# COMMUNICATION SCIENCES
# AND
# ENGINEERING

# IX. SPEECH COMMUNICATION[*]

## Academic and Research Staff

Prof. K. N. Stevens
Prof. M. Halle
Prof. W. L. Henke
Prof. A. V. Oppenheim

Dr. Mary C. Bateson
Dr. Margaret Bullowa
Dr. A. W. F. Huggins

Dr. R. D. Kent
Dr. D. H. Klatt
Dr. Paula Menyuk
A. R. Kessler

## Graduate Students

R. W. Boberg
D. E. Dudgeon
R. W. Hankins
Emily F. Kirstein

R. M. Mersereau
B. Mezrich
J. S. Perkell
M. R. Sambur
J. S. Siegel

H. A. Sunkenberg
R. N. Weinreb
M. L. Wood, Jr.
V. W. Zue

## A. TRANSLATION OF DIVERS' SPEECH USING DIGITAL FREQUENCY WARPING

### 1. Introduction

The objective of this study is to investigate the possibility of applying the technique of spectral warping to correct divers' speech. The technique of spectral warping is discussed in greater detail elsewhere.[1,2] Consequently only those results that are pertinent to this study will be mentioned here. The term "divers' speech" is used specifically to mean the speech produced by divers breathing helium-rich gas mixtures at ambient pressures higher than atmospheric pressure. It is well known that divers' speech is usually highly unintelligible. Preliminary results indicate that the technique of spectral warping may be applicable to improve the intelligibility of such speech.

### 2. Theory

The technique of spectral warping is based on the transformation of a sequence (sampled time function) to a new sequence so that the Discrete Fourier Transform (DFT) of the new sequence is equal to unequally spaced samples of the z-transform of the original sequence around the unit circle. This transformation corresponds to expansion of the original sequence in terms of a set of linearly independent sequences $\{\psi_k(n)\}$. That is,

$$f(n) = \sum_{k=-\infty}^{\infty} g(k)\, \psi_k(n), \tag{1}$$

---

[*] This work was supported in part by the U. S. Air Force Cambridge Research Laboratories under Contract F19628-69-C-0044; and in part by the National Institutes of Health (Grant 5 RO1 NS 04332-08).

where

f(n) ~ original sequence

g(k) ~ new sequence.

If we require this transformation to satisfy certain properties, then the DFT of f(n) and g(n) can be related by a nonlinear transformation in frequency. If we denote by $\omega$ and $\Omega$ the frequency variables of the original and new sequences, respectively, the relation between $\omega$ and $\Omega$ is given by

$$\Omega = \theta(\omega) = \tan^{-1}\left[\frac{(1-a^2)\sin\omega}{(1+a^2)\cos\omega - 2a}\right], \tag{2a}$$

or equivalently

$$\Omega = \theta(\omega) = \omega + 2\tan^{-1}\left[\frac{a\sin\omega}{1-a\cos\omega}\right]. \tag{2b}$$

It can be shown that the inverse of Eq. 2a is given by

$$\omega = \theta^{-1}(\Omega) = \tan^{-1}\left[\frac{(1-a^2)\sin\Omega}{(1+a^2)\cos\Omega + 2a}\right] \tag{3}$$

which corresponds to replacing a with −a in (2a). The z-transform of $\psi_k(n)$ is

$$\psi_k(z) = \left[\frac{z^{-1}-a}{1-az^{-1}}\right]^k. \tag{4}$$

We shall assume henceforth that f(n) and g(k) are zero for $n < 0$ and $k < 0$, respectively. From Eq. 4, the following relationship holds:

$$\sum_{n=0}^{\infty} n\psi_r(n)\,\psi_k(n) = k\delta_{rk}. \tag{5}$$

From Eqs. 1, 4 and 5, the sequence g(k) can be found by

$$g(k) = \frac{1}{k}\sum_{n=0}^{\infty} n\psi_k(n)\,f(n) \qquad k = 1, 2, \ldots$$

$$g(0) = \sum_{n=0}^{\infty} f(n)\,a^n. \tag{6}$$

In practice, g(k) can be obtained by passing f(−n) through a linear, shift-invariant digital network as shown in Fig. IX-1, where $g(k) = \tilde{g}_k(0)$.

Note that the g(k) are coefficients of an infinite sum. In practice, however, the summation is truncated to obtain an approximation of the actual expansion. As we have said, detailed treatment can be found elsewhere.[1,2]

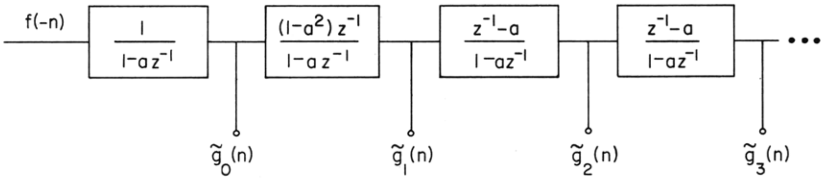


Fig. IX-1. Digital network used to implement spectral warping.

In the rest of this report, the symbols N, K, and a will be used to denote the length of input sequence, output sequence, and the parameter of the digital network, respectively. The parameter a is assumed to be real and between −1 and +1 throughout this study.

This system has been implemented on the PDP-9 computer of the Speech Communication Group. Figure IX-2 includes spectra of some speech before and after being processed. In this case the parameter a was so chosen that the original spectrum is scaled downward in frequency.

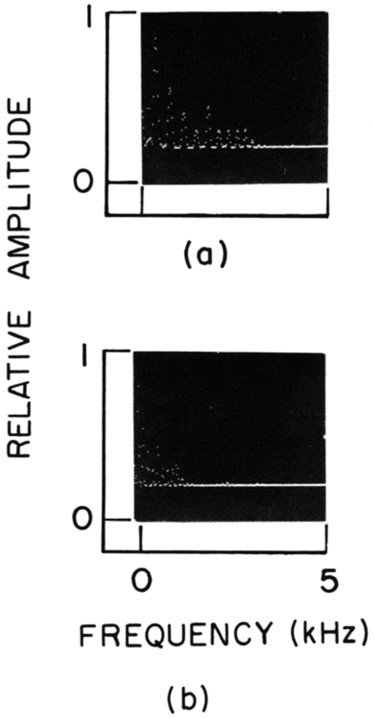


Fig. IX-2. Example of spectral warping on a sample of speech: (a) original; (b) warped spectrum with a = −1/2.

### 3. Divers' Speech Translation

Extensive studies of divers' speech have been made.[3-7] It is known that the ambient pressure and breathing mixture tends to shift the vocal tract resonant frequencies of diver's speech nonlinearly, the nonlinearity being more severe at low frequency. At higher frequencies, the frequency translation is almost linear. Resonance

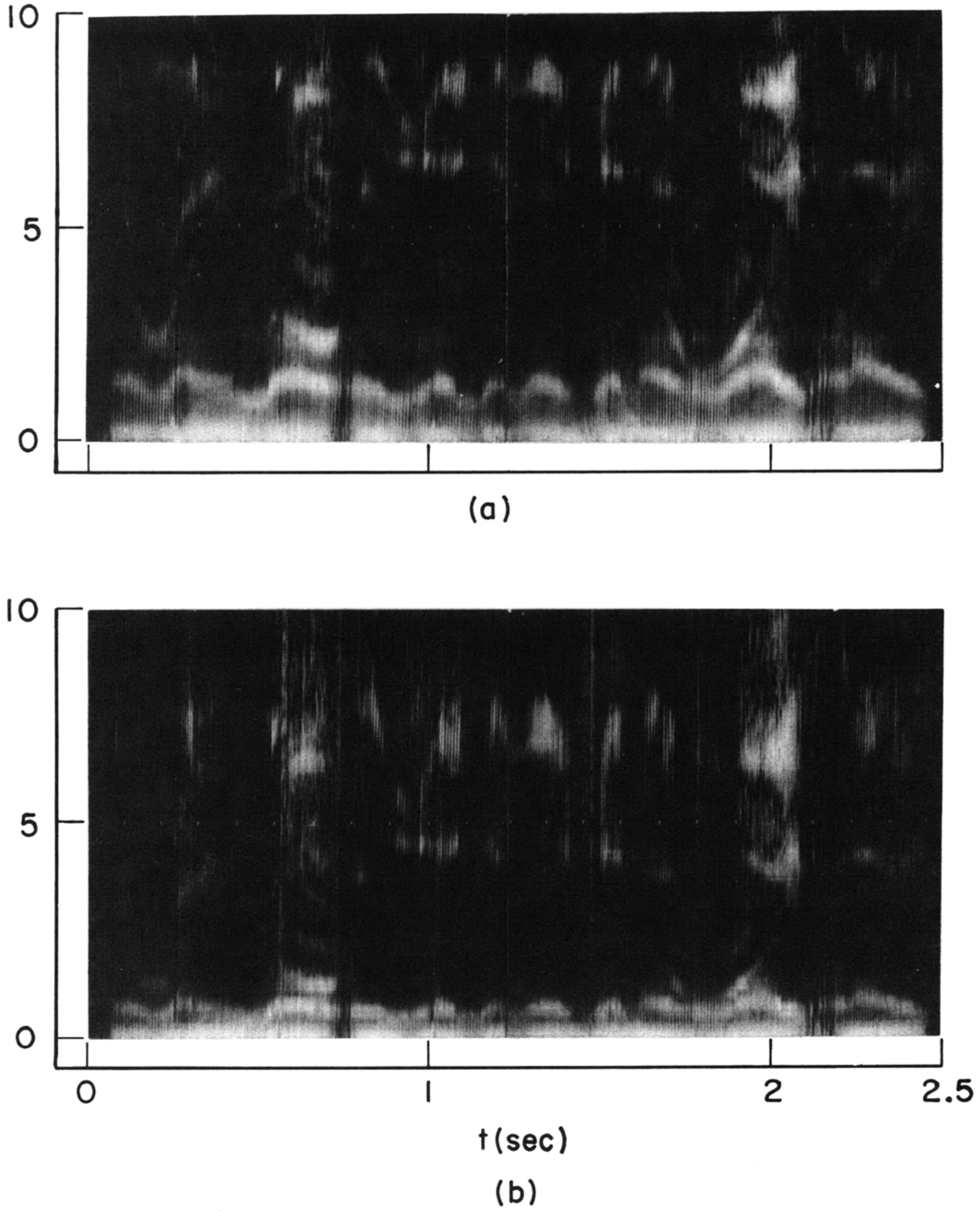


Fig. IX-3. Spectrograms of speech: (a) before being processed; (b) after being processed (with a = −5/16).

bandwidths remain relatively unchanged which gives rise to an amplitude reduction for higher formants. It is also known that the change in fundamental frequency due to pressure and gas mixtures is not significant, if at all.

The basic properties of diver's speech seem to suggest that any attempt to correct such speech must be able to scale down the vocal-tract resonant frequencies while retaining pitch information. In the present attempt, this end is achieved by processing individual pitch periods separately through the spectral warping network shown in Fig. IX-1, and then reconvolving with a train of impulses, which carries the original pitch information, to obtain translated speech.

For the data used in this study, the composition of breathing mixture was 96% helium and 4% oxygen. The pressure depth was 800 ft, which corresponds roughly to 25 times atmospheric pressure. Data were first sampled at 20 kHz and band-limited to 10 kHz. Individual pitch periods were hand-marked. The start of each period was arbitrarily defined to be the last zero crossing of waveform before the maximum peak within each period. The D-A clock was set to 20 kHz. Figure IX-3 includes spectrograms of original and processed versions of the same sentence. In comparing the two spectrograms, the nonlinear translation of formants is apparent. In Fig. IX-3b the upper formants appear to have been shifted by an undesirable amount. Also, in many places a considerable amount of spectral energy can be found above 5 kHz. Although improvement of intelligibility is readily observable, Fig. IX-4 indicates that the present system does not agree favorably with Gerstman's equation.[7]

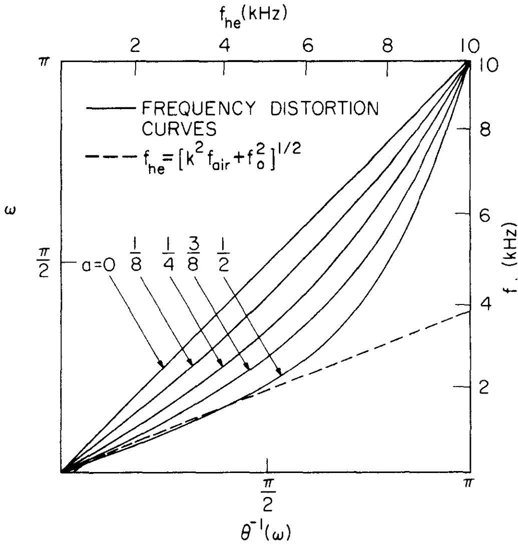


Fig. IX-4. Relations of frequency variables, Eq. 3, for several values of a. Also included is the ideal equation relating the frequency variables.[7] For these data k = 2.57 Hz and $f_o$ = 361 Hz.

