2, 71-77.

Massaro, D., and M. Cohen. (1976) "The contribution of fundamental frequency and voice onset time to the /zi/-/si/ distinction," *Journal of the Acoustical Society of America* 60, 3, 704-717.

McCasland, G. P. (1977) "English stops after /s/ at medial word-boundary," *Phonetica* 34, 218-228.

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

Reeds, J. A., and W. S.-Y. Wang. (1961) "The perception of stops after *s*," *Phonetica* 6, 78-81.

Schane, S. (1973) *Generative Phonology*. (Englewood Cliffs, N.J.: Prentice-Hall, Inc.)

Stanley, R. (1967) "Redundancy rules in phonology," *Language* 43, 2, 393-436.

Summerfield, Q., and M. Haggard. (1977) "On the dissociation of spectral and temporal cues to the voicing distinction in intital stop consonants," *Journal of the Acoustical Society of America* 62, 2, 435-448.

Trubetzkoy, N. S. (1939) *Principles of Phonology*, trans. by C. Baltaxe (1969). (Berkeley, Calif.: University of California Press)

Umeda, N., and C. H. Coker. (1974) "Allophonic variations in American English," *Journal of Phonetics* 2, 1-5.

Zlatin, M. A. (1974) "Voicing contrast: perceptual and productive voice onset time characteristics of adults, " *Journal of the Acoustical Society of America* 56, 3, 981-994.

The linguistic use of different phonation types*

Peter Ladefoged
University of California, Los Angeles

The theme of this paper is that what is a pathological voice quality in one language may be phonologically contrastive in another. Putting it more colloquially, one person´s voice disorder is another person´s phoneme. This is evidenced by the fact that there are several languages that contrast not only voiced and voiceless sounds, but also additional phonation types, such as "breathy" voice (in Gujerati) or "creaky" voice (in Hausa), which would be considered strongly stylistically marked or pathological if used by speakers of a language such as English.

The difficulties in classifying abnormal voice qualities are commonly referred to by speech pathologists. Van Riper and Irwin (1968) devote several pages to a description of "harsh" voice quality, for example, mentioning that it can be said to contain a "large amount" of "vocal fry," which itself can be thought of as a "rhythmic beat or scraping noise" (pp. 232-234). Citing other research, these authors go on to report that the sound is easy to demonstrate but difficult to describe, likening it to the sound made by youngsters imitating a motor boat or like vigorously popping corn. Similarly, Zemlin (1968) notes that voice quality is a controversial issue with terminology as the basis of much of the disagreement. In his review, Perkins (1971), too, observes that vocal disorders have "unfortunately persistently resisted exact definition" (p. 285). Zemlin (1968) selects four "abnormal" voice qualities (e.g. "breathy," "hoarse," etc.) from a clinical population for spectrographic display, despite the problems of terminology.

I would like to suggest that speech scientists and speech pathologists will find many benefits from studying voice qualities in other languages. When studying a clinical population a frequent problem is that the data are not available from a uniform group. But in some languages a particular type of voice quality is required to differentiate meanings between words; when this occurs all the speakers of the language will use it consistently. The voice quality can then be categorized, and its acoustic and physiological correlates noted. In contrast when dealing with a clinical population, the individual patients may vary their behavior from moment to moment (they are, after all, not trying to sound deviant).

The set of phonation types that occur contrastively in different languages does not completely intersect with the set of phonation types that are observable in clinical situations. This is because languages can be expected to use only those variations in phonation type that are functionally controllable. Obviously clinicians also have to deal with organic malfunctions. But languages use a far wider range of glottal actions than is usually supposed. Many of these overlap with the set of "functional disorders" commonly observed in speech pathology.

In surveying the phonation types that occur in the languages of the world, the first difference to observe is that between voiced and voiceless sounds. As is well known, many pairs of sounds are distinguished by the presence or absence of vocal cord vibration. Thus, in each of the pairs /b/ vs /p/, /d/ vs /t/, and /z/ vs /s/, the first sound may be said to be voiced and the second voiceless. It is worth noting that all languages have some sounds that are voiced and some that are voiceless; but not all have contrasts between these two possibilities. For example Maori and Hawaiian have no pairs of sounds that are distinguished by voicing alone. Voiceless sounds in different languages may vary in the degree of abduction of the vocal cords. During the closure for a French unaspirated /p/ the vocal cords are closer together than they are for an English aspirated /p/. Apart from noting this fact, differences in voice onset time (VOT) will not be considered in this paper.

From an acoustic point of view, voiceless sounds are all very similar, differing only in the fricative noise that may be present. But it is by no means certain that voicing is the same in all languages. Some languages, such as Zulu, may have a particular voice quality due to their choosing a particular mode of vibration of the vocal cords. The same may be true in certain dialects of English. Thus Trudgill (1974) has suggested that working class speakers from the East Anglian city of Norwich have a distinctive voice quality that he associates with a raised larynx.

Many languages have phonemic contrasts involving other kinds of phonation. Some languages exploit a breathy voice quality which linguists call murmur (see, e.g. Pandit, 1957; Ladefoged, 1973), in which the vocal cords vibrate without coming completely together. Murmur is exemplified noncontrastively in English intervocalic [h] as in "behind" and contrastively in many Indo-Aryan languages in /bh, dh, gh/ etc. For example, Nepali, an Indo-Aryan language, has a four way contrast:

pal	p^hal	bal	b^hal
"rear"	"throw away"	"burn"	"forehead"

Languages may differ in the type of murmured voice that they use. Nepali and Hindi /gh/ seem to be more breathy -- have a larger glottal opening -- than the /gh/ in the Niger-Kordofanian (West African) language, Igbo. It may be that speakers of different languages can not only use murmur distinctively, but can also choose, as a group, to use one degree of murmur rather than another. In other words, speakers of these languages can hear not only the difference between murmured and regularly voiced sounds, but also the difference between two degrees of murmur, so that they can use the degree that is appropriate for their own language. Presumably, if this is correct, a Nepali speaker using his own /gh/ sound when speaking Igbo would have a foreign acccent that other speakers of Igbo could detect.

Other murmured consonants include nasals and liquids, as exemplified by the following:

Newari (Sino-Tibetan)			
ma:	"garland"	m^{h}a:	"to be unwilling"
Marathi (Indo-Aryan)			
ma:r	"beat"	m^{h}a:r	"a caste"
Tsonga (Niger-Kordofanian)			
màkálá	"embers"	m^{h}àká	"matter"

[In these and subsequent examples ´indicates high tone and ` low tone; vowel length is shown by :]

There seem to be phonetic differences in the murmured nasals in these different languages. Marathi murmur is much more breathy than Newari murmur, and Tsonga murmur causes a large decrease in the fundamental frequency.

Murmur is usually a property of either the vowels or the consonants in a given language. Contrasts seldom occur at more than one place in a syllable. A possible exception to this is Gujerati, which has developed surface phonetic contrasts between plain and murmured vowels, in addition to the more common Indo-Aryan contrasts between plain and murmured stops, e.g.:

tara	"stars"	ta̤ra	"yours"

[The subscript „ as in a̤ denotes murmur.]

Another reasonably common phonation type is a form of creaky voice that linguists call laryngealization, which is found in approximately 10% of the world´s languages (data from UPSID, the UCLA Phonological Segment Inventory Database, described by Maddieson, 1980). A good example occurs in Mpi, a Tibeto-Burman language with six tones, each of which may occur with a plain or a laryngealized vowel. So that the same segments, /si/, have 12 different meanings:

		Plain	Laryngealized
1.	low contour	"to be putrid"	"to be dried up"
2.	low	"blood"	"seven
3.	mid contour	"to roll"	"to smoke"
4.	mid	(a color)	(classifier)
5.	high contour	"to die"	(man´s name)
6.	high	"four"	(man´s name)

A lesser degree of laryngealization occurs in Nilotic languages such as Nuer. These languages have contrasting vowel harmony sets, one being slightly laryngealized, the other being slightly breathy. A contrast between slightly murmured and slightly laryngealized vowels also occurs in Bruu, a Mon Khmer language, in words such as:

ki:t	"fear of crowds"	ki̤:t	"to sharpen"

So far we have been considering laryngealization only in vowels, but it also occurs in stop consonants in Chadic languages such as Hausa:

bá:bè	"locust"	b’á:b’è	"to quarrel"

[The raised comma ’ denotes laryngealization.]

A slight degree of laryngealization occurs in the Korean so-called fortis stops:

pul	"fire"	p*ul	"horn"

[The asterisk * denotes the fortis nature of the stop.]

Laryngealized stops, nasals, liquids and approximants are quite common in Native American languages. Laryngealized semi-vowels also occur in Hausa, Margi, and other Chadic languages, as for example, in Bura:

wáskí	"each"	w’álá	"big"
jà	"give birth"	j’áhà	"doctor"

The Danish glottal catch, which can occur on consonants or vowels, may also be considered as an additional phonation type. Gårding (personal communication) has also suggested that Swedish tonal word accents are accompanied by slight differences in voice quality. From all these examples we may conclude that the human larynx can be used to produce much more than voiced-voiceless distinctions. It is regularly used in other languages to produce sounds that would be considered somewhat deviant if produced by English speakers.

As a final demonstration of this point, we will consider in somewhat more detail one of the Khoisan languages spoken by Bushmen in Southern Africa, which have some additional unusual phonation types. These languages distinguish voiced and murmured vowels, each with and without possible added laryngealization. In addition, as has been conclusively demonstrated by the x-ray studies of Traill (1981), they have "epiglottalization" -- a form of extreme pharyngealization that functions like a phonation type, forming strident (or "pressed") vowels. There are thus six contrasting phonation types in !Xóõ.

Table 1. Six contrasting phonation types in !Xóõ.

		(a)	(b)
		Voiced	Murmured
1.	Plain	//áa "camelthorn tree"	!a̤o̤ "slope"
2.	Laryngealized	g/à’je "bend"	/a̤’je "wait for him"
3.	Strident	qà̰a "long ago"	!à̤̰o "base"

[The symbols / ! and // denote dental, alveolar, and lateral clicks, respectively; epiglottalized vowels are marked by a subscript ~ .]

Recordings were made of ten speakers of !Xóõ saying the six words of !Xóõ listed in Table 1.These recordings were then digitized and stored on a computer. Figure 1 shows the waveforms of the voiced and murmured vowels of one speaker. The voiced vowel in the upper part of the figure has sharp regular voicing pulses, with the first formant frequency clearly evident in the damped exponential part of the waveform. The waveform of the murmured vowel in the lower part of the picture looks much more like a sine wave at the fundamental frequency, with additional components corresponding to fairly high frequencies. There is very little evidence of the first formant in the lower waveform. It is also clear that the murmured vowel has a lower pitch.

The same points can be seen in the spectra of voiced and murmured sounds. Figure 2 shows the DFT and LPC spectra of a pair of vowels. The sampled waveform was first differenced before making these spectra, so there is a high frequency boost of 6db/octave. As a first measure for quantifying the amount of breathiness in a vowel, we considered the difference in db in these spectra between the amplitude of the largest harmonic in the first formant (indicated by a dashed line on the figure) and the amplitude of the fundamental (indicated by a solid line). The difference in intensity (the distance between the lines) is much greater for the regular voice than it is for the murmur. In the murmured vowel, the solid line indicating the intensity of the fundamental is only about 10 db below the dashed line indicating the intensity of the major harmonic in the first formant.

Figure 2 also shows that the murmured vowel has more irregular energy evident in the higher frequencies, and a slightly less falling spectrum. The degree of regularity of the vocal cord excitation difficult to quantify in the case of vowels such as these, because they are fairly short, and not said on a steady fundamental frequency. The spectral tilt is also a difficult measure to use in these contrasts, because it is affected by two opposing factors: the less sharp glottal pulses that occur during murmur produce smaller amounts of energy in the higher frequencies; but there are also higher airflow rates because the vocal cords are vibrating without coming completely together, and as a result there is more turbulence, which causes more noise excitation of the higher frequencies. The previously mentioned measure, the difference between the intensity of the first formant and the intensity of the fundamental is a much more useful measure of the differences in voice quality. For all ten speakers, on all occasions, this measure was greater for the voiced sounds than for the corresponding murmured ones. It is a reliable and, by any statistic, a highly significant measure of the phonemic differences.

Spectra of the laryngealized parts of the voiced vowels of two speakers are shown in figure 3. Note the great difference between the amplitude of the first formant and the amplitude of the fundamental in each case. The narrow bandwidths of the formants are also evident -- in so far as bandwidth can be assessed from LPC spectra. The same features can be seen in figure 4, which shows spectra at two different moments in time in the same vowel, the first one in the murmured part, and the second one in the laryngealized part just before the glottal stop.

For each of the ten speakers, measurements were made of four spectra, one in the middle of the voiced vowel, one in the middle of the murmured vowel, one near the beginning of a murmured laryngealized vowel, and one near the end, in the laryngealized part of the same vowel. Figure 5 shows, for each speaker, on the ordinate, the degree of breathiness in his murmured vowels, measured as the difference between the amplitude of the first formant (A1) and that of the fundamental (A0), and, on the abcissa, the degree of breathiness in his voiced or laryngealized vowels. There is a weakly significant correlation ($p<.05$) between the murmured vowels and the voiced or laryngealized vowels taken together. Those speakers who have a high degree of breathiness in their regular voice tend to have a higher than normal degree of breathiness in their murmured vowels. It seems that breathiness, like pitch, is a relative matter. What counts as a high pitch for one person may be a low pitch for another. Similarly, when a language uses breathiness to form a phonemic contrast between voiced and murmured vowels, what counts as voiced for one person may be very like what is murmured for another. Thus the speakers producing the vowels marked A and B on the graph have a difference of 10-15 db between the amplitude of the first formant and the amplitude of the fundamental in their murmured vowels. A difference of the same magnitude occurs in the voiced vowels of speakers C and D.

The final row in Table 1 illustrates an even more intriguing pair of sounds. These sounds will be fully described in Trail and Ladefoged (forthcoming). Here we will note only that they are pronounced with a major contraction just above the vocal cords, near the root of the epiglottis. When accompanied by vibrations of the vocal cords, the effect is one of extreme pharyngalization. As in the case of pharyngalized sounds in Semitic languages such as Arabic, the voicing sometimes (but not always) has a laryngealized, creaky voice, quality. In the case when they are accompanied by murmur there is considerable fricative noise, which can probably be associated with turbulent airflow in the neighborhood of the epiglottis. These vowels sound are very different from anything that ever occurs in the normal speech of a speaker of any European languages.

As has been shown by Traill (1981) all six phonation types illustrated in Table 1 can be used to produce phonemic distinctions in !Xóõ. From a narrow, ethnocentric, point of view they might be considered to be voice disorders. But it seems that if you want to sound like a Bushman, you have got to be prepared to have what we sometimes naively consider to be a pathological voice quality.

Acknowledgements

Many members of the UCLA Phonetics Laboratory have contributed to this paper. Their help is gratefully acknowledged.

Footnote

[*] A casette tape recording illustrating some of the linguistic examples cited in this paper is available from: The Phonetics Laboratory, Linguistics Department, UCLA, Los Angeles, Ca 90024. Price $10.00.

References

Ladefoged, P. 1973. The features of the larynx. *Journal of Phonetics* 1:73-83.

Maddieson, I. 1980. UPSID - the UCLA Phonological Segment Inventory Database. *UCLA Working Papers in Phonetics* 50:4-56.

Pandit, P. B. 1957. Nasalization, aspiration and murmur in Gujarati. *Indian Linguistics* 17:165-172.

Perkins, W. H. 1971. *Speech Pathology: an Applied Behavioral Science.* Mosby, St. Louis.

Traill, A. 1981. Phonetic and phonological studies of !Xóõ Bushman. Ph. D. Thesis, University of Witwatersrand, Johannesburg.

Traill, A. and Ladefoged, P. (forthcoming) *Phonation types in !Xóõ*

Trudgill, P. 1974. *The Social Differentiation of English in Norwich* Cambridge University Press.

Van Riper, C. and Irwin, J. V. 1958. *Voice and Articulation.* Prentice-Hall, Englewood Cliffs.

Zemlin, W. R. 1968. *Speech and Hearing Science: Anatomy and Physiology.* Prentice-Hall, Englewood Cliffs.

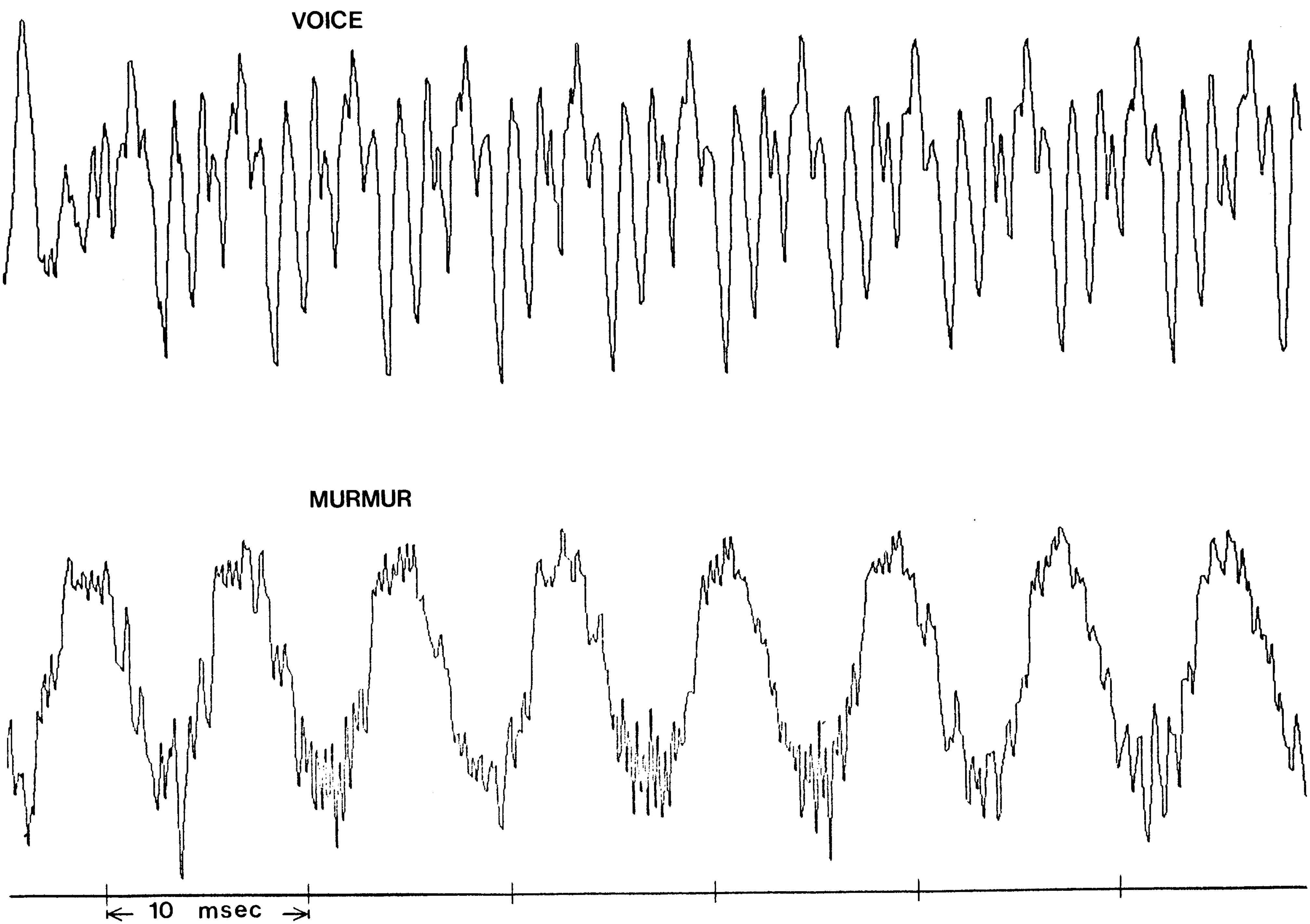


Figure 1. Voiced and murmured vowels. Parts of the waveform near the midpoints of the vowels in the !Xóõ words //áa and !a̤o̤ from Table 1.

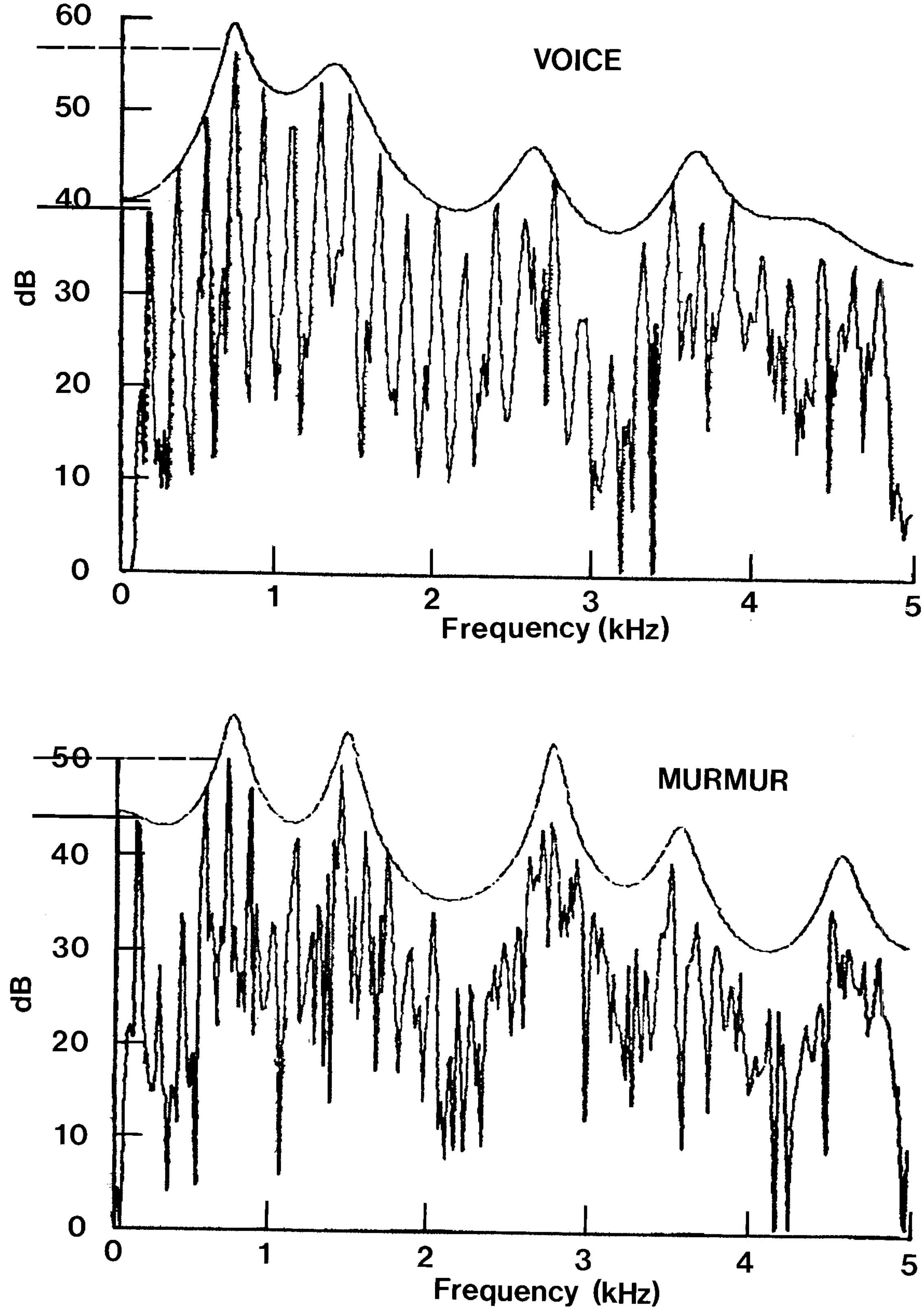


Figure 2. Spectra of the voiced and murmured vowels.

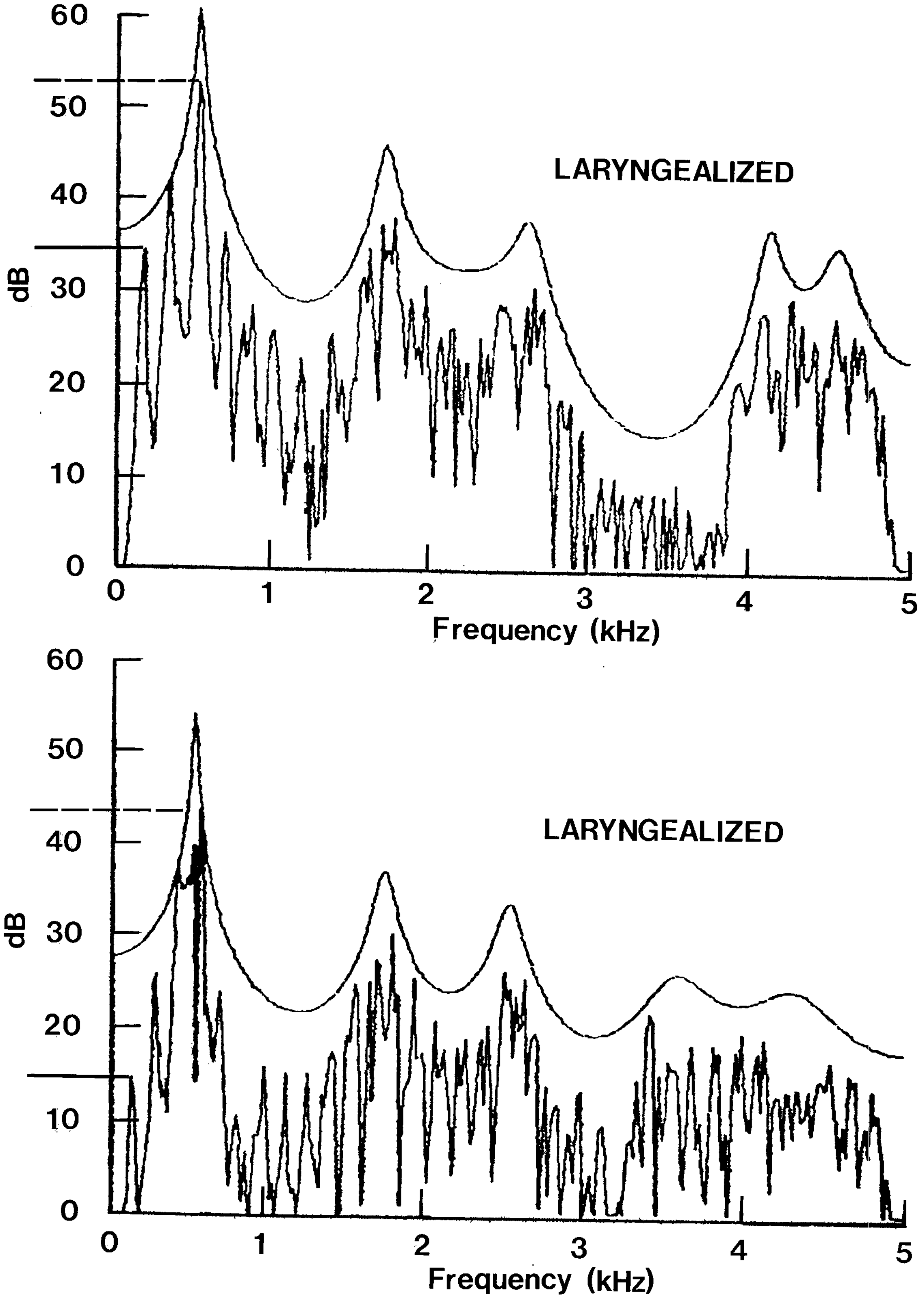


Figure 3. Spectra of the laryngealized vowels in the !Xóõ word g/à'je from Table 1, as produced by two speakers.

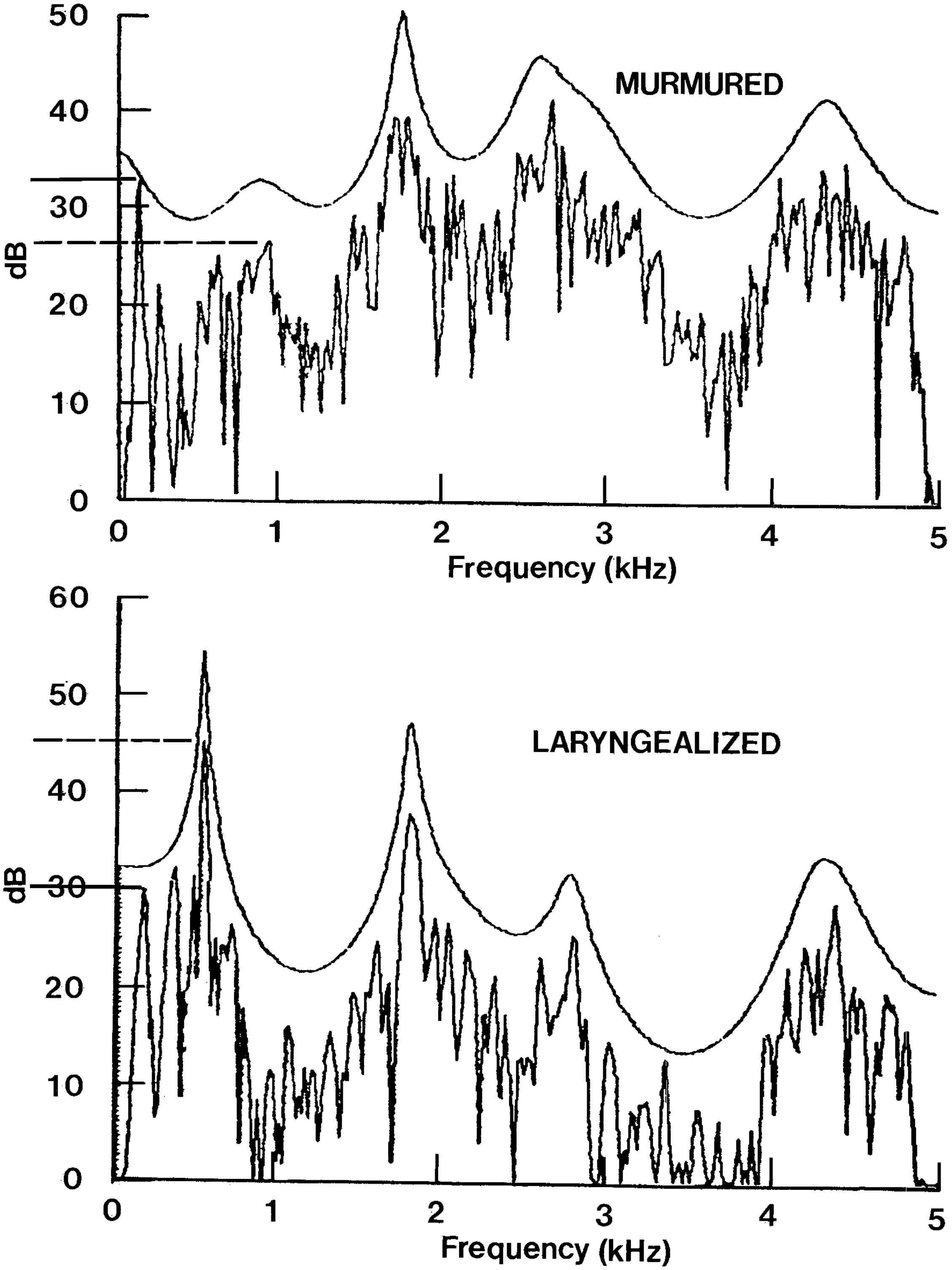


Figure 4. Spectra from the murmured part of the vowel in the first part of the !Xóõ word /a̤'je and from the laryngealized vowel in the second part of this word.

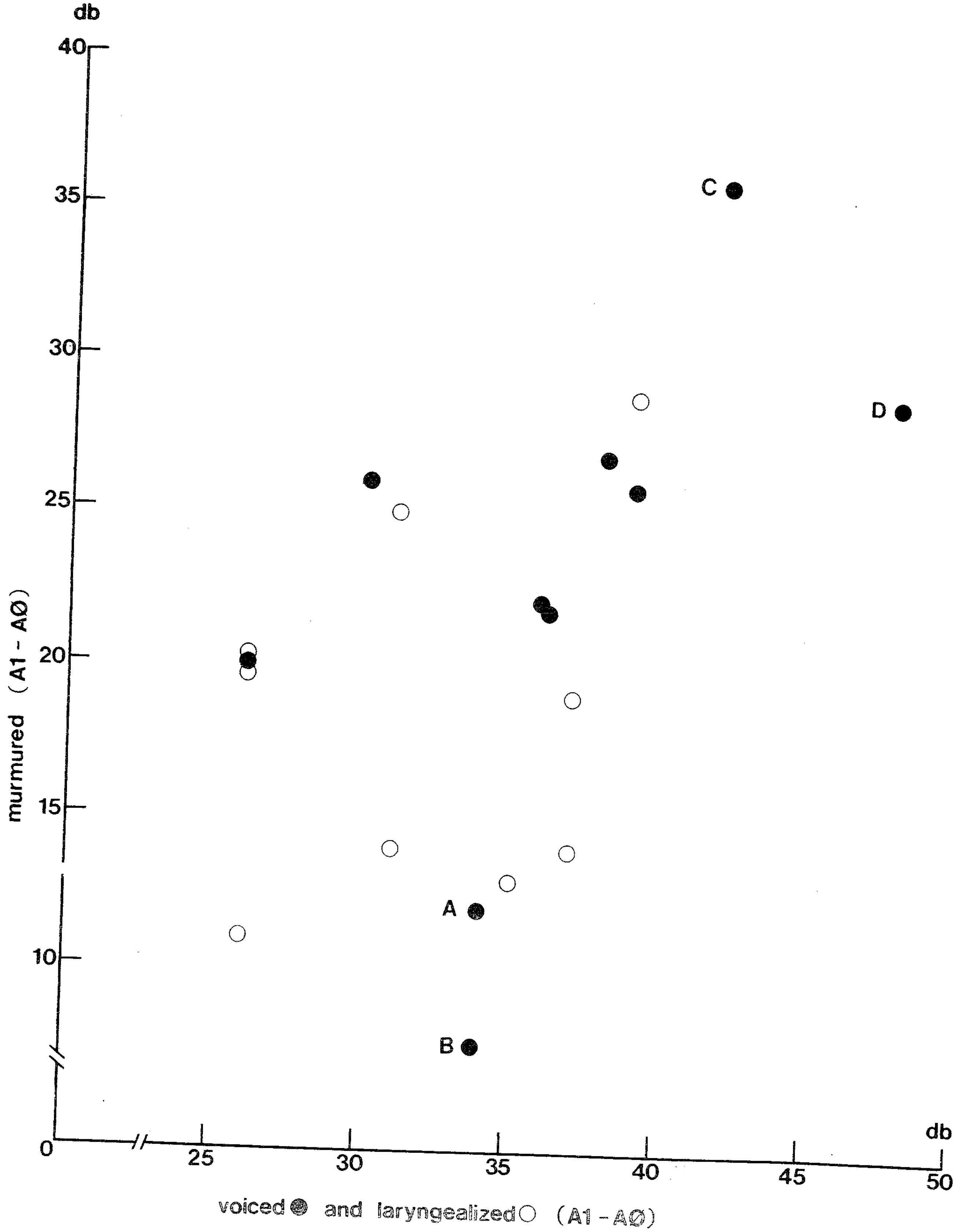


Figure 5. The relation, for 10 speakers, between (a) the murmured and voiced vowels in the !Xóõ words //áa and !a̤o̤ (closed circles), and (b) the beginning (murmured) and end (laryngealized) vowels in the !Xóõ word /a'je (open circles).

Attempts by Human Speakers to Reproduce Fant´s nomograms

Peter Ladefoged and Anthony Bladon

Fant´s nomograms, which constitute a simulation of the acoustic resonances which result from different vocal tract shapes, were checked against natural vowels by two phoneticians carefully producing a full range of vowel articulation in which (a) the location of the minimum aperture in the vocal tract and (b) the lip aperture were separately varied. Spectrographic measurements of formant frequencies were made, some with synchronised lip photographs. Ultrasonic and radiographic methods were used to verify vocal tract constriction size. In many major respects our data agreed with Fant´s predictions, but three main kinds of discrepancies were noted: firstly, we could not produce as great a range of variations as occurs in the nomograms; secondly, when an [i]-like constriction is moved as far forward as possible, the frequency of F2 does not decrease; and thirdly, variations in lip rounding affect high front vowels and high back vowels differently with no gradual change from one class to the other as in Fant´s data.

Probably the most well known diagrams representing articulatory-acoustic relations are Figures 1.4-11a,b from Fant´s The Acoustic Theory of Speech Production [1,p.82], reproduced here (with some additional material) as Figure 1. These diagrams show the formant frequencies that are produced when an electrical analog of the vocal tract is varied so as to simulate the acoustic output of a resonator system approximating the vocal-tract. Fant specifies the shape of the vocal tract in terms of the size and location of the minimal cross-sectional area, and the area of a 1 cm long lip section, when present. In the upper part of Figure 1 the size of the constriction is 0.65 cm^2, which corresponds to a rather small aperture, comparable to that in the vowel [i] as in "heed." In the lower panel of the figure it is 2.6cm^2, which is more like [ɛ] as in "head." In each panel of the figure the place of this constriction is varied, the abcissa showing its distance from the glottis. There are five curves for each of the first five formants. The solid lines denote a fairly large lip opening (8cm^2), comparable with, or even lower than, that in [ɑ] as in "father." The other curves show increasing lip rounding up to 0.16 cm^2, which is a smaller aperture than occurs in any English vowel.

As far as we are aware, the validity of this diagram has never been checked by comparing it with measurements of a similar set of articulations produced by human speakers. Broad and Wakita [2] have reported measurements of a speaker producing a set of vowels some of which would be similar to those specified by Fant. But no one has tried to produce a wide range of vowels that vary in the way simulated by Fant´s electrical analog. It is, of course, difficult if not impossible to obtain measurements of a set of exactly comparable observations. But it is possible to check the validity of a number of points in the diagram.

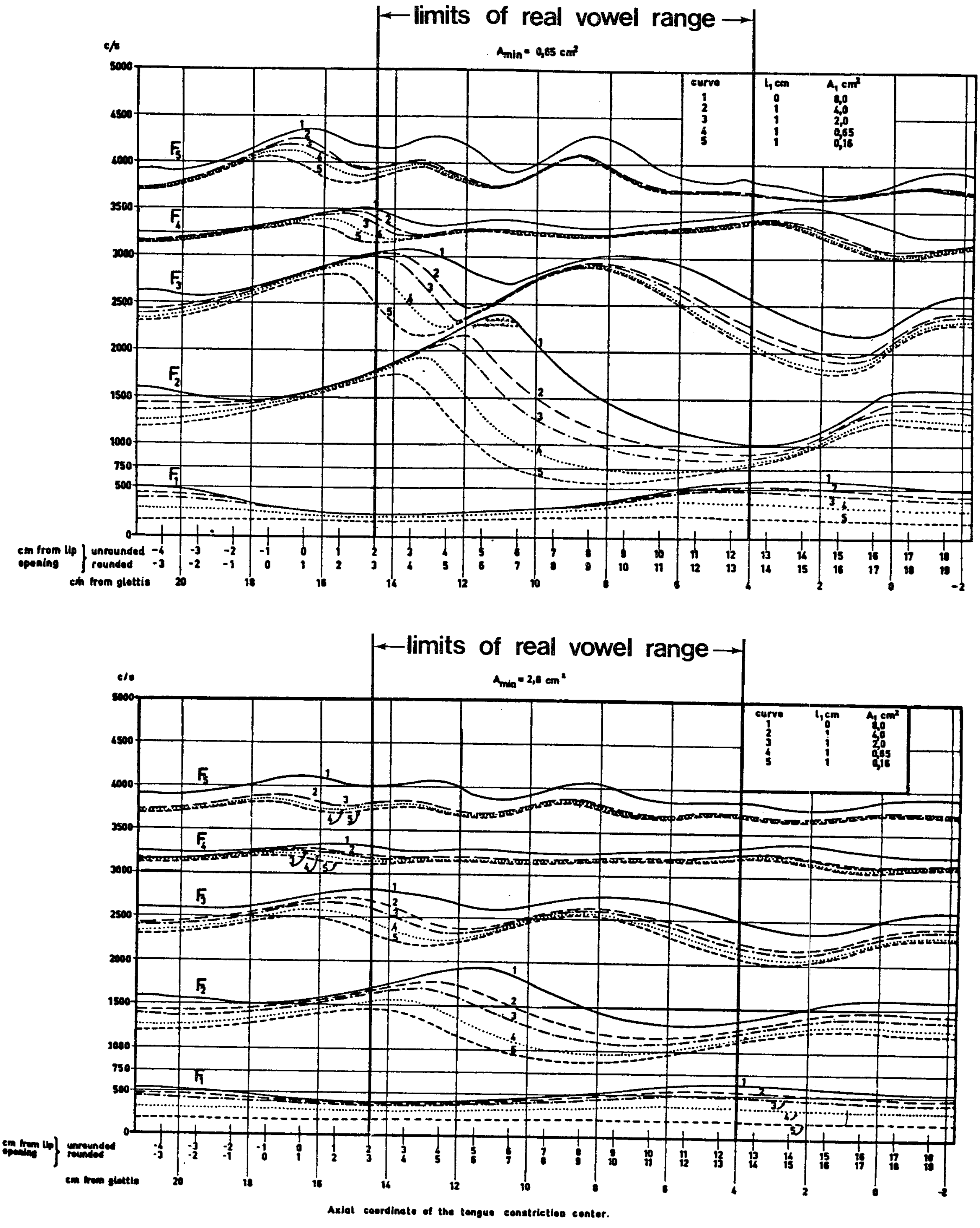


Figure 1. Nomograms (from Fant, 1960) showing the formant frequencies when a tongue constriction of .65 cm^2 (upper diagram) or 2.6 cm^2 (lower diagram) is moved along the vocal tract.

Procedure

1. Varying the constriction location

Each of the authors made a series of vowels in which the position of the lips was kept steady while the location of the minimum aperture was moved progressively further from the glottis. The vowels produced included vowels similar to the low back vowel [ɑ] and the high front vowel [i], as well as vowels with tongue positions comparable with those in the IPA vowels [ɯ] (a high back unrounded vowel) and [ɨ] (a high central unrounded vowel). At any given moment we cannot be sure of the location of the constriction with reference to the glottis. But we are sure that the location of the constriction was moved progressively from slightly below that for [ɑ] to slightly beyond that for [i], passing through a large number of points in between. The position for the lips was kept steady by the speaker observing himself in a mirror. In addition, each author carefully observed the other when he was speaking, to ensure that in his opinion also the correct series of vowels had been produced. We found it easier to keep a constant lip opening when the position of the mandible was fixed by positioning a small bite block between the molars. Each of the authors recorded sets of vowels with four different lip apertures. The position of the larynx was monitored by the speaker holding his neck so that he could feel the location of the thryoid cartilage. Both speakers managed to maintain a fairly constant larynx position throughout the production of each set of vowels, except when producing very low back vowels. They each felt that there may have been some lowering of the larynx when they said vowels with a minimum aperture closer to the glottis than that in [ɑ].

2. Area of minimum aperture

We do not know the precise area of the minimum aperture. But we know that when it was in the appropriate part of the pharynx the vowel produced was [ɑ] as in "father," a sound for which we have published x-ray data for one of the authors [3]. It is possible that the vocal tract shape was substantially different on the occasion when the x-ray photograph was taken, but this is unlikely considering the similarity of the formant structure of the two sounds (each of the first three formants of the vowel for which there is x-ray data is withing 10 Hz of the corresponding formant in the current data). When the constriction was nearly as far from the glottis as possible the tongue position was very comparable with that in [i] as in "see," for which we also have previous x-ray data to which the same considerations apply. The size of the constriction in all the intermediate positions was maintained as constant as possible. We know that it cannot have varied to any great extent because the sounds being produced were always close to being fricatives. When making vowels of this kind, with a constant rate of flow of air as judged by the constant vocal effort, the size of the constriction defines fairly precisely the point at which the airflow becomes turbulent. When each of a set of sounds is nearly, but not quite, fricative, we can deduce that the whole set has a fairly constant minimal aperture.

For the second author (AB), a more thorough determination of the size of the minimal vocal tract aperture was carried out using an ultrasonic technique, as follows. An ultrasound sector scan of 90° arc showing the tongue surface from the blade to the root was made for half of this author's vowels. (The assistance of Dr.J. Hamill and Dr. C. Woodham of the Radiology Department, John Radcliffe Hospital, Oxford, is gratefully acknowledged). The ultrasound data had to be obtained on a later occasion than the original experiment but we have confidence in the replication of the vowel sounds since (a) controls of the original recordings (lip aperture, bite block size, larynx monitoring) were replicated with care, (b) the production of the vowels in the ultrasound run was accompanied by headphone presentation to the subject of the vowels he originally recorded and (c) replicability of a wide range of sounds is part of the phonetician's training.

The ultrasound transducer was held at the skin surface comfortably above the thyroid cartilage and angled upward and posteriorly so that the 90^{o} arc sector scan was bisected by the radius upon which a tongue-uvula contact was aligned. For each vowel, the following measurements were made using an on-screen caliper: (a) the distance to the tongue surface at the constriction during the vowel and (b), from a different record, the distance to the tongue surface when held in contact with the upper vocal tract surface (i.e. the palate, velum or pharynx wall) - both measurements being made along the same scanning radius. Data on the relation between the vocal tract cross dimension and vocal tract area, for a given location along the tract, have been published by Ladefoged et al. [4], and these were used to derive the estimated vocal tract area (area of minimum aperture). These areas are consistent with the area data suggested by Catford [5]; but the literature contains somewhat incompatible results (compare for instance Figure 10 of Gauffin and Sundberg [6, p.157]) on the factors involved in transforming vocal tract cross dimensions into areas. However, for our present purposes, we regard these differences as irrelevant to our concluding that the vowels produced by AB in our experiment are highly comparable to the area ($0.65cm^2$) simulated by Fant in the upper panel of Figure 1.

3. Varying the lip aperture

As a second way of obtaining data comparable with that in Figure 1, each of the authors recorded vowels in which the position of the tongue was kept constant while that of the lips was varied continously. We are sure that the only change during the pronunciation of these vowels was in the degree of lip rounding. The procedure in these recordings of lip variation was as follows.

Each of the authors recorded vowels in which the tongue positions were comparable with those in high vowels such as IPA cardinal vowel [i], [i] as in "heed," IPA [ɨ], [u] as in "who," and IPA cardinal [u]. While maintaining each of these tongue positions the lip position was varied from an extremely spread position with the corners of the mouth retracted to a position with a near maximal degree of lip rounding. For each of the authors, the measurements of lip positions reported below were obtained in one of the two recording sessions of this type, in which the lips were photographed with a motorized 35 mm camera that took pictures at aproximately 350 msec intervals. The click of the camera shutter was evident on the recordings, so that we were able to measure the lip aperture at known moments in the acoustic record.

4. Determination of formant frequencies

Our principal technique for determining the acoustic structure of all these vowels was the analysis of sound spectrograms. We opted for this technique because we believe that we can make more reliable measurements of formant frequencies from spectrograms than from any of the other more sophisticated techniques available to us, such as computerized FFT or LPC analysis. This is because of the problems inherent in formant identification. Fant's nomograms show continuously varying resonances of the vocal tract. As will become apparent when we present some of the data, no known computer algorithm (except, perhaps, analysis-by-synthesis in terms of poles and zeros, which was not available to us) could determine these formant frequencies. It seems that the comment that one of us made several years ago in still true: "without knowing the sound that is being investigated, and without some previous knowledge of where the formants of such a sound might be expected to be, it is impossible to make valid measurements of the formant frequencies." [7, p.85].

One problem in determining formant frequencies is illustrated by the spectrogram in Figure 2. This vowel occurred just after the mid point of a series in which the tongue constriction was moved progressively further from the glottis; it corresponds approximately to IPA [ʉ]. During this vowel there was no noticeable articulatory movement of any kind. Neither of the authors nor any other skilled observers can detect any perceptual change in vowel quality during this vowel. The changes in the acoustic structure are almost certainly due to variations in the glottal source function. There is a slight perceptual change from the beginning to the end, which can best be described as one of loudness. But any algorithm that we know would say that the frequencies of F3 and F4 were approximately 2700 Hz and 3400 Hz at the beginning of the vowel, and 2070 Hz and 2650 Hz at the end. When this variation is produced on a speech synthesizer there is a very distinct change in vowel quality, even when an additional formant (F5) is added to the second vowel.

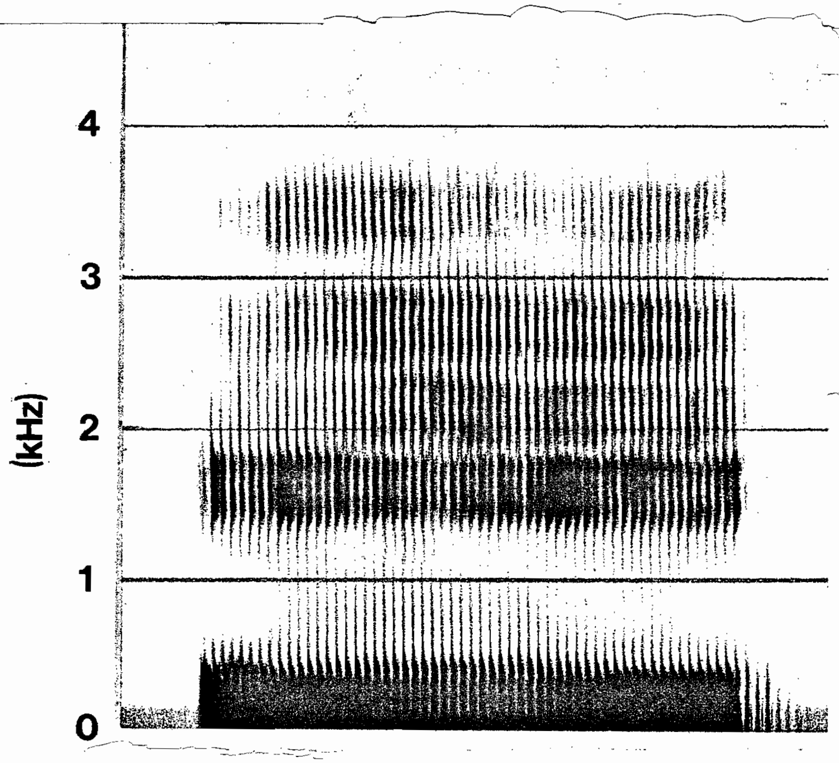


Figure 2. Wide band spectrogram of PL series 2, vowel 8.

Other problems in determining formant frequencies are illustrated in Figure 3, which shows the central parts of a complete series of vowels in which the constriction is moved progressively forward. On the original recording each of these vowels was about 400 msec long, and separated by about 300 msec from the next vowel in the series. We had no valid way of determining the frequencies

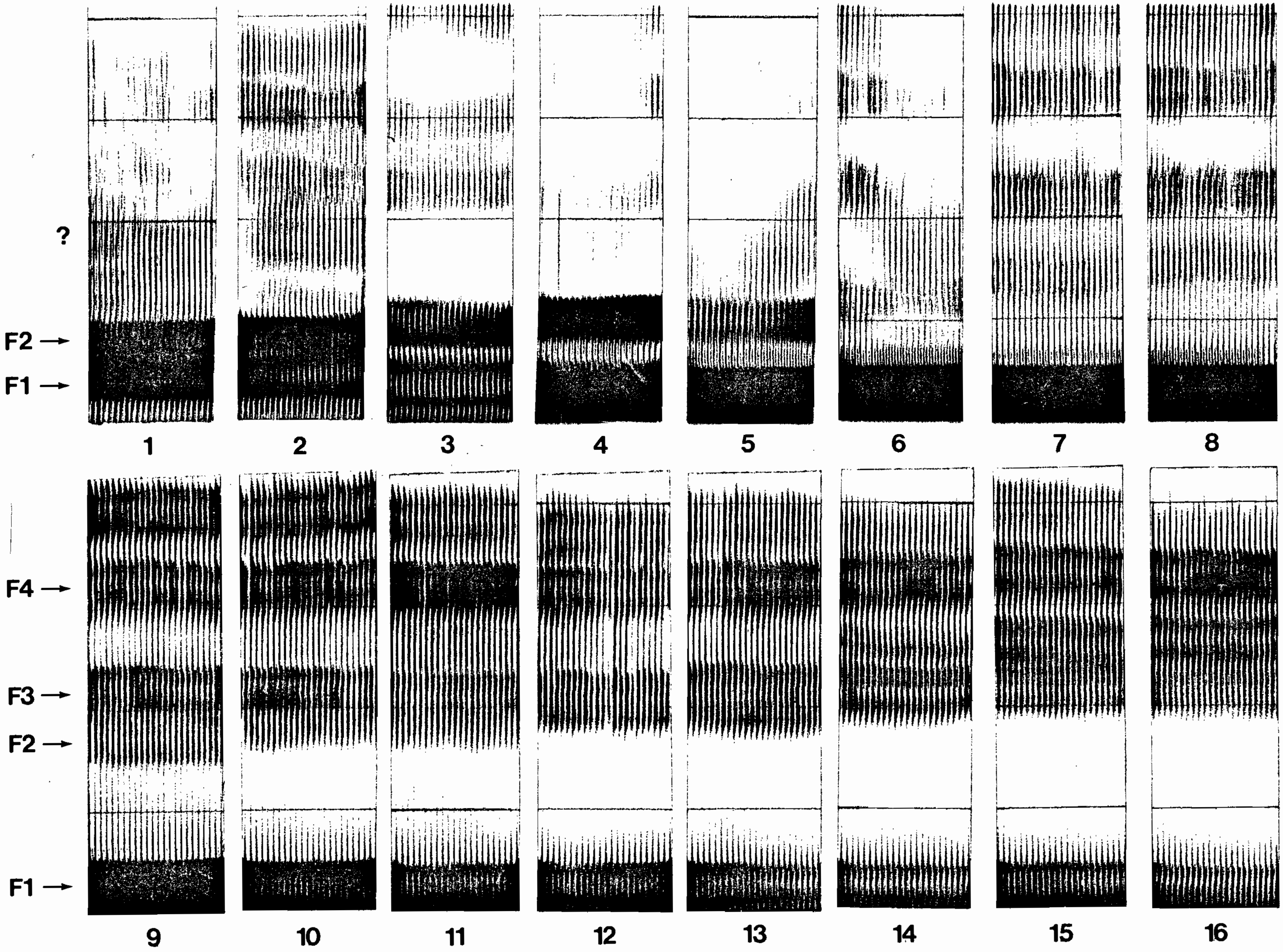


Figure 3. Wide band spectrogramsof PL series 3.

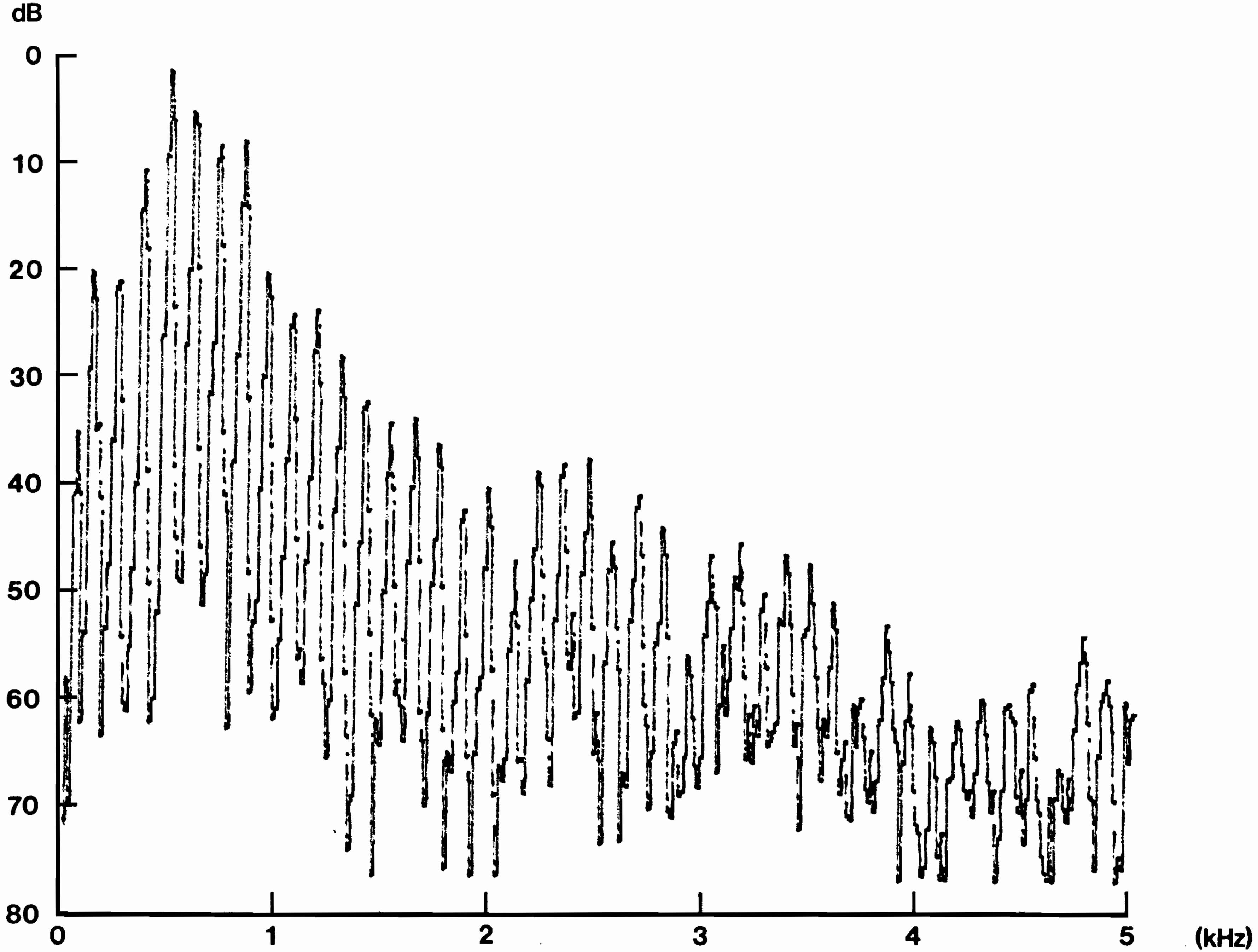


Figure 4. Power spectrogram of the first vowel in Figure 3.

corresponding to the upper resonances of the vocal tract in the first few vowels in this series. For both authors (and for many other speakers whom we have observed) the acoustic energy in low back vowels such as [ɑ] is distributed over a wide range of frequencies. A power spectrum of the first vowel in Figure 3 is shown in Figure 4. It is apparent that this kind of analysis does not make it any easier to determine the true formant frequencies. Considering the other vowels in this series, we might expect to find resonances at about 2300 Hz corresponding to F3 and 3175 Hz for F4. But any algorithm that leads to the selection of peaks in these regions would also lead to peaks being found at 1950 Hz and 2600 Hz. Peak picking in such circumstances is a very arbitrary procedure, and cannot be guaranteed to reveal the true resonances of the vocal tract. In the analysis of our data we followed a rule enjoining us: when in doubt, leave it out. The frequencies of the higher formants were not noted for the vowel in Figure 4.

Further difficulties in locating formant frequencies can be illustrated by reference to vowels 9 through 16 in Figure 3. In vowel 9 the formants are fairly definitely as shown by the arrows to the left of the figure, F3 and F4 having virtually the same frequencies as they had in the previous vowel, and F2 being a little higher. The increase in F2 can be followed through vowels 10 and 11, But in vowel 12 there is only a single formant visible at about 2100 Hz, and F3 would appear to have suddenly assumed a value comparable to that of F4 in the previous vowel. There is also a sudden lowering of the highest visible formant in this vowel. It is not until vowel 14 that the previous progression becomes readily apparent again. It is difficult to relate acoustic behaviors such as these sudden formant discontinuities to the articulatory states which produced them, namely moving the tongue progressively in small steps along the upper surface of the vocal tract. Note also the constancy of the resonance (F3 near the middle of the

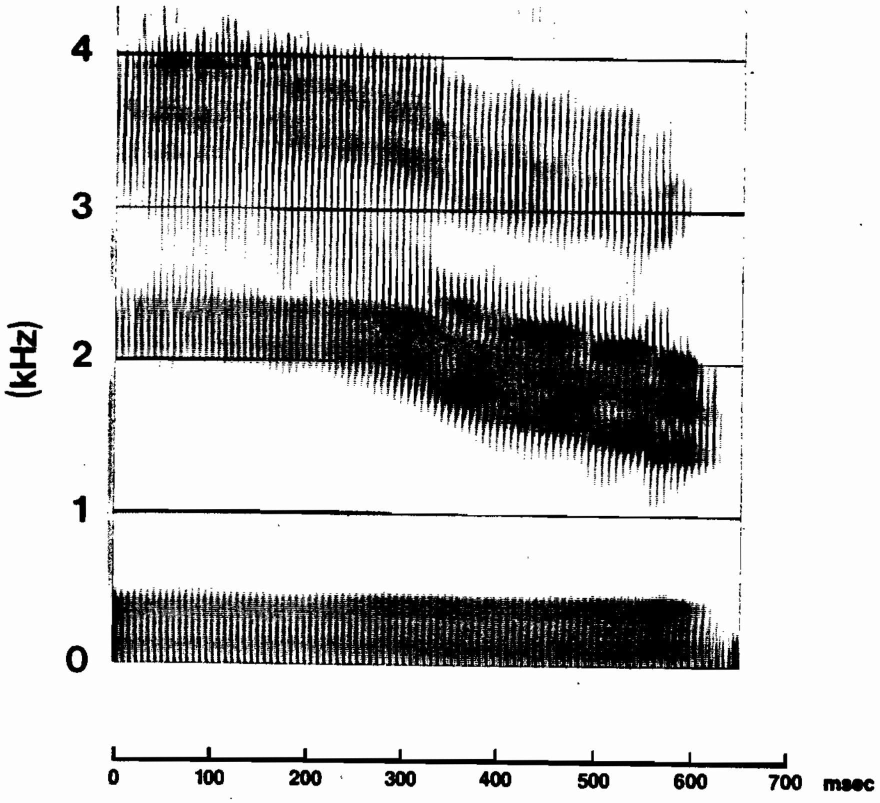


Figure 5. Wide band spectrogram of AB [i - y].

series and F2 towards the end) at about 2100 Hz. This frequency has a wavelength of about 16.8 cm. It probably corresponds to a half wave resonance of the back cavity; but if this is so, it must mean that the back cavity has in effect a fairly constant length throughout a considerable range of articulations. For vowels in the neighborhood of [ɨ] it is the source for F3, and for vowels more like [i] it is the source for F2. We will return to the discussion of this comparatively constant resonance in the next section.

The articulatory basis of the formant changes becomes a little (but only a little) clearer when we study the effect of a continuous change in lip position of the kind shown in Figure 5. This vowel starts with a quality similar to that of [i] in "heed." At the beginning both F3 and F4 are falling, while F2 remains relatively constant. The changes in F3 and F4 are dependent on the changes in size of the front cavity as the corners of the lips come forward, an articulatory change which is not directly simulated in the model of the vocal tract used to generate the nomograms in Figure 1. F2 does not vary during the first part of the vowel, because it is simply a half wave resonance of the back cavity. As the lip rounding increases so that the vowel becomes more like [y] as in French "tu," F3 goes down into the F2 range, and both of them become affected by the increasing lip rounding. During this part of the vowel F4 is relatively constant. Again in Figure 5 however the progressive (and now continuous) change in articulation is accompanied by certain abrupt shifts in apparent formant frequencies.