The question just raised can be answered by noting, in Fig. IX-4, that for $\Omega > \cos^{-1} a$ spectral energy is accentuated, and for $\Omega < \cos^{-1} a$ it is attenuated. With a = −5/16, this critical frequency, i.e., the frequency at which the curves in Fig. IX-4 have unity slope, is approximately 4 kHz. This means that above 4 kHz spectral-energy under formants are increased and the effect becomes more and more severe as frequency increases.

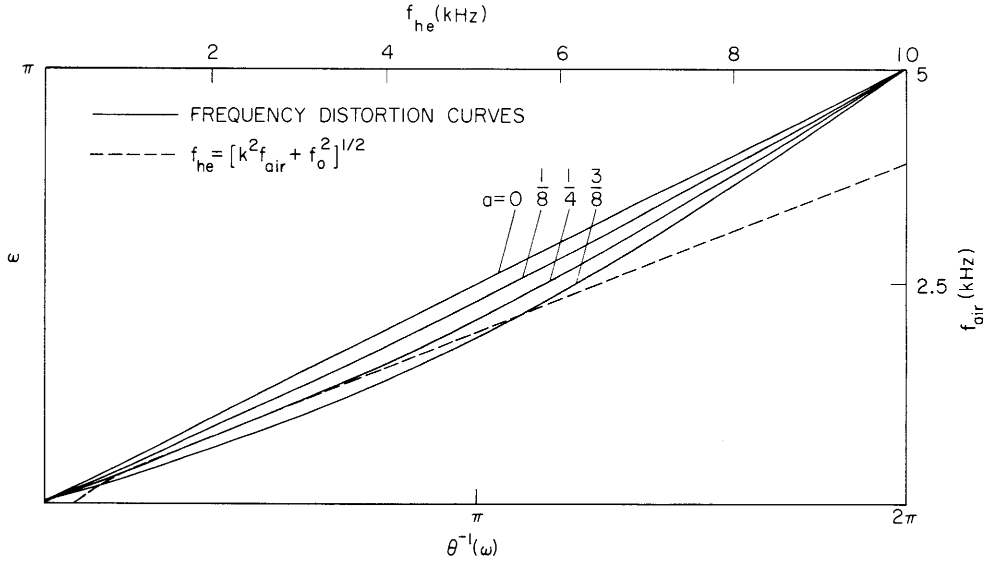


Fig. IX-5. Relations of frequency variables, Eq. 7, for several values of a.

This system is then modified as follows: The data are processed through the spectral warping network as before, except that the processed pitch periods are now convolved with an impulse train where the distance between impulses corresponds to half the original pitch period, and the D-A converter clock rate is set to 10 kHz. This can be shown to correspond to the following frequency variable transformation

$$\omega = \theta^{-1}(\Omega) = \Omega + \tan^{-1} \frac{(1-a^2) \sin \Omega}{(1+a^2) \cos \Omega + 2a}. \tag{7}$$

In Fig. IX-5 Eq. 7 is compared with Gerstman's equation. The two sets of curves agree much more favorably in this case.

Figure IX-6 shows spectrograms of synthetic helium speech before and after processing. Figure IX-7 presents spectrograms of processed data for different values of a. Subjective judgment by a panel of listeners indicates that intelligibility increases greatly from original to a = 0 to a = -1/4.

This research does not attempt to correct all adverse effects of divers' speech. The present study does indicate, however, that this procedure is feasible in improving the intelligibility of such speech. At the present time, a 2.5 sec sentence will take

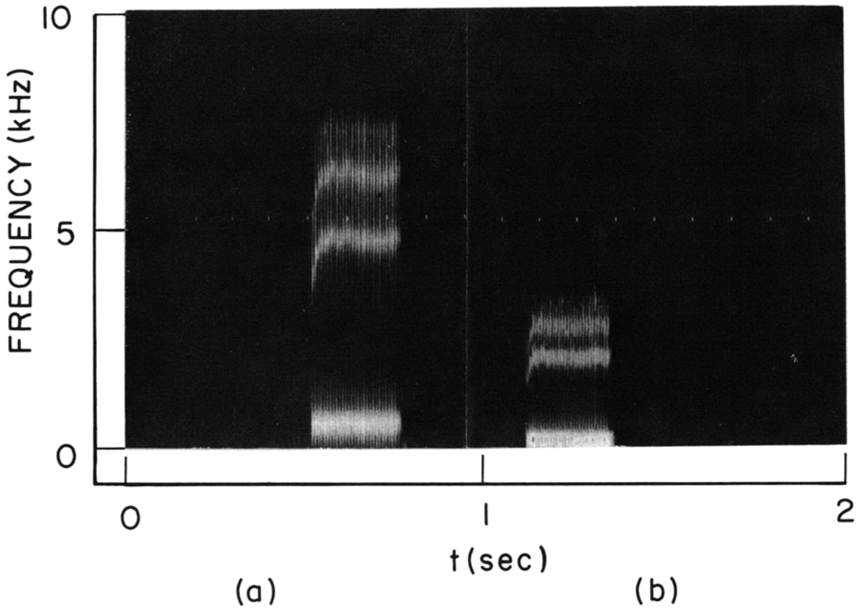


Fig. IX-6. Spectrograms of some synthetic speech: (a) before being processed; (b) after being processed through the second system (with a = $-1/16$).

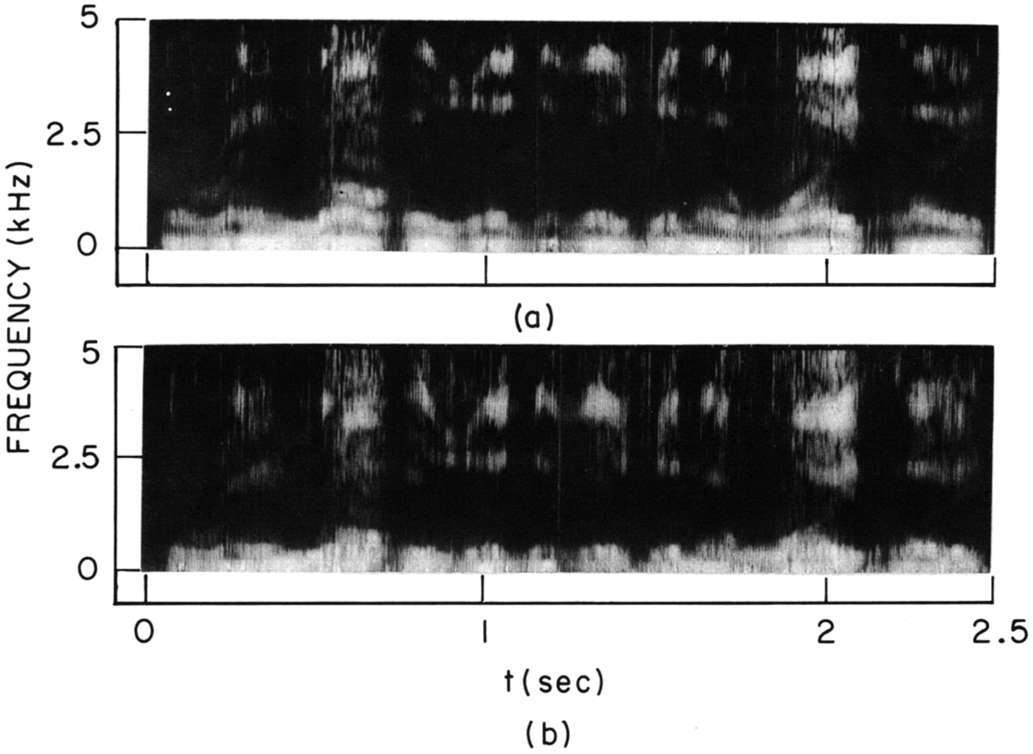


Fig. IX-7. Spectrograms of speech after being processed through the second system: (a) a = 0; (b) a = $-1/4$. The speech material is the same for Figs. IX-3 and IX-7.

approximately 30 min to process. With special digital hardware it is anticipated that the processing time can be reduced to within the same order of magnitude as real time. The relative simplicity of implementation is one of the advantages of this system.

Research in automatic pitch extraction and pitch smoothing is under way.

V. W. Zue

### References

1. A. Oppenheim, D. Johnson, and K. Steiglitz, "Computation of Spectra with Unequal Resolution Using the Fast Fourier Transform," Proc. IEEE 59, 299-301 (1971).
2. A. Oppenheim and D. Johnson, "Discrete Representation of Signals" (being prepared for publication).
3. V. Nedzelnitsky, "Analysis of Speech Produced in Pressurized Helium-Rich Environments," S.M. and E.E. Thesis, M.I.T., June 1969.
4. F. Quick, "Helium Speech Translation Using Homomorphic Techniques," S.M. Thesis, M.I.T., June 1970.
5. G. Fant and B. Sonesson, "Speech at High Ambient Air Pressure," Quarterly Progress and Status Report, Speech Transmission Laboratory, Royal Institute of Technology (KTH), Stockholm, July 1964.
6. G. Fant and J. Lindquist, "Pressure and Gas Mixture Effects on Diver's Speech," Quarterly Progress and Status Report, Speech Transmission Laboratory, Royal Institute of Technology (KTH), Stockholm, April 1968.
7. L. J. Gerstman, "Breathing Mixture and Depth as Separate Effects on Helium Speech," a paper presented at the Seventy-second Meeting of the Acoustical Society of America, November 1966.

## B. DIRECTION OF LANGUAGE PROCESSING

In a recent study the sentence and nonsense-syllable repetition abilities of children with and without language disorders has been compared.[1] Medical and psychological examinations of children with a language disorder gave no indication of why their language was deviant. They showed no evidence of damage to the peripheral auditory or vocal mechanisms; hearing thresholds were normal for both pure tones and speech, and motor development was normal including activities of the tongue, lips, and jaw. Neurological examinations when available revealed no obvious central nervous system dysfunctions, and psychological tests indicated that measured intelligence was average or above average These children had been categorized as having a specific language disorder because their language development alone was retarded. Since children with a specific language disorder do not suffer from any obvious peripheral damage or from any intellectual retardation, they give us an opportunity to study the organization of central encoding and decoding in language processing in the following manner. It is possible that we might be able to trace the connections between components of the grammar by observing the sequences of difficulties encountered by these children. Using normal-speaking control subjects, we can also attempt to observe whether this behavior is unique or if it represents some extreme instances of average behavior.

One question asked in the study was whether or not phonological processing takes place with the same accuracy and in the same manner when processing is dependent and when it is independent of semantic-syntactic processing. The accuracy and pattern of recall of meaningful and meaningless phonological sequences was tested and measured by asking both groups of children, deviant-speaking and normal-speaking, to repeat phonological sequences that comprise words in sentences and nonsense syllables in a sequence. The words appeared in sentences that were 3-5 words in length and were either active-declarative, imperative, negative or question sentences. Every American-English singleton consonant appeared at least once in initial, medial, and final position in the words. The nonsense syllables were given for repetition in a series of three and were of two types. The first type was repetition of the same consonant in each of three syllables with the vowel altered, and the second type was repetition of the same vowel (/a/) with the consonant altered. In the first type of stimuli all of the consonants in American English were given for repetition in initial, medial, and final position. As an example of the stimuli presented for repetition, the following are the three triplets presented for the consonant /b/: (i) /bi/, /ba/, /bu/; (ii) /abi/, /aba/, /abu/; (iii) /ib/, /ab/, /ub/. In the second type of stimuli consonants in the triplet were varied according to place of articulation (for example, /ba/, /da/, /ga/) and according to manner of articulation (/ba/, /pa/, /ma/). Repetitions of words and nonsense syllables were recorded and then transcribed and an analysis was made of consonantal errors. Some of the results of that analysis are presented here.

The deviant-speaking children (hereafter referred to as the Experimental group) made a significantly greater percentage of errors in repetition of consonants than the normal-speaking children (hereafter referred to as the Control group) when repeating both nonsense syllables and words. The children in the experimental group made a significantly greater percentage of errors, however, when repeating type 1 nonsense syllables (varying in vowel) than when repeating words, and the children in the control group exhibited a tendency to do the same. Meaningless and meaningful phonological sequences appeared to be recalled differently by both groups of children but this difference was significantly more marked in the experimental group. Table IX-1 presents the percentage of errors made with consonants in the repetition of type 1 nonsense syllables and words by both groups of children.

No significant differences were found in the percentage of consonantal errors made as a function of the position of the consonant in the words by either group. A significantly greater percentage of errors were made, however, with medial and final

Table IX-1. Percentage of errors made with words and nonsense syllables by both groups.

| Words | | | | Nonsense Syllables | | | |
|---|---|---|---|---|---|---|---|
| Experimental | | | Control | Experimental | | | Control |
| % Error | $X^2$ | p value | % Error | % Error | $X^2$ | p value | % Error |
| 15% | 5.0 | .05 | 5% | 43% | 16.1 | .01 | 13% |

| Control | | | | Experimental | | | |
|---|---|---|---|---|---|---|---|
| Words | | | Nonsense Syllables | Words | | | Nonsense Syllables |
| % Error | $X^2$ | p value | % Error | % Error | $X^2$ | p value | % Error |
| 5% | 3.56 | .10 | 13% | 15% | 13.52 | .01 | 43% |

consonants than with initial consonants in type 1 nonsense syllables by the children in the experimental group. No significant differences in percentage of errors because of consonant position in type 1 nonsense syllables were found in the control group. The control group did, however, make a significantly greater percentage of errors with syllable triplets that varied in initial consonant than they did with type 1 nonsense syllables. This was not the case in the experimental group. The difference in performance of the control group with type 1 and type 2 nonsense syllables is, perhaps, due to the fact that they did not expect and were therefore unprepared for a change in the system – that is, a change in consonants. Interestingly, this effect was not observed in the experimental group. Both types of nonsense syllables were treated, on the whole, as being equally difficult. The fact that position of consonant plays a more important role in the repetition of nonsense syllables than words again makes it appear that meaningful and meaningless sequences are processed differently at least by the children in the experimental group. The percentage of errors attributable to position and type of stimuli are presented in Table IX-2.

All of the results indicated that the children in the experimental group were not as capable of analyzing phonological sequences in terms of segmental or syllabic phonological features as the children in the control group. Since a great many of the errors made in repetition were substitution errors, one other aspect of the performance of the two groups that was examined was the percentage of

Table IX-2. Percentage of errors according to position in sequence and type of nonsense triplet.

| | Nonsense Syllables Type 1 | | Words | |
|---|---|---|---|---|
| Position | Experimental | Control | Experimental | Control |
| | % Error | % Error | % Error | % Error |
| Initial | 26% | 10% | 14% | 4% |
| Medial | 46% | 13% | 20% | 7% |
| Final | 60% | 15% | 15% | 5% |
| Nonsense Syllables Type 2 | | | | |
| Type | | | | |
| Place | 54% | 47% | | |
| Manner | 50% | 43% | | |

maintenance of original features in the repetition and substitution of consonantal segments in the words and two types of nonsense syllables. The results of this analysis for words and type 1 nonsense syllables are presented in Fig. IX-8. The similarity in the

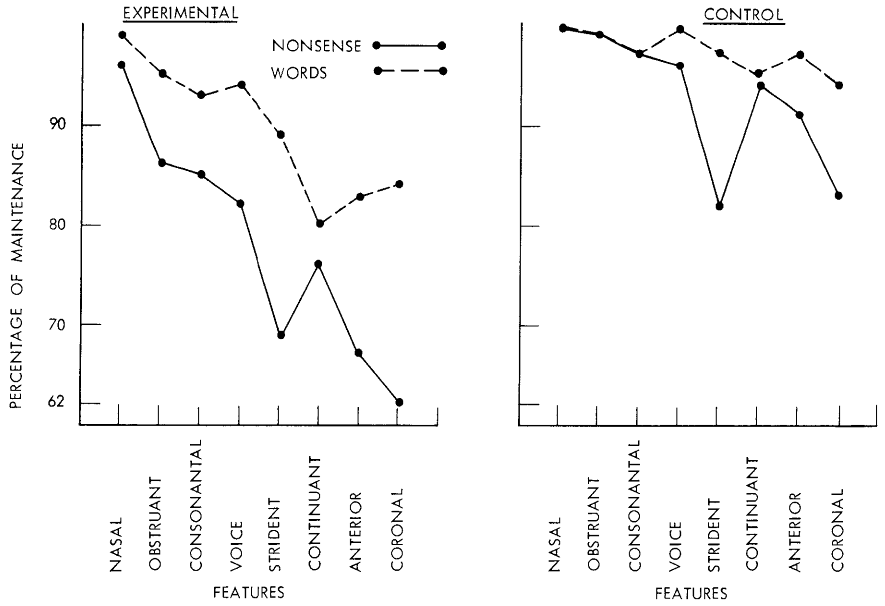


Fig. IX-8. Percentage of feature maintenance.

pattern of feature maintenance for both types of stimuli can be seen.

The percentage of errors in feature maintenance was markedly below the percentage of errors in consonantal repetition, regardless of context. Furthermore, the rank order correlations of feature maintenance between groups and between words and type 1 nonsense syllables was very high. The rank order correlation of feature maintenance between nonsense syllables of type 1 and type 2 syllables that varied in manner was not so high. These results are given in Table IX-3.

These data indicate that despite the fact that the accuracy of recall of consonants

Table IX-3. Percentage of errors in feature maintenance and rank order correlations of feature maintenance.

% Errors in Feature Maintenance

| Words | | | | Nonsense Syllables | | | |
|---|---|---|---|---|---|---|---|
| Experimental | $X^2$ | p value | Control | Experimental | $X^2$ | p value | Control |
| 10% | 2.77 | .10 | 3% | 22% | 7.76 | .01 | 7% |

| Control | | | | Experimental | | | |
|---|---|---|---|---|---|---|---|
| Words | $X^2$ | p value | N. S. | Words | $X^2$ | p value | N. S. |
| 3% | .9 | | 7% | 10% | 4.50 | .05 | 22% |

Rank Order Correlations of Feature Maintenance

| Group | Correlation | Spearman R. O. |
|---|---|---|
| Experimental | N. S. and Words | .82 |
| Control | N. S. and Words | .82 |
| Words | Experimental and Control | .86 |
| N. S. | Experimental and Control | .92 |
| Experimental | N. S. Type 1 and Manner Type 2 | .50 |
| Experimental | N. S. Type 1 and Place Type 2 | .86 |
| Control | N. S. Type 1 and Manner Type 2 | .66 |
| Control | N. S. Type 1 and Place Type 2 | .83 |

N. S. = Nonsense Syllables.
R. O. = Rank Order.

was depressed when the stimuli were meaningless and that this was especially the case with language-disordered children, the same pattern of recall in terms of what aspects were remembered could be observed with both nonsense syllables and words. On the whole, features of original consonants best remembered in recalling words were also best remembered in recalling nonsense syllables. Therefore, one might speculate that descriptions of language processing that represent lexical categorization as including semantic properties and phonological features, rather than segmental or syllabic characterization of words, have some psychological reality. Furthermore, one might speculate from these data that speech sound processing takes place in much the same manner as lexical processing and can be independent of it, but that, overall, accuracy of recall is affected by whether or not lexical categorization can be employed.

One other aspect of these children's phonological processing was observed which also indicates that the children in the experimental group tended to rely heavily on the lexical categorization of sequences as words rather than as segments or syllables to reproduce them. The children in the experimental group recalled final consonants in words in a manner that depended on the syntactic role of these consonants. If the consonants marked tense or plurality, they were usually omitted. Nevertheless, if they were part of the word, they were usually repeated or substituted by another consonant. For example, the verb "makes" ($3^{rd}$ person singular) was most frequently repeated as "make," whereas the word "dress" was most frequently repeated correctly or with /t/ substituted for the final /s/. The verb "going" was most frequently repeated as "goin" or "go," whereas the word "nothing" was most frequently repeated as "nutting" or "nudding." The plural nouns "leaves," "shoes," "toys" were frequently repeated with the final /z/ omitted, whereas the word "nose" was most frequently repeated with the final strident unvoiced. The final tense marker in the verbs "gazed," "named," "examined" was most frequently omitted, whereas the final /d/ in "friend" was most frequently repeated correctly. These differences in phonological recall occurred in addition to having many of the so-called function words completely changed in repetition. Changes such as "he" recalled as "him," "I" as "me," "his" as "him," "she" as "her," and so forth were considered separately. The data are intriguing but very limited. Therefore, an experiment has been planned in which many instances of this contrast between word ending and syntactic marker will be given for repetition.

In summary, normal-speaking children tend to recall words more accurately than nonsense syllables, even though the words appear in longer sequences than the nonsense syllable triplets. Deviant-speaking children recall words and word stems minus their grammatical markers significantly more accurately than nonsense syllables in terms of number of segmental errors. Given these facts, it seems reasonable to suppose that the direction of processing is that phonological sequences are identified first in terms of their lexical form (that is, in terms of semantic properties <u>and</u> phonological features) before being analyzed in terms of actual speech sound sequences – syllabic or segmental. This sequence of analysis is probably true in the development of language, as well as in listening to and understanding sentences, since the normal-speaking children tend in this direction. But there is some evidence that phonological sequences are processed in much the same manner independent of their lexical categorization – that is, in terms of phonological features.

Paula Menyuk

References

1. This study was carried out in collaboration with Patricia L. Looney of the Children's Hospital Medical Center, Hearing and Speech Division, Boston, Massachusetts, and will be reported in detail in a forthcoming publication, entitled "A Model of Language Processing: Evidence from Children with and without Language Disorders."
2. The method used to determine percentage of feature maintenance is described in Paula Menyuk, "The Role of Distinctive Features in Children's Acquisition of Phonology," J. Speech Hearing Res. <u>11</u>, 844-860 (1968).

## C. THE ROLE OF FORMANT TRANSITIONS IN THE VOICE-VOICELESS DISTINCTION FOR STOPS

The acoustic cues for the distinction between voiced and voiceless stop consonants in initial position in English have been examined in detail through measurements on spoken consonant-vowel syllables (Lisker and Abramson[1]), and through experiments in which the acoustic parameters of synthetic stimuli are manipulated (Liberman, Delattre, and Cooper[2]). The data have suggested that the acoustic characteristic providing the simplest and most direct indication of whether a stop consonant is voiced or voiceless is the time from the release of the closure to the onset of vocal-cord vibration, called the voice-onset time (VOT). If this time is greater than ~40 ms, the consonant is voiceless; for smaller voice-onset times, the consonant is voiced. The data from natural utterances show that the voice-onset times for the two classes of consonants are clustered around 0-10 ms and about 50 ms or more, and intermediate voice-onset times are rarely found, especially for stressed consonant-vowel syllables. [These numbers apply to labials and apicals; they are somewhat greater for velars.] Furthermore, identification and discrimination judgments for synthetic stimuli indicate that there is a sharp perceptual boundary between voiced and voiceless responses. Discrimination of pairs of stimuli within a given class is poor, while discrimination for stimuli near the phoneme boundary is good (Liberman, Harris, Kinney, and Lane[3]).

Recent experiments on the discrimination of synthetic syllables by infants (Eimas et al.[4]) has drawn further attention to the importance of the voice-onset time as a property that the auditory mechanism uses to categorize speech sounds into classes. The Eimas experiments suggest that infants cannot discriminate changes in VOT within a region that adults would classify as one phoneme, but can readily discriminate changes of similar magnitude spanning a phoneme boundary.

A conclusion that one would tend to make from all of these findings is that the auditory system is, as it were, "wired" to place a boundary at some fixed absolute noise duration. Study of a variety of phonetic features indicates, however, that the acoustic differences between minimal pairs of phonemes in a given phonetic environment appear to be characterized by distinctively different properties rather than by differences in the magnitude of a given acoustic parameter that takes on a range of values (Jakobson, Fant, and Halle[5]; Stevens[6]). As far as we know, there is no evidence to suggest that a 40-ms noise duration preceding a vowel-like sound represents a natural perceptual boundary.