Results

In our analysis of the acoustic data we always kept in mind the fact that we were trying to find the resonances of the vocal tract. Whenever possible we noted peaks in the regions predicted by Fant's nomograms, disregarding other peaks which might be due to other factors, such as the characteristics of the vocal cord source. Our object was to show to the maximum extent possible the agreement between Fant's theoretical calculations and observed real vowels.

Curves showing the effect of varying the location of the constriction are shown in Figure 6. Each panel shows a different condition of jaw opening (as determined by bite blocks) and of lip aperture. As we do not know the exact location of the constriction at any given moment in time, the abcissa cannot be quantified in these graphs. The points have been arranged on this axis so as to give a smooth variation in F2 which is maximally in accord with that in Fant's nomograms. Vowel symbols have been added below the first and last graph for each author to give some indication of the different vowel qualities involved. In the previously cited data [3], in the case of PL the constriction is 6.7 cm from the glottis for [ɑ] and 15 cm for [i].

Perhaps the first thing to note about Figure 6 is the limited extent of the variations. Our estimate of the area of Fant's diagram that is realizable in human speech is indicated by the bold lines in Figure 1. Both the authors are experienced phoneticians, capable of producing a wide range of articulations. But neither of them feel that they can produce vowels with constrictions much closer to the glottis than that in [ɑ]. Nor can they produce lingual constrictions at a greater distance than that in IPA [i]. This in accord with the definitions of cardinal vowels [i] and [ɑ] which are described as the vowels produced as high and far front as possible, and as low and far back as possible [8].

One can, of course, produce consonants with constrictions in more extreme regions. Pharyngeal consonants of the kind that occur in Semitic languages have

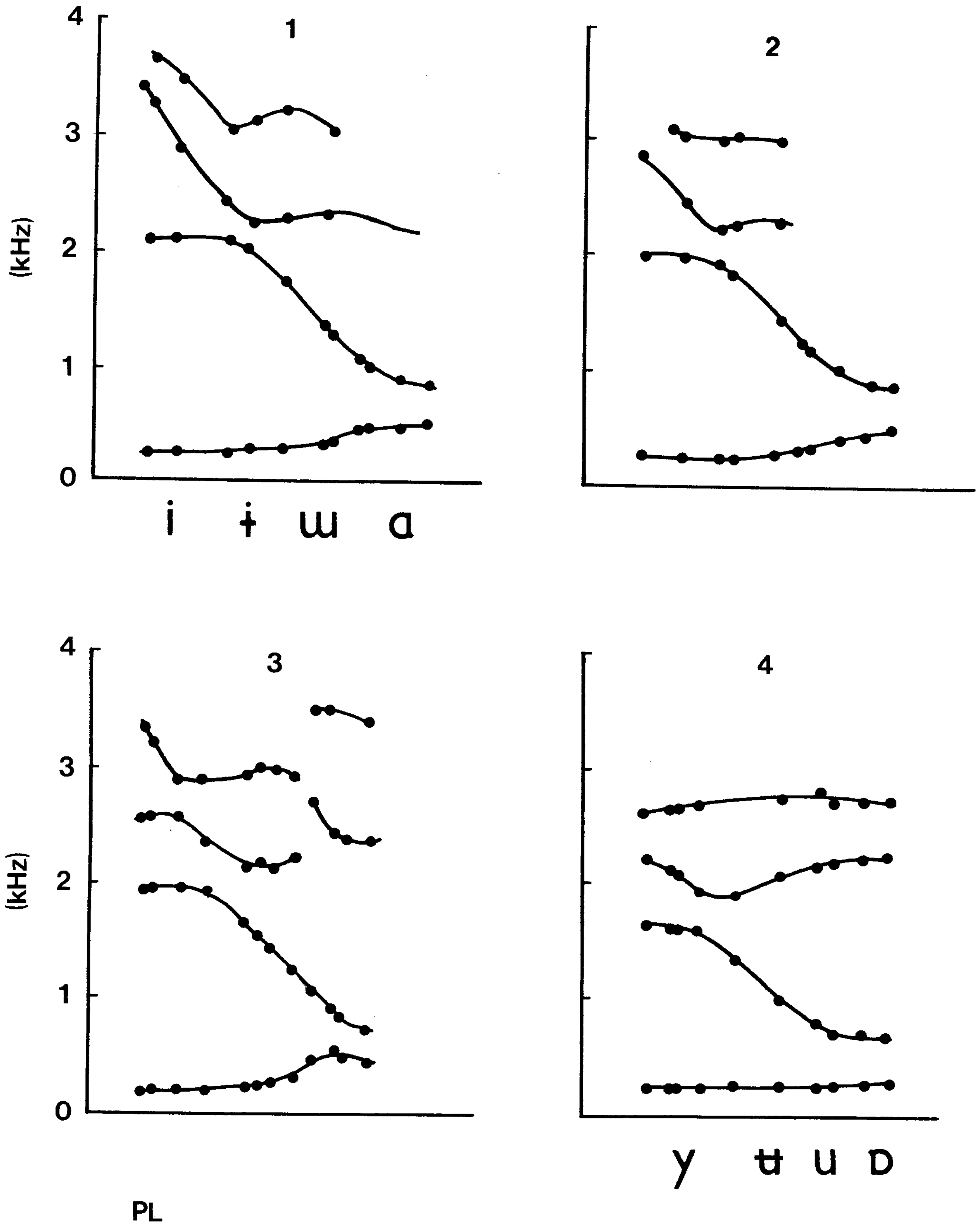
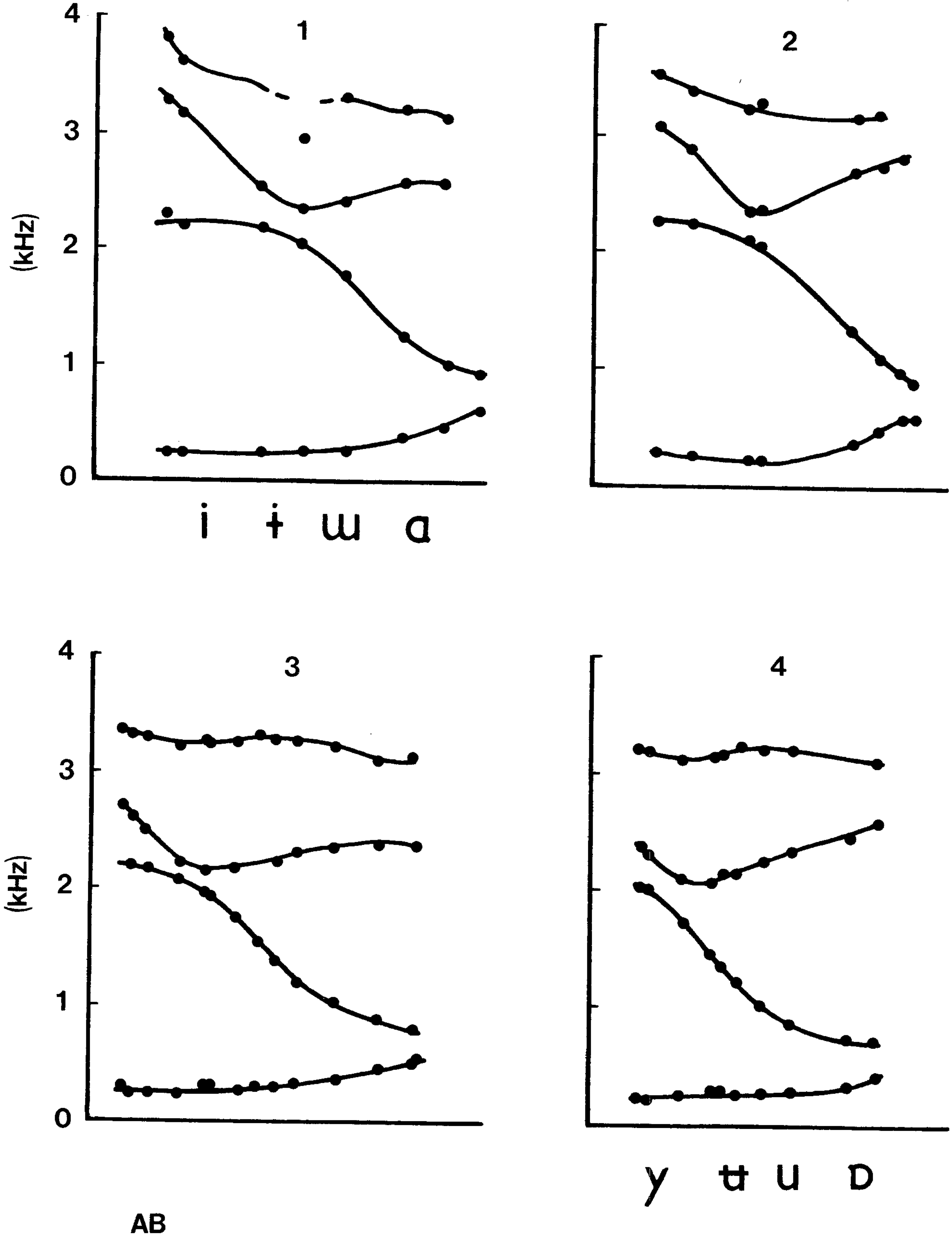


Figure 6. Nomograms corresponding to four sets of vowels produced by each of the authors in a manner similar to that modeled by Fant.


4
3
2
1
0
(kHz)
1
2
i ɨ ɯ ɑ
3
4
y ʉ u ɒ
AB

constrictions nearer to the glottis. Accordingly, to allow for the possibility of vowel-like sounds in this area we have drawn our right hand limit in Figure 1 at 4 cm from the glottis, rather than 6 cm which would be appropriate for a very low form of the vowel [ɑ]. We have set the left hand limit in Figure 1 at 14.5 cm from the glottis. Labial, dental, and alveolar consonants have constrictions beyond this point. No doubt Fant thought that his nomograms would be useful in the interpretation of vocal tract shapes approaching those in these consonants, as indeed they are. But we should view these interpretations with caution, as none of these consonants can be made with vocal tract shapes similar to those specified for the nomograms. The articulatory model used in the nomograms requires a convex curvature of the tongue which can be made only in the range bounded by the bold lines in Figure 1. Thus only about half of each of Fant's diagrams pertains with more than a rough approximation to vocal tract shapes that can really occur in human speech.

Within this area there are discrepancies between our data and Fant's calculations. The most notable is that we do not find any tendency for F2 to become lower as the constriction moves forward in the region corresponding approximately to that shown in the nomograms as 11-14 cm from the glottis. We know that the articulations we were making extended into the region where F2 might have been expected to fall because our articulations did yield the concomitant rise in F3 (see Figure 6) which the nomograms predicted. In these [i]-like vowels, we have never seen anything comparable to the 500 Hz lowering that occurs in this part of Fant's nomograms. In part, this is due to Fant's having modelled, by his extreme values of F2 in the 10-12 cm region, high vowels with improbably large lip aperture. In part also, we suspect that this discrepancy arises from the fact that F2 in these forward articulations is largely associated with the length of the cavity behind the constriction. Because of the curvature of the vocal tract, the length of this cavity does not increase when the constriction moves closer to the alveolar ridge and the tongue body flattens. In Fant's model the vocal tract is considered to be a straight tube with a fixed length. In his model, when the place of constriction is moved further back the length of the the front cavity increases and the length of the back cavity decreases to a corresponding extent. But in human speakers the effective length of the vocal tract itself will vary with the location of the constriction. If the tongue is raised up so that the constriction is near the center of the hard palate the back cavity will be as long as when the tongue is lower and the constriction is near the alveolar/palatal boundary. Conversely, as the constriction is moved back up and back in the vicinity of the alveolar/palatal boundary the length of the front cavity increases, and the frequency of F3 decreases. But the effective length of the back cavity does not decrease in a corresponding manner when the constriction is moved in this way, so the frequency of the second formant does not vary very much.

We should note that there is an extreme anterior area for vowels, corresponding somewhat to the region shown in the nomograms as perhaps 14-15 cm from the glottis, where it is nevertheless reasonable that F2 should lower. This is due to the possibility of apical or alveolarized vowel sounds made with the raising of the tongue/tip blade and a lowered tongue body. Naturally occurring examples of these vowels are not common, but appear to include Chinese [ɿ^{z}] and dialectal Swedish "Visby-i". Such vowels would, in one sense, exhibit a much more anterior location of the constriction than IPA cardinal [i]. Formant data (Wang [9]; Lindblom [10]) do illustrate an F2 lowering to about 1500-1600 Hz as would be predicted by Fant. But Lindblom's data also indicate a lowering of F3 for the Swedish vowels; and formant data for two speakers of Mandarin (Ren Hong-Mo, personal communication) also have values for F3 that are lower than in the corresponding [i] vowels. So Fant's nomograms, which have F3 going up while F2 goes down, do not predict what happens in these naturally occurring vowels. It may be that they are produced with vocal tract shapes that differ significantly from those of the model used to produce the nomograms. They may, for instance, have a more complex curvature of the tongue. Alternatively the lack of real vowels in this area in which F2 goes down while F3 goes up may be simply due to the failure of Fant's model to take into account the curvature of the vocal tract, as discussed above.

The changes in F3 and F4 are more nearly as in the nomograms. The fall in F3 due to the increasing size of the front cavity can be seen on the left of each graph in Figure 6. This is followed, as in the nomograms, by a small increase in the frequency of F3. In high central and back vowels F3 is more nearly a full wave resonance of the tract as a whole . When the tongue is moved backward through vowels of the [ɨ] type, the continuing increase in the size of the front cavity is associated with a decrease in the frequency of F2. As the movement from the central to the back vowels continues, F3 comes closer to F4.

For both authors there are occasional discrepancies between our analyses and Fant's nomograms which we cannot resolve, despite our best efforts to interpret ambiguous records in a way that is maximally in agreement with Fant's data. Thus one of AB's vowels in series 2 has an unambiguous value for F4 that is clearly different from that of the rest of the series. Similarly in PL's series 3 there are abrupt breaks which can be associated with a change in the cavity affiliations of F3 and F4.

The graphs in Figure 6 also indicate, by their separate panels, the effect of variations in the degree of lip rounding. It is, however, easier to assess the effect of lip rounding by considering the data which we recorded specifically for this purpose. In addition, Fant's nomograms do not show the effect of changes in lip rounding in a way that is easy to read. They have separate curves for the different degrees of lip rounding. We find it preferable to consider curves that show what happens when the lip rounding is varied while the tongue position remains constant. Accordingly we have replotted the data in his Figure 1.4-11a in the form shown in Figure 7. As the principal changes are in F2 and F3 we have simply plotted the frequencies of these two formants against each other. Figures 8 and 9 show comparable plots for our data. We have included our data on vowels made with different tongue apertures because they enable us to make some useful points concerning articulatory-acoustic relations of this kind.

Again, the first point to note is the wide range of Fant's data in comparison with ours. This is partly due to Fant having modeled a very large lip aperture, 8.0 cm^2, in one curve of his nomograms. The largest lip aperture in a high vowel in our photographs was 5.78 cm^2 which was produced with a very exaggerated hyper-spread lip width of 6.6 cm. A more typical maximum lip aperture in a high vowel in our photographs was 4.2 cm^2. This helps to explain why in Figures 8 and 9 we do not attest the extreme extension into the high F2 region of the curves for 10.5 and 11.5 constrictions. Correspondingly in Figure 1, we suggest that there is an area of F2 values, shaded by us in the center of Fant's diagram, which does not represent a realistic nomogram for human speakers. We have also noted previously the fact that F2 does not lower in our speakers at very forward locations, and this explains why in Figures 8 and 9 we do not attest curves such as Fant's replotted (Figure 7) ones for tongue constrictions at 14.5 cm and 13.5 cm.

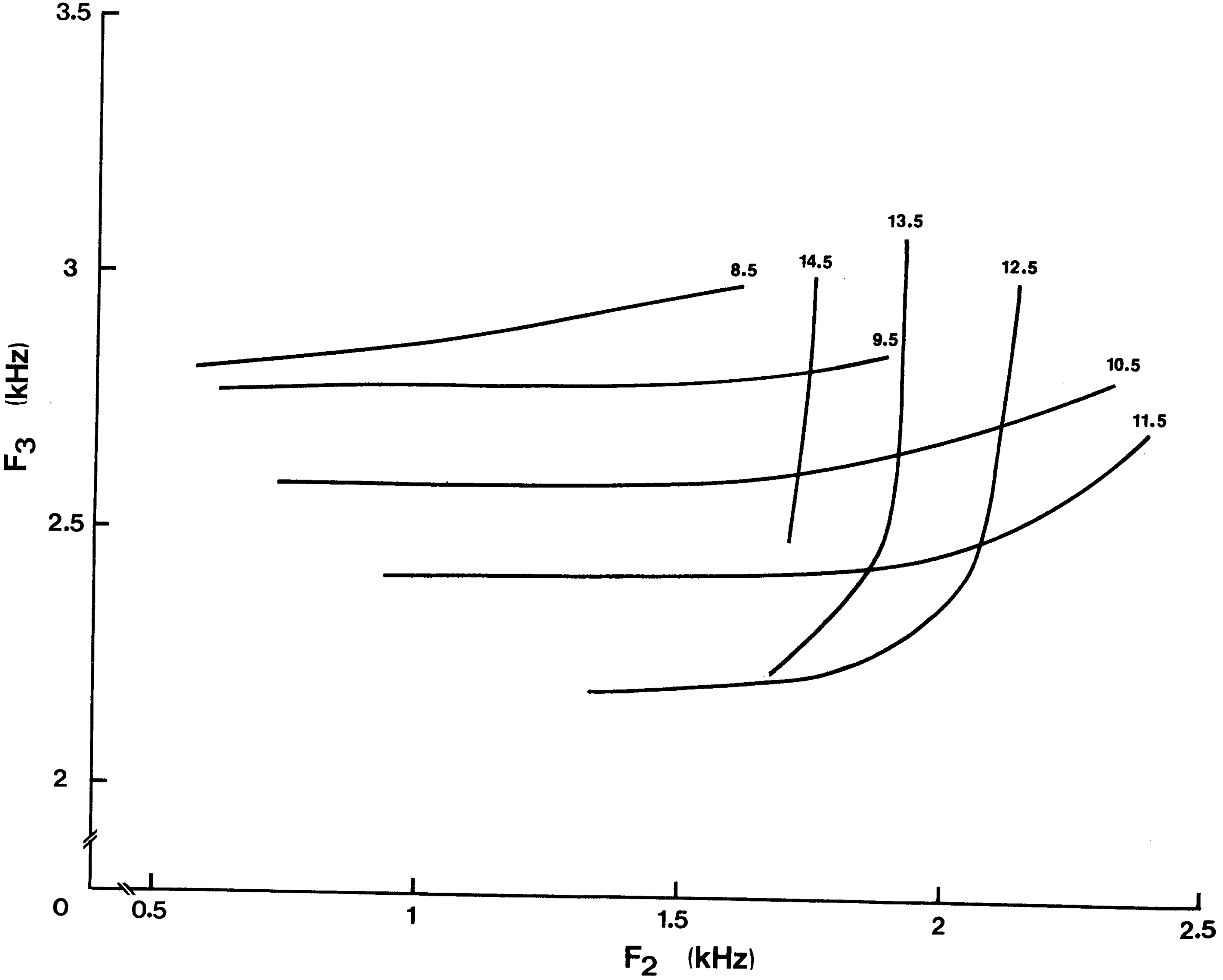


Figure 7. The effect of lip rounding on F2 and F3 replotted from the data in Fant's nomograms. The different curves correspond to the variations in the distance of the constriction from the glottis.

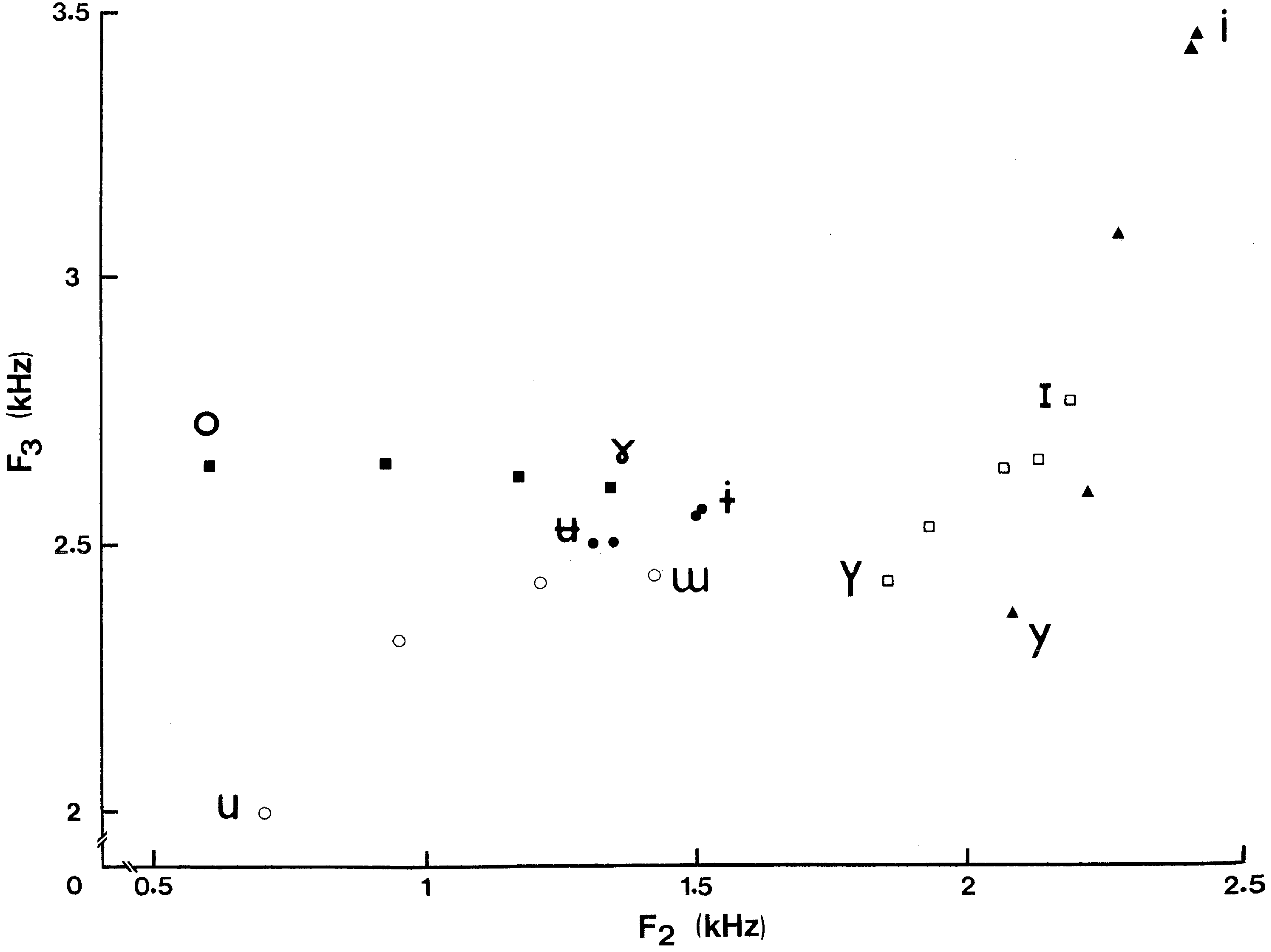


Figure 8. The effect of lip rounding on F2 and F3 in sets of vowels pronounced by AB in which the position of the tongue remained constant.

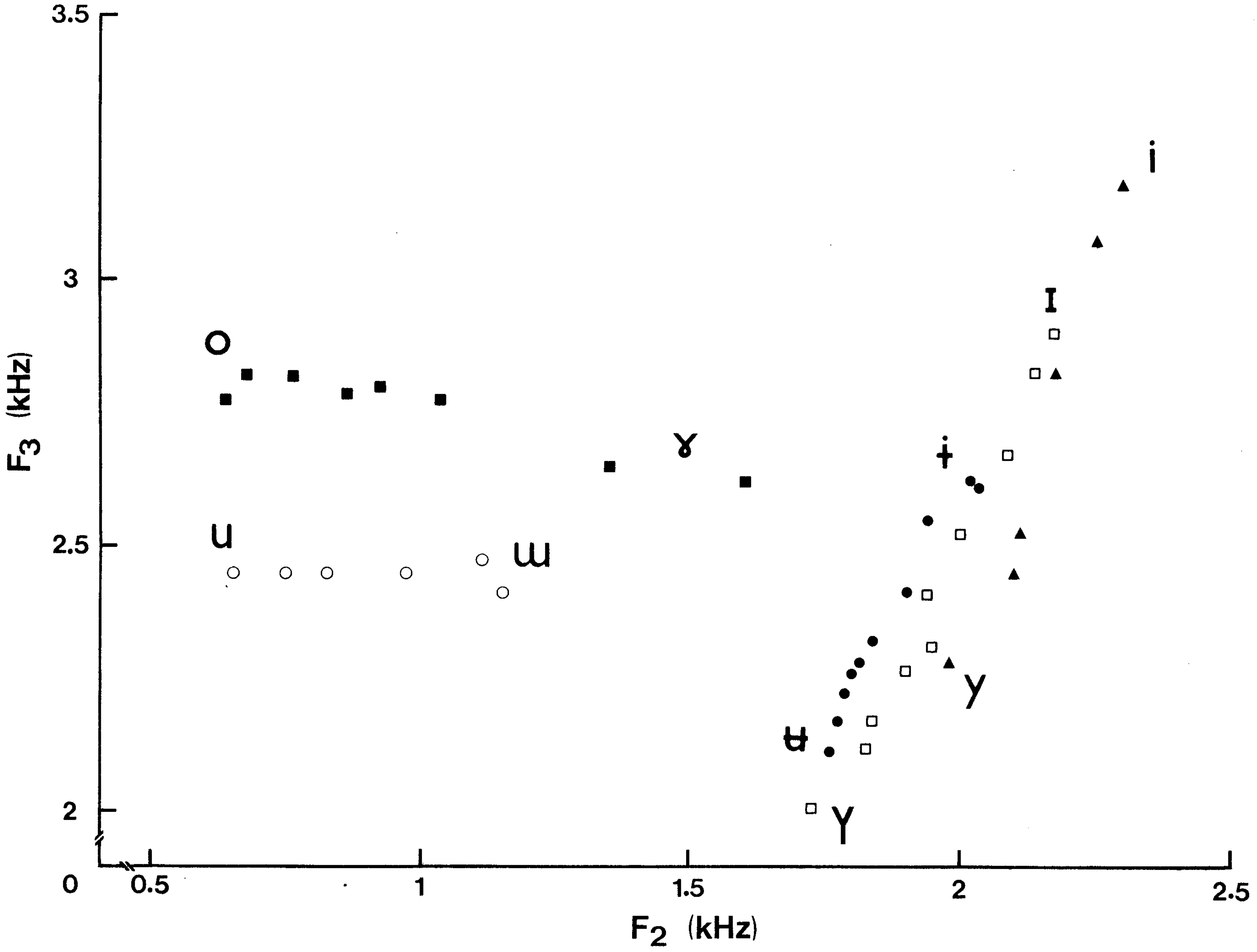


Figure 9. The effect of lip rounding on F2 and F3 in sets of vowels pronounced by PL in which the position of the tongue remained constant.

Fant's data also differ from ours in that his relation between F2 and F3 alters continuously as the constriction moves from the front to the centre of the mouth. But in our data there are two distinct sets of lines, one corresponding to high front vowels, and the other to high back vowels. In our front vowels the main change is in F3, with the frequency of F2 also decreasing slightly; in our back vowels the frequency of F3 is more or less constant, while F2, which is associated with the front cavity for these vowels, changes considerably. Our high central vowels fall into one group or the other, depending, as far as we can feel, on a very small change in tongue position. If the constriction is just behind a certain point, then added changes in lip rounding produce changes only in F2 (see[ɨ-ʉ] in Figure 8.). In front of this point changes in lip rounding affect F3 more significantly (See [ɨ-ʉ] in Figure 9.). A change in lip rounding produces a change in both F2 and F3 only when we are producing more open central vowels such as [ʌ]. But when pronouncing vowels of this more open type neither of the authors was able to produce large changes in the formant frequencies by adding lip rounding. This may be correlated with the fact that no language known to us uses an opposition of this kind [11].

Finally, summarising all our findings with respect to Fant's data, we may note that there are three main kinds of discrepancies that occur when we as phoneticians try to produce vowels similar to those simulated in the nomograms. Firstly, when varying vocal tract shapes while maintaining curvatures of the kind similar to those specified by the model, we cannot produce the range of variations that occur in the nomograms. Secondly, when the tongue body constriction is moved forward the frequency of F3 increases (as predicted by Fant), but the frequency of F2 does not decrease. Thirdly variations in lip rounding affect high front and back vowels differently with no gradual change from one class to the other as in Fant's data. These discrepancies, however, are minor in comparison with the major points of agreement between Fant's predictions and the data observed in real speech. Fant's nomograms remain as one of the major achievements of his acoustic theory of speech production.

References

[1] G. Fant, Acoustic Theory of Speech Production, Mouton, The Hague 1960.

[2] D.J. Broad and H. Wakita, "Piecewise-planar representation of vowel formant frequencies" Journal of the Acoustical Society of America Vol.62, No.6, 1977, pp1467-1473.

[3] P. Ladefoged, Elements of Acoustic Phonetics, Chicago University Press, Chicago, 1962.

[4] P. Ladefoged et al., "Direct measurement of the vocal tract" UCLA Working Papers in Phonetics, Vol 19, 1971, pp. 4-13.

[5] J.C. Catford, Fundamental Problems in Phonetics, Edinburgh University Press, Edinburgh, 1977.

[6] J. Gauffin and J. Sundberg, "Pharyngeal constrictions" Phonetica, Vol. 35, 1978, pp157-168.

[7] P. Ladefoged, Three Areas of Experimental Phonetics, OUP, London, 1967.

[8] D. Jones, Outline of English Phonetics, 8th edn., Heffer, Cambridge, 1956.

[9] W. S.-Y. Wang, "The basis of speech" Project on Linguistic Analysis, No. 4, 1968.

[10] B. Lindblom, "Phonetic universals in vowel systems", In Ohala J.J. ed., Experimental Phonology, Academic, London, Forthcoming.

[11] I. Maddieson, "UPSID Data and Index" UCLA Working Papers in Phonetics Vol. 53, 1981.

Physiological motivations for phonetic naturalness

Patricia A. Keating and Wendy Linker
UCLA

Paper presented at the Conference on Causality and Linguistic Change
UCLA, May 1982

In this paper we will describe work we have been doing which addresses the question, "What can phonetics contribute to the study of grammar, and in particular, what are the nature and role of phonetic constraints in the phonology?". Our interest has been purely a synchronic one, and our paper will reflect this orientation. However, we feel that our research may be relevant for historical linguists interested in causes of change, to the extent that those causes are thought to include a tendency toward "natural" patterns. Our study of phonetic constraints can be thought of as bearing on the basis of phonetic naturalness. We will not make any claims about naturalness as a diachronic causal mechanism, leaving that issue to historical linguists. Rather, we simply hope to provide a more precise way to think about the contribution of the articulatory mechanism to considerations of overall naturalness.

One way in which this has been described in articulatory terms has been "ease of articulation", also known as the principle of least effort. For example, de Saussure (1959) said:

> "The cause of phonetic changes has also been ascribed to the law of least effort by which two articulations are replaced by one or a difficult articulation by an easier one. This idea, regardless of what is said about it, is worth examining. It may clarify the cause of phonetic changes or at least indicate the direction that the search for it must take. The law of least effort seems to explain a certain number of cases: (...) But we might mention just as many instances where exactly the opposite occurs: (...) In fact, we can scarcely determine what is easiest or most difficult for each language to pronounce... The law of least effort would require extensive study. It would be necessary to consider simultaneously the physiological viewpoint (the question of articulation) and the psychological viewpoint (the question of attention)." (Part Three, Ch. II, p. 148).