These considerations have prompted a further investigation of the acoustic properties of voiced and voiceless stop consonants. Examples of spectrograms of the syllables /da/

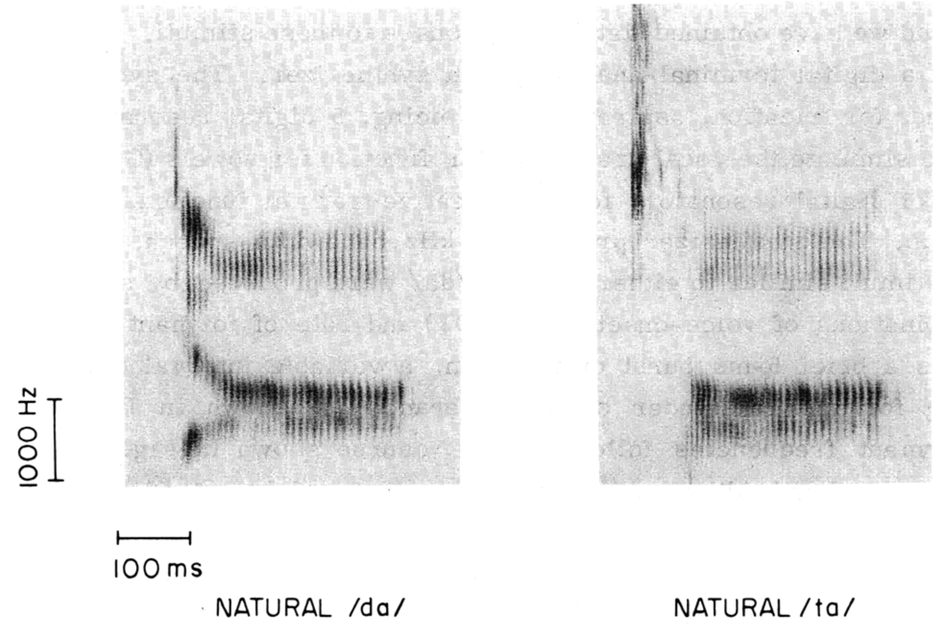


Fig. IX-9. Spectrograms of the syllables /da/ and /ta/ as spoken in isolation by an English talker.

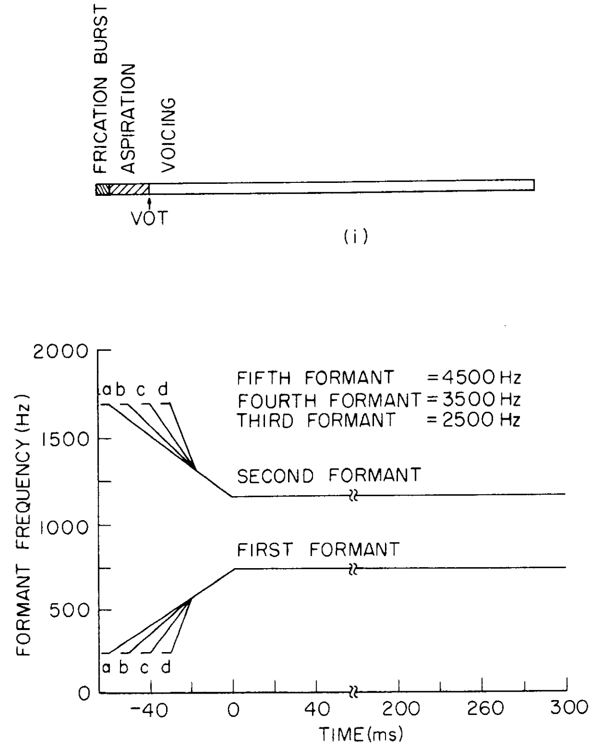
and /ta/ are shown in Fig. IX-9. The difference in VOT is clearly seen in the spectrograms. Another difference between the sounds, however, is that the voiced stop has a well-defined transition in the first formant (as well as in the second formant) after the onset of voicing, whereas the $F_1$ transition for the voiceless stop is essentially nonexistent after the onset of voicing. The lack of formant transitions after voicing onset for the aspirated consonants indicates that the rapid movements of the supraglottal articulators (the tongue tip in the case of Fig. IX-9) are essentially complete before the vocal cords are in a configuration where voicing begins. It is known, in fact, that the duration of these transitions is of the order of 40 ms, based both on synthesis experiments (Liberman, Delattre, Gerstman, and Cooper[7]), and on cineradiographic data (Kent and Moll[8]). (The durations are slightly longer for velars and shorter for labials.)

These examples suggest that one of the possible cues for the voiced-voiceless distinction for stops is in the presence or absence of a significant and rapid spectrum change at the onset of voicing. According to this hypothesis, absence of such a spectrum change would be a requirement for perception of a well-formed voiceless stop. It is known, of course, that a rapid spectrum change always occurs at the release of a (nonglottal) stop consonant. We are suggesting here that, in the case of a voiceless stop consonant, another rapid spectrum change at the onset of voicing is a negative cue for voicelessness.

In order to test this hypothesis, we have generated a series of synthetic consonant-vowel stimuli in which voice-onset time and transition durations were manipulated independently, and we have obtained listener responses to these stimuli. The stimuli were produced on a digital terminal-analog speech synthesizer. The synthesizer comprises sources for frication, aspiration and voicing, 5 digital resonators connected in cascade to simulate the vocal-tract transfer function for vowels (Gold and Rabiner), another set of digital resonators for the fricative transfer function, and a radiation characteristic. The synthesizer produces 5-kHz bandwidth speech.

Sixteen stimuli similar to either /ta/ or /da/ were produced by selecting a number of combinations of voice-onset time (VOT) and rate of formant motion. Each stimulus was a brief 5-ms burst of frication, a variable interval of aspiration, and then voicing for the remainder of the utterance, as shown in Fig. IX-10 (Upper trace). Formant frequencies follow a time course shown in Fig. IX-10 (Lower trace). Formant bandwidths were set to constant values of 50, 80, 120, 180, 300 Hz for the first through fifth formants, respectively.

Four formant trajectories are defined in Fig. IX-10, corresponding to moderate (a) through rapid (d) consonant-vowel transition times. Four values of voice-onset time were selected for each trajectory. All transition times are within the range that yields a reasonable /da/ stimulus for the shortest VOT's. Preliminary



(i)

(ii)

Fig. IX-10. A synthetic utterance consists in a frication burst, a variable interval of aspiration, followed by onset of voicing, as shown in part (i). The two lowest formant frequencies follow one of four possible trajectories that are labeled (a) through (d) in part (ii) of the figure.

tests were performed in order to place the /d-t/ phoneme boundary within the range of voice-onset times for each trajectory. VOT values are given in Table IX-4 for all stimuli. Spectrograms of an acceptable synthetic /ta/ (stimulus #4), an acceptable /da/ (stimulus #5) are shown in Fig. IX-11. There are significant formant transitions for the synthetic /da/, whereas the transitions are essentially completed for /ta/ (although a small residual transition remains in this case).

The values of other synthesis parameters as a function of time have an important influence on the outcome of the experiment, so the synthesis strategy will be described in some detail. Control parameters were updated at discrete 5 ms intervals. The fundamental frequency was held fixed a 130 Hz from voicing onset to 80 ms and fell linearly to 100 Hz from 80 ms to 350 ms. The first voicing pulse in an utterance occurred precisely at the onset of voicing.

The frication burst spectrum was produced by exciting a single resonator set to the formant frequency and bandwidth values of the 5th formant, as is to be

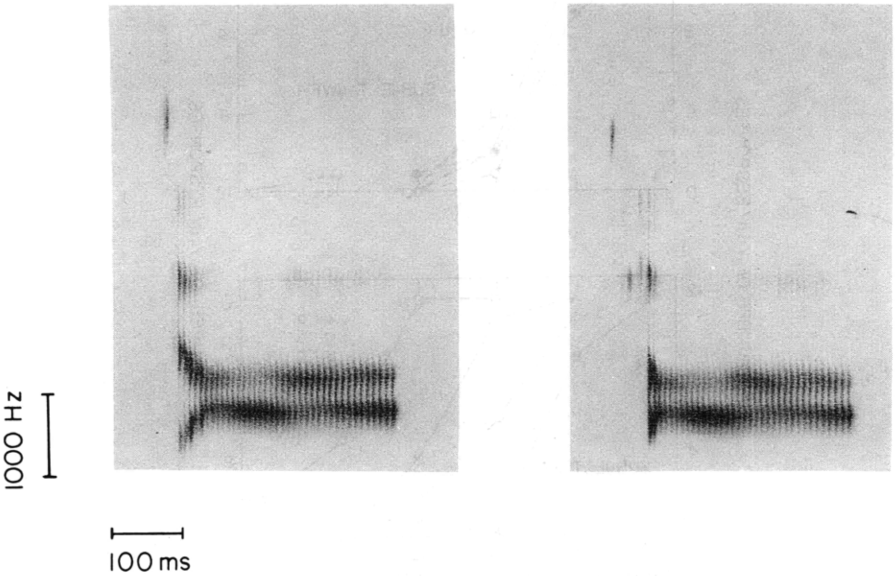
Table IX-4. The 16 stimuli are defined in terms of the formant transition types shown in Fig. IX-11 and the voice-onset time with respect to frication burst onset.

| Stimulus Number | Formant Transition Type | Voice Onset Time (VOT) |
|---|---|---|
| 1 | a | 25 |
| 2 | a | 35 |
| 3 | a | 45 |
| 4 | a | 55 |
| 5 | b | 15 |
| 6 | b | 25 |
| 7 | b | 35 |
| 8 | b | 45 |
| 9 | c | 15 |
| 10 | c | 25 |
| 11 | c | 35 |
| 12 | c | 45 |
| 13 | d | 15 |
| 14 | d | 25 |
| 15 | d | 35 |
| 16 | d | 45 |

expected from analysis of natural speech (Stevens[10]). Frication source amplitude was set to produce a 5th-formant level in the burst approximately 6 dB greater than the 5th-formant level during the following steady vowel.

The amplitude control for the aspiration source was set to a constant level during the aspiration interval and was off otherwise. The amplitude was set so that the excitation in the region of F3 and F4 remained at the same level in going from aspiration to voicing (Stevens[11]); i.e., there was no discontinuity in level in this frequency region in passing from noise to buzz excitation. There is a slight difference between the level of normal aspiration and the synthetic aspiration at aspiration onset, as can be seen by comparing the /ta/ examples of Figs. IX-9 and IX-11 because the first formant always began at 250 Hz in the synthesis, whereas the open glottis condition applying to normal aspiration production results in a higher F1 with a concomitant higher level of aspiration.

The amplitude of voicing was set to a constant level during the voicing interval and was off otherwise. Again this produced a slight difference between normal



SYNTHETIC /da/ SYNTHETIC /ta/

Fig. IX-11. Spectrograms of two of the synthetic utterances. Stimulus #5 was unanimously identified as /da/ and stimulus #4 was unanimously identified as /ta/ by the listeners.

voicing onset and synthetic onset, as can be seen by comparing the /ta/ examples of Figs. IX-9 and IX-11 because the first few periods of voicing in a voiceless-voiced transition of natural speech are not as intense as in a steady-state vowel (even though the first formant has already attained a high value before voicing onset in /ta/).

Tape recordings of some replications of these 16 stimuli in random order were prepared, and were presented over monaural headphones to 5 listeners, who were asked to identify each stimulus as /da/ or /ta/. Each listener made 12 responses to each of the stimuli. Results for two of the listeners are shown in Fig. IX-12. These data represent, in a sense, two extremes of performance of the listeners. For one listener (AWFH), the boundary between /da/ and /ta/ responses occurred when the onset of voicing was at a fixed time relative to completion of the transitions, independent of the time of release of the consonant. In the case of the other listener (VZ), there was a range of voice-onset times relative to the completion of the formant transition. The voice-onset time relative to the consonantal release can be obtained by adding 65 ms to the abscissa for stimulus type (a), 55 ms for (b), 45 ms for (c), and 35 ms for (d). Thus for listener VZ, the absolute voice-onset

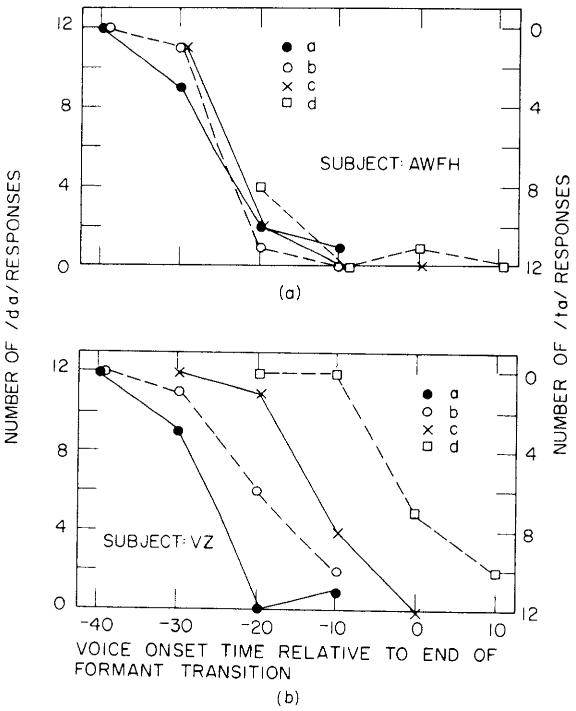


Fig. IX-12. Identification data for two subjects. The number of /da/ responses is plotted as a function of time of onset of voicing, where zero on the scale corresponds to the completion of the formant transition shown in Fig. IX-10. A family of 4 curves is drawn to indicate how the responses change as a function of rate of formant motion.

times at the 50% points are 39, 35, 32, and 33 ms, respectively.

For all listeners there was a tendency for the absolute voice-onset times to be longer for the stimuli with slower transitions, and there appeared to be a minimum VOT that was acceptable as /t/ (~32 ms in the case of VZ).

Average data for the 5 listeners are shown in Fig. IX-13. Each point represents the average VOT at the 50% response point for the five listeners. For this figure, the abscissa is the voice-onset time relative to the consonantal release. The horizontal line indicates a constant voice-onset time; the data points would lie along this line if the absolute VOT provided the only cue for the voiced-voiceless distinction. The line sloping up to the left represents the locus of constant duration of the transition from voicing onset to termination of the transition. Data points would lie along a line with this slope if the cue for the voiced-voiceless distinction were the presence or absence of substantial formant transitions after voicing onset. The line in Fig. IX-13 is drawn, in fact, for a time of voicing onset of 23 ms, as indicated by the arrow in the inset. The amount of spectrum change occurring as a result of the rising transition of the first formant during this interval is relatively small compared with the spectrum change resulting from the entire transition from 250-750 Hz. The average rise in intensity of the higher formants due to the $F_1$ transition is about 5 dB in the former case and approximately 20 dB in the latter (Fant[12]; Stevens and House[13]).

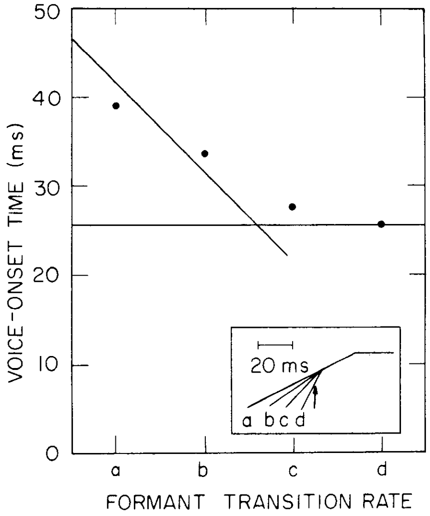


Fig. IX-13.

The identification data of 5 subjects is summarized by plotting the average VOT at the phoneme boundary for each rate of formant motion. The VOT in this case is the time from consonantal release to onset of voicing. The letters on the abscissa are identified with different rates of transition, as indicated in the inset. A horizontal line could be fitted to the data if the phoneme boundaries were based on the delay of voicing with respect to plosive release. A sloping line could be fitted to the data if the phoneme boundaries were based on the presence or absence of a detectable formant transition (or of a fixed duration of residual transition) during the initial portion of the voicing. The sloping line corresponds to a time of voicing onset indicated by the arrow in the inset.

The data points of Fig. IX-13 show that subjects require a VOT in excess of 25 ms, on the average, if the stimulus is to be accepted as a /t/, even if there is a negligible $F_1$ transition after onset of voicing. On the other hand, for longer $F_1$ transitions a longer VOT is necessary is a /t/ response is to be obtained. As Fig. IX-12 has indicated, subjects differ somewhat in the degree to which they rely on the lack of an $F_1$ transition as a cue for /t/.

In order to examine further the role of the $F_1$ transitions following voicing onset as a negative cue for voicelessness, stimuli with a number of different VOT's were synthesized, with and without a rising $F_1$ transition after onset of voicing. Subjects were asked to identify the initial consonants as /t/ or /d/ and to rate the adequacy of the consonants on a four-point scale. Stimuli with long VOT's (40 ms or more) were usually identified as /t/, but if there was an $F_1$ transition, these stimuli received a low rating.

It can be concluded, therefore, that for adult listeners both the VOT and the degree of $F_1$ transition are cues for the voiced-voiceless distinction for stop consonants in English. These stop consonants are all characterized by a rapid spectrum change in the initial 20-40 ms following the release. Events in this brief time interval provide cues for place of articulation (Stevens[11]). The interval is entirely voiceless for an aspirated stop in English and, for the most part, is voiced for an unaspirated stop.

The findings of this experiment are consistent with the data of Liberman, Delattre, and Cooper,[2] who examined the responses of listeners to a series of consonant-vowel stimuli that differed only in the amount of cutback of the first formant relative to the onset time. In other words, the stimuli were produced by starting with a sound like that of /da/ in Fig. IX-10, and by "blanking out" the initial part of the $F_1$ transition in varying amounts. Their data show that the stimuli with larger $F_1$ cutback were identified as beginning with /t/. The boundary between /d/ and /t/ responses occurred when the cutback was at a point where most of the $F_1$ transition was completed.

The results of the present experiments, as well as those of Liberman, Delattre, and Cooper, suggest a way of interpreting the results of Eimas et al.[4] on the response of infants to the t-d distinction. The auditory system provides one kind of response when there is a significant transition in $F_1$ after onset of voicing, thereby creating a rapid spectrum change or transient, and another kind of response when there is no such transient. Stimuli that differ in this property yield different responses; if both members of a pair of stimuli have either one property or the other, they are identified as being the same. It may well be that this transient provides the primary cue for the infant with little previous exposure to speech. At a later age when he learns to use the voiced-voiceless distinction, he finds that other cues, such as absolute duration of VOT, accompany this primary cue, and he may give some weight to these subsidiary cues.

K. N. Stevens, D. H. Klatt

## References

1. L. Lisker and A. S. Abramson, "A Cross-language Study of Voicing in Initial Stops: Acoustical Measurements," Word 20, 384-422 (1964).
2. A. M. Liberman, P. C. Delattre, and F. S. Cooper, "Some Cues for the Distinction between Voiced and Voiceless Stops in Initial Position," Language and Speech 1, 153-167 (1958).
3. A. M. Liberman, K. S. Harris, J. A. Kinney, and H. Lane, "The Discrimination of Relative Onset-Time of the Components of Certain Speech and Non-Speech Patterns," J. Exp. Psychol. 61, 379-388 (1961).
4. P. D. Eimas, E. R. Siqueland, P. Jusczyk, and J. Vigorito, "Speech Perception in Infants," Science 171, 303-306 (1971).
5. R. Jakobson, C. G. M. Fant, and M. Halle, Preliminaries to Speech Analysis (The M.I.T. Press, Cambridge, Mass., 1963).
6. K. N. Stevens, "The Quantal Nature of Speech: Evidence from Articulatory-Acoustic Data," in E. E. David, Jr. and P. B. Denes (eds.), Human Communication: A Unified View (to be published by McGraw-Hill, New York).
7. A. M. Liberman, P. C. Delattre, L. J. Gerstman, and F. S. Cooper, "Tempo of Frequency Change as a Cue for Distinguishing Classes of Speech Sounds," J. Exp. Psychol. 52, 127-137 (1956).
8. R. D. Kent and K. Moll, "Vocal-Tract Characteristics of the Stop Cognates," J. Acoust. Soc. Am. 46, 1549-1555 (1969).

9. B. Gold and L. R. Rabiner, "Analysis of Digital and Analog Formant Synthesizers," IEEE Trans. on Audio and Electroacoustics, Vol. AU-16, pp. 81-94, 1968.
10. K. N. Stevens, "Acoustic Correlates of Place of Articulation for Stop and Fricative Consonants," Quarterly Progress Report No. 89, Research Laboratory of Electronics, M.I.T., April 15, 1968, 199-205.
11. K. N. Stevens, "The Role of Rapid Spectrum Changes in the Production and Perception of Speech" (to appear in *Form and Substance* (Festschrift for Eli Fischer-Jørgensen)).
12. C. G. M. Fant, "On the Predictability of Formant Levels and Spectrum Envelope from Formant Frequencies," in M. Halle and others, *For Roman Jakobson* (Mouton and Co., The Hague, 1956), pp. 109-120.
13. K. N. Stevens and A. S. House, "An Acoustical Theory of Vowel Production and Some of Its Implications," J. Speech Hearing Res. 4, 303-320 (1961).

## D. A NOTE ON LARYNGEAL FEATURES

In this report we investigate the mechanisms that underlie various phonetic features such as voicing, aspiration, and glottalization, which for want of a better term we may designate by the adjective "laryngeal." Our purpose here is to give wider currency to certain results of recent acoustic investigations (Stevens[1, 2] and Kim[3]) and and to support modifications in the universal phonetic feature framework which seem to us to be indicated by these new results.

### 1. Review of Acoustical and Mechanical Aspects of Vocal-Cord Operation

The acoustical analysis is based on a model that represents each vocal cord as a mass that can change in shape and which forms a flexible wall for the glottal opening, as shown by the lateral section through the glottis in Fig. IX-14. When a subglottal pressure $P_s$ is applied, the glottis assumes a configuration that may resemble that shown in Fig. IX-14, with an average static opening $w_s$ between the vocal cords. The pressure $P_g$ in the glottis, which arises from the glottal airflow and from the pressures $P_{sup}$ and $P_{sub}$ above and below the glottis, causes an outward force on the vocal cords that in the static situation is exactly balanced by the restoring force, because of the stiffness of the cords.

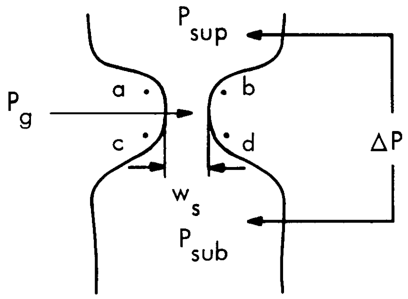


Fig. IX-14. Lateral section through glottis for a typical vocal-cord configuration. $P_g$ represents the average pressure in the glottis, $P_{sub}$ is the subglottal pressure, and $P_{sup}$ is the supraglottal pressure. $P_{sup}$ is zero and $\Delta P = P_{sub}$ in the case of a nonobstruent. The upper edges of the vocal cords are identified by points ab, and the lower edges by cd. The average glottal width is $w_s$.

Under certain conditions of stiffness of the glottal walls, the static opening $w_s$, and the pressure $\Delta P$ across the glottis, the system is unstable, and in-and-out vibration of the vocal cords occurs. In general, the oscillations are not in phase across the thickness of the glottis: the outward and inward displacement of the lower edges (at points c, d in Fig. IX-14) occur slightly ahead of the oscillatory displacement of the upper edges (at points a, b). In fact, analysis shows (Ishizaka and Matsudaira[4]) that this out-of-phase vibration of the upper and lower margins of the vocal cords is the mechanism primarily responsible for the transfer of energy from the steady glottal airflow to the vibrating vocal cords in the mode of vibration normally used in speech. This energy transfer is essential for the maintenance of oscillation.

The various acoustic effects that occur as a result of larynx manipulations are: (i) no vocal-cord vibration, negligible airflow, and hence little or no generation of turbulence noise at the glottis; (ii) no vocal-cord vibration, appreciable airflow and turbulence noise generation at the glottis; and (iii) vocal-cord vibration, which may or may not be accompanied by turbulence noise generation. When vocal-cord vibration occurs, the frequency of vibration and the waveform of the airflow pulse that passes through the glottis for each vibratory cycle can be changed through appropriate muscular adjustments. The first question that we seek to answer in our analysis of vocal-cord operation is, What are the ways in which the configuration of the vocal cords can be manipulated by the laryngeal musculature in order to produce distinctive acoustic end products that are potentially useful for the formation of phonetic categories?