One way we have looked at articulatory motivations for naturalness is in a study of the allophones of voiced and voiceless stops. "Voicing" here includes both differences between aspirated and unaspirated stops as occurs in English, and differences between unaspirated and fully voiced stops as occurs in French, that is, any differences along the Voice Onset Time dimension. Our study has had two parts: first, the collection of synchronic data on allophone distribution, and second, use of an articulatory model to describe mechanical constraints on voicing. We will discuss each in turn, and then we will apply the results to the question of phonetic naturalness.

There have been many interesting observations made about frequency of occurrence across languages of certain sounds and sound sequences. A good example is the UCLA Phonetics Lab's UPSID project -- UCLA Phonological Segment Inventory Database -- which comprises phonetic feature characterizations of surface phonemes in 317 languages (Madddieson, 1981). Statistical generalizations from this database have also been provided (Maddieson, 1980); of interest to us is the volume by Jonas Nartey (1979) on stops and fricatives. From such segment surveys one finds statistical support for the common claim that more languages have voiceless stops than have voiced stops, and one finds the accompanying claim that the more frequent type is articulatorily easier. However, such data on phoneme occurrence cannot serve as the basis for articulatory modeling, for a phoneme is by definition an abstraction from any particular surface phonetic variant, while an articulatory model must be context specific. Therefore, if we want to assess articulatory demands, we must work from segment data which is also context specific.

We have concentrated on positional allophones in our data collection, since this factor appears to have the strongest effect on allophonic variation in voicing. In addition to classic allophonic statements, we derive data from surface variation of stops in languages with no voicing contrast, surface forms of stops in positions of neutralization, and morphophonemic alternations.

Our data collection, which is still in progress, currently consists of information on obstruents in about 30 languages, drawn from the literature, from the Phonetics Lab archives, and in some cases from acoustic analysis. A number of phonetic generalizations concerning stop consonant voicing emerge from this corpus, which we report here for the first time.

1) We might expect that languages with no voicing contrast either overall or in a particular environment would show more allophonic variation -- or conversely, that a contrast constrains otherwise free variation. Of particular interest are our two languages without a voicing contrast, Alyáwarra and Hawaiian. They have voiceless unaspirated stops in all environments showing only slight variation in Voice Onset Time as a function of stress, vowel, place of articulation, etc., in the usual way.

2) The exclusive use of voiceless unaspirated stops by these languages is a specific instance of the general phenomenon that virtually all languages in all positions have at least voiceless unaspirated stops (except in certain consonant clusters).

3) The one case that we expect to be an exception to this generalization is that medial stops should preferably be voiced. There is a strong intuition among linguists (including us) that between two vowels or two sonorant consonants there should be no interruption of voicing. Again, such an expectation is not supported by the data, with regard to stops alone. There is a Korean medial voiced allophone, and the marginal case of the American English neutralization of /t/ and /d/ to a tap (which is not a stop) before stressless vowels. There are also several cases like English, which in medial position has fully voiced /b d g/ and voiceless unaspirated /p t k/. While this is a case of allophonic medial voicing, it does not indicate any preference for medial voicing over medial voicelessness. (It does indicate a preference for a phonetic voicing contrast over an aspiration contrast in medial position.) An example of medial allophonic stop voicing in the literature (Corsican, in Dinnsen and Eckman 1977) is not included in our sample. On the other hand, there is a case of neutralization to medial voiceless unaspirated stops, in Barra Island Scots Gaelic, of voiceless aspirated and unaspirated stops (Borgstrøm, 1937). It seems that the best we can say in support of our intuition is that medial position does not work against voicing, since voiced medial stops are not actually more frequent than voiceless medial stops.

4) Languages with a traditional voicing contrast not involving aspiration, such as Polish, French, Dutch, and Tagalog, show very little difference between stops in initial and intervocalic positions; if the final stops are not neutralized and are released, they also are similar.

5) On the other hand, languages like English whose contrasts involve aspiration show a great deal of allophonic variation. Both position and stress appear to affect surface realizations. A hypothesis currently under investigation is that such languages will usually show such variation.

6) Languages with three or four contrastive categories seem to show less allophonic variation but more neutralization. Initial position is the site of the greatest number of contrasts; final position is the site of the fewest, with both devoicing and deaspiration being common.

7) Final devoicing is quite common, cutting across contrast types. However, minimal pairs with final devoiced and voiceless stops are not necessarily acoustically identical. Vowel duration and/or the amount of voicing during closure distinguish the surface variants of words containing underlying voiced and voiceless stops in German, Russian, Polish, and Bulgarian, even though such differences may not be perceptually discriminable. Recent work on this topic include Port et al. (1981) and various UCLA class papers. When closure voicing differences are involved, we can say that the underlying voiced and voiceless stops are not phonetically neutralized, although they may not be perceptually discriminable. This appears to be the case in Polish. Westbury and Keating (1980), citing a paper by Giannini and Cinque and drawing on a collection of Polish recordings, indicated that at least a few more pitch periods of closure voicing occur in underlyingly voiced final Polish stops than in underlyingly voiceless final Polish stops. Vowel duration does not vary according to consonant voicing in Polish. While the relevant data is not generally available for all the many languages with final devoicing, in those cases known to us (German, Russian, Bulgarian, Polish), minimal pairs of this sort which are neutralized phonologically are not neutralized phonetically, and in at least the case of Polish it is clearly the final stops themselves which are not phonetically neutralized.

8) Assimilation of voicing in obstruent clusters is very common across contrast types. However, languages vary in directionality, output, and input restrictions.

It is also the case that data on child speech supports these various trends, with perhaps a somewhat greater preference for medial voicing, although this shows more variation across individuals than other trends do.

In sum, we expect to find voiceless unaspirated stops initially, medially, and finally. We expect to find fully voiced and voiceless aspirated stops initially somewhat less frequently than the voiceless unaspirated ones, but each as often as the other. Medially, fully voiced stops may be as preferred as voiceless, but the occurrence of aspiration will depend on stress. Finally we expect perhaps to find two kinds of voiceless unaspirated stops, those from underlying voiced stops, and those from underlying voiceless stops.

Given such consistent variation in stop consonant allophones, we go on to ask what role articulatory constraints play in such patterns. Following a direction in linguistic phonetics taken by Ohala (e.g. Ohala 1975), we have pursued the use of an aerodynamic model of articulation -- that is, a model of variations in air pressure and air flow through the vocal tract for different articulatory gestures and states. Unlike Ohala's original aerodynamic model, but like his more recent version, the model implemented at MIT by Pat Keating and John Westbury is derived from an electrical circuit analog of the vocal tract due to Rothenberg (1968), but that technical detail is irrelevant here.

To see why such a model can tell us about variation in stop consonant voicing, let us review some of the basics of voicing mechanics. Figure 1 shows a schematic representation of the vocal tract, after Lieberman (1977), that indicates the path of air flow.

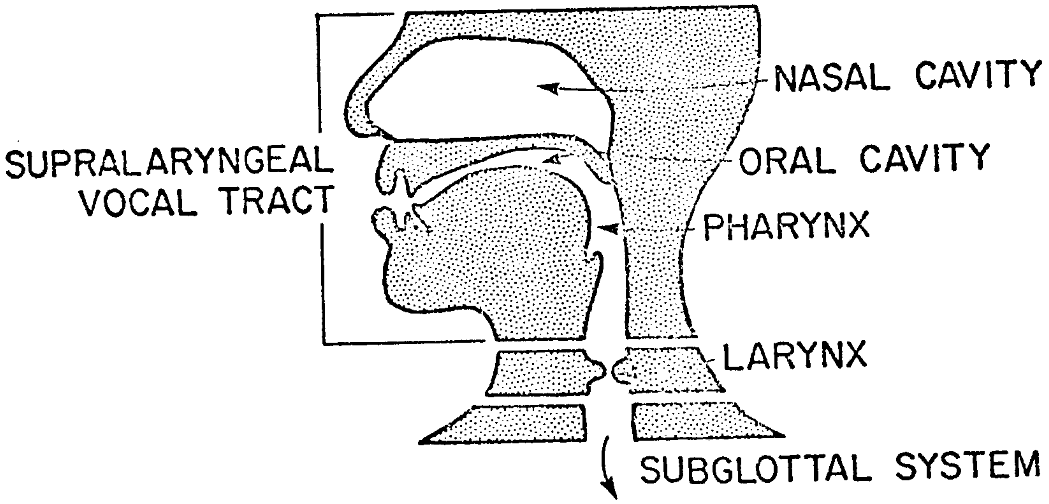
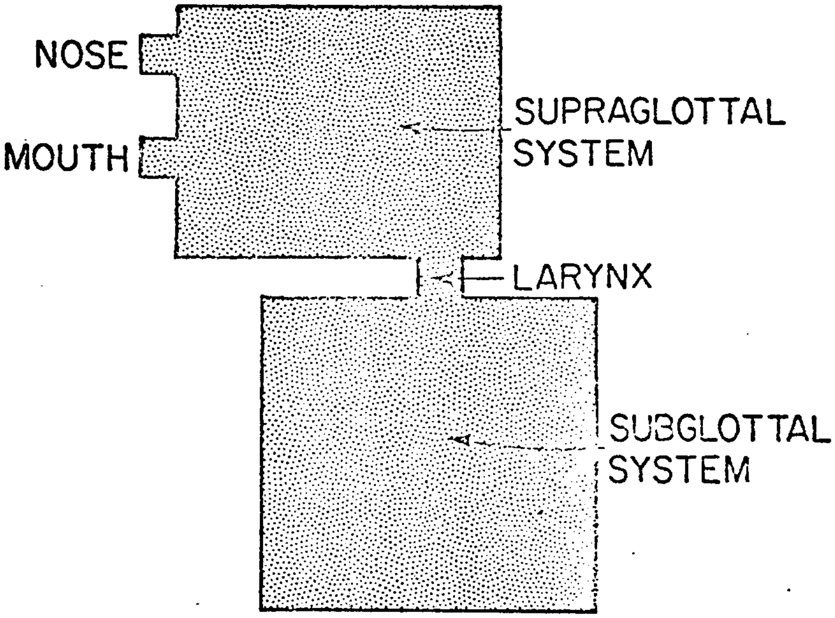


Figure 2 shows a more schematized representation of the elements crucial to voicing: a cavity below the larynx, a cavity above the larynx, and the larynx between them.



There are two conditions necessary for voicing: first, that the vocal cords be in a position and state that allows vibration, and second, that there be sufficient airflow between them to set off the vibratory cycle. Obviously, if we simply pull the vocal cords far apart we won't get voicing no matter what else we do. It is thus more interesting to consider what happens when the vocal cords remain ready, but other conditions change. Therefore most of our modeling to date has been for cases where the vocal cords will vibrate whenever high airflow is maintained. Sufficient airflow will occur when the pressure above the larynx is less than the pressure below. An easy way to guarantee low pressure above is to provide an escape for the air, through the nose and/or mouth. If air flowing into the oral cavity is matched by air flowing out, pressure will never build up, and voicing will be possible. But an oral stop consonant by definition closes off both oral cavity exits, trapping the air so that as more air passes through the larynx it is simply added to the air already there, increasing pressure until voicing becomes impossible even if the vocal cords are ready. This is the usual explanation for why the unmarked obstruent phoneme is voiceless: stops, which cut off airflow, contradict voicing, whereas sonorants, which involve free flow, provide optimal conditions for voicing.

But this explanation won't suffice entirely, since languages generally do have at least one phonetically voiced stop. If stops really meant that voicing was difficult, we wouldn't find voiced stops so frequently. A fair amount of recent research has been directed at determining the articulatory gestures which will allow voicing in various stop allophones. Let's turn to using our articulatory model to see what's involved in voicing production, keeping in mind the requirement that vocal cord vibration depends on air flow which in turn depends on air pressure.

While it might be expected that such a model would include only a few parameters that influence voicing, in fact there are several, each of which is important and must be represented. To give you an idea of what the parameters are, and what it means to provide values for each one, we will illustrate them by showing the values used for an utterance final labial stop. These values will be given in conventional cgs units which will not be named.

First, there are parameters which have to be taken into account but whose values cannot be varied by the speaker. These parameters are all part of the representation of the subglottal system. One of these is the stiffness of the walls of the trachea, bronchi, etc., which includes what we normally describe as "elastic recoil". Our sample value for this parameter is 0.1. Other parameters in this category and their sample values are mass of the walls, 0.037; mechanical heat loss in the walls, 3.3; other subglottal losses, 3.0.

Second, there are parameters whose values are selected by the speaker, but over a stretch of utterance larger than a segment. Again, these are part of the representation of the subglottal system. One is a term which includes the volume of air in the lungs and a constant, the net term being 0.0034. The other is the respiratory muscular force, which causes inhalation before an utterance, and later in that utterance counters a drop in subglottal pressure.

Third, there are parameters whose values are selected by the speaker, over a segment-sized stretch of the utterance. These represent the glottal and supraglottal systems. One is the mean glottal opening as a product of three dimensions: .3 X 1.8 X 0.018. The choice of a place of articulation affects the values of several parameters because it determines the size of the oral cavity, the surface area of the oral cavity, and the dimensions of the constriction itself. Thus we have losses of energy in the walls due to their mass (0.012) and heat loss (8.48), and the volume of air in the cavity, with a constant (0.000072). The parameter representing the stiffness of the walls reflects both the place of articulation and the tenseness of the walls (0.00056). The dimensions of the oral constriction as the labial stop is begun are 0.2 X 2.0 X 0.125. There is assumed to be no active expansion of the vocal tract and no nasal opening.

Given these values, the model generates a partly voiced, largely devoiced final labial stop. In fact, this arrangement of settings, chosen independently, gives an output that agrees well with the acoustics of devoicing in Polish and in child speech. The devoicing is due to the fall in subglottal pressure at the ends of utterances.

Suppose we want to produce a fully voiced stop instead: there are at least two ways. One is to relax the cheeks and other surfaces of the vocal tract in an extreme way, giving some additional msec of voicing. Another is to actively expand the size of the oral cavity. Here a small amount of energy input into the system goes a long way -- since not much more voicing during consonant closure is required, almost any expansion suffices to change the stop. And if we want to produce a completely voiceless stop, extreme tensing of the cheeks can eliminate a few msec of voicing during closure, while separation of the vocal cords will of course also prevent voicing.

Results of such modeling have shown that in utterance-initial position we will also tend to have voiceless unaspirated stops. To produce aspiration in initial position, an appropriately timed apart-and-together movement of the vocal cords is required. To produce prevoicing in an oral stop in initial position, active expansion of the oral cavity is required. In inter-sonorant position, we will tend to have voicing through most of the stop, more than enough for us to hear a voiced stop unambiguously. These results accord well with observations about word initial and final stops in various languages and in child speech. In medial position, they accord more with our intuitions than with observations. The model can also be used to look at differences with place of articulation, but we will not discuss these here.

However, let us step back for a moment and consider what is involved in these approximations. Recall how we generated the final voiceless unaspirated labial stop: by providing a value or function for each element in the model. Where did those values come from? Were those the only values that would produce this "natural" output? Some of the elements are not controlled by the speaker and are constant across all segment types; their values are derived from the literature. Others reflect the place of articulation; their values are taken from our own physiological studies. Finally, stiffness of the walls can have any of a range of values derived from nonspeech physiological studies. We have chosen values which are segment neutral and thus have aimed to describe a speech ready state. It turns out that by using these values we are able to approximate our data on segment naturalness. In contrast, as an example of a value which could <u>not</u> be neutral, consider vocal cord spreading, utterance-finally, as a mechanism of devoicing. Studies at Haskins Labs have shown that the vocal cords stay together after the last segment, even reclosing after a voiceless obstruent, so vocal cord spreading must be treated as a gesture specific to voiceless segments (T. Baer, personal communication).

Another interesting point to be made from the results of this model concerns the stiffness of the vocal tract walls: if they expand in response to increasing pressure, voicing will continue for a longer time. There has been some debate over whether this parameter by itself can form the basis of linguistic segments. Our modeling indicates that by itself tenseness can affect the duration of voicing or of aspiration in a way that may be important for cross-speaker or cross-language phonetic differences, but that extreme values are required to effect a categorial distinction. A model such as ours could prove useful in distinguishing features which are readily available for contrastive effect, from those such as tenseness which are not.

Another possible use could be to look at asymmetries between articulatory inputs and acoustic outputs. Several different combinations of values can give the same output in some context: a change in one element can be compensated for by a change in another, so that two different sets of motor gestures or states will have the same effect. However, in some other context these same motor differences could produce different results. As an example, there are various possibilities for producing voiceless unaspirated stops in initial position. Used intervocalically, these will no longer be equivalent. Conversely, to get the same acoustic output in two different environments may require two different sets of gestures. Thus, compare the inputs required to give voicing in initial vs. medial positions.

Taken together, these results and the results from the allophone database suggest that there are many possible differences between sounds, that could in principle influence the changes they might undergo. On the one hand, close inspection showed acoustic differences between sounds that appear to be the same. There were both obvious differences across positional allophones that simply were not noted, and subtler differences as in the final devoicing case. On the other hand, we found through modeling that acoustically identical allophones may be produced by different articulations either in a single environment (that is, with free choice), or in two environments (that is, by necessity, to produce the same acoustic result).

It is often the case that sound changes do not apply context freely to all the allophones of a sound, or to all the sounds in a class of sounds. We are suggesting that sounds within a language can differ in rather subtle phonetic ways. If a rule or change is phonetically specific, it may apply to some but not all variants in ways that may appear idiosyncratic. The model helps to predict different outputs due to contextual effects.

There is another implication of our research that returns us to the question of phonetic naturalness. We have proposed that some sounds are given automatically by the articulatory device, given certain neutral values for the various elements in the system. To the extent that those values are common to all segment types, that is, are "default" inputs to the system, then we can talk about the "default" outputs of the system. And to the extent that sound systems preferentially contain those outputs, we can say that languages are influenced by articulatory considerations of the sort we are discussing here. Thus, for initial and final stops, we can conclude that the most commonly found segment types can be entirely accounted for by articulatory constraints of the type we have described. On the other hand, we can conclude that the facts about medial position do not derive entirely from articulatory considerations of this sort. Other factors must also be involved. However, note that we are not saying that the vocal tract imposes some conditions on outputs. Rather, we have indicated how the most neutral values as normally *chosen by speakers* influence outputs. When a sound system does contain some sound other than the default output of the model, we can use the model to see what articulatory combinations could produce it. Such articulations can be described in terms of their departure from the default parameter settings described earlier. The elements of the articulatory model, and their ranges of possible values, in a sense define the degrees of freedom given to a speaker in the production of sounds. Sometimes languages will not exercise any choices other than to choose the basic values: but we stress that this is still a choice being made. Other times values of many elements may be varied, producing less usual outputs. Although we have not worked out a specific proposal, it seems reasonable to suggest the following. The similarity between a sound (in a sequence) and a default output of the model (for that sequence) could be a metric of the phonetic naturalness of that sound. The differences in parameter settings required for a sound and for a default output of the model could be a metric of the sound's articulatory cost.

Ohala (1974 and elsewhere) has proposed that sound changes are explained in part by physical limits imposed by the speaking machine. He has given several cases where changes are unidirectional and has offered explanations for them. These examples have had the appearance of showing sounds at the outer limits of speeh production abilities changing to ones more within the capabilities of the system. It always seems that the state prior to the change was an unnatural, tenuous one, and one feels that the better the explanation of why the change should have occurred, the greater the burden to justify the prior state of affairs, a paradox of explanation.

Our account is intended to have a somewhat different character. Research has shown that most of speech operates well within the limits imposed by the articulatory apparatus. We are not constantly challenging the system or running it at full throttle. Rather, we run it in an efficient, convenient way. In our account, some sounds are the most convenient of all: the default outputs. Other sounds are somewhat less convenient, requiring more articulatory adjustments. But their production does not generally strain the system in any way. All things being equal, a language may choose the luxury of "easiest" articulation. But all things are rarely equal; languages may indulge themselves in all sorts of sound sequences.

At the 1969 UCLA conference on historical linguistics, Schane (1972) expressed the belief that modern experimental phonetics, particularly as done at UCLA, would provide an explicit characterization of ease of articulation. We hope that work such as ours on the role of phonetic constraints may be of interest to historical linguists in that regard.

REFERENCES

Borgstrøm, C. H. (1937). "The Dialect of Barra in the Outer Hebrides". *Norsk Tidskrift for Sprogvidenskap VIII*:71-242.

de Saussure, F. (1959/1966). *Course in General Linguistics*. McGraw-Hill Paperback Edition.

Dinnsen, D. A., and Eckman, F. (1977). *Some Substantive Universals in Atomic Phonology*. Indiana University Linguistics Club and *Lingua 45*:1:14.

Giannini, A. and Cinque, U. (1978). "Phonetic Status and Phonemic Function of the Final Devoiced Stops in Polish". *Speech Laboratory Report I,* Laboratorio di Fonetica Sperimetale, Napoli.

Lieberman, P. (1977). *Speech Physiology and Acoustic Phonetics*. New York: Macmillan Publishing Co., p. 4.

Maddieson, I. (ed.) (1981). *UPSID*. *UCLA Working Papers in Phonetics 50*.

Maddieson, I. (1981). *UPSID: Data and Index*. *UCLA Working Papers in Phonetics 53*.

Nartey, J. (1979). *A study in Phonemic Universals, especially concerning fricatives and stops*. *UCLA Working Papers in Phonetics 46*.

Ohala, J. (1974). "Experimental historical phonology". In Anderson and Jones (eds.), *Historical Linguistics II. Theory and description in phonology*. Amsterdam: North Holland Publishing Co.

----- (1975). "A Model of Speech Aerodynamics". Paper presented at the Eighth International Congress of Phonetic Sciences, Leeds.

Port, R., Mitleb, F., and O´Dell, M. (1981). "Neutralization of obstruent voicing in German is incomplete". Paper presented at the 102d meeting of the Acoustical Society of America.

Rothenberg, M. (1968). *The Breath-Stream Dynamics of Simple-Released-Plosive Production*. Basel: S. Karger.

Schane, S. (1972). "Natural Rules in Phonology". In Stockwell and Macauley (eds.), *Linguistic Change and Generative Theory,* Indiana University Press.

Westbury, J. and Keating, P. (1980). "A model of stop consonant voicing and a theory of markedness". Paper presented at the Linguistic Society of America.

Phonetic differences in glottalic consonants

Mona Lindau

Paper presented at the 30th meeting of the Acoustical Society of America in Chicago, May 1982.

This study forms part of the UCLA project on phonetic differences between languages. We have shown in many instances that reliable differences can be found between what have been classified as "the same sounds" in different languages. Glottalic consonants in a number of languages were studied. These consonants are produced with an airstream that is created by rapid vertical movement of the glottis. A rapid downward movement of the glottis results in implosives, like [aɓa]. A rapid upward movement of the glottis results in ejectives, like [ak'a]. Studies of these non-English sounds give us important insights into the potentialities of the vocal mechanism.

The data consist of tape recordings of several speakers for each language. Most of the recordings were made on a good reel to reel tape recorder in the field. The glottalic consonants were in intervocalic position between a-type vowels in real words, said in a frame. These consonants were analyzed using a computer system for displaying the waveforms and spectra.

Implosives

The list below states the languages concerned, their linguistic classification, and the number of speakers from whom data was collected. According to Maddieson (1981) about 10 % of the world's languages have implosives, and they are very common in certain geographical areas, like West Africa. All the languages in this list are spoken in Nigeria.

	Languages		Number of subjects
NIGER-CONGO	Ijo	Kalabari	7
		Okrika	4
		Bumo	5
	Edoid	Degema	4
AFROASIATIC	Chadic	Hausa	14

Figure 1 shows the waveforms of a regular plosive [b] and an implosive [ɓ] spoken by a speaker of Degema.

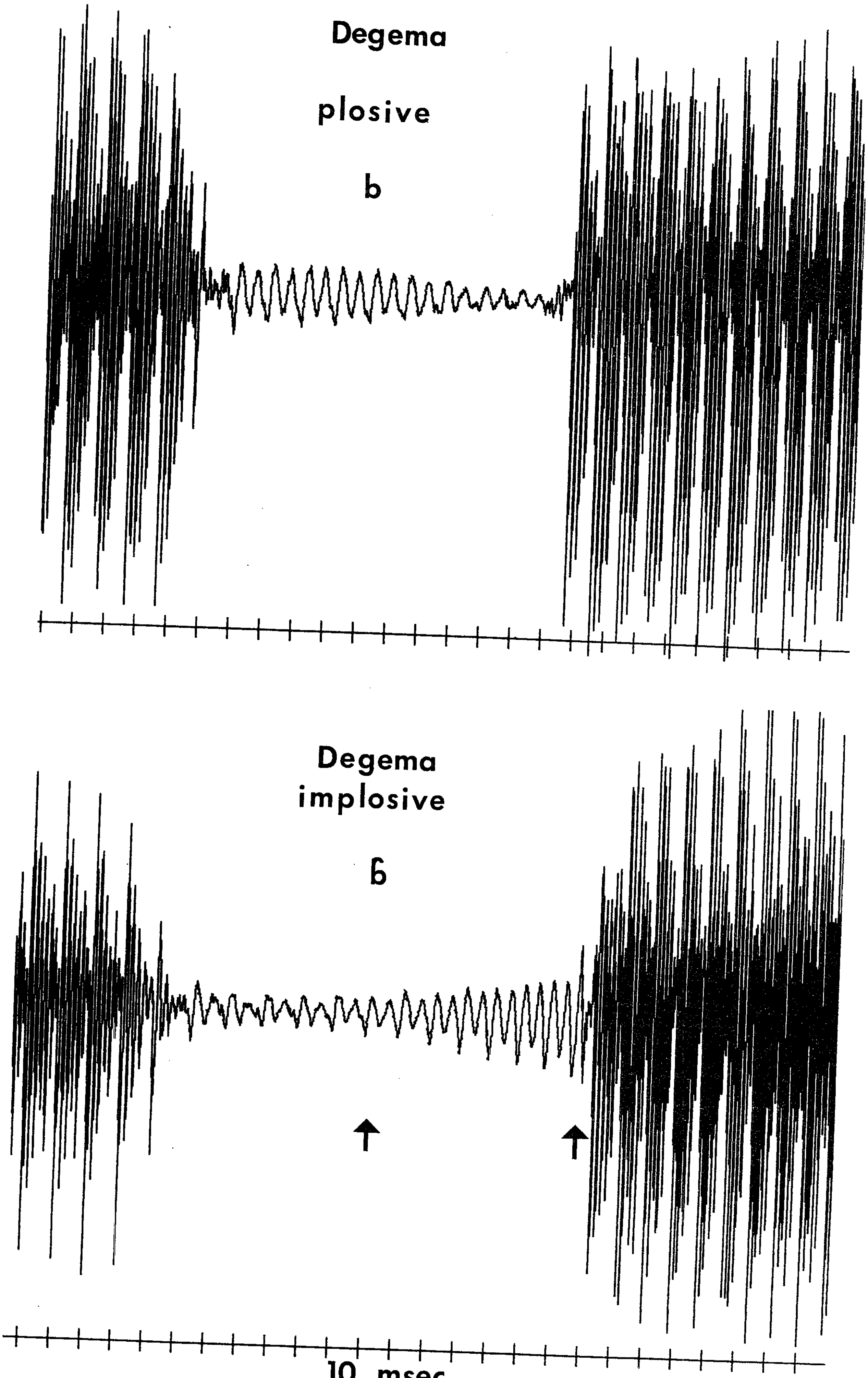


FIGURE 1

In the plosive, the amplitude of the vocal cord vibrations decreases gradually throughout the closure. As the supraglottal pressure increases, airflow through the glottis decreases, and eventually voicing dies out. Implosives, on the other hand, typically have either a gradual increase in the amplitude of the voicing - as in the figure - or, in other cases, a level, fairly large amplitude. This is due to an increase in the size of the vocal tract, as the larynx is lowered, and, also, the tongue body behind the place of articulation is typically lowered. The enlarged vocal tract volume keeps the supraglottal pressure from increasing, and voicing can be maintained at the same amplitude or at an increasing amplitude throughout the closure.

Note, too, that the first part of the implosive sound wave contains a certain amount of higher frequencies.

To describe some of the differences between the implosives in the selected languages, voicing amplitude and closure duration were measured, and characteristics of the waveform were examined. Peak-to-peak amplitude of voicing was measured in the middle and at the end of the closure at points marked by arrows on the figure. A ratio of these two amplitudes was calculated by dividing the final amplitude by the medial one. An implosive with level or increasing amplitude will have a ratio of 1 or more. In a regular plosive this ratio will be much less than 1. This measurement provides an indirect indication of the amount of cavity expansion in a voiced implosive.

Closure duration was also measured between points on the waveform display which, because of sharp changes in amplitude, were taken to be the offset and onset of the surrounding vowels. Thirdly, periodicity of the waveform and spectra at selected points were studied as an indication of phonation type.

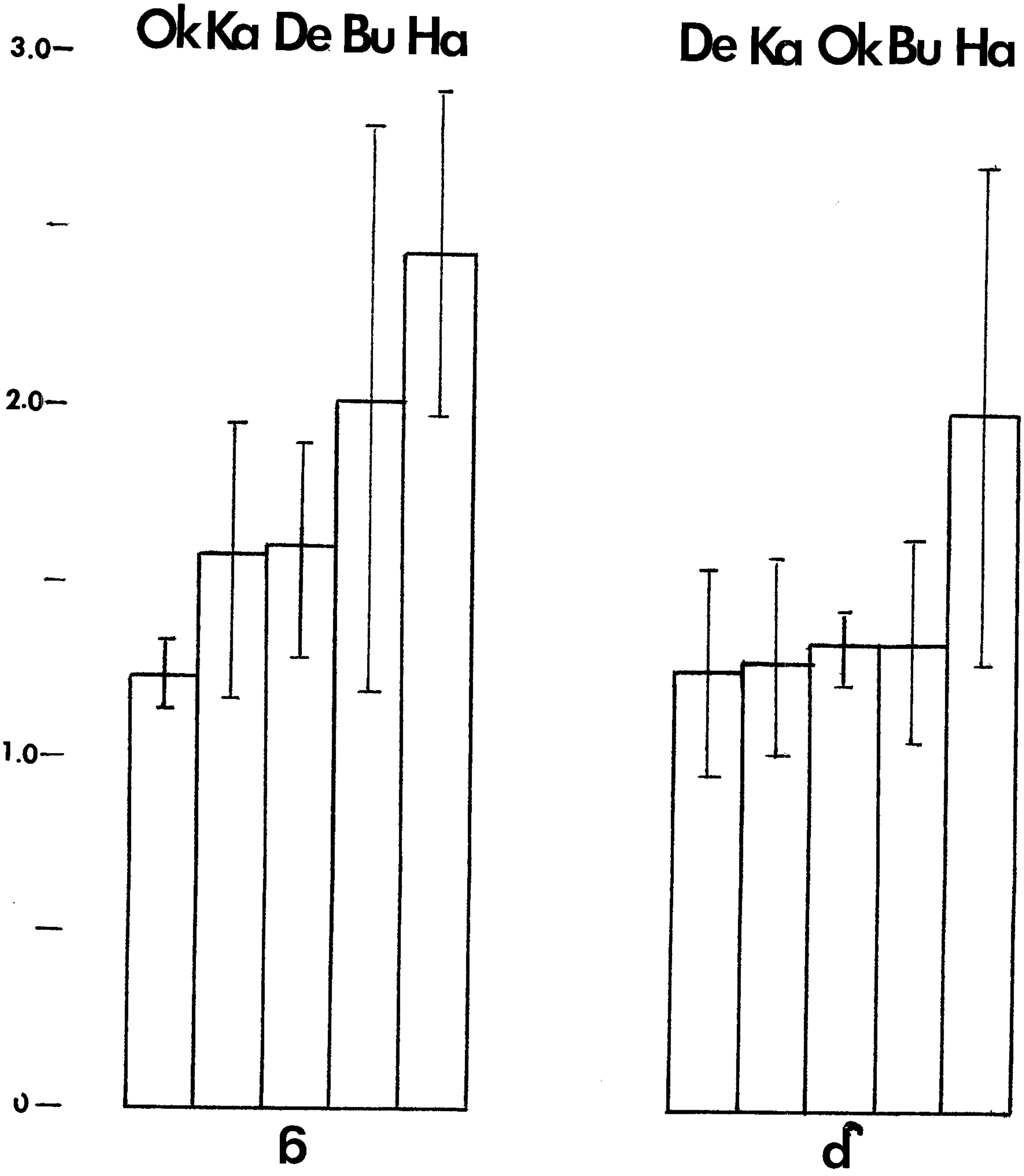
Five out of the fourteen Hausa speakers produced implosives with voiceless closures, so these were excluded in the measurements of amplitude ratios and closure duration.