The principal manipulations of the musculature that are available to produce these acoustic effects are: (i) adduction or abduction of the vocal cords by appropriately positioning the arytenoid cartilages, thereby changing the static glottal opening $w_s$ in Fig. IX-14; and (ii) stiffening or slackening of the vocal cords through adjustments of the thyroarytenoid and cricothyroid muscles, thereby changing the flexibility of the glottal walls. Theoretical analysis shows that stiffening of the vocal cords tends to raise the frequency of vibration, whereas slackening lowers the frequency. Slackening or stiffening of the vocal cords also has an influence on whether or not there is vocal-cord vibration. With slack vocal cords, there is a relatively wide range of values of glottal opening over which glottal vibration occurs, and the pressure across the glottis can be quite small. When the vocal cords are stiff, the range of glottal widths over which vibration occurs is greatly reduced, and a larger pressure across the glottis is required.

Figure IX-15 illustrates the regions of vocal-cord configurations and pressure drops over which glottal vibration can occur. For the most part, the curves in this figure are derived from a theoretical analysis of the vocal cords, based on a representation of each vocal cord as two masses corresponding roughly to the upper edges a (or b) and to the lower edges c (or d), respectively, in Fig. IX-14. These masses in the model are coupled together by a "spring" whose stiffness is determined by the stiffness or slackness of the tissues that form the glottal walls. On the abscissa is plotted the glottal opening $w_s$ that is obtained if the vocal cords were to assume a static position without beginning to oscillate. If oscillations do occur, this value of $w_s$ can be interpreted as the initial glottal opening that is the starting point for the build-up of oscillations. This glottal width $w_s$ may be slightly different from the static opening that would occur if no subglottal pressure were applied and hence if there were no airflow. The ordinate in Fig. IX-15 is the pressure $\Delta P$ across the glottis. In the case of nonobstruent vocal-tract configurations, $\Delta P$ is simply the subglottal pressure $P_{sub}$; for obstruent consonants, $\Delta P$ is the difference between $P_{sub}$ and the supraglottal pressure $P_{sup}$. Under most circumstances, the maximum value of $\Delta P$ is $P_{sub}$, corresponding to zero supraglottal pressure (although there is the possibility of producing a negative $P_{sup}$ during implosive stop consonants).

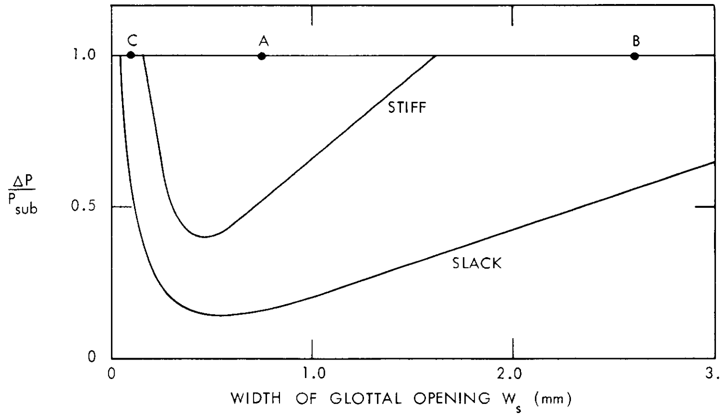


Fig. IX-15. Sketch showing approximate ranges of conditions under which vocal-cord vibration occurs. $\Delta P$ is the pressure across the glottis, $P_{sub}$ is the subglottal pressure, and $w_s$ is the static width that would be assumed by the glottis if there were no vibrations. If the values of $\Delta P$ and $w_s$ give rise to a point above the curve labeled "slack," then vocal-cord vibration is initiated when the vocal-cord stiffness is small. Likewise the curve labeled "stiff" represents the boundary of vocal-cord oscillation for relatively stiff vocal cords. Below these lines, the vocal cords remain in a static position with no oscillations. Points A, B, and C represent glottal widths that lie within regions of "normal" glottal vibration, spread glottis and constricted glottis for nonobstruents, for which $\Delta P = P_{sub}$. The portion of the chart corresponding to obstruent configurations is well below the line $\Delta P/P_{sub} = 1$. The regions are based on an assumed subglottal pressure of ~8 cm $H_2O$. The shapes of the curves for $w_s > 0.5$ mm are derived from theoretical analysis of a two-mass model of the vocal cords (Ishizaka and Matsudaira[4] and Stevens[2]). For smaller values of $w_s$ the curves are estimated.

Two contours are drawn in Fig. IX-15: one represents the limiting conditions for oscillation of the vocal cords when they are relatively slack (relatively small coupling between upper and lower edges of the vocal cords in Fig. IX-14); the other represents the threshold of oscillation for relatively stiff vocal cords. When the values of pressure drop $\Delta P$ and glottal opening $w_s$ lead to a point below one of these curves, then vocal-cord oscillation for the stiffness condition identified by the curve does not occur. When, on the other hand, the point represented by $\Delta P$ and $w_s$ lies above the curve, there are oscillations of the vocal cords.

The horizontal line at $\Delta P = P_{sub}$ corresponds to nonobstruent configurations, i.e., vowels and glides. Three regions of glottal opening can be identified along this line, centered on points A, B, and C. In the region surrounding point A, vocal-cord vibration can occur whether the vocal cords are stiff or slack, and the stiffness adjustment simply alters the frequency of vibration. Point B identifies a spread glottal configuration for which vocal-cord vibration occurs only for the slack condition but not for the stiff condition. When the vocal cords are stiff, there is a rapid flow of air through the glottis with this glottal opening, giving rise to turbulence noise or aspiration. For narrow glottal openings around point C, oscillations also occur for slack vocal cords but not for stiff vocal cords. With a constricted glottis and stiff vocal cords, there is essentially no airflow through the glottis and hence no turbulence noise is generated. When the pressure across the glottis is reduced below the subglottal pressure by forming a supraglottal constriction and hence increasing the supraglottal pressure, there is a decreased range of glottal openings for which vocal-cord vibration occurs. Thus, for example, if $\Delta P$ is reduced below about $1/2\ P_{sub}$, oscillations can no longer be initiated when the vocal cords are stiff, regardless of the value of $w_s$.

### 2. Proposed Laryngeal Features

In view of the theoretical findings summarized schematically in Fig. IX-15, we postulated that there are two independently controlled parameters of the model: the stiffness of the vocal cords, and the static glottal opening. Manipulation of these parameters to particular ranges of values gives rise to distinct and well-defined acoustic consequences. These parameters are the outcome of adjustments of specific groups of intrinsic laryngeal muscles.

The four features listed below represent adjustments in these glottal parameters that appear to be relevant to the classification of phonetic segments, and we propose that they should be included in the universal phonetic feature framework, to replace such traditional features as voicing, aspiration, glottalization, vowel pitch, and so forth.

1. Spread glottis. By rotation and displacement of the arytenoid cartilages, the vocal cords can be displaced outward relative to their positions for normal voicing, leaving a large glottal width. If the vocal-cord stiffness is sufficiently large, the combination of wide glottis and stiff glottal walls inhibits vocal-cord vibration. On

the other hand, slackening of the glottal walls by reducing the stiffness can lead to a condition in which vocal-cord vibration will occur, even with a relatively wide glottal opening.

2. Constricted glottis. Adduction of the arytenoid cartilages relative to the position for normal voicing (accomplished, perhaps, by fibers of the thyroarytenoid muscles, as well as by the lateral cricoarytenoid muscles) can cause the vocal cords to be pressed together and the glottis to narrow or to close. When the vocal-cord stiffness is large in this situation, vocal-cord vibration does not occur, and no air passes through the glottis. For a lower coupling stiffness, vocal-cord vibration can be initiated, probably with relatively narrow, peaked pulses.

3. Stiff vocal cords. Increasing the stiffness of the vocal cords makes the coupling between upper and lower edges of the vocal cords larger. Stiffening of the vocal cords affects glottal vibration, regardless of the size of the glottal aperture. When the vocal cords are in a configuration for normal voicing (neither spread nor constricted), the rate of vocal-cord vibrations increases with increasing stiffness. Increased stiffness of the vocal cords will inhibit vocal-cord vibration under the following circumstances: (a) when an obstruction in the vocal tract causes the intraoral pressure to build up and hence the pressure across the glottis to decrease; (b) when the glottis is spread to cause a wide aperture or when it is constricted. Thus an increased stiffness of the vocal cords tends to narrow the range of transglottal pressures and glottal apertures over which vocal-cord vibration occurs.

4. Slack vocal cords. The vocal cords can be made more slack by decreasing the coupling between upper and lower edges of the vocal cords. This is probably accomplished by a decrease in the tension of the vocal cords, as well as by a decreased stiffness of the walls of the glottis. Slackness of the vocal cords can allow glottal vibration to occur even with a spread or constricted glottis. When the vocal cords are slackened, there is a decrease in the frequency of glottal vibration.

These four features are not completely independent. The combinations [+spread, +constricted] and [+stiff, +slack], are, of course, logically and physiologically excluded. The 4 features proposed thus yield 9 distinct phonetic categories of segments. We shall study separately the categories as they pertain to sonorants and to obstruents because these two classes of segments are affected rather differently by the proposed modification of the feature framework.

As with all muscular adjustments, the changes in vocal-cord configuration corresponding to initiation of one of these features do not occur instantaneously, but extend over an interval of time that may be 100 ms or more. Thus evidence for a gesture corresponding to a feature may be distributed over an appreciable duration of the sound wave.

We have listed the nine possible feature combinations in Table IX-5. Each column

Table IX-5. Classification of obstruent, glide, and vowel features in terms of proposed glottalized features.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| obstruents | $b_1$ | b | p | $p_k$ | $b^h$ | $p^h$ | ɓ | ʔb | $p^ʔ$ |
| glides | w, y | | | | ɦ | h, W, Y | | ʡ | ʔ, ʔw, ʔy |
| vowels | V | V̀ | V́ | voiceless vowels | breathy vowels | | | creaky voice vowels | glottalized vowels |
| spread glottis | - | - | - | + | + | + | - | - | - |
| constricted glottis | - | - | - | - | - | - | + | + | + |
| stiff vocal cords | - | - | + | - | - | + | - | - | + |
| slack vocal cords | - | + | - | - | + | - | - | + | - |

is identified in terms of a phonetic category for obstruents and sonorants. The sonorants are further subdivided into glides and vowels. We now consider each of these broad classes in detail.

## 3. Laryngeal Feature Combinations for Obstruents

In the first row of Table IX-5, we have identified the nine possible combinations for labial stops. We note first that the features spread glottis and constricted glottis produce a tri-partite division of the obstruents into plain [-spread glottis, -constricted glottis], aspirated [+spread glottis, -constricted glottis], and glottalized [-spread glottis, +constricted glottis]. There is little to be said about the fact that glottalized consonants – ejectives and implosives – are produced with a constricted glottis. At least since Catford's[5] contribution this has been accepted as standard doctrine in phonetics. We attribute aspiration to [+spread glottis] on the basis of the recent findings of Kim[3] which appear to us wholly convincing and are in agreement with our own investigations of glottal features. We may observe in this connection that among aspirated stops we include not only the voiceless [$p^h$] found, e.g., in modern English as well as the voiced [$b^h$] of Hindi or Sinhalese, but also the moderately aspirated (opposed to the fully aspirated), stop of Korean, symbolized here by [$p_k$] (see Kim[3]). Finally, the third class of obstruents that are neither aspirated nor glottalized are naturally characterized as [-spread glottis, -constricted glottis].

The sounds in each of these three classes are further subdivided into 3 groups by means of the features *stiff vocal cords* and *slack vocal cords*. When the vocal cords are [+stiff, -slack], Fig. IX-15 has shown that they will not vibrate when the pressure across the glottis is reduced. This observation suggests that [+stiff, -slack] is the appropriate feature assignment for the traditionally voiceless consonants. We recognize 3 distinct obstruents of this type (cf. Table IX-5). The normal voiceless unaspirated stop consonant [p] is identified by the combination of features in column 3. The glottis is neither constricted nor spread, and the vocal cords are stiff. Evidence for this combination of features includes the cineradiographic observations of Kim,[3] which have shown that the vocal-cord position remains more or less unchanged for an intervocalic [p]; the almost immediate voicing onset following release of this consonant in initial position (Lisker and Abramson[6]); and the observation that this consonant, in common with other voiceless consonants, tends to cause a rise in fundamental frequency of a preceding vowel (L. Maran and D. H. Klatt[7]).

Evidence from the same sources indicates that the voiceless aspirated stop [$p^h$] (see column 6) has a spread glottis and delayed voice onset time in initial prestressed position, and tends to increase the fundamental frequency of an adjacent vowel (House and Fairbanks[8]), as the feature [+stiff] would suggest. The voice-onset time for the voiceless aspirated stop is at least as great as the duration of the formant transitions following the consonantal release, so that voicing is initiated after the rapid spectrum change is completed (Klatt and Stevens[9]).

The third member of the class, the ejective or checked stop consonant [$p^?$] is represented by the combination of features in column 9 of Table IX-5. The combination of constricted glottis and stiff vocal cords means that the glottis remains tightly closed, without vibration, during the closure interval. There is no airway between the trachea and the supraglottal space. In order to obtain sufficient acoustic energy at the instant of release of the supraglottal closure (the lips in the case of the labial consonants under discussion here), pressure in the supraglottal cavities is built up by decreasing the volume of these cavities during the closure interval. This volume reduction appears to be accomplished by raising the larynx (and probably also by a contraction of the lumen of the lower pharynx), and produces a supraglottal pressure that is well in excess of the normal subglottal pressure.[10] After release of the supraglottal closure, there is a delay of 50-odd ms before the adducting glottal musculature can be relaxed and the glottis can assume a configuration appropriate for the onset of vocal-cord vibration. It is worth noting that ejective stops tend to have points of articulation that are more posterior, the palatal and retroflex alveolar consonants being more common than labials (Greenberg[11]). This can, perhaps, be explained by the fact that it is easier to achieve a pressure build-up in the supraglottal cavities by raising the larynx (less laryngeal displacement being required) if the initial cavity volume is smaller.

Vocal cords that are [-stiff] are capable of vibrating for smaller transglottal pressures than are [+stiff] glottal cords, as Fig. IX-16 has shown. Vibrations are further facilitated when the glottal cords are not only [-stiff], but also [+slack], and, in fact, the traditional voiced obstruents are in the present framework represented as [+slack, -stiff]. Among this class of obstruents we again find three distinct types. For the (usually) voiced stop [b] represented in column 2 of Table IX-5 the vocal cords are in a position that is neither spread nor constricted, with a glottal width that is probably slightly greater than that for normal voicing due to the pressure build-up in the glottis. If voicing is to continue throughout the closure interval, some mechanism must prevent the supraglottal pressure from becoming too high, i.e., keep the transglottal pressure sufficiently large. This condition is achieved through a continuing increase in the volume of the supraglottal space (Perkell[12]) – presumably a muscular adjustment in the supraglottal walls in response to a pressure increase.[13]

A voiced stop that contrasts with [b] in some languages is the aspirated consonant [b$^h$], identified in column 5 of Table IX-5. The features are the same as those for [b] except that the glottis is [+spread] rather than [-spread]. For this consonant there is some voicing during the closure interval, since presumably there is an increase in the supraglottal volume to provide some glottal airflow, but this voicing cannot continue for many periods before there is a build-up of supraglottal pressure, since the airflow during breathy voicing is rather large. Thus the voicing ceases during the latter part of the closure interval unless the adducting maneuver is delayed until the end of this interval. Following the release, there is an interval of aspiration, probably consisting of breathy voicing (as in the [ɦ] of Table IX-5, to be discussed below), before the vocal-cord configuration and stiffness return to the normal condition appropriate for a vowel.

The third obstruent having the features [+slack, -stiff] is glottalized. Its feature configuration is given in column 8 of Table IX-5. In discussing obstruents of this type, Greenberg[11] writes: "As Ladefoged (1968) correctly indicates, there are no less than three related phonetic possibilities: truly implosive sounds in which the larynx is lowered and ingressive air follows the oral release, sounds with laryngealized voicing, and preglottalized sounds." Ladefoged[15] notes, however, that he is not able "to distinguish consistently between voiced consonants with an accompanying glottal stop (preglottalized-KS/MH) and similar consonants marked by laryngealization." As this distinction does not appear consistently in any language known to us, we shall follow Ladefoged and distinguish only two types of nonejective obstruents: "voiced implosives, in which there is always a downward movement of the glottis – and there may or may not be laryngealized voicing; and ...

laryngealized consonants (as in Hausa) in which there is always a particular mode of viration of the vocal cords – and there may or may not be a lowering of the larynx".[16] Ladefoged represents the former by the letter ɓ and the latter by the digraph ʔb, a notational convention that we have followed here. The [+slack vocal cords] of [ʔb] allow for vocal-cord vibration without the decrease in supraglottal pressure produced by a lowering of the glottis. This sound differs, therefore, from the true implosive, which is [–slack, –stiff] (column 7, Table IX-5) and is produced with an active lowering of the larynx during the closure interval, probably accompanied by an enlargement of the pharyngeal width – gestures that increase the supraglottal cavity volume and allow air to flow through the glottis to maintain voicing. Furthermore, the true implosive [ɓ] apparently does not cause a lowering of the tone of an adjacent vowel, as would normally be expected for a voiced ([+slack]) consonant like [b] (Greenberg[11]), since the vocal cords are not slack. Another difference is the laryngealized or creaky voicing that apparently follows the release for [ɓ], but not for [b]. In languages with true implosives, there appears to be a preference for more anterior points of articulation, presumably because such positions result in a larger supraglottal cavity which in turn is capable of being more readily expanded in volume.[10]

The class of [–slack, –stiff] obstruents includes, in addition to the true implosive [ɓ] just discussed, the sounds labelled $[b_1]$ (column 1, Table IX-5) and $[p_k]$ (column 4). The $[b_1]$, which probably represents what has sometimes been called a lax voiceless stop, appears in Danish; for example (Fischer-Jørgensen[17]), and may occur in initial position for many speakers of English. The vocal cords are neither spread nor constricted, and the [–slack, –stiff] configuration results in a cessation of vocal-cord vibration, particularly if there is little or no expansion of the supraglottal cavities during the closure interval.

The third stop consonant which is [–slack, –stiff] is the "partially aspirated" consonant $[p_k]$ that has been described in Korean (Lisker and Abramson[6] and Kim[18]). The widening of the glottis for this stop consonant has been observed in anterior-posterior cineradiographs by Kim,[3] who notes that it is not as great as the abduction maneuver observed in a contrasting voiceless aspirated stop. Since the glottis is abducted during the closure interval, there will be a brief interval of aspiration following the release of the stop, but since voicing onset occurs more readily with a [–stiff] configuration (i.e., voicing is initiated with a wider glottal opening) than with a [+stiff] configuration, this interval preceding the onset of voicing should be somewhat shorter than that which follows a voiceless unaspirated stop. Measurements of voice-onset time support this conclusion (Lisker and Abramson,[6] Kim,[18] and Han and Weitzman[19]). Kim[18] has observed that the fundamental frequency in the vowel following release of a stop tends to be higher for the voiceless stops [p] and $[p^h]$ than for the partially aspirated $[p_k]$. This finding supports the classification of the first two of these stops as [+stiff] and the last as [−stiff], since vocal-cord stiffness has an influence on the frequency of vibration. The Korean [$p_k$] is voiceless during the closure interval in initial position, indicating that there is little or no expansion of the supraglottal cavity volume. Such an expansion probably occurs when the stop consonant is in intervocalic position, since voicing continues during the closure interval in this phonetic environment. (Alternatively, the vocal cords are adjusted to be [+slack] in intervocalic position, in order to maximize the contrast with [p] and [$p^h$].)

Various combinations of values of the four proposed features can be used to classify fricatives, as well as stop consonants. As in the case of stops, there are 3 voiceless fricatives with the feature [+stiff]. These are the aspirated fricative [$s^h$], which occurs, for example, in Korean or Burmese, the "normal" fricative [s], and the ejective [$s^?$] (in Hausa). The ejective is produced with a closed glottis, air being expelled through the constriction by raising the glottis and narrowing the pharynx, thereby creating an increased pressure in the mouth. Voiced fricatives are generated with vocal-cord vibration, as well as turbulence noise at a supraglottal constriction, the intraoral pressure being maintained at a value less than the subglottal pressure. Thus implosive fricatives with a reduced intraoral pressure, with the features [+constricted glottis, −stiff], are not possible. The combination [−constricted glottis, +slack] is, however, a possibility, and gives rise to the voiced fricative [z]. To our knowledge, there are no languages that contrast an unaspirated and an aspirated voiced fricative, i.e., columns 2 and 5 in Table IX-5. The feature combinations in columns 1 and 4 for fricatives are apparently either not utilized or, what is more likely, do not give acoustic outputs that are significantly different from those represented by columns 3 and 6.

### 4. Laryngeal Feature Combinations for Nonobstruents

The second and third rows in Table IX-5 represent the different categories of nonobstruent nonconsonantal sounds that are created with the help of the four features under discussion. Among these nonobstruents we distinguish the syllabic *vowels* from the nonsyllabic *glides* as indicated. The four laryngeal features that are proposed subcategorize the vowels, as well as the glides, in much the same fashion as they have been seen to subcategorize the obstruents. As in the case of the obstruents, the two features *spread glottis* and *constricted glottis* define 3 classes of sounds that will be designated here as *plain*, *aspirated*, and *glottal*. The plain sounds which are characterized as [−constricted, −spread] include the familiar types of vowels and glides. The *aspirated* sounds are produced with a glottis that is [−constricted, +spread]; these include voiceless, as well as "breathy voice," vowels and the various types of aspirated glides. The glottal sounds are produced with a glottis that is [+constricted, -spread]. We encounter here various kinds of laryngealized vowels ("creaky voice"), as well as glottalized glides. Each of these three classes is further subdivided into 3 sets by the features *stiff vocal cords* and *slack vocal cords*. We recall that [+stiff vocal cords] made voicing impossible for the obstruents, whereas [+slack vocal cords] facilitated voicing. In the case of nonobstruents, voicing is impossible when vocal cords are stiff and the glottis is either constricted or spread, but vocal-cord vibration *can* occur for the combination [+stiff] and [-constricted, -spread]. All nonobstruent combinations with the feature [+slack vocal cords] are voiced, as are the corresponding obstruents.

As the glides resemble in many ways the obstruents, we shall begin the discussion with them. We shall assume that in the case of the aspirated ([+spread glottis]) and the glottalized ([-constricted glottis]) glides, the unmarked situation calls for [+stiff vocal cords] (columns 6 and 9 in Table IX-5) because the unmarked glides ([h] and [?]) are both voiceless. Occurrence of these features in combination with the feature [+high] gives the voiceless glides (symbolized by Sapir as [W, Y]) found, for instance, in Southern Paiute and other American languages, and the glottalized glides [?w, ?y] of Nootka, for example, cf. Sapir[20]). These glides with the feature [+stiff vocal cords] contrast with their [+slack vocal cords] congeners (columns 5 and 8 in Table IX-5) much like voiceless and voiced obstruents. Thus, we have both a voiceless [h] and a voiced [ɦ], the latter occurring, for example, in some dialects of Arabic, and in Bengali. The parallel contrast between the voiceless [?] and the voiced [ʡ] glottal stop appears to be attested in Jingpho (personal communication from LaRaw Maran, and see also below). For plain glides, the neutral configuration is [-stiff vocal cords, -slack vocal cords], as in column 1 of Table IX-5.

The various classes of glides are summarized in Table IX-6. It is noted that the aspirated and glottalized glides are subdivided into two instead of the logically possible three classes. Whether this reflects a shortcoming in our framework or whether it is merely due to ignorance on our part about what varieties of glides are to be found in the languages of the world cannot be established at this point.