The amplitude ratios and closure durations were averaged for all the speakers of each language. T-tests were used to assess the significance levels of the differences between the languages.

Figure 2 shows histograms of the means of the amplitude ratios for the bilabial and alveolar implosives in each of the five languages. The histograms have been plotted in an order of increasing ratios, that is in an order of increasing amount of cavity expansion. The bars indicate one standard deviation above and below the mean.

For both implosives Hausa has the highest ratio, indicating that it has the highest degree of implosion.

For the bilabial implosive the amplitude ratios are very significantly different between Hausa, on the one hand, and Degema, Kalabari, and Okrika on the other. Okrika is also significantly different from Bumo and Degema. For the alveolar implosive this measurement is significantly different only between Hausa on the one hand, and Kalabari and Degema on the other.



end / middle amplitude ratio

FIGURE 2

Figure 3 shows histograms of mean closure durations in the five languages in increasing order. Note that this measurement orders the languages in the same way for both bilabial and alveolar implosives. Hausa has significantly shorter closure durations than the other languages.

The Hausa implosives from the nine speakers are thus produced with both a relatively short closure duration and an amplitude ratio indicating a higher degree of implosion than occurs in the other languages.

However it is not true that a shorter closure always implies a greater degree of cavity expansion. For the languages apart from Hausa there is a tendency towards the opposite relationship between amplitude ratio and closure duration. Particularly for the bilabials, shorter closure durations are associated with lower amplitude ratios. Thus closure duration and degree of cavity expansion are independently variable phonetic parameters of voiced implosives.

In addition phonation types were studied. Figure 4 shows a waveform of the bilabial implosive /ɓ/ in Bumo. It is typical of the voiced implosives in the four Niger-Congo languages. The first part of the closure displays a considerable amount of high frequency energy. Below the waveform there is a spectrum of the first 50 milliseconds of the closure, centered at the arrow, showing a clear formant structure. The high frequency energy in the waveform is thus an indication of the upper formants. This is probably due to the vocal cords vibrating with a relatively sharp closure while they are being held tightly together in the descending larynx, and this results in cavity resonances. The first part of the closure in these languages is thus typically produced with a form of laryngealization.

Hausa differs considerably from the Niger-Congo languages in the waveform pattern during the closure (cf. Ladefoged 1964). There is also considerable individual variation here. Five out of the fourteen speakers produce a voiceless beginning of the closure, presumably from a glottal closure as the larynx descends. One speaker has an implosive just like those in the Niger-Congo languages. The eight remaining speakers produce an implosive as seen in figure 5. The closure displays highly aperiodic vibrations. The spectrum shows no clear formant structure but there is a peak around 3500 Hz. This peak cannot be due to cavity resonance from sharp closures in the vocal cord vibrations. If it were, the lower formants would be apparent as well as one at this high frequency. The peak is possibly instead due to noise from INcomplete closures in the vocal cord vibrations, and possibly also to noise generated by perturbations of the vocal tract walls as the larynx descends. The incomplete closures would also explain why the Hausa implosives last a shorter time, as there will be leakage of air through the descending glottis. The Hausa implosives are thus produced with aperiodic, inefficiently closing vocal cord vibrations. This is usually also labelled "laryngealization". Apparently, what we label laryngealization may involve several different mechanisms.

The voiced plosive [b] in Hausa has periodic voicing vibrations during the closure phase, so the voicelessness or aperiodicity in Hausa may serve to keep the implosives apart from the voiced plosives.

Summarizing the data on implosives, it is apparent that there are several independent phonetic parameters distinguishing segments that have been called implosives in different languages. Some languages use one combination of values of these parameters and others another combination. It is also evident that there

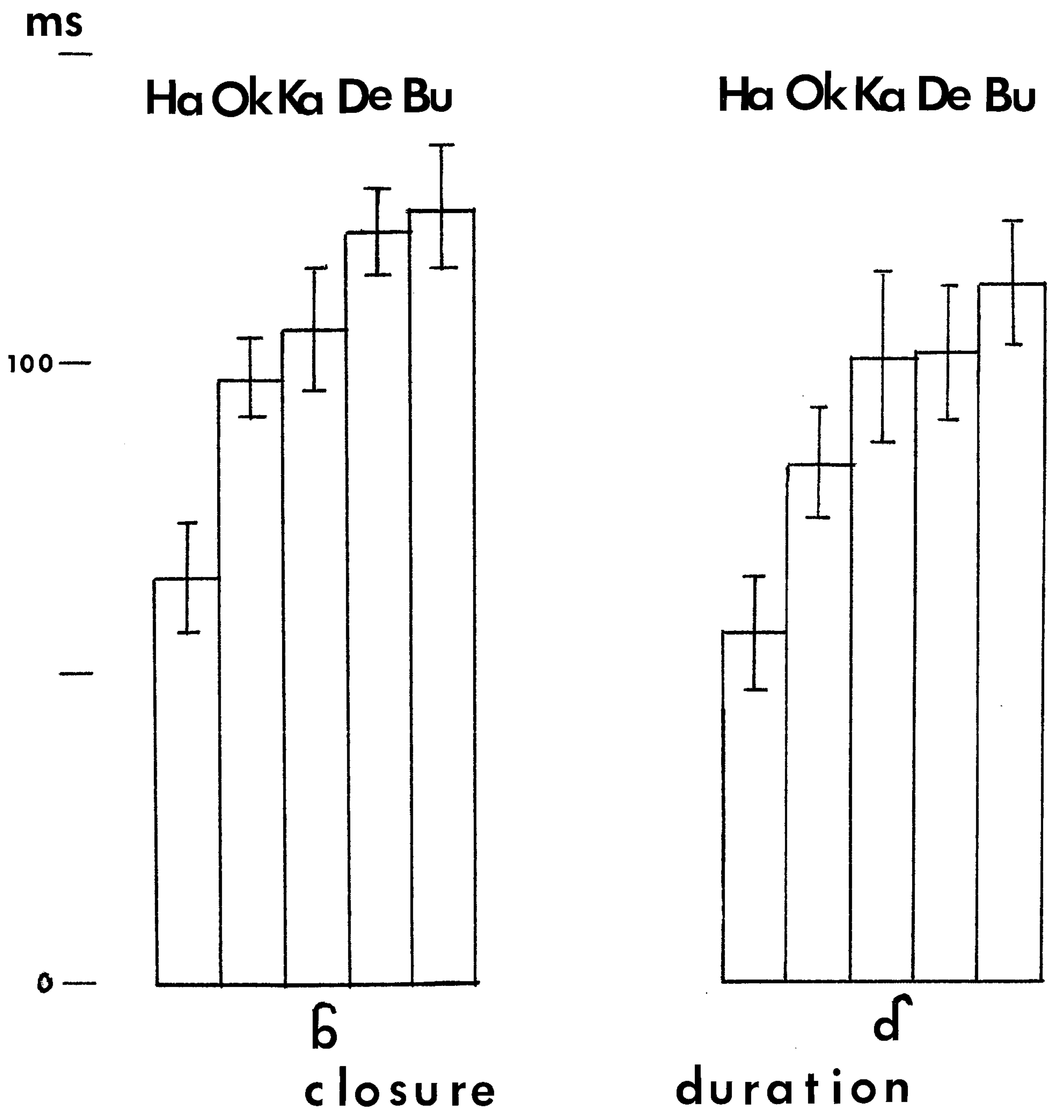


FIGURE 3

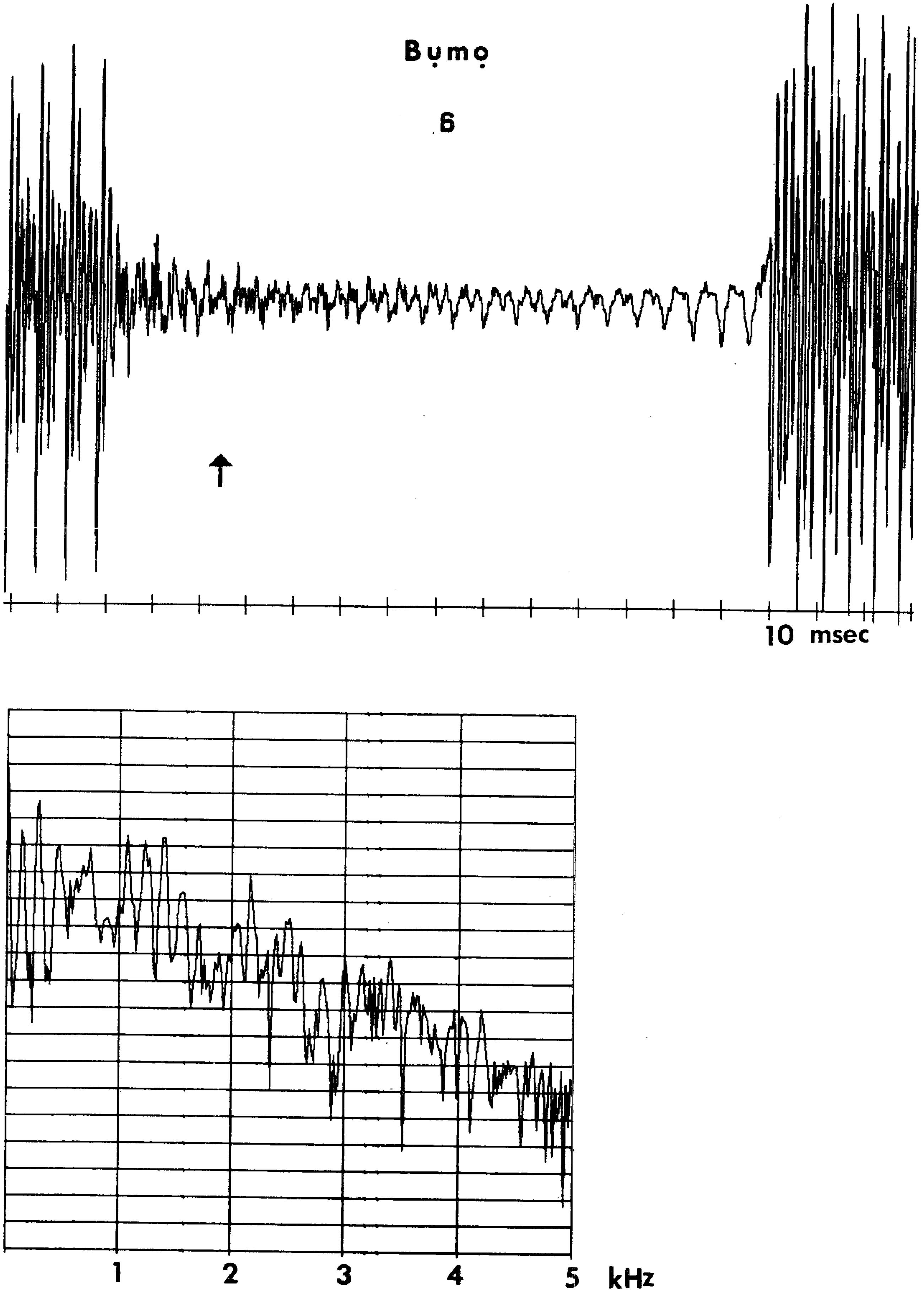


FIGURE 4

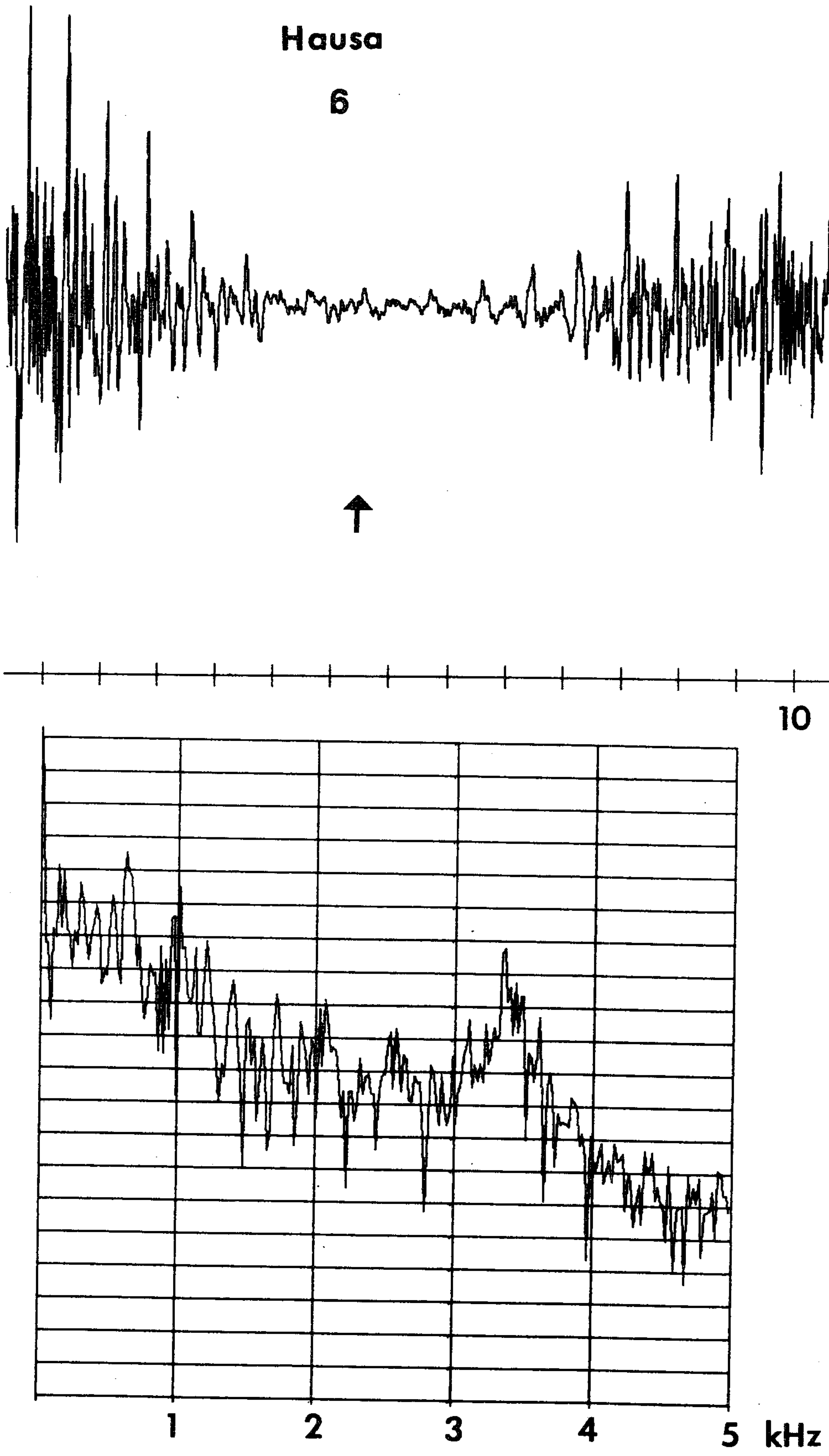


FIGURE 5

is considerable speaker to speaker variation between implosives in languages, and that languages may differ in the way that they maintain distinction between implosives and the corresponding voiced plosives.

Ejectives.

Ejectives are more common in the languages of the world than implosives, about 18% of the world's languages have ejectives. These stops are produced with a closed glottis moving rapidly upwards, followed by glottal and oral releases. My data on ejectives consist of velar ejectives from twelve Hausa speakers and nine speakers of Navaho, an Athabascan language spoken in the Southwestern United States. These data were also analysed from displays of the waveform.

For each ejective, the total duration - that is, the sum of the closure duration and the VOT - was measured and a ratio of the closure duration to the VOT was calculated. A qualitative examination of the waveform was made to reach conclusions about phonation type.

The results show that here too variation between speakers was prominent in Hausa. Four out of the twelve Hausa speakers realise the ejective /k'/-phoneme, not as an ejective but as an unaspirated [k] or as a voied [g]. All nine Navaho speakers have a true ejective [k'].

Figure 6 shows histograms of the means of the total duration and the closure/VOT ratio. Standard deviations are shown by the bars. For both these measures, the differences between Hausa and Navaho are highly significant. The ejective in Navaho has more than twice the total duration of the one in Hausa. This difference is not due to a slower speaking rate in Navaho. The rate of speaking in both languages was measured as number of syllables per second, and the difference was non-significant (both 4.3 - 4.5 syllables per second).

As the difference in the ratio indicates, these languages also differ in relative durations of the different parts of the ejectives. The closure duration in Hausa is about twice that of the VOT part, while the closure duration in Navaho is only 1.5 times as long as the VOT.

In addition, the vowel onset differs considerably in the two languages. This is illustrated in figure 7. In Navaho the glottal release coincides with the vowel onset, so the vowel starts with a sharp, large amplitude. In Hausa, the glottal release occurs together with the oral release, and the vowel begins gradually, with aperiodic vibrations. This aperiodicity of the vowel onset after the ejective could be an important cue for differentiating voiceless plosives and ejectives in Hausa, since the Hausa velar plosive is followed by periodic onset of the vowel.

In conclusion, this study shows that both types of glottalic consonants can vary in a number of ways. Some of this variation is reliably associated with the particular language concerned. From this it follows that contrasts between similar pairs of segments, such as implosives and voiced plosives, may be maintained by different values of particular phonetic parameters in different languages. These facts suggest that the phonetic component of the grammars of languages must be much more specific and detailed than is provided for in most current theories.

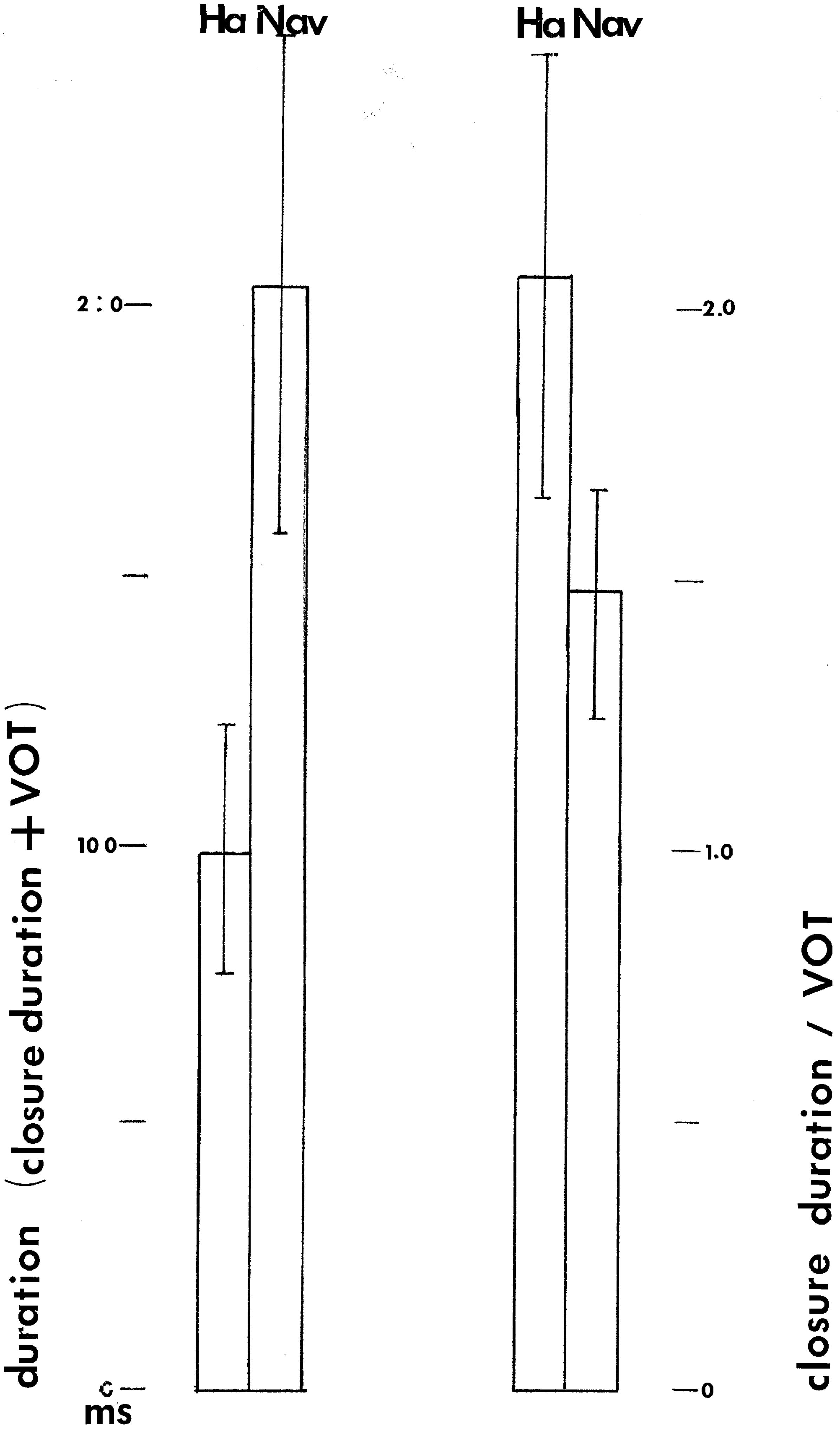

Ha Nav
Ha Nav
2:0
100
c
ms
duration (closure duration + VOT)
2.0
1.0
0
closure duration / VOT

FIGURE 6

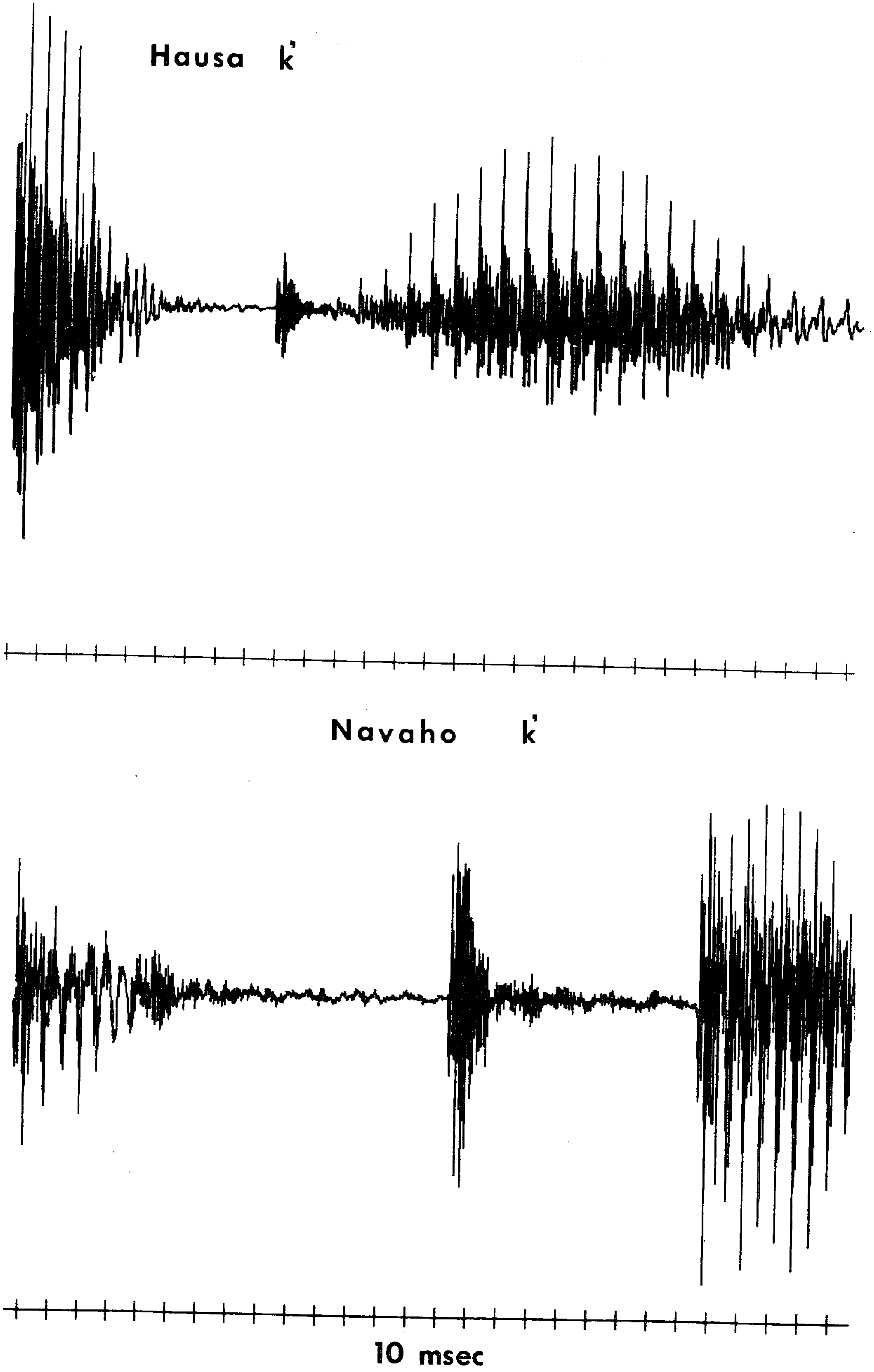


FIGURE 7

References.

Ladefoged, P. (1964) Phonetic differences in West African languages. Cambridge: Cambridge University Press.

Maddieson, I. (1981) UPSID. UCLA Phonological Segment Inventory Database: Data and Index. UCLA Working Papers in Phonetics 51.

Many, many thanks to Peter Ladefoged and Ian Maddieson for all their helpful comments. Thanks are also due to Kay Williamson, C. Jenewari, S.Okpara, and M.Garba for their help in collecting the data.

STRESS EVALUATION AND VOICE LIE DETECTION: A REVIEW

Sandra Ferrari Disner

1. INTRODUCTION

Scientists have long sought to learn whether it is possible to determine from a person's voice whether he or she is under psychological stress. Many different answers have been put forward over the past four decades, and even today there is little consensus among researchers. This diversity of opinion may be attributed, on the one hand, to the many different experimental conditions set up, and on the other hand, to the many different speech characteristics studied.

Unfortunately, there is no "best" way of eliciting stress, nor is there any obvious candidate for the most likely correlate of stress in the voice. Thus, the assembled literature reports a wide variety of responses to a wide variety of experimental stimuli which, taken together, present a rather confusing picture of how stress may affect the voice. But if each result is viewed in the context of the particular type and degree of stress that produced it, the picture becomes a good deal clearer. This review will attempt to place each study in such a context.

2. TYPES OF STRESS

In response to a physiologically or psychologically stressful situation, the body's pituitary-adrenal system is activated to produce hormones to protect the body from further damage (Levine 1971). These in turn may cause changes in the pulse rate, blood pressure, and respiration, along with other, less apparent changes, all of which have at least the potential of affecting the delicate laryngeal interactions which are responsible for speech. However, it is not clear whether conditions as disparate as bodily injury, drugs, temperature extremes, fear, and time pressure are equally likely to give rise to measurable changes, nor is it clear whether the effects on the voice are the same across conditions. For practicality, the present discussion will be limited to forms of psychological stress. Yet even within this context it is not at all obvious whether, for example, task-related stress (such as is experienced when adding numbers under time pressure) is characterized by the same vocal effects as fear-related stress. Moreover, individual speakers may differ markedly in their response to psychological stress. With such potential variability, the task of detecting stress in the human voice is not a simple one.

If different sorts of psychological stress prove to have different vocal manifestations, it is perhaps too much to ask a single device to recognize them all. This is of particular importance because devices are already being designed to recognize the effects of psychological stress on the human voice. Certainly, researchers involved in developing voice lie detectors are primarily interested in the correlates of deception-related stress. Those who are involved in developing speech recognition devices, on the other hand, may be more interested in other sorts of psychological stress. For example, a speech recognition device used in a potentially hazardous field, such as air traffic or space flight, must be robust enough to recognize fear-related speech. And a speech recognition device used in a field with periods of heavy trade volume, such as check or credit card validation, must be robust enough to recognize time pressure-stressed speech. It may be unnecessary, however, for such specialized devices to be capable of recognizing other types of psychological stress.

3. EXPERIMENTAL CONDITIONS

Let us set aside for the moment the question of whether or not different stressors have different effects on the voice, and turn to a more basic consideration: how to go about collecting a sample of speech which is similar to that found in real-life situations when people are under psychological stress. To narrow the field down to the general topic of "deception", for example, is just a beginning. Deception can take many different forms. Various parameters enter into consideration, such as whether the subject is in a state of arousal, whether he knows the experimenter to have sanctioned the deception, whether he is sufficiently involved in the task to take an interest in its outcome, or, as an extreme case of the latter, whether the outcome might somehow place him in jeopardy. These parameters undoubtedly bear on the overall intensity of the psychological stress -- creating low, medium, or high levels of stress -- and, therefore, on the likelihood of the stress being detected in the subject′s voice.

3.1 Deception

A deception task with no penalty attached, one which is sanctioned -- in fact, requested -- by the experimenter, and which does not particularly arouse the subject creates a low-stress condition. In fact, this is one of the most rigorous tests for voice lie detection, since it has been shown that the less involved a subject is with his lie, the more difficult it is to detect (Gustafson & Orne 1963). There may well be some sort of threshold, a certain level of stress which must be reached in the individual before changes in the voice occur (Barland 1973). Indeed, some of the low-stress experiments reported in the literature seem to verge on such a limit. For example, one common type of low-stress experiment requires the subject to choose a number and then to respond "no" to all the experimenter′s questions regarding the number chosen (Friedhoff et al. 1963; Alpert et al. 1963; Barland 1973; McGlone 1975). Some of the subjects tested by Streeter et al. (1977) were also given a low-stress task: to falsify, in an interview with fellow students, selected answers on topics such as religion, politics, personal future, and values. (The choice of these topics may well have induced a slightly higher level of arousal than the absolutely neutral "numbers" task described above.)

Somewhat more involving is the task described by Kubis (1973): a simulated theft of money by a pair of subjects, one of whom acted as a lookout, the other as thief. An "innocent" subject was also used as a control. In the ensuing interrogation, designed for "yes" or "no" answers, the "guilty" pair was instructed to deny all involvement in the theft. Kubis′ experimental design has been challenged (Dektor 1974b) on the grounds that, though fairly involving, this task is not sufficiently realistic for stress to reach a detectable level in the voice; the use of actual criminal suspects is recommended instead. It should be noted, however, that the source of this particular criticism is a company with a financial interest in the successful recognition of stress under these circumstances.

In the same vein, Barland (1973) sought to increase his subjects' involvement in his "numbers" task by following up with two more trials: one which carried with it a small wager, and another which was preceded by a discussion of the ethics of lying. The results of these additional trials were found not to differ from those of the original, low-stress condition. However, Streeter et al. (1977) did find significant differences between their low-stress group and a more motivated group, to whom the task had been presented as a particularly challenging and creative one.

In a study involving more stressful conditions, Scherer and his colleagues (Ekman, Friesen & Scherer 1976; Scherer 1977) showed a particularly unpleasant medical film to a group of student nurses and asked them to deceive the interviewer about their feelings. The task was made particularly ego-involving by pointing out to the subjects that their nursing careers would require precisely this sort of ability to hide their feelings from their patients. Scherer (1978) acknowledges that it is impossible to separate the arousal due to aversive stimulation from that due to the deception itself, but he makes the valid point that this merging of stimuli is what often happens in real-life situations in which negative affect has to be concealed.