This brings us to the last class, the vowels. Following a suggestion made to us by LaRaw Maran, we propose that in the plain vowels, [+stiff vocal cords] is the articulatory correlate of *high pitch*,[21] whereas [+slack vocal cords] is the articulatory correlate of *low pitch*. *Neutral pitch* for the vowels is produced by the configuration [-slack, -stiff]. We observe that these feature assignments are compatible with the well-known fact that voiceless – i.e., [+stiff] – obstruents cause an upward shift in pitch in the adjacent vowel, whereas voiced – i.e., [+slack] – obstruents cause a downward shift in pitch. Evidently the gesture of stiffening or slackening the vocal cords is relatively sluggish (possible requiring roughly 100 ms), and hence

Table IX-6. Classification of glides according to the four proposed laryngeal features.

| Feature | Stiffness of Vocal Cords | | |
|---|---|---|---|
| | +stiff<br>-slack | -stiff<br>-slack | -stiff<br>+slack |
| +spread glottis<br>-constricted glottis<br>(aspirated) | h, W, Y<br>(voiceless) | | ɦ<br>(voiced) |
| -spread glottis<br>-constricted glottis<br>(plain) | | w, y | |
| -spread glottis<br>+constricted glottis<br>(glottalized) | ?, ?w, ?y<br>(voiceless) | | ʡ<br>(voiced) |

has an influence on the fundamental frequency over an appreciable time interval.

The effect of the obstruents on the pitch of the adjacent vowel has been studied in considerable detail in connection with the evolution of the tonal systems of the languages of East Asia (cf. Haudricourt[22] and L. Maran, personal communication). It is in one of these languages – Jingpho – that we find two types of glottal stop, the [+stiff] [?] and the [+slack] [ʡ], as noted above. The former parallels the voiceless obstruents in producing a rise in the pitch of the preceding vowel, whereas the latter, like voiced obstruents, produces a drop in pitch. Informal measurements on one informant (LaRaw Maran) give the following results. The numbers represent the change in fundamental frequency (in Hz) from the central portion of the vowel to the end of the vowel.

| | | | | | |
|---|---|---|---|---|---|
| kat | 125 → 155 | ka? | 170 | di? | 140 → 160 |
| kad | 110 → 85 | kaʡ | 110 → 75 | diʡ | 100 → 90 |

These data are especially noteworthy, as in the more familiar languages the glottal stop causes only a rise in pitch, indicating that it is [+stiff].

The features stiff and slack produce somewhat different results when superimposed on the other classes of vowels. Among the aspirated vowels we distinguish those that are voiceless (whispered) from those that are produced with "breathy voice" (columns 4 and 5 of Table IX-5). The former have [-slack vocal cords], whereas the latter have [+slack vocal cords]. Voiceless vowels[23] are found in a variety of languages of North America, e.g., Comanche (Canonge[25]), Acoma (Miller[26]) and Southern Paiute (Sapir[27]); vowels with breathy voice occur in some of the languages of South-East Asia, e.g., Meo and Kuy (Smalley[28]). We do not possess evidence for a third category of "aspirated" vowels (column 6), and we assume the "aspirated" vowels that are [-slack] are also [-stiff].

Among the glottalized or the laryngealized ([+constricted glottis]) vowels there again seem to exist only two of the three logically possible categories, identified as columns 8 and 9 in Table IX-5. These are, on the one hand, the vowels pronounced with "creaky voice" and, on the other hand, the glottalized vowels of Vietnamese, Nez Perce and Acoma. We propose to characterize the former as [+slack, -stiff] and the latter as [-slack, +stiff].

Among the glottal vowels we propose to include also the Danish stød. In terms of the framework sketched here, the stød is produced with a constricted glottis. Presumably near the onset of the vowel, the vocal cords are [+stiff], since the pitch is high (Lauritsen[29]). A fixed time later the vocal cords become [-stiff] and the glottis apparently changes to [-constricted].[30] This change of the vocal cords from [+stiff] to [-stiff] accounts readily for the fact that the cognate of the stød in Swedish is the so-called "first tone," which at least in some dialects (e.g., Stockholm) consists in a high tone followed by a low tone, i.e., [+stiff vocal cords] followed by [+slack vocal cords] (cf. Kacnel'son[34] and Öhman[31]).

Phenomena quite similar to the Danish stød have been reported in various quite unrelated languages. For instance, the so-called "broken tone" of the Hanoi dialect of Vietnamese greatly resembles the stød (cf. Haudricourt[33] and Han[35]). The Keresan languages of North America also possess sounds of this type, as has been noted by I. Davis: "The Keresan sound system ... involves certain complexities not generally characteristic of American Indian languages. For example, the system of tonal accents, if not unique, is at least of a type not often recognized. Although the contrast between a level pitch and a falling pitch on stressed syllables is reminiscent of Swedish, the Acoma system is more complex. In addition to level and falling accent, it involves a third type which Miller terms "glottal accent." This is characterized by falling pitch followed by a light glottal catch and a subsequent rearticulation of the vowel. It might be noted that the Acoma glottal accent apparently represents a merging of two types of accent, glottal and breathy, observed in some Eastern Korean dialects ..." (Davis[36]).

Table IX-5 suggests that the nonsyllabic glides are characterized either by the feature [+high] or by a + on two of the laryngeal features, or both. The kind of articulatory gesture and acoustic attribute that distinguishes syllabic from nonsyllabic segments must await further study of the nature of the feature syllabic. In the acoustic domain, a [+syllabic] segment presumably gives rise to a peak in intensity or in loudness. In the articulatory domain, segments that are nonsyllabic may be distinguished from their syllabic cognates by having, in some cases, a more extreme gesture. In the case of the laryngeal feature, it is possible that for nonsyllabic segments the false vocal cords form a partial or complete constriction in addition to that formed by the true vocal cords, whereas this is not the case for syllables. Some evidence for the involvement of the false vocal cords in glottal stops has been reported by Lindquist,[37] but further experimental data on this point are clearly needed.

The comments and suggestions of Kenneth Hale and Mary Louise Kern are gratefully acknowledged. In its early stages, this work was stimulated by discussions with LaRaw Maran.

M. Halle, K. N. Stevens

## Footnotes and References

1. K. N. Stevens, "Aerodynamic and Acoustic Events at the Release of Stop and Fricative Consonants (in preparation).
2. K. N. Stevens, "An Analysis of Glottal Acitivity during Consonant Production" (in preparation).
3. C. W. Kim, "A Theory of Aspiration," Phonetica 21, 107-116 (1970).
4. K. Ishizaka and M. Matsudaira, "What Makes Vocal Cords Vibrate?," Reports of the 6th International Congress on Acoustics IIB, pp. B-9 to B-12, 1968.
5. J. C. Catford, "On the Classification of Stop Consonants," Le Maître Phonétique, Third Series, Vol. 65, pp. 2-5, 1939.
6. L. Lisker and A. Abramson, "A Cross-language Study of Voicing in Initial Stops: Acoustical Measurements," Word 20, 384-422 (1964).
7. L. Maran and D. H. Klatt, Personal communication, 1970.
8. A. S. House and S. Fairbanks, "Influence of Consonant Environment upon the Secondary Acoustical Characteristics of Vowels," J. Acoust. Soc. Am. 25, 105-121 (1953).
9. See Section IX-C.
10. We leave open for the present the question whether the raising of the glottis in the ejective and the lowering of the glottis in the implosive should be attributed to separate features in the universal framework.
11. J. H. Greenberg, "Some Generalizations Concerning Glottalic Consonants, Especially Implosives," Int. J. Am. Linguistics 36, 123-145 (1970).
12. J. S. Perkell, Physiology of Speech Production: Results and Implications of a Quantitative Cineradiographic Study, Research Monograph No. 53 (The M.I.T. Press, Cambridge, Mass., 1969).
13. An alternative mechanism that has been suggested is a small velopharyngeal opening which provides a high resistance leak from the supraglottal cavity (Rothenberg[14]).
14. M. Rothenberg, "The Breath-Stream Dynamics of Simple-Released-Plosive Production," Bibliotheca Phonetica No. 6 (S. Karger, Basel, 1968).
15. P. Ladefoged, A Phonetic Study of West African Languages (Cambridge University Press, London, 1968), see p. 17.

16. *Ibid.*, p. 16.

17. E. Fischer-Jørgensen, "Voicing, Tenseness and Aspiration in Stop Consonants, with Special Reference to French and Danish," Annual Report of the Institute of Phonetics, University of Copenhagen, No. 3, 1968, pp. 63-114.

18. C. W. Kim, "On the Autonomy of the Tensity Feature in Stop Classification (with Special Reference to Korean Stops)," Word **21**, 339-359 (1965).

19. M. S. Han and R. S. Weitzman, "Studies in the Phonology of Asian Languages V: Acoustic Features in the Manner-Differentiation of Korean Stop Consonants," Acoustics Phonetics Research Laboratory, University of Southern California, Office of Naval Research Technical Report NR 049-183, Nonr 228 (28) (1967).

20. E. Sapir, "Glottalized Continuants in Navaho, Nootka and Kwakiutl," *Selected Writings* (University of California Press, Berkeley, 1949), pp. 225-250.

21. We use the term *pitch* to mean fundamental frequency, recognizing that this is somewhat inconsistent terminology, since pitch is normally considered as a subjective attribute of tones. The term pitch has come to be used in this context, however, and hence we shall follow this convention.

22. A. G. Haudricourt, "De l'origine des tons en Vietnamien," J. Asiatique **242**, 69-82 (1954).

23. Greenberg[24] has suggested that voiceless vowels are usually unstressed, and hence cannot be regarded as having high pitch. Thus we have assigned the features [-stiff, -slack] to the voiceless vowels rather than [+stiff, -slack], since the feature [+stiff] would imply high pitch.

24. J. H. Greenberg, "Some Methods of Dynamic Comparison in Linguistics," in *Substance and Structure of Language*, J. Puhvel (Ed.) (University of California Press, Berkeley, 1969).

25. E. D. Canonge, "Voiceless Vowels in Comanche," Int. J. Am. Linguistics **23**, 2, 63-67 (1957).

26. W. R. Miller, *Acoma Grammar and Texts*, University of California Publications in Linguistics, Vol. 40 (University of California Press, Berkeley, 1965).

27. E. Sapir, "Southern Paiute," Proc. Am. Acad. Arts Sci. **60**, 1-3; 1273-1393 (July-September 1930 and May 1931).

28. W. A. Smalley, *Outline of Khmu? Structure* (American Oriental Society, New Haven, Conn., 1961).

29. M. Lauritsen, "A Phonetic Study of the Danish stød," Project on Linguistic Analysis Reports Second Series, **7**, University of California at Berkeley, Department of Linguistics, 1968.

30. During an interval of time in which the supraglottal vocal-tract configuration is relatively open (i.e., during a sequence of one or more nonobstruent nonconsonantal segments), a variety of sequences of gestures can be traversed by the laryngeal musculature, and these gestures may be timed in different ways to yield distinctive acoustic outputs. Thus a sequence of feature combinations or "segments" for the laryngeal features need not necessarily bear a one-to-one relationship to the sequence of segments expressed in terms of supraglottal features types. Furthermore, as Öhman[31] has shown, the precise timing of a given sequence of laryngeal gestures may differ from one dialect to another in a given language. The possibility of a nonsynchrony between segments relating to supraglottal gestures and those relating to laryngeal gestures has been discussed by Kovitz[32] in connection with the tones of various Chinese dialects.

31. S. Öhman, "Word and Sentence Intonation, a Quantitative Model," Speech Transmission Laboratory Quarterly Progress and Status Report, Royal Institute of Technology, Stockholm, 15 October 1967, pp. 20-54.

32. Johanna Kovitz, "Chinese Tone Sandhi and the Phonology of Tone (unpublished manuscript, Massachusetts Institute of Technology, 1970).

33. A. G. Haudricourt, "Bipartition et tripartition des systèmes de tons dans quelques langues d'extrême orient," Bull. Sociètè de Linguistique de Paris 56, 163-180 (1961).

34. S. D. Kacnel'son, "Germanskaja akcentologija," Sravnitel'naja grammatika germanskix jazykov II, 160-220 (1962).

35. M. S. Han, "Studies in the Phonology of Asian Languages VIII: Vietnamese Tones," Acoustics Phonetics Research Laboratory, University of Southern California, Office of Naval Research Technical Report NR 049-183, Nonr 228 (28) (1969).

36. I. Davis, Review of: Acoma Grammar and Texts by Wick R. Miller, Language 44, 185-189 (1968), see p. 186.

37. J. Lindquist, "Laryngeal Mechanisms in Speech," Speech Transmission Laboratory Quarterly Progress and Status Report, Royal Institute of Technology, Stockholm, 15 October 1969, pp. 26-32.