Also quite stressful was Barland's (1973) second study involving fourteen criminal suspects, whose voices were taped during the course of a polygraph examination. The polygraph results indicated that all had lied when they denied involvement in a crime; significantly, the voices of eight of the suspects also showed elevated stress levels at relevant points in the questioning. (Tests on the remaining six subjects were merely inconclusive.) These results point to the reliability of the voice as an indicator of short-term psychological stress, but Barland avoids equating this stress directly with the deception. While the polygraph results were completely confirmed in six of the cases, and while there was no particular cause to doubt the remaining eight, Barland stresses that "the issue of whether the polygraph examiners' decisions were all correct" -- i.e., the very fact of deception -- "is not important. [...] The question explored in this study is the extent to which autonomic changes recorded by the polygraph will be reflected by changes in the voice" (Barland 1973:9).

Two investigations similar in design to Barland's have been conducted by researchers associated with one of the leading manufacturers of voice lie detectors (Kradz 1972, Dahm 1974). Kradz tested a larger sample of criminal suspects (n=43) than did Barland, and he claims a 100% correlation between changes detected in the voice by the Dektor Psychological Stress Evaluator and "independent" corroboration. In making this evaluation, however, Kradz steps beyond Barland's tentative criteria; he accepts independent police investigations as proof of whether deception had actually occurred. In fact, only 13 of the 16 "guilty" suspects admitted guilt, while the guilt of the remaining three (and the innocence of 21 of those deemed not guilty) was determined by independent investigations.

The Dahm (1974) study, based on several thousand simultaneous polygraph and voice lie detection studies, found a 99.85% correlation between the two modalities. It also indicated that voice lie detection had invariably been accurate, although only 20% of the cases were independently corroborated.

Hiller (1975) views all of the simultaneous polygraph/voice studies as "tainted from the presence of too many variables in the study" (p. 310). She notes that the same examiner would be likely to have access to charts from both machines, and might well have a vested interest in obtaining identical results from both. She further states that the uncorroborated figures ought to be disregarded, since "the mere fact that the examiners did not later learn of their mistakes did not mean that mistakes were not made" (ibid.). Moreover, she notes that the nature of the corroboration, where reported, is often unexplained.

3.2 Other induced stress

Studies of the effects of psychological stress on the voice have not been exclusively concerned with deception. For example, Hecker et al. (1967) did a detailed study of task-induced stress. Their subjects were assigned increasingly stressful tasks, all of which involved reading a set of meter displays. The first task was simply to read the meter display aloud. The second was to perform a somewhat complicated addition of the readings from six different meters, and to report the sum aloud. The third task introduced time pressure as well, by limiting the amount of time available to do the addition. This third task was tailored to each individual subject: in order to create a stressful situation for the subject without his ever losing interest in the task, the experimenters progressively decreased (or if necessary increased) the display duration, taking into consideration such factors as lessened accuracy, increased anxiety, and requests for more time. By varying the difficulty of the task in this manner, the experimenter was able to accommodate subjects with widely different tolerances for stress.

A more moderate level of stress was induced by the Stroop color-word conflict test in a study by Borgen and Goodman (1976). Once in this condition, half of the subjects received doses of an anti-anxiety drug. The resultant speech was evaluated with a Dektor Psychological Stress Evaluator. For validation, the results were compared with four physiological measures known to fluctuate with the subject's state of arousal (heart rate, blood pressure, skin potential, and forearm blood flow).

Other studies have examined the effects of physiological stress on the voice. Such stressors, which are somewhat beyond the scope of the present review, include periods of sensory isolation in a darkened, soundproof room (Rubenstein 1966), exposure to an unpleasant smell (ammonium chloride) (Ostwald 1963), and electric shocks of increasing intensity, randomly administered during the reading of a neutral passage (McGlone 1975).

3.3 Real-life stress situations

Rather than set up stressful conditions in the laboratory, a number of researchers have chosen to take advantage of stressful conditions in real life. Highly involving or even life-imperiling situations, in which the stress level is higher than normally experienced, stand a greater likelihood of evincing measurable changes in the human voice. But one should be cautious in generalizing the results of such studies; a stress-identification procedure gauged on such samples of speech may well have too high a threshold for practical use in everyday situations.

One group of experiments centers on the stress experienced by test pilots (Williams & Stevens 1969; Simonov & Frolov 1973; Kuroda et al. 1976), astronauts (Popov et al. 1971; Simonov & Frolov 1973, 1977; Older & Jenney 1976), and even parachute jumpers (Simonov, Frolov & Taubkin 1975) under stressful conditions, including equipment malfunctions and extra-vehicular activity. Especially in the case of prolonged space flight, it was necessary to differentiate emotional stress from physiological stress and fatigue; the use of recognition theory proved particularly valuable in this regard (Simonov & Frolov 1977).

Streeter et al. (1978) studied the effects of a unique and highly stressful incident -- the 1977 New York blackout -- on the voice of the system operator in charge and on that of his immediate superior. Wiegele (1978) studied the effects of international crisis situations, such as the Cuban missile crisis and the capture of the Pueblo, on the recorded voices of three U.S. presidents. It should be pointed out that these latter studies, while based on undeniably high-stress situations, nevertheless lack the element of physical peril that characterizes many of the aviation studies.

Less highly stressed, too, are the real-life studies of stage fright (Brenner 1974) and of the anxiety experienced by students during final exam periods (Brockway 1977). As always, the stress level should be kept in mind when the results of different studies are being compared.

3.4 Simulation studies

As an alternative to taking advantage of known stressful situations, or to creating parallel conditions in the laboratory, a number of researchers have simply asked their subjects to act out a particularly stressful situation. It is difficult to judge the degree of stress elicited in this manner, however. In some studies the subjects′ personal involvement is quite low, as when the voice characteristics associated with stress are merely overlaid on some neutral subject matter. Subjects have, for example, been asked to read a series of numbers, or a semantically neutral sentence, or even a single word, in a variety of emotional tones (Steer 1974; Huttar 1967; Lieberman 1961; Lieberman & Michaels 1962; Bluhme 1971; Osser 1964). It should be noted that a number of these "emotional" tones, such as pain, grief, joy, and embarrassment, are not comparable to the sort of psychological stress we have been discussing up to this point, and should therefore be disregarded. A similar, but somewhat more involving task is the reading of dramatic passages (Lynch 1934; Williams & Stevens 1972) or the detailed acting out of a robbery, as described above (Kubis 1973).

Another variable is the subjects′ level of dramatic training. The studies cited above rely on naive subjects; however, a number of others (Fairbanks & Pronovost 1939; Fairbanks & Hoaglin 1941; Williams & Stevens 1972; Scherer et al. 1972; Simonov & Frolov 1973; Simonov, Frolov & Taubkin 1975) utilize trained actors, who are "presumably able to portray clear and unambiguous emotions" (Williams & Stevens 1972:1238). Williams and Stevens′ study is particularly germane to this discussion, since they not only measured the acoustic characteristics of a trained actor′s voice during a dramatic reading, but also independently tested the validity of their experimental method. The latter was done by comparing a recording of the historic, highly emotional news broadcast of the crash of the Hindenburg with a trained actor's simulation of that event. The authors found the actor's simulation to be "not inconsistent" with the original recording, though the similarities between the two are hardly striking.

Those who are involved in developing stress-recognition devices should perhaps be cautioned against relying too heavily on simulation studies. In the absence of the complex physiological changes which are brought on by genuine stress, actors -- whether trained or naive -- must rely to some extent on a set of conventions to convey their message. These conventions include particular speech patterns which have come to be associated with various stresses by the public at large, even though these patterns may not be uniquely or even reliably associated with stress in real life. Thus, just as a cowboy with a black hat is recognized as a villain by theatre audiences, certain speech patterns may be recognized as denoting a state of stress, even though neither is quite realistic. The nature of such dramatic conventions certainly merits further study.

4. VOCAL INDICATORS OF STRESS

Once having obtained an appropriate sample of speech, the researcher must decide how and where to look for indications of stress. Not surprisingly, there is little more agreement on the likeliest indicator of stress in the voice than there is on the best way of obtaining a sample of stressed speech. Some studies focus on the fundamental frequency, others on amplitude, rate, or various combinations of acoustic or articulatory characteristics. A sizable number compare the output of one of the many commercial stress-analysis devices to known patterns of stress in the speech sample.

4.1 Prosodics

4.1.1 Fundamental frequency -- Some of the earliest investigations focused on changes in the average fundamental frequency of the voice. Fairbanks & Pronovost (1939) measured the means and ranges of F0 and found these to differ from one (simulated) emotion to another; of the emotions examined, fear was found to have the highest mean F0 and the widest F0 range. Lynch (1934) obtained quite similar results.

In several more recent simulation studies, the F0 contour, rather than its mean value, appears to be reliably associated with emotional stress. Williams & Stevens (1972) found that "the aspect of the speech signal that appears to provide the clearest indication of the emotional state of a talker is the contour of F0 vs. time" (1972:1249). Different emotions were found to produce very different contours, and some, notably fear, were in fact found to have no clear and consistent correlates. Greenberg (1969) also observed that the F0 contour is the primary indicator of emotional attitudes, with rate and voice quality (e.g. creak, rasp) playing a somewhat lesser role.

Proceeding from simulated to real emotions, Williams and Stevens (1969) found that the mean fundamental frequency of test pilots' voices was consistently higher during serious flight difficulties than it was during routine operations. Simonov, Frolov and Taubkin (1975) reached a similar conclusion, based on the voices of amateur parachute jumpers.

With regard to deception, Streeter et al. (1977) found that their subjects' mean F0 was significantly higher ($p<.05$) when lying than when telling the truth; moreover, the magnitude of this difference was marginally greater when the subjects were also in a mild state of arousal (i.e., when told that their performances would be evaluated by a psychiatrist). Yet in spite of this correlation, it was found that listeners will not ordinarily rely on F0 differences as a cue to deception. In a companion study to the above, a panel of listeners were asked to rate the truthfulness of the subjects' responses; their judgments were reliably correlated with changes in F0 only when other cues were unavailable (i.e., when the semantic content had been filtered out).

Ekman, Friesen & Scherer (1976) and Scherer (1977) also observed that the mean F0 of the student nurses in their study was significantly higher ($p<.01$) when lying than when telling the truth. However, the disparity between the two conditions quite likely was heightened by the aversive stimulation which was present only during deception. Thus, all that can be said with certainty is that when the subject is concealing negative affect, the fundamental frequency is reliably increased. Streeter et al. (1977) point out that the nurses in the Scherer study were never asked to tell the truth about their unpleasant experience; such a cross condition would have made it possible to partial out the effects of deception alone.

4.1.2 Amplitude -- Friedhoff et al. (1963) and Alpert et al. (1963) examined the effects of stress on amplitude alone. They found that the mild stress brought on by a simple deception task was reliably marked by a change in amplitude. However, Friedhoff et al. noted that the direction of this change varied from speaker to speaker: some gave consistently softer responses when lying, others consistently louder responses.

4.1.3 Voice quality -- Voice quality characteristics, such as rasp, creak, or breathiness, have been found to correspond fairly well (though not as well as pitch) with a range of emotions produced in three different simulations studies (Osser 1964; Greenberg 1969; Bolinger 1978).

4.1.4 Survey studies of prosodic effects -- Hecker et al. (1967) looked at a broad range of acoustic features, including mean F0 and F0 contour, amplitude, rate, the shape and regularity of the glottal pulse, and spectral characteristics. While most of these were influenced to a degree by stress, no single feature could be taken as a reliable indicator of stress for all the subjects. For example, stress was indicated in some subjects by an increase in the mean F0, and in others by a decrease in the mean F0. (Those with the most markedly lowered F0 also tended to have lowered amplitude under stress, perhaps due to reduced lung pressure.) A number of subjects displayed a characteristic F0 contour, beginning lower than average and then rising sharply. Some subjects, but not all, showed a reduction in high-frequency spectral energy. Most subjects also exhibited a degree of voicing irregularity under stress; some even exhibited this irregularity when relaxed. Significantly, the manifestations of stress did not merely vary from subject to subject; there were within-subject differences as well. Some individuals' voices showed few, if any, consistent acoustic indications of stress.

Streeter et al. (1978) examined the speech rate, utterance length, fundamental frequency, and amplitude of the recorded voices of two technicians involved in the 1977 New York blackout. The results were surprisingly dissimilar between the two. The voice of the system operator had progressively decreasing mean F0 and maximum amplitude, while the voice of his immediate superior had progressively increasing values along these parameters.

Huttar (1967) also looked at a range of acoustic and perceptual features. He found the measured fundamental frequency and, to a lesser extent, also the perceived pitch, rate, and amplitude, to be correlated with the emotions expressed by the participants in his simulation studies. But the results were not uniform across the range of emotions studied. For example, the expression of anger was directly related to F0 in one experiment, and inversely related to it in another. And the expression of fear gave rise to ambivalent intensity characteristics.

Lieberman and Michaels (1962) designed a forced-judgment test to determine the effects of (simulated) emotions on the voice. They presented their subjects with various prosodic components (F0 only; amplitude only; F0 and amplitude; F0 (smoothed)) of sentences read in emotional tones, and asked them to identify the emotions being expressed. The subjects correctly identified the emotions in about half the sentences when given the amplitude and F0 together, and slightly less when given these components singly; in comparison, the unfiltered speech yielded an identification score of 85%. Evidently, none of these prosodic features can be considered the unique acoustic correlate of emotional stress. Lieberman and Michaels also noted that the ranking of these features varied from emotion to emotion. Each different emotional state had a slightly different acoustic manifestation; for example, fear was better predicted by the overall amplitude.

4.2 Temporal aspects of speech

Several researchers have investigated the effects of stress on the rate or continuity of speech. Fairbanks and Hoaglin (1941) ranked the speech of actors simulating a variety of emotions along the parameter of speech rate; fear was characterized by the most rapid speech, grief by the slowest, and anger in between. Cook (1969) chose instead to induce stress in his subjects, and obtained rather different results. In the course of a series of emotionally involving interviews regarding work and sex, his subjects showed no clear effects of stress on the rate of speech. Cook concluded that "some people react to transient anxiety by speaking more slowly, while others react by speaking more quickly."

The few studies on speech disruption present conflicting results. Mahl (1956) and Kasl (1957) found the ratio of filled (with hesitation noise) pauses in speech to correlate significantly with the subjects′ feelings of anger. However, in a subsequent study, Feldstein and Jaffe (1962) found neither filled nor unfilled pauses to be correlated with the results of an anger-provoking interview.

4.3 Individual differences

There have been ample indications of the variability of speech patterns under stress in the studies reviewed thus far. It has been reported that some individuals speak more loudly, others more softly; some raise the pitch of their voice, others lower it; some speak more rapidly, others more slowly, and so on. There is some evidence, moreover, that these differences are related to aspects of the speaker′s personality and cultural background (Scherer 1979) and even sex (Steer 1974). One might conclude, with Lieberman (1967:122) that "all we can say is that emotion is marked by a departure from the normal speaking habits of the individual." Yet even this statement is too specific, for some speakers show little consistency from utterance to utterance. Still, listeners are usually able to detect stress in speech, and even to differentiate various emotions. Are they able to do so in the case of these latter speakers as well? It would be informative to find out how well correlated listeners′ judgments are with the consistency of the acoustic cues presented to them.

Individual differences tend to be reported less often in the simulation studies outlined here than in the studies based on genuinely stressful situations. It is at least possible that these studies tap different sources; the rather consistent correlates of simulation-study stress may well correspond to the sort of dramatic conventions suggested earlier. Indeed, most people simply assume that, in response to stress, speakers raise the pitch of their voice and increase their speech rate, even though studies reveal that a sizable percentage of speakers do just the opposite. The prevalent view must be carefully distinguished from the reality of behavior under stress.

4.4 Voice analysis devices

It has been claimed (Dektor 1974) that one very reliable indicator of stress in the human voice is the suppression of a characteristic tremor of between 8 and 14 Hz. This tremor may be associated with the normal, physiological tremor that is readily detected in outstretched limbs or closed eyelids (Graham 1945; Yap & Boshes 1966). However, it is much more difficult to measure or even to prove the existence of tremor in the small and rapidly-moving muscles of the larynx. McGlone (1975) and Inbar and Eden (1976) have sought evidence of such voice tremor with electromyography. McGlone found no convincing evidence, but Inbar and Eden noted a significant correlation between EMG activity in the vocal tract and the low-frequency response modulated on F3.

Several devices now on the market claim to be able to measure this so-called "micro-tremor". These have been used by a number of researchers to determine whether tremor is indeed reduced under stress, and if so, what specific types of stress may be detected in this manner. Other researchers have extended the application of voice analysis devices to fields as disparate as pharmacology and political science.

In the field of voice lie detection the validity of voice analysis devices is open to some question. As we have seen, different researchers have reached very different conclusions from simultaneous tests of the polygraph and the Dektor Psychological Stress Evaluator (PSE) (Barland 1975; Kubis 1973; Kradz 1972 and Dahm 1974).

In the medical field the outlook appears more promising. Researchers have tested the PSE for possible application in diagnosing psychologically disturbed patients (Smith 1975, 1977; Reeves 1976; Rockwell & Hodgson 1976), in the clinical assessment of anti-anxiety drugs (Borgen & Goodman 1976), and in the evaluation of nursing techniques (Brockway et al. 1976). In most cases the PSE results were found to correspond to the subjects′ state of psychological stress, and to the physiological correlates of such stress. The exceptions appeared in low-stress situations. Reeves and Rockwell and Hodgson thus suggest that the diagnostic use of voice analysis devices may have to be limited to patients showing high levels of anxiety.

Researchers in the field of social psychology have also made use of voice analysis devices. For example, Brenner (1974) used the Dektor PSE to measure the stress associated with stagefright. Not only was voice stress found to be greater before an audience than it was in the control condition, but this stress was found to be proportional to the size of the audience. Brockway (1977) measured stress associated with final exams; the PSE results were in accordance with students′ self-ratings of anxiety, and were resistant to acclimatization over a period of several days.

Wiegele (1976, 1978) tested the efficacy of the PSE as a tool of political analysis. He found evidence of stress in the voices of foreign-policy decision-makers which appeared to be associated with crisis situations. However, Wiegele points out the difficulty of finding adequate controls for such unique occurrences.

Older and Jenney (1976) used the PSE to compare the stress experienced by astronauts under routine and trying flight conditions. While in-flight difficulties would certainly be expected to give rise to an appreciable amount of stress, no significant differences were detected between these conditions. This reported failure of the PSE under optimal, high-stress conditions is surprising, for it is at variance with the results obtained even under less favorable (i.e. lower stress) conditions.

5. CONCLUSIONS

A number of different experiments have been reviewed, and their results summarized. The effects of stress have been noted in the fundamental frequency, amplitude, spectrum, rate, continuity, and characteristic tremor of the voice. Some stress-related changes, such as increased F0, are noted more often than others, and would seem to be likely candidates for a stress-recognition strategy. But there are also indications to the contrary; for example, certain speakers consistently decrease their F0 under stress. In view of these facts, the effects of stress have been described as simply a departure from one′s normal speaking patterns. Yet, as it turns out, not even this broad definition is quite adequate.

Some light is shed on this matter when the different experimental conditions that give rise to stress are examined in detail. Three parameters emerge as likely predictors of the vocal output.

The first is the <u>degree</u> of stress, ranging from mild to intense, which seems to have a bearing on the robustness of vocal effects. It has been suggested that stress may have to reach a certain threshold before its effects are detectable in the voice. This observation is most frequently made in regard to vocal tremor, but it applies to other vocal characteristics as well.

The second parameter is the <u>authenticity</u> of the stress. This is a binary feature, differentiating simulated stress from stress actually experienced by the subject. Simulated stress appears to be characterized by a more reliable and consistent set of vocal changes, which, in turn, are consistently recognized by the listener. Whether this is an elegant model of real-world stress, or merely a convenient dramatic device, remains to be explored.

The third parameter is the <u>type</u> of stress (e.g. deception, fear, frustration). It has been noted by a number of researchers that different types of psychological stress may have quite different vocal characteristics. This fact is of great importance in the development of stress-recognition devices. For example, a voice lie detector need not be attuned to the correlates of task-related stress; these may be unrelated, or even orthogonal, to the correlates of deception. By properly constraining the task of such devices, their reliability may be significantly enhanced.

REFERENCES

Alpert, M., Kurtzberg, M., and Friedhoff, A. 1963. Transient voice changes associated with emotional stimuli. Archives of General Psychiatry 8:362-365.

Apple, W., Streeter, L., and Krauss, R. 1979. Effects of pitch rate on personal attributions. Journal of Personality and Soc. Psychol. 37:715-727.

Barland, Gordon. 1973. Use of voice changes in the detection of deception. Read at the 86th Meeting of the Acoustical Society of America. Abstract published in J. Acoust. Soc. Am. 52 (Supplement 1):63.

Barland, Gordon. 1975. Detection of deception in criminal suspects: a field validation study. Unpublished dissertation, University of Utah.

Blair, D., Glover, W., Greenfield, A., and Roddie, I. 1959. Excitation of cholinergic vaso-dilator nerves to human skeletal muscles during emotional stress. J. Physiol. 148:633-647.

Bluhme, Taka. 1971. L'Identification de différentes attitudes émotionnelles par l'intonation. Travaux de l'Institut de Phonétique de Strasbourg 3:248-260.

Bolinger, Dwight. 1978. Intonation across languages. Universals of Human Language, ed. by Joseph H. Greenberg, 474-516. Stanford: Stanford Univ. Press.

Borgen, Lowell A., and Goodman, Lowell I. 1976. Voice analysis of anxiolytic drug effects: preliminary results. Presented at the American Society of Clinical Pharmacology and Therapeutics. Abstract published in Clin. Pharmacol. Therap. 19:104.

Borland, R.G., Cannings, R., and Nicholson, A.N. 1977. Pitch and formant analysis of the voice in the investigation of stress in pilots. J. Physiol. 270:158-168.

Brenner, M. 1974. Stagefright and Stevens' law. Read at the Spring convention of the Eastern Psychological Association.

Brockway, Barbara. 1977. Situational stress and temporal changes in self report and vocal (PSE) measurements. Read at the Annual meeting of the American Association for the Advancement of Science.

Brockway, B.F., Plummer, O.B., and Lowe B.M. 1976. The effects of two types of nursing reassurance upon patient vocal stress levels, as measured by a new tool, the PSE. Nursing Research 25:440-446.

Brod, J., Fencl, V., Hejl, Z., and Jirka, J. 1959. Circulatory changes underlying blood pressure elevation during acute emotional stress in normotensive and hypertensive subjects. Clin. Sci. 18:269-279.

Cook, M. 1969. Anxiety, speech disturbances, and speech rate. British Journal of Social and Clinical Psychology 8:13-21.

Dahm, Anthony E. 1974. Study of the field use of the Psychological Stress Evaluator. Hearings on the use of polygraphs and similar devices by federal agencies, before the Subcommittee on Foreign Operations and Government Information of the House Committee on Government Operations. 93rd Congress, Second Session 255-267. Washington: U.S. Govt. Printing Office.

Davitz, Joel R. 1964. The Communication of Emotional Meaning. New York: McGraw Hill.

Davitz, Joel R. 1969. The Language of Emotion. New York: Academic Press.

Deane, G. 1961. Human heart rate responses during experimentally induced anxiety. J. Exp. Psych. 61:489-493.

Dektor Counterintelligence and Security, Inc. 1974. The Kubis Report of 1973: an invalid study. Hearings on the use of polygraphs and similar devices by federal agencies, before the Subcommittee on Foreign Operations and Government Information of the House Committee on Government Operations. 93rd Congress, Second Session, 224. Washington: U.S. Govt. Printing Office.

Diehl, C.F., White, R., and Burk, K.W. (1959) Voice quality and anxiety. J. Speech & Hearing Res. 2(3):282-285.

Edwards, A.S. 1949. Involuntary movements following perceived and recalled emotional situations. Journal of General Psychology 41:233-238.

Ekman, Paul, Friesen, Wallace, and Scherer, Klaus. 1976. Body movement and voice pitch in deceptive interaction. Semiotica 16:23-27.

Fairbanks, Grant,and Hoaglin, LeMar. 1941. An Experimental study of the durational characteristics of the voice during the expression of emotion. Speech Monographs 8:85-90.

Fairbanks, G., and Pronovost, W. 1939. An Experimental study of the pitch characteristics of the voice during the expression of emotions. Speech Monographs 6:87-194.

Feldstein, S., and Jaffe, J. 1962. The relationship of speech disruption to the experience of anger. J. Consult. Psychol. 26:505-510.

Fonagy, Ivan. 1976. La Mimique buccale: aspect radiologique de la vive voix. Phonetica 33:31-44.

Friedhoff, A., Alpert, M., and Kurtzberg, M. 1963. An electro-acoustic analysis of the effects of stress on voice. Journal of Neuropsychiatry 5:266-272.

Fuller, Fred. 1973. Results of the investigations of two speech parameters in the detection of emotional stress/tension. Read at the 86th Meeting of the Acoustical Society of America. Abstract published in J. Acoust. Soc. Am. 52 (Supplement 1):63.

Graham, J.D.P. 1945. Static tremor in anxiety states. J. Neurol., Neurosurg. & Psychiat. 8:57-60.

Greenberg, S.Robert. 1969. An Experimental study of certain intonation contrasts in American English. UCLA Working Papers in Phonetics 13. Los Angeles: Dept. of Linguistics, UCLA.

Gustafson, L.A., and Orne, M.T. 1963. Effects of heightened motivation on the detection of deception. J. Appl. Psych. 47: 408-411.

Hecker, M., Stevens, K.N., von Bismarck, G., and Williams, C.E. 1967. The effects of task-induced stress on speech. Air Force Cambridge Research Laboratories Monograph.

Helfrich, H., and Scherer, K. 1977. Experimental assessment of antidepressant drug effects by spectral voice analysis. Read at the 94th Meeting of the Acoustical Society of America. Abstract published in J. Acoust. Soc. Am. 62 (Supp.1):26.

Hiller, Deborah Lewis. 1975. The Psychological Stress Evaluator. Cleveland State Law Review 24:299:340.

Hollien, H., Majewski, W., and Hollien, P. 1974. Perceptual identification of voices under normal, stress, and disguised speaking conditions. J. Acoust. Soc. Am. 56 (Supp.1):53.

Hollien, H., Michel, J., and Doherty, E.T. 1973. A Method for analyzing vocal jitter in sustained phonation. Journal of Phonetics 1: 85-91.

Huber, H.P. 1965. Mikrovibration als Stressindikator bei Neurotikern und Gesunden. Arch. Ges. Physiol. 117:166-187.

Huttar, George. 1967. Some relations between emotions and the prosodic parameters of speech. SCRL Monograph No. 1.

Inbar, G.F., and Eden, G. 1976. Psychological Stress Evaluators: EMG correlation with voice tremor. Biol. Cybernetics 24:165-167.

Kasl, V. 1957. The relationship of speech disruption to experimentally induced states of emotion. Unpublished dissertation, Yale University

Krauss, R.M., Geller, V., and Olson, C.T. 1976. Modalities and cues in the detection of deception. Read at the American Psychological Association, Washington, D.C.

Kradz, Michael. 1972. The Psychological Stress Evaluator: a study. Prepared for the Howard County Police Dept., Maryland. Also in Hearings on the use of polygraphs and similar devices by federal agencies, before the Subcommittee on Foreign Operations and Government Information of the House Committee on Government Operations. 93rd Congress, Second Session 243-254. Washington: U.S. Govt. Printing Office.

Kubis, Joseph F. 1973. Comparison of voice analysis and polygraph as lie detection procedures. Report prepared for U.S.Army Warfare Laboratory Aberdeen Proving Ground, Maryland. Also in Hearings on the use of polygraphs and similar devices by federal agencies, before the Subcommittee on Foreign Operations and Government Information of the House Committee on Government Operations. 93rd Congress, Second Session 503-555. Washington: U.S. Govt. Printing Office.

Kubis, Joseph F. 1973. Physiological and voice indices of stress in a simulated theft experiment. Read at the 86th Meeting of the Acoustical Society of America. Abstract published in J. Acoust. Soc. Am. 52 (Supplement 1):63.

Kuroda, I., Fujiwara, O., Okamura, N., and Utsuki, N. 1976. Method for determining pilot stress through analysis of voice communication. Aviat. Space Environ. Med. 47:528-533.

Ladd, D. Robert. 1980. The Structure of Intonational Meaning: Evidence from English. Bloomington: Indiana University Press.

Levine, Seymour. 1971. Stress and behavior. Sci. Amer. 224(1):26-31.

Lieberman, P. 1961. Perturbations in vocal pitch. J. Acoust. Soc. Am. 33:597-603.

Lieberman, P. 1967. Intonation, Perception, and Language. Cambridge, Mass.: MIT Press.

Lieberman, P., and Michaels, S.B. 1962. Some aspects of fundamental frequency and envelope amplitude as related to the emotional content of speech. J. Acoust. Soc. Am. 34:922-927.

Lynch, J. 1934. A Phonophotographic study of trained and untrained voices reading factual and dramatic material. Archives of Speech 1:9-25.

Mahl, G.F. 1956. Disturbances and silences in the patient′s speech in psychological therapy. J. Abnorm. Soc. Psychol. 53:1-15.

McGlone, R.E. 1975. Tests of the Psychological Stress Evaluator (PSE) as a lie and stress detector. Proceedings of the 1975 Carnahan Conference on crime countermeasures, pp. 83-86.

Older, H.J., and Jenney, L.L. 1976. Psychological stress measurement through voice output analysis. Report prepared for NASA Lyndon B. Johnson Space Center.

Osser, Henry A. 1964. A "Distinctive features" analysis of the vocal communication of emotion. Unpublished dissertation, Cornell Univ.

Ostwald, P.F. 1963. Soundmaking. Springfield, Ill.: Thomas.

Papçun, George. 1973. The Effects of psychological stress on speech: literature survey and background. Read at the 86th Meeting of the Acoustical Society of America. Abstract published in J. Acoust. Soc. Am. 52 (Supplement 1):62-63.

Pike, Kenneth. 1945. The Intonation of American English. Ann Arbor: University of Michigan Press.

Popov, V.A., Simonov, P.V., Tischenko, A.G., Frolov, M.V., and Khachaturyants, L.S. 1966. Analysis of intonational characteristics of speech as a criterion of the emotional state of man under conditions of space flight. Zhurnal Vysshei Nervnoi Deyatelnosti Imeni I. P. Pavlova 16:974-983.

Popov, V.A., Simonov, P.V., Frolov, M.V., and Khachaturyants, L.S. 1971. Chastoti spektr rechi kak pokazatel stepeni i kharaktera emotsionalnogo napryazheniya u cheloveka. Frequency spectrum of speech as an indicator of the degree and nature of emotional stress in man. Zhurnal Vysshei Nervnoi Deyatelnosti Imeni I. P. Pavlova 21:104-109.

Reeves, T.E. 1976. The measurement and treatment of stress through electronic analysis of subaudible voice stress patterns and rational-emotive therapy. Unpublished dissertation, Walden University.

Rice, Berkeley. 1978. The New truth machines. Psych. Today 12(1): 61-78.

Rockwell, D., and Hodgson, M. 1976. Psychological stress evaluator: a validation study. Read at the Annual meeting of the Society of Biological Psychiatry, San Francisco.

Ross, Mark, Duffy, Robert, Cooker,Harry, and Sargeant, Russell. 1973. Contribution of the lower audible frequencies to the recognition of emotions. American Annals of the Deaf 118:37-42.

Rubenstein, Leonard. 1966. Electro-acoustical measurement of vocal responses to limited stress. Behav. Res. and Therapy 4:135:38.

Scherer, Klaus. 1977. The effect of stress on the fundamental frequency of the voice. Read at the 94th Meeting of the Acoustical Society of America. Abstract published in J. Acoust. Soc. Am. 62 (Supp.1):25-26.

Scherer, Klaus. 1978. Nonlinguistic vocal indicators of emotion and psychopathology. Emotions in Personality and Psychopathology, ed. by C.E. Izard, 495-529. New York: Plenum.

Scherer, Klaus. 1979. Personality markers in speech. In Scherer and Giles, eds., Social Markers in Speech. London: Cambridge Univ. Press.

Scherer, Klaus, Koivumaki, Judy,and Rosenthal, Robert. 1972. Minimal cues in the vocal communication of affect: judging emotions from content-masked speech. Journal of Psycholinguistic Research 1:269-285.

Scherer, Klaus, and Oshinsky, James. 1977. Cue utilization in emotion attribution from auditory stimuli. Motivation and Emotion 1:331-346.

Shipp, Thomas, and McGlone, Robert. 1971. Laryngeal dynamics associated with voice frequency change. Journal of Speech and Hearing Research 14:761-768.

Shipp, Thomas, and McGlone, Robert. 1973. Physiologic correlates of acoustic correlates of psychological stress. Read at the 86th Meeting of the Acoustical Society of America. Abstract published in J. Acoust. Soc. Am. 52 (Supplement 1):63.

Simonov, P.V., and Frolov, M.V. 1973. Utilization of human voice for estimation of man´s emotional stress and state of attention. Aerospace Med. 44:256-258.

Simonov, P.V., and Frolov, M.V. 1977. Analysis of the human voice as a method of controlling emotional state: achievements and goals. Aviat. Space Environment. Med. 48:23-25.

Simonov, P.V., Frolov, M.V., and Taubkin, V.L. 1975. Use of the invariant method of speech analysis to discern the emotional state of announcers. Aviat. Space Environ. Med. 46:1014-1016.

Smith, G.A[lan]. 1974. The Measurement of anxiety: a new method by voice analysis. IRCS (Res. Biomed, Technol., Psychiat. and Clin. Psychol.) 2:1707.

Smith, [G.] Alan. 1975. Secret lie detection in the lab. New Scientist 67:476-478.

Smith, G.A[lan]. 1977. Voice analysis for the measurement of anxiety. British Journal of Medical Psychology 50:367-73.

Sogon, Shunya. 1975. A Study of the personality factor which affects the judgment of vocally expressed emotions. Kyoiku Shinrigaku Kenkyu (The Japanese Journal of Psychology) 46:247-54.

Steer, Angela. 1974. Sex differences, extraversion, and neuroticism in relation to speech rate during the expression of emotion. Language and Speech 17:80-86.

Stephens, J.A., and Taylor, A. 1972. Continuous on-line spectral analysis of muscle tremor. Abstract published in Proceedings of the Physiological Society.

Streeter, Lynn, Krauss, Robert, Geller, Valerie, Olson, Christopher, and Apple, Wm. 1977. Pitch changes during attempted deception. Journal of Personality and Social Psychology 35:345-350.

Streeter, Lynn, Krauss, Robert, Apple, Wm., and Macdonald, Nina. 1978. Acoustic consequences and perceptual indicators of stress. Read at 96th Meeting of the Acoustical Society of America. Abstract published in J. Acoust. Soc. Am. 64 (Supplement 1):115.

Trenner, D.J. 1978. With emotion: the question of lie detectors. New Law Journal 128(5860):665-666.

Uldall, E. 1964. Dimensions of meaning in intonation. In Honour of Daniel Jones, ed. by D. Abercrombie, D.B.Fry, et al, 271-279. London: Longmans.

Voronin, L.G., Konovalov, V.F., and Senina, R.Y. 1973. Indicators of emotional stimuli in the EEG spectrum and GSR in humans with normal and impaired memory. Zhurnal Vysshei Nervnoi Deyatelnosti Imeni I. P. Pavlova 23:34-41.

Wendahl, R.W. 1963. Laryngeal analog synthesis of harsh voice quality. Folia Phoniatrica 15:241-250.

Wiegele, Thomas C. 1976. Voice stress analysis: the application of a physiological measurement tecnique to the study of the Cuban missile crisis. Read at the Annual convention of the International Studies Association.

Wiegele, Thomas C. 1978. The Psychophysiology of elite stress in five international crises: a preliminary test of a voice measurement technique. International Studies Quarterly 22:467-511.

Wiggins, S.L., McCranie, M.L., and Bailey, P. 1975. Assessment of voice stress in children. J. Nerv. Ment. Dis. 160:402-408.

Williams, Carl E., and Stevens, Kenneth N. 1969. On determining the emotional state of pilots during flight: an exploratory study. Aerospace Med. 40:1369-1372.

Williams, Carl E., and Stevens, Kenneth N. 1972. Emotions and speech: some acoustical correlates. J. Acoust. Soc. Am. 52:1238-1250.

Wood, C.A. and Michel, J.F. 1969. A Study of jitter in frequency glides. Read at 45th Natl. Convention of the American Speech and Hearing Assn.

Yap, C.B., and Boshes, B. 1966. The Frequency and patten of normal tremor. Electroenceph. & Clin. Neurophysiol. 22:197-203.

LIQUIDS IN CHAGA

Anthony Davey, Lioba Moshi and Ian Maddieson

Chaga is a cluster of divergent dialects spoken in areas surrounding Mount Kilimanjaro in Tanzania. This language has a rather diverse* and extensive group of sounds that belong to the general class of liquids. In the KiVunjo dialect of Chaga, spoken by the second author, there seem to be at least ten sounds in this class which, phonetically speaking, can be fairly easily distinguished. This is an unusually rich variety of such sounds, particularly for a Bantu language where the range of such sounds is normally rather limited. We will show below that KiVunjo Chaga has four phonemic liquids, and discuss the phonetic distinctions between their allophones, and the historical origin of these sounds.

I. Phonological and Phonetic Analysis.

By the conventional criterion of contrast in minimal pairs, Kivunjo Chaga has two rhotic (r-sound) phonemes and two lateral phonemes. The rhotics are an alveolar continuant, usually with friction, and an alveolar trill. In a convenient phonemic/orthographic transcription, these might be written as /zr/ and /r/. The laterals are a dental approximant and an alveolar flap. These might be written as /l/ and /d/ respectively. An example of a near-minimal contrast between all four of these segments is given below:

/ízrâ/	"to smell bad"
/íìrâ/	"to carry"
/íláà/	"to lie down"
/ídà/	"leaf"

In addition to the four phonemes represented in these examples, there is a palatalized alveolar lateral continuant with a restricted distribution. This might be regarded as an allophone of the dental lateral approximant, or as a separate phoneme. More will be said on this below. An example of the sound concerned, which we will write with the digraph /ly/, occurs in the word:

/ílyà/ "to eat"

There are some quite marked allophonic variants of certain of these liquid phonemes. This variation adds to the overall initial impression of a great variety of liquids in Chaga.

The phonetic nature of these consonants was analyzed spectrographically from tokens recorded from the speech of the second author, a female native speaker of Chaga. Selected spectrograms are reproduced as Figures 1-16. The spectrographic analysis, supplemented by auditory and palatographic investigations, is the basis for the more detailed descriptions of these segments which follows.

* This fact has been commented on before: Hinnebusch and Nurse (1981 fn. 7) note that:

> "There are a variety of "r" sounds in Chaga: [R], a retroflexed liquid;
> [r], a trill; [r̲] ([ɾ]), a flap; and [ɹ], an alveolar fricative"

They indicate that no more than three of these are found in any one dialect. Their transcription indicates that they observed only one lateral.

The alveolar continuant /zr/.

The most typical allophone of this phoneme is a voiced apical alveolar (or post-alveolar) nonsibilant fricative [ɹ̲^]. This sound is illustrated in Figures 1 and 2. There is generally a noisy excitation of F_2 (and F_3) and some high frequency noise above 4000 Hz. However, the noise is not high and intense enough for this sound to be like the more common sibilant alveolar fricative [z]. Note that the amount of friction may vary, as it does between the two similar segments on Figure 2 (/ízrézrémà/ "to shake"), where the first occurrence of [ɹ̲^] has more intense friction and weakly defined formants but the second has relatively weak friction and large formant amplitudes. On occasion, the segment can be produced with no friction at all, i.e. as the approximant [ɹ]. No conditioning environment for this allophonic variation was discovered but all allophones of this phoneme are continuants, whereas all allophones of the next phoneme discussed are interrupted.

The voiced alveolar trill /r/.

A phonetic alveolar trill is illustrated on Figures 3-5. There are 3 distinct contacts in the initial trill on Figure 3 (/ríkò/ "hearth"). Two taps can be identified in the other two spectrograms illustrating this sound. The initial trill on Figure 4 (/rùùnâ/ "asthma") has very weak energy, much less than that in Figure 3. In Figure 3, where formant frequencies in the trill can be seen, the onset frequency of F_3 is relatively low - around 2000 Hz - and F_2 is at about 1400 Hz. Both formants rise rapidly toward the values in the vowel /i/ which follows, where F_2 is about 2450 Hz and F_3 is about 2800 Hz. This token thus contains one of the acoustic characteristics which has sometimes been associated with r-sounds, namely a low F_3 (Ladefoged, 1975, Lindau, 1981). However, this characteristic cannot be observed in the other spectrograms which include trills. For example, on Figure 5 (/íìrâ/ "to carry") no sign of a lowered F_3 can be seen in the transitions between the trill and either the preceding or the following vowel. The alveolar location of the contacts in an intervocalic trill was confirmed by palatography.

There is apparent free variation between a simple trill and one that is accompanied by noticeable friction. Examples of fricated trills are provided by Figures 6 and 7. The words in Figures 6 and 4 contain the same root, but only in Figure 6 (/írúúnà/ "to snore") is there noticeable friction, whereas Figure 4 shows a non-fricated initial trill (/rùùnâ/ "asthma"). In Figure 7 (/ímrì/ "root"), there is a fairly broad spectrum of noise (with minima at about 900 and 2600 Hz) visible between the two contacts of the trill. The other sample of this allophone, on Figure 6, has more sustained friction within rather more constricted bands centered on about 1400 and 3300 Hz. On the spectrogram, it is difficult to determine that this is a trill, but the auditory impression made by this token is of a trill with two contacts with accompanying friction noise.

A variant of the trill is a segment with a single tap. This is illustrated on Figure 8 (/írùmíshà/ "to extinguish"). There is some lowering of both F_2 and F_3 in the transition into this consonant from the initial /i/.

On occasion, a voiceless trill was observed. This was only heard as a sporadic variant in words in which a voiced trill (or tap) was normally produced. No tokens were produced during the recording session from which the spectrograms were made.

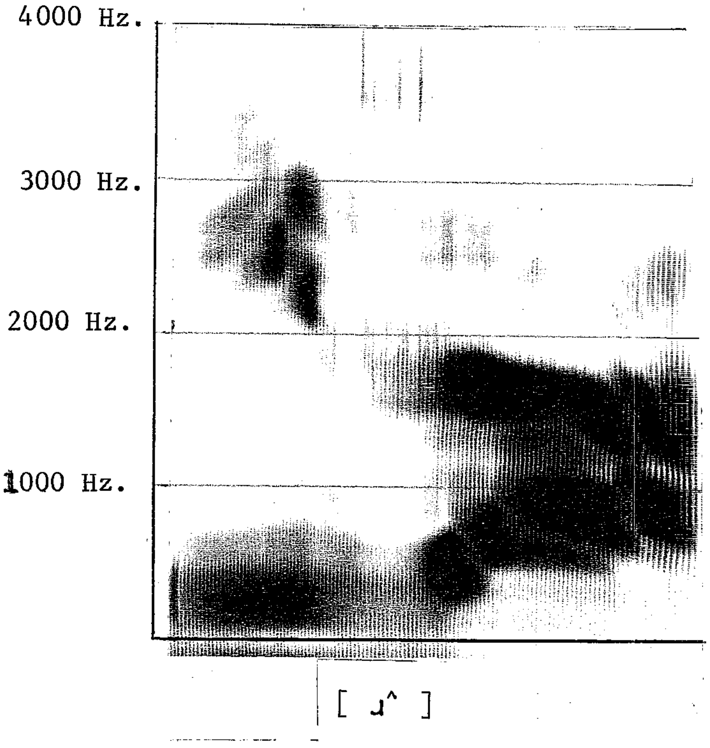


Figure 1.
/izra/ "to smell bad"

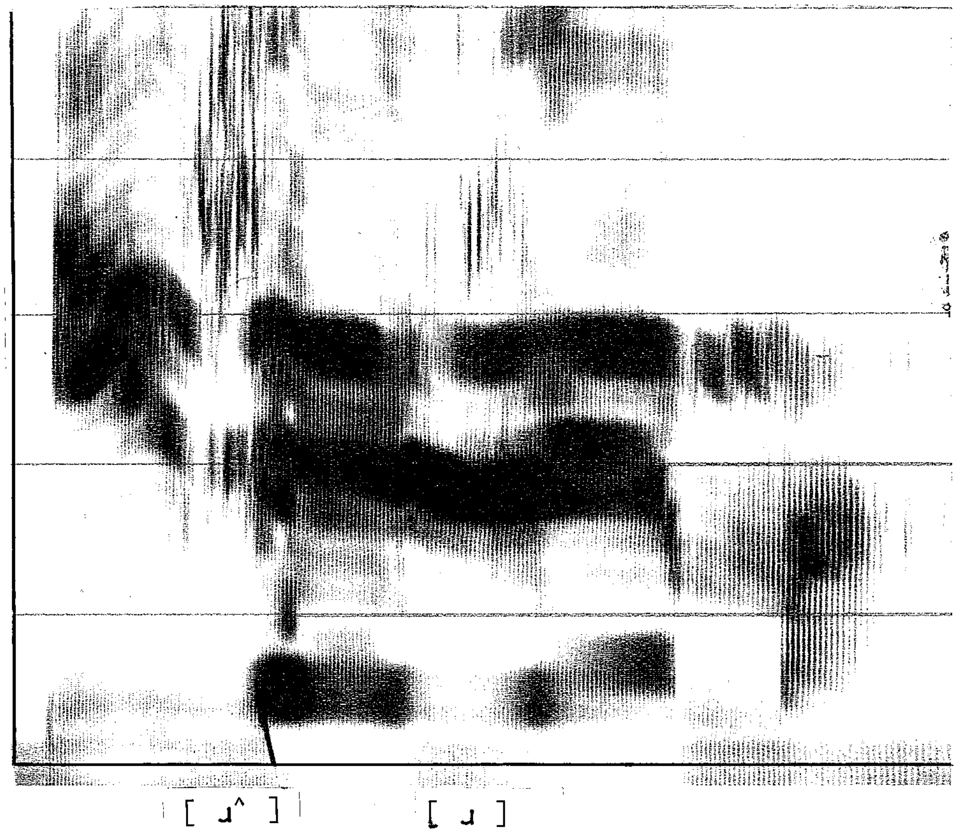


Figure 2.
/izrezrema/ "to shake"

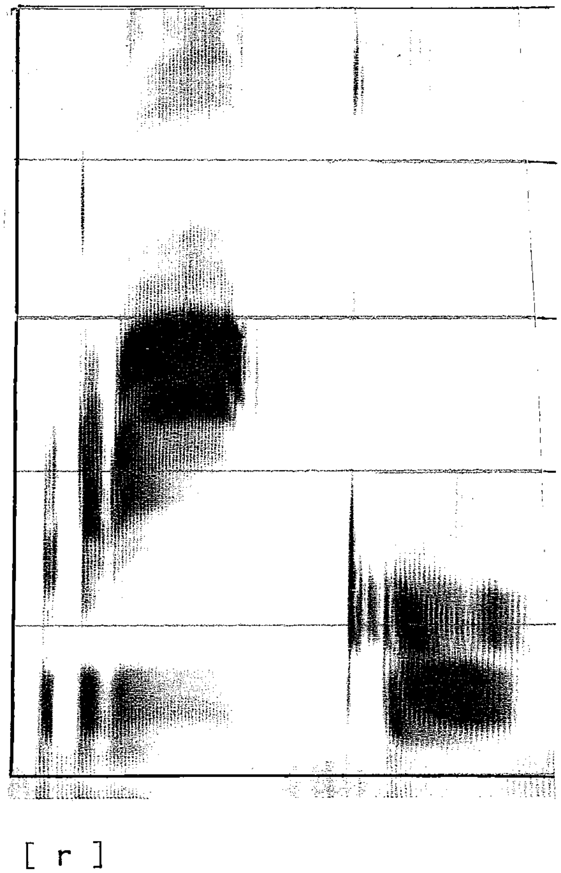


Figure 3.
/ríkò/ "hearth"

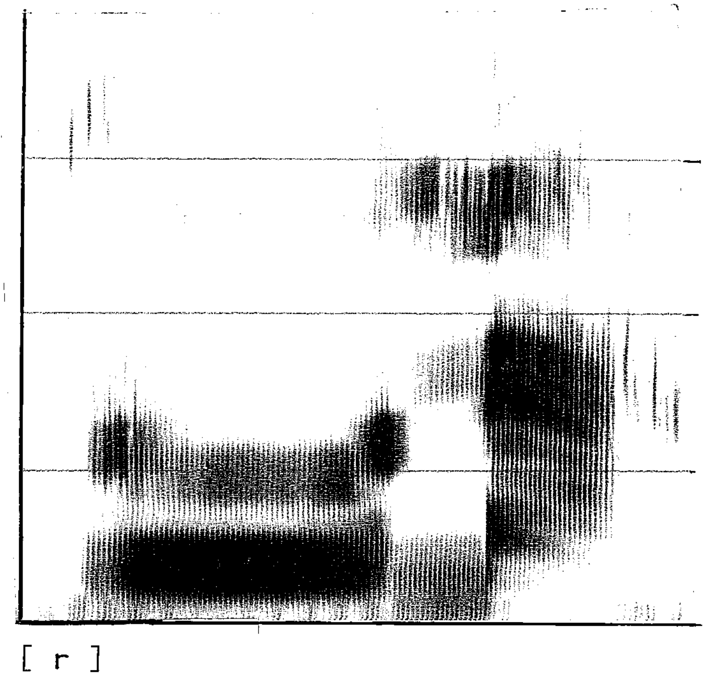


Figure 4.
/rùùnâ/ "asthma"

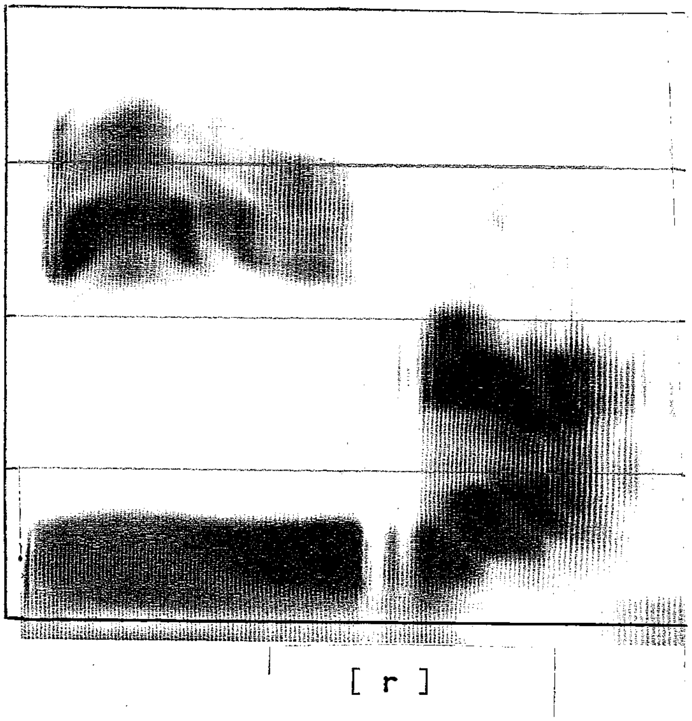


Figure 5.
/íirâ/ "to carry"

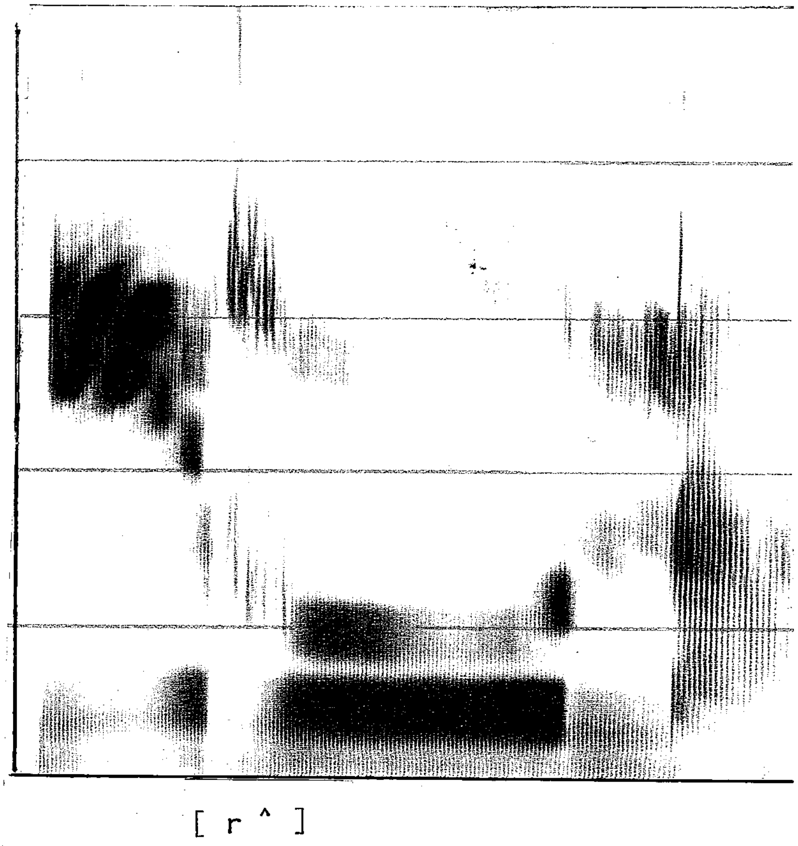


Figure 6.
/írúúnà/ "to snore"

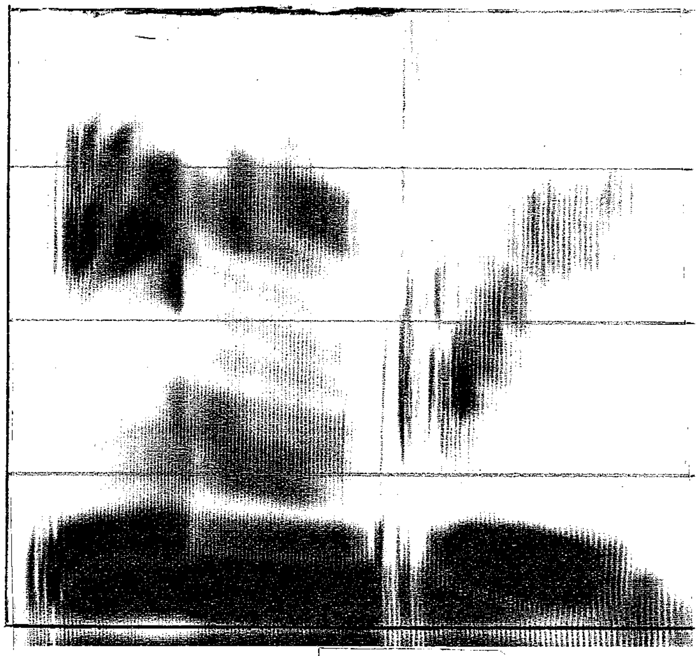


Figure 7.
/ímrì/ "root"

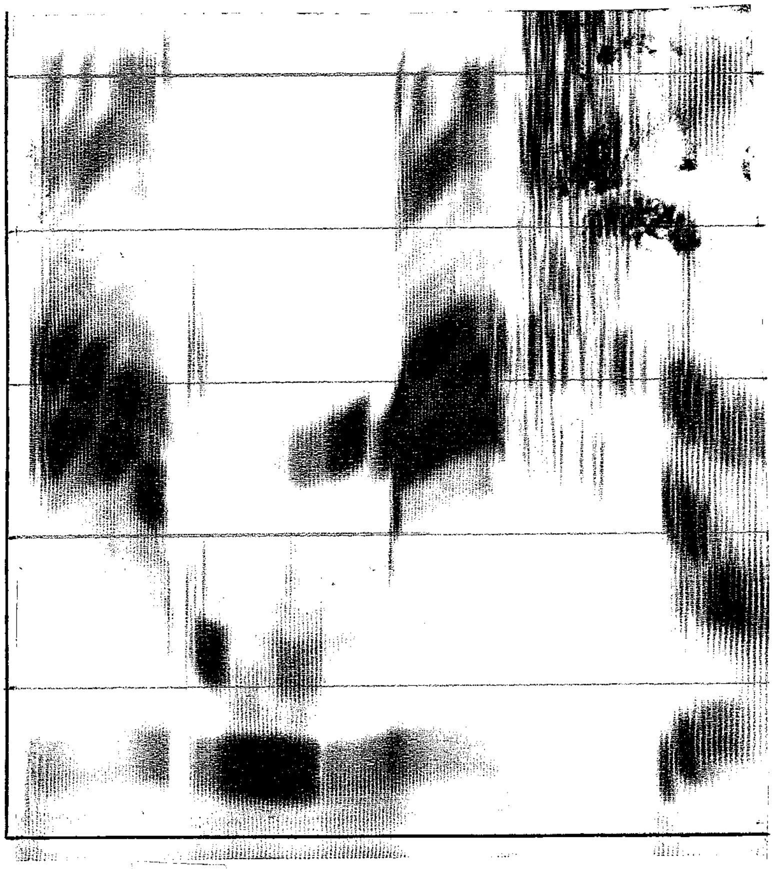


Figure 8.
/írùmíshà/ "to extinguish"

The voiced dental lateral approximant /l/.

One of the two lateral phonemes is a continuant. The most common allophone is a voiced velarized dental lateral approximant. This allophone appears before all vowels except /i/ and may follow any vowel. An example of this sound is shown in Figure 9, and there are also examples in Figures 10 and 15 where there are also other liquids in the words. As is normal with lateral approximants, there is little amplitude in higher formants but the segment has a regular formant pattern. This sound is characterized by a very sharp transition of F_2 when /i/ precedes, from about 2500 Hz in the vowel to about 1600 Hz in the lateral. When surrounded by back vowels, as the first lateral on spectrogram 1 (/íyɔ̀lɔ́líyà/ "to be cold") is, F_2 may be as low as 1350. The auditory impression is of quite a "dark" /l/ although the F_2 observed here is considerably higher than that reported for velarized alveolar lateral approximants in English (Lehiste, 1964, reports averages of about 820 Hz in 3 male speakers; Bladon, 1979, reports F_2 at about 1100 Hz for males), and in the "hard" /l/ of Russian, which is basically a velarized dental (for which Fant (1960) reports F_2 of 850 Hz for a male speaker). The F_2 in the Chaga sound is more comparable to that of the palatalized lateral in Russian (1600 Hz), although 1600 Hz in a female voice might be considered equivalent to only about 1400 Hz in a male voice (cf. the male/female differences in English vowels reported by Peterson and Barney, 1952).

The dental articulation in Chaga was confirmed palatographically, there being a complete wipe-off from the teeth. There was also a wipe-off area on the alveolar ridge. It seems likely that rather than having a velar/pharyngeal narrowing like that in the English and Russian velarized laterals - which produces a low F_2 similar to that for a truly back vowel [u] - the Chaga segment may have a secondary constriction which is further from the glottis, perhaps in a location similar to that for the central vowel [ɨ].

There is a markedly different set of allophones of this phoneme in the environment before /i/. These are palatal lateral continuants, both approximant and fricative. The approximant is exemplified in Figure 11 (/ílíyà/ "to cry"). In this segment, F_2 is at about 2200 Hz. Palatography showed that there is a large area of wipe-off on the palate. In this allophone, which has no secondary constriction behind the primary articulation, F_2 is expected to correspond to a half wavelength of the cavity behind the constriction, as it does in a high front vowel such as [i] (Fant 1960: 167). As may be seen from the figure, F_2 in this segment is indeed similar to that in [i] and the palatal lateral allophones of /l/ can be considered in both an articulatory and an acoustic sense to have been assimilated to the place of articulation of the following vowel.

Friction often accompanies the palatal laterals. This can be clearly seen in the second lateral in Figure 10. There is a clear voicing bar, and a noisy excitation of F_2, as well as a band of noise above 4000 Hz. The amount and intensity of the friction is quite variable, as may be seen by comparing Figure 11, the approximant, with Figure 12, which shows a repetition of the same word but with a fricative lateral. Unlike with the fricative rhotic phoneme /zr/, the friction in the palatal lateral is only found in one variant of a positional allophone of a phoneme whose other allophones do not have friction. Hence the friction in this segment is not considered a defining characteristic of the phoneme to which it belongs.

Figure 12 also shows that these palatal laterals may have a burst-like onset, as if there is a brief period of total occlusion before the lateral escape of air

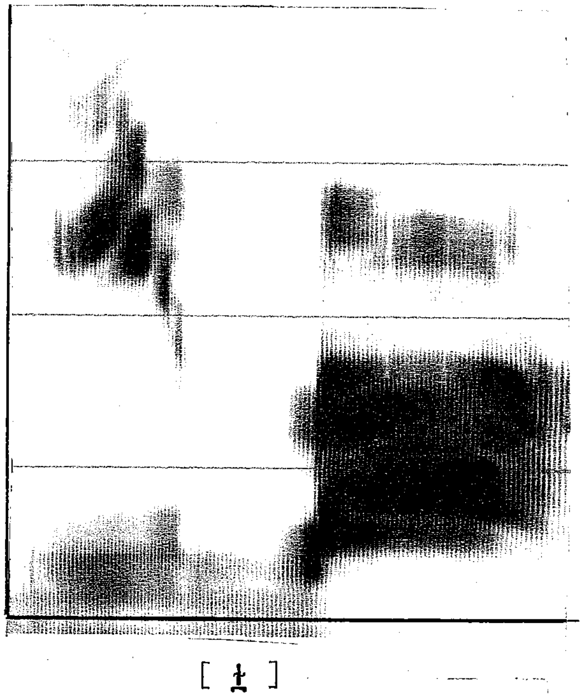


Figure 9.
/íláà/ "to lie down"

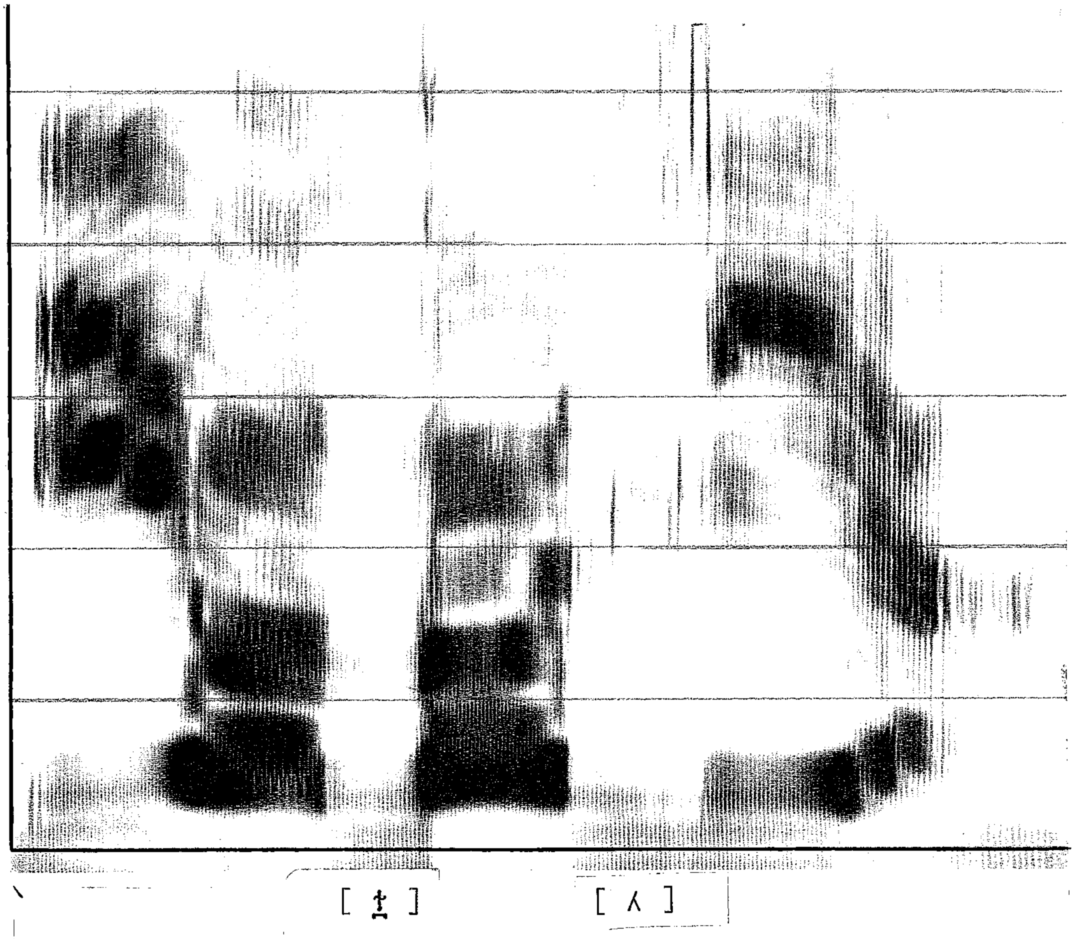


Figure 10.
/íyòlólíyà/ "to be cold"

is initiated. The example shown is rather an extreme case of this phenomenon, as in this token there is a closure lasting about 30 msec. In such a case one might talk of a pre-stopped lateral; there is no percept of a lateral affricate.

The major allophonic differences within this phoneme are both auditorily and acoustically quite distinct, but a rule which assimilates a lateral to a following high front vowel and optionally fricates it is a natural one, and this segment may be viewed as being "underlyingly" a dental lateral. It is interesting to note that a preceding high front vowel does not cause the same assimilation as a following /i/ does. That is, the co-articulation follows the conventional syllabification rule which places an intervocalic consonant in the same syllable as the vowel which follows (see Bell & Hooper, 1978, for some discussion of the evidence for this as the universal syllabification rule).

The alveolar palatalized lateral approximant [l^j].

In addition to the palatal laterals discussed above, there is a palatalized alveolar lateral approximant in Chaga. This sound is rare, seeming only to occur when a lateral precedes a historic /i/ which has become non-syllabic before a low vowel, as in the word /ílyà/ "to eat" shown on Figure 13. There is a very marked palatal offglide to this segment, but the primary place of articulation indicated by the palatographic examination is alveolar. Although the higher formants are not very distinct on the spectrogram, it appears that F_2 and F_3 are closer together in this sound than they are in the palatal laterals, apparently because F_3 is lower.

This segment presents an interesting problem of analysis. For historical reasons and for the sake of economy it seems desirable to relate this lateral to the phoneme /l/, but if words with /ly/ are treated as containing the segment /i/ then the general rule that /l/ --> [ʎ] before /i/ is violated. In this case, the rule must be "bled" so that words in which /i/ is desyllabified do not undergo the rule, but instead are subject to a special rule converting /l/ to a palatalized alveolar. The reason why the output should have an alveolar place of articulation is not transparent. However, if this rule complexity is rejected, the alternative is to propose a phoneme of highly limited distribution yet with obvious phonetic similarity to another.

The alveolar lateral flap /d/.

In addition to the various lateral segment types discussed above, there is also a lateral flap in Chaga. Examples of this segment are provided in Figures 14-16. It is obviously characterized by its brevity. Palatographic examination showed that there is a central contact on the alveolar ridge and a bilateral wipe-off at the molars, but between these two there is an area of zero wipe-off. This confirmed that this segment is a lateral. Although there is some coarticulation with the adjacent vowels, this segment is much less variable than the lateral approximant phoneme. Even between two /i/ vowels, as in Figure 16, F_2 is still only just above 2000 Hz; between two vowels with considerably lower F_2's (/o/ and /a/), as in Figure 15, F_2 in the flap is at about 1500 Hz.

Because the phoneme /r/ has a tap allophone, there is thus a possibility of minimal contrast between a (central) tap and a (lateral) flap at the alveolar place of articulation: compare Figures 8 and 14. What cues might serve to disambiguate such a contrast cannot be demonstrated with the available data, nor indeed that the two could be reliably distinguished.

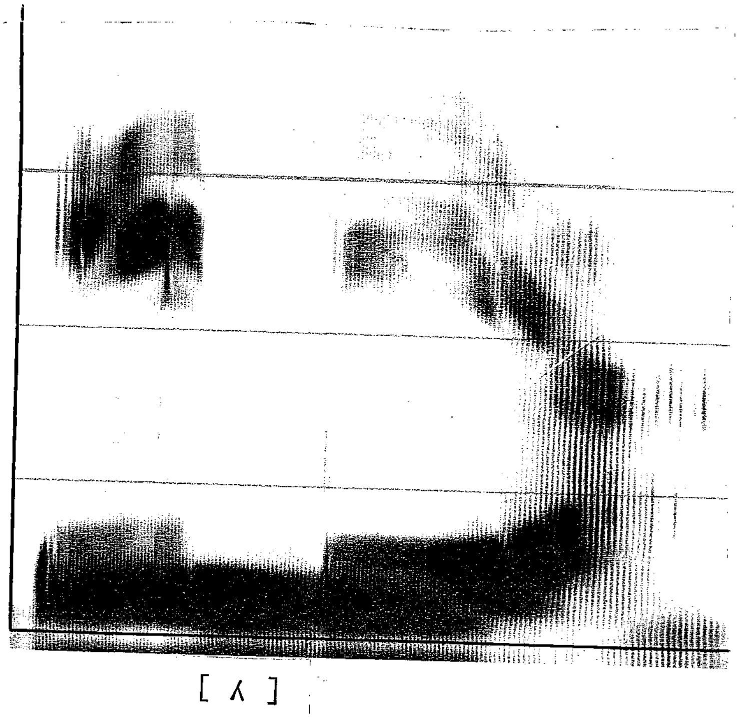


Figure 11.
/ílíyà/ "to cry"

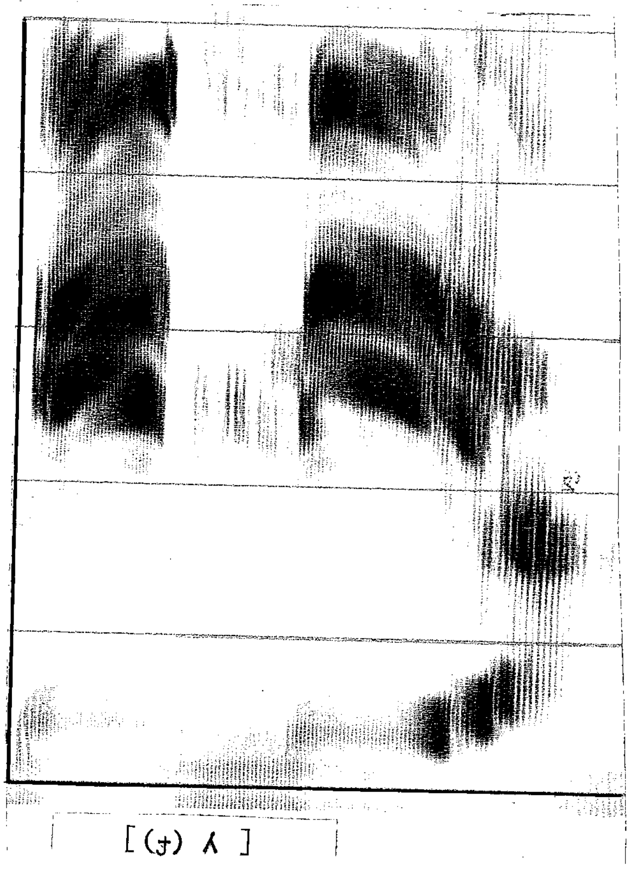


Figure 12.
/ílíyà/ "to cry"

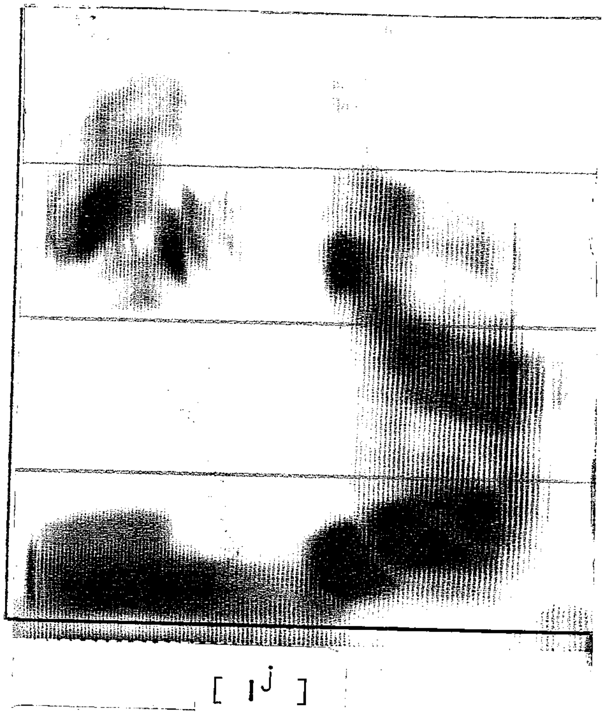


Figure 13.
/ílyà/ "to eat

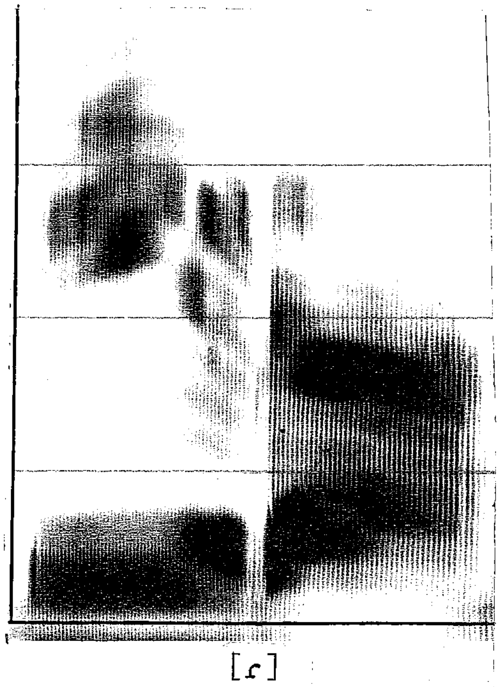


Figure 14.
/ídà/ "leaf"

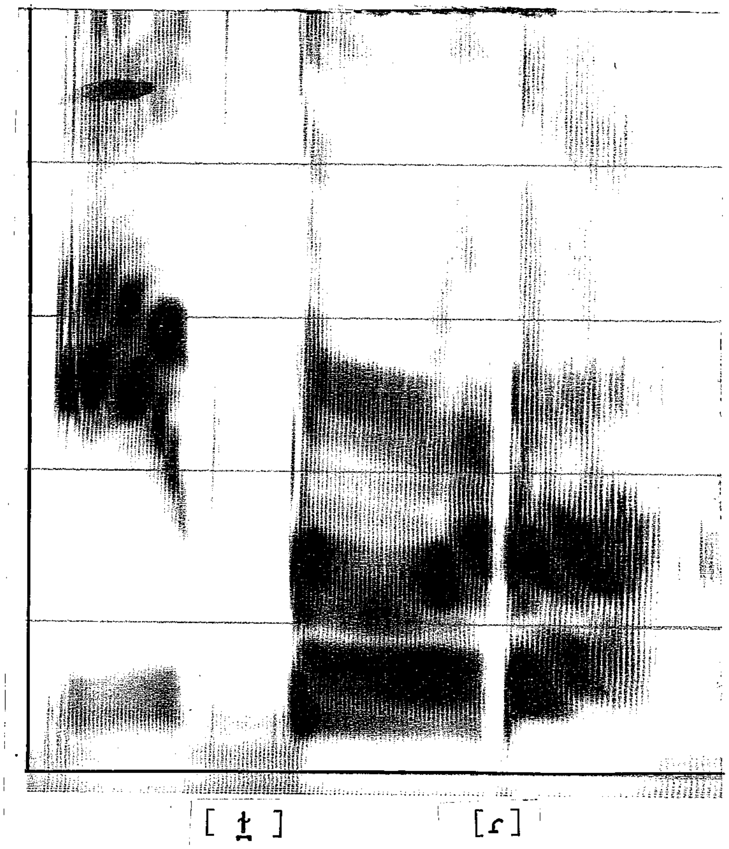


Figure 15.
/ílódà/ "to show"

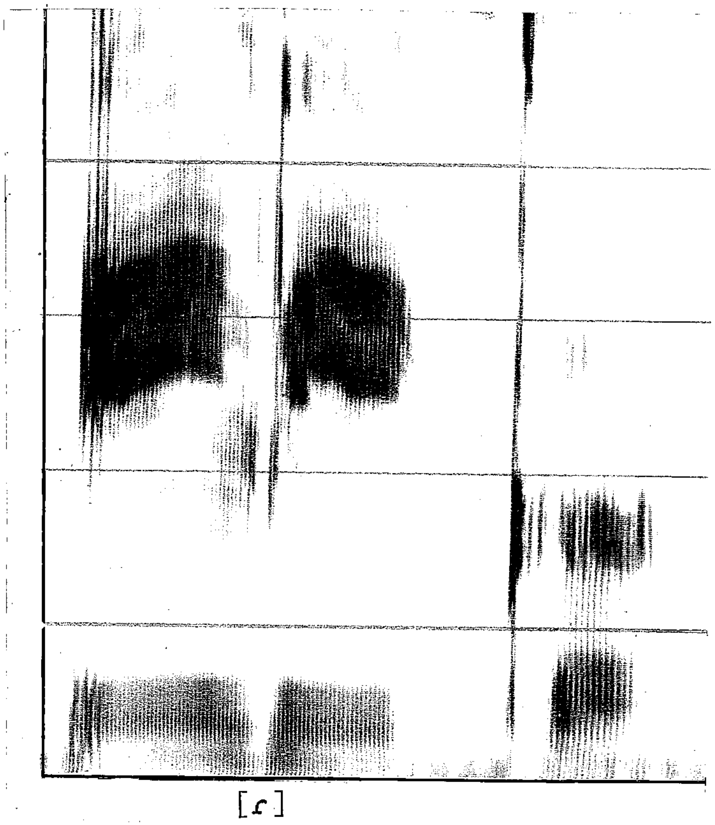


Figure 16.
/ídíkà/ "to hide"

Summary of phonetic liquids observed.

The range of sounds within the general category of liquids that we observed in the second author's speech included the following:

r-sounds (rhotics)

1 A voiced apical post-alveolar nonsibilant fricative [ɹ̠^]
2 A voiced central alveolar approximant [ɹ]
3 A voiced apico-alveolar trill [r]
4 A voiceless apico-alveolar trill [r]
5 A voiced apico-alveolar trill with friction [r^]
6 A voiced apico-alveolar tap [ɾ]

laterals

7 A voiced velarized dental lateral approximant [ɫ̠]
8 A voiced palatal lateral approximant [ʎ]
9 A voiced palatal lateral fricative [ʎ^]
10 A voiced alveolar palatalized lateral approximant [l^j]
11 A voiced apico-alveolar lateral flap [D]

II. Historical development.

Proto-Bantu, as reconstructed by Meinhof (1906) and other earlier authorities, had a single liquid /l/. Meinhof suggests that this was fricative and perhaps "half-plosive". This may be interpreted to mean that he viewed this segment as a lateral affricate. His Proto-Bantu (P-B) phonology also includes /d/. Later views have suggested that there were lateral and plosive allophones of a single underlying segment, usually represented as /*d/ (e.g. Guthrie, 1967: 62). This proto-segment has more reflexes which are lateral in nature than reflexes which are stops, and Guthrie suggests that it may have been a lateral in P-B at least between vowels. The symbol /*d/ was chosen partly for the sake of symmetry with the voiceless stop series, which contains /*p, *t, *k/, and partly because [d] reflexes occur after nasals and in a few other environments. In any case, it is generally agreed that there were no rhotic segments in P-B.

In his summary of the principal sound-shifts between Proto-Bantu and Chaga, Guthrie (1971: 46) suggests that /*d/ is normally deleted, but becomes a flap /ɾ/ before /i/ or /e/, a trill /r/ before the "superclose" vowels /*ɨ/ and /*ʉ/ and a lateral approximant /l/ before long vowels. He suggests that P-B /*t/ has become a uvular fricative /ʁ/, except before /*ɨ/ where it has become the trill /r/. In addition, he suggests that the reflex of /*g/ before /*ɨ/ is also /r/.

His analysis of Chaga therefore differs from the one offered here, in that we observe that the flap is lateral and that there is no uvular fricative but rather a non-sibilant alveolar fricative. In addition,, the historical sources of the Chaga liquids that he proposes seem to differ in certain respects from the correspondences we observed, as do those mentioned by Hinnebusch and Nurse (1981). Clearly, however, the various Chaga dialects or languages differ somewhat from one another with respect to the development of liquids.

We compared KiVunjo Chaga cognates with forms in Guthrie's list of comparative stems to verify the historical developments represented in the dialect under study. The items below were found for this comparison. In each list, the first item is the comparative stem number assigned to the item by Guthrie. Next the form in which the stem is given by Guthrie follows, then the

Chaga cognate and the gloss. If there is a divergence of meaning between the posited P-B gloss and the meaning in Chaga, two glosses are given separated by a slash. Correspondences followed by S are uncertain and should be treated as suspect.

Words with the fricative /zr/

CS 620	*-dímb-	wàzrìmâ	"forget"	
CS 633	*-dó	zrò	"sleep" (n)	
CS 695	*-dúk	zrékà	"vomit"	
CS 673	*-dóót-	íyózrà	"to dream"	
CS1243	*-kùtú	kúzrû	"ear"	
CS1629	*-tá	úzrà	"spittle"	
CS1630	*-tá	shízrà	"war"	
CS1650	*-tákò	zrákò	"buttock"	
CS1651	*-tákųn-	zrápfúnà	"chew"	
CS1681	*-táp-	zráhíà	"draw water"	
CS1696a	*-tédik-	zréhíà	"put pot on fire"	
CS1698	*-tég-	zrégà	"set a trap"	
CS1719	*-tét-	zrézrà	"speak"	
CS1726	*-tètim-	zrézrémà	"shake"	
CS1729	*-tí	zrí	"tree, medicine, stick."	
CS1736	*-tííd-	zríchà	"run away"	S
CS1738	*-tímà	zrímà	"liver, heart"	
CS18xx	*-tú	ḿzrò	"head"	
CS2018	*-yítik-	zrícíyà	"agree, answer"	
CS2138	*-yótò	múzrò	"fire"	

Words with the trill /r/.

CS 289	*-cátù	sárù	"python"
CS 591	*-dį̀	íimrì	"root"

CS 603	*-dɨ̀bà	ḿrà	"pool, deep water/rainwater stream"	
CS 604	*-díbà	rùwâ	"milk"	S
CS 826	*-gɨ̀dò	ḿrìmù	"taboo"	S
CS 828	*-gɨ́kò	ríkò	"hearth"	
CS 831	*-gɨ́nà	rínà	"name"	
CS 852	*-gòn-	rùùnâ	"snore"	
CS1664	*-tànd-	rànduò	"tear" (vb)	

Words with the lateral continuant /l/.

CS 252	*-cáádè	ḿshàlê	"arrow"	
CS 455	*-dààd-	làdâ làa	"sleep; lie down"	
CS 507	*-dè	lèyâ	"(be) long"	
CS 510	*-dèd-	lèlâ	"look after child/look after mother after childbirth"	
CS 529	*-dèm-	lèmâ	"become too difficult for"	
CS 551	*-díá	úlíà	"that"	
CS 552	*-dìàngò	ḿlàngô	"door"	
CS 561	*-dìd-	lìyâ	"cry"	
CS 568	*-dìm-	lìmà	"cultivate"	
CS 571	*-dímɨ̀	úlìmí	"tongue"	
CS 641	*-dòd-	lòdà	"show"	S
CS 644	*-dòg-	lògà	"bewitch"	
CS 677	*-du′a′d-	lùwô	"be sick"	
CS 699	*-dúmɨ̀	úlùmì	"tongue"	
CS1639	*-tád-	tàlâ	"count"	

Words with the lateral flap /d/

CS 245a	*-càdud-	sàmbúdà	"choose"	S
CS 308	*-cèdid-	sédémkà	"go down, skid"	

CS 442	*-dà	ḿdà	"intestines"
CS 455	*-dáád-	làdâ	"sleep, lie down"
CS 615	*-dɨ̂ɨ̂k-	dìkà	"bury, hide"
CS 619	*-dímu	wádùmù	"spirit (bad)"
CS 624	*-dɨ́ng-	díngà	"surround"
CS 641	*-dòd-	lòdà	"show"
CS 642	*-dòòdɨ̂	dòdî	"whistling"
CS 690	*-dúdùè	dúdúwè	"gall bladder"
CS 701	*-dùmb-	dùmìshâ	"praise, thank"
CS1112	*-kódɨ-	kódà	"light (a fire)"
CS1189	*-tátù	dádù	"three"

Words with /ly/

CS 550	*-dí-	ílyà	"to eat"

Our data suggest that the regular source of the fricative rhotic /zr/ is the P-B alveolar plosive /*t/. Note that the /zr/ reflex of /*t/ is not found to occur before the "superclose" vowels /*ɨ/ and /*ʉ/, where the reflexes are instead voiceless affricates (/tʃ/ and /pf/ respectively). There are also some sporadic /zr/ reflexes from /*d/.

The sources of the trill /r/ are hard to identify. The most regular development is of the sequence /*gɨ/ to /ri/, of which we have found three instances and no ready counterexamples. Note that the sequence /*gi/ develops into /y(i)/ (CS 809 *-gí íyàyí "egg"; CS 811 -gɨ̂- yà "go") and the sequence /*ngɨ/ develops into /nzi/ (CS 819 *gɨ̂ nzî "fly" (n.); CS 827 *-gɨ̂gè nzîyè "locust"). An additional source of /r/ is /*dɨ/, although the two cases of this development are outnumbered by the four examples of the development of /*dɨ/ to /d(i)/ (CS 615, 624, 642, 1112). We do not find that /*tɨ/ is a source of /r/, as Guthrie suggests, and none of the dialects reported on by Hinnebusch and Nurse (1981) show such a development.

As for the two laterals /l/ and /d/, both of them develop from the P-B /*d/ phoneme, but which of the two Chaga reflexes appears does not seem to be governed by any obvious regularity. There is some preference for the approximant /l/ if the segment is stem initial and for the flap /d/ elsewhere (cf. CS 455 and 641 for clear instances), but CS 510, and 1639 with non-initial /l/, and CS 615, 624, 642, 690 and 701 with initial /d/ show that this is not invariable. There is no support for Guthrie's suggestion that the approximant /l/ develops before long vowels in any regular fashion (compare CS 455 with 615 and 642).

For the phonetician interested in explanations for sound change we find three main processes to account for:

i) the voicing and frication of /*t/ (in the absence of a truly high vowel vowel following),

ii) the development of an alveolar trill from /*g/ (and sometimes /*d/) before /*ɨ/,

iii) the tendency for /*d/ to become a lateral approximant stem-initially but a lateral flap elsewhere.

As Hinnebusch and Nurse would surely suggest, (i) should probably be related to a more general process of spirantization in Chaga and other Eastern Bantu languages. However, the fact that voiceless affricates result from /*t/ before the highest vowels /*ɨ/ and /ʉ/ but a voiced fricative occurs elsewhere as the reflex of /*t/ may be related to aerodynamic factors discussed by Jaeger (1978) (but see Javkin, 1979). Compare also the tendency towards later voice onset with higher vowels reported by, for example, Klatt (1975).

As for (ii), a trill might result from /*gɨ/ through an intensification of the phenomenon of a multiple burst which can often be observed in the release of velar and palatal obstruents (cf. Fischer-Jørgensen, 1954). This has been explained as a consequence of a Bernouilli force acting to reclose the oral passage. This occurs because of the relatively slow rate at which the articulators move apart in such consonants, a fact attributed to the greater area of the contact made in their production. Fant (1960: 179), commenting on the Russian syllable [xa], suggests that

> "the repetition of this process would cause a vibrational force on the posterior part of the velum and the uvula, in an extreme case causing a uvular trill."

If such a development had occurred in Chaga it would still be necessary to posit some process, perhaps perceptually based, by which the segment would change from a dorsal uvular articulation to an apico-alveolar one.

The third point, concerning the distinction between /l/ and /d/, may well be related to the dynamics of word structure. It is not uncommon for word-structure and stress to be related in languages in such a way that the stem-initial position is also the beginning of a more strongly stressed syllable. If this was also true of Chaga, then we could suggest that in this position the longer approximant was retained, whereas in less strongly stressed positions the abbreviated flap frequently developed as the regular pronunciation. There are obvious similarities to the distribution of the stop and flap allophones of /t/ and /d/ in American English if this account is correct (cf, for example, Kahn, 1976).

III. Summary.

This paper has shown that there are four phonemic liquids in the variety of Chaga discussed. These have a variety of free and conditioned allophones so that a large number of different phonetic liquids can be heard in the speech of a single individual. The liquids have mainly developed from the alveolar stops of Proto-Bantu, although the sources of the trill /r/ are heterogeneous. The processes involved in the the historical developments provide an opportunity for consideration of the role of phonetic factors in sound change.

References

Bell, Alan & Joan Bybee Hooper. 1978. Issues and evidence in syllabic phonology. *Syllables and Segments,* ed. by A. Bell & J.B. Hooper: 3-24. Amsterdam: North-Holland.

Bladon, R.A.W. 1979. The production of laterals: some articulatory properties and their acoustic implications. *Current Issues in the Phonetic Sciences,* ed. by Harry and Patricia Hollien, Part 1: 501-508. Amsterdam: John Benjamins.

Fant, Gunnar. 1970. *Acoustic Theory of Speech Production.* The Hague: Mouton.

Fischer-Jørgensen, E. 1954. Acoustic analysis of stop consonants. *Miscellanea Phonetica* II: 42-59.

Guthrie, Malcolm. 1967. *Comparative Bantu, Volume 1: The Comparative Linguistics of the Bantu Languages.* Farnborough: Gregg.

---------- 1971. *Comparative Bantu, Volume 2: Bantu Prehistory, Inventory and Indexes.* Farnborough: Gregg.

Hinnebusch, Thomas J. & Derek Nurse. 1981. Spirantization in Chaga. *Sprache und Geschichte in Afrika* 3: 51-78.

Jaeger, Jeri J. 1978. Speech aerodynamics and phonological universals. *Proceedings of the Fourth Annual Meeting of the Berkeley Linguistics Society*: 311-329.

Javkin, Hector R. 1979. Phonetic explanations for the devoicing of high vowels. *Proceedings of the Fifth Annual Meeting of the Berkeley LinGuistics Society*: 413-418.

Kahn, Daniel. 1976. *Syllable-based generalizations in English Phonology.* (Ph.D. dissertation, M.I.T.) Bloomington: Indiana University Linguistics Club.

Ladefoged, Peter. 1975. *A Course in Phonetics*. New York: Harcourt, Brace and Jovanovitch.

Lehiste, Ilse. 1964. *Acoustical Characteristics of Selected English Consonants.* Bloomington: Indiana University.

Lindau, Mona 1980. The story of r. *UCLA Working Papers in Phonetics* 51: 114-119.

Peterson, Gordon E. & Harold L. Barney. 1952. Control methods used in a study of vowels. *JASA* 24: 175-184.

Postcript to
Attempts by human speakers to reproduce Fant's nomograms

During a recent visit (18 June 82) to the UCLA Phonetics Lab, Professor Gunnar Fant recorded the sounds from which the spectrograms in Figure 1 were made. The qualities were as noted in the legend, and the point of maximum constriction might be presumed to be about (a) 12 cm, (b) 12.5 cm, and (c) 14 cm from the glottis. The size of the opening was comparable for all three vowels, except that it was probably slightly greater for the first vowel. Note that F2 decreases throughout this series, being slightly higher in Swedish [e] than it is in Swedish [i]. It is not easy to see exactly where F3 is in Figure 1(a), but it seems that it increases slightly in going from (a) to (b), and then falls slightly in going to (c). These formant movements are very comparable to those illustrated in the nomograms, the only difference being in the magnitude of the changes.

Accordingly we are left with a puzzle. What is it that Ladefoged and Bladon are doing that produces a clearly rising F3 without a falling F2, and how does it differ from Fant's articulatory behavior in producing the vowels in Figure 1? Part of the answer may be that Ladefoged and Bladon did not make articulations sufficiently far from the glottis. But, as noted in the paper, we felt that we could not go farther forward without making a different kind of tongue shape. Another part of the answer may be in variations in the size of the constriction. Fant made no claims about the size of the constriction for the vowels in Figure 1, and there are variations in F1 that indicate possible differences of this kind. Nevertheless it seems reasonable to assume that Fant made a series of vowels with the point of maximum constriction moving progressively farther from the glottis. And it is certainly a fact that in Fant's vowels F2 fell considerably while F3 first rose and then fell (as in the nomograms). But it is also a fact that in Ladefoged's and Bladon's vowels F2 remained level while F3 rose considerably (which is not what happens in the nomograms). As the stock answer goes: "More research is needed."

PL

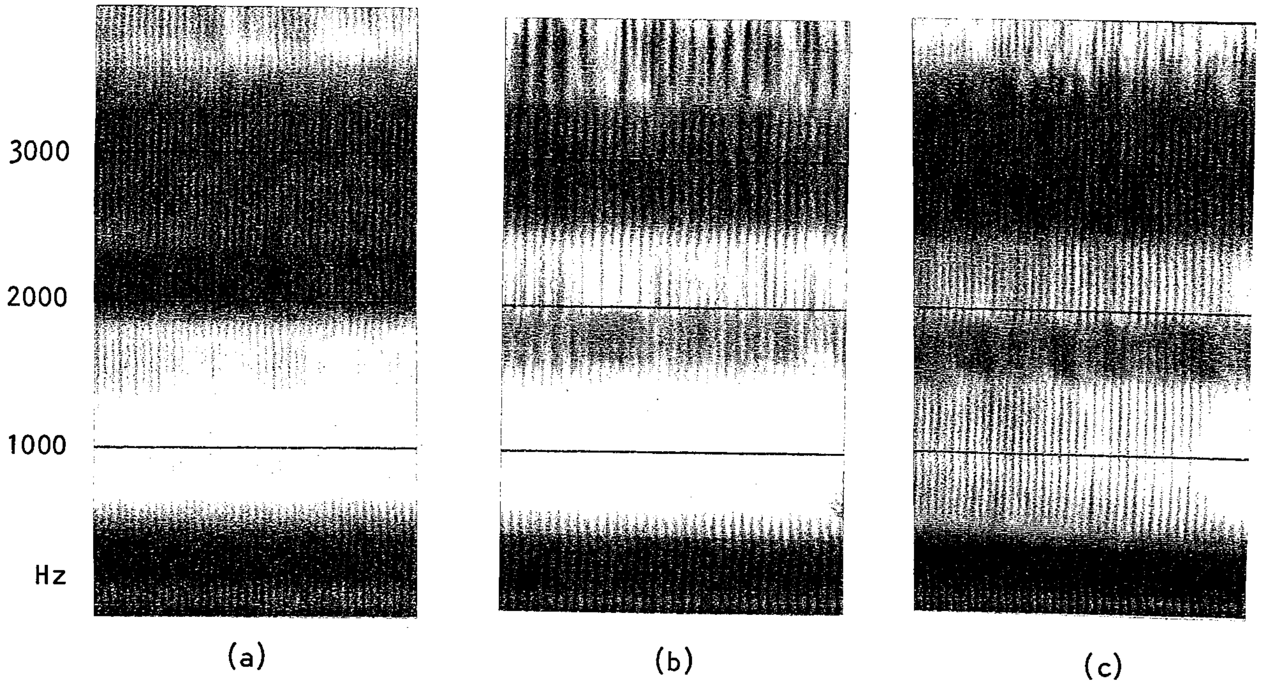


Figure 1. Three vowels produced by Gunnar Fant: (a) Swedish [e]; (b) Swedish [i]; (c) an approximation to Swedish Visby [ɨ].