DOCUMENT RESUME

ED 088 289 | FL 005 221

AUTHOR	Lea, Wayne A.
TITLE	Influences of Phonetic Sequences and Stress on Fundamental Frequency Contours of Isolated Words.
PUB DATE	Nov 72
NOTE	10p.; Paper presented at the Meeting of the Acoustical Society of America (84th, Miami Beach, Fla., November 1972)
EDRS PRICE	MF-$0.75 HC-$1.50
DESCRIPTORS	Acoustic Phonetics; Articulation (Speech); Charts; Computers; Consonants; *Distinctive Features; Graphs; Language Patterns; *Language Research; Pattern Recognition; Phonemics; *Phonetics; *Phonology; Structural Analysis; Syllables; *Vocabulary; Vowels

ABSTRACT

At vowel onset following unvoiced consonants in /h-cvc/ utterances spoken by two talkers, Fo began high and fell about seven percent in the first five centiseconds. At closure of voiced oral obstruents, Fo suddenly dipped about ten percent, remained flat, suddenly rose about twenty five percent at opening of closure, and, after vowel onset, gradually rose. The high/low feature of the vowel, and manner and place of prevocalic consonant articulation had progressively less effects on vowel Fo values. The final consonant had no apparent effect on Fo values in the vowel. As previous synthesis work has suggested, the fall or rise of Fo in the initial portion of a vowel appears to be a cue to the state of voicing of previous consonants. Initial and peak Fo values in the vowel also can indicate state of consonant voicing. However, Fo contours in bisyllabic words with contrasting stress patterns and similar phonemic sequences showed that: (1) an initially-falling Fo in a vowel may indicate either previous unvoiced consonant or an unstressed vowel, and (2) a rising Fo contour may indicate either a word-initial vowel, a preceding voiced consonant, or a stressed vowel. (Author/DD)

INFLUENCES OF PHONETIC SEQUENCES AND STRESS

ON

FUNDAMENTAL FREQUENCY CONTOURS OF ISOLATED WORDS

Wayne A. Lea
Univac DSD
P.O. Box 3525
St. Paul, Minn. 55165

ABSTRACT

At vowel onset following unvoiced consonants in /həCV́C/ utterances spoken by two talkers, F_0 began high (thirty percent higher than in /ə/), and fell about seven percent in the first five centiseconds. At closure of voiced oral obstruents, F_0 suddenly dipped about ten percent, remained flat, suddenly rose about twenty five percent at opening of closure, and, after vowel onset, gradually rose (an average of eight percent in the first ten centiseconds). The high/low feature of the vowel, and manner and place of prevocalic consonant articulation had progressively less effects on vowel F_0 values. The final consonant had no apparent effect on F_0 values in the vowel. As previous synthesis work has suggested, the fall or rise of F_0 in the initial portion of a vowel appears to be a cue to the state of voicing of previous consonants. Initial and peak F_0 values in the vowel also can indicate state of consonant voicing. However, F_0 contours in bisyllabic words with contrasting stress patterns and similar phonemic sequences (e.g., <u>per</u>mit, per<u>mit</u>) showed that: (1) an initially-falling F_0 in a vowel may indicate either previous unvoiced consonant or an unstressed vowel, and (2) a rising F_0 contour may indicate either a word-initial vowel, a preceding voiced consonant, or a stressed vowel.

Presented at the 84th Meeting, Acoustical Society of America, Miami Beach, Florida, Nov., 1972.

INFLUENCES OF PHONETIC SEQUENCES AND STRESS
ON
FUNDAMENTAL FREQUENCY CONTOURS OF ISOLATED WORDS

Wayne A. Lea

1. Introduction

Voicing of consonants cannot always be detected simply from whether or not the fundamental frequency (F_o) contour is continuous, since at least some devices for tracking fundamental frequency do not yield continuous contours throughout consonants, and since phonemically unvoiced consonants are sometimes voiced, at least for their initial several centiseconds (Snow and Hughes, 1969; Lea, 1972, Ch. 4). It has been suggested (Haggard, Ambler, and Callow, 1970) that pitch is a voicing cue, in that voice fundamental frequency (F_o) will progressively increase in the beginnings of vowels after voiced consonants, while F_o starts high and progressively decreases following unvoiced consonants (cf. also Lehiste, 1970, p. 74).

The present study showed that the fall or rise of F_o is indicative of consonant voicing, but that the consonant and vowel identities, and the stress pattern, demand qualifications to the notion of pitch as a voicing cue. The rising F_o after voiced consonants was not expected to occur following voiced strident fricatives, (Haggard, Ambler, and Callow, 1970, p. 616). Data given in this paper show no such differences between strident and nonstrident fricatives.

2. F_o Contours in Isolated [hə ˈCVC] Utterances

A computer program (Snow and Hughes, 1969) was used to extract and plot F_o contours from acoustic data for over 320 isolated [hə ˈCVC] utterances, each spoken by two talkers. All symmetric $C_1V_2C_2$ sequences (with C_1=C_2) were included, where C_1 was any English consonant which can occur in both positions (i.e.,/p, t, k, b, d, g, f, θ, s, ʃ, tʃ, v, ð, z, dʒ, m, n, l/). Each of the twelve vowels /i, ɪ, e, ɛ, ɝ, æ, ɑ, ɔ, ʌ, o, ʊ, u/ occurred as the stressed vowel ("V_2") in each symmetric consonantal context. Asymmetric CVC combinations were also recorded, for cases where C_1 was /j, r, w, h/ or silence and C_2 was one of the six stops/p, t, k, b, d, g/. Other asymmetric CVC combinations paired stops in the C_1 position with final C_2 being either /ŋ, ɝ, ʒ/ or silence (no final C_2).

Contour parameters as defined in Figure 1 were selected to measure the effects, on F_o contours, of the state of voicing and manner and place of articulation of C_1 and C_2. Also considered were the effects of vowel V_2 identity on such parameters. A judgment was made of the value $F_o(V1)$ of F_o representative of the unstressed [ə] and the value $F_o(C1)$ representative of F_o values within voiced medial consonants. The initial F_o within the stressed vowel, $F_o(i)$, was determined as that value at voicing onset after phonetically unvoiced consonants, or at the position of most rapid energy and formant structure change after voiced consonants. Peak F_o in the stressed vowel, $PF_o(V2)$, was usually immediately after voicing onset following unvoiced consonants, but nearer the center of the vowel after voiced consonants. The differences Δ_5F and $\Delta_{10}F$, between F_o five or ten centiseconds after vowel onset and those at vowel onset, were positive if F_o were rising in the initial part of the vowel, and negative if F_o were falling in the initial part of the vowel.

Central F_o values within obstruents were about 15% lower than those in non-vowel sonorants (Chistovich, 1969). Peak F_o in the vowel was generally 10 to 15% higher with surrounding unvoiced consonants than with surrounding voiced consonants. Effects of manner of consonant articulation on F_o values are shown in Figure 2, where peak F_o in Hertz is plotted for each vowel, with consonants pooled into each manner class. Thus, unvoiced fricatives usually gave the highest peak F_o in vowels, especially when spoken by talker ASH, while, for some low vowels, talker GWH showed highest F_o in the environment of unvoiced stops. Similar curves were obtained for initial F_o in the stressed vowels (Lea, 1972), and for this limited data from two talkers, 98% of all stops and fricatives could be correctly categorized as unvoiced or voiced by a simple hypothesis that "if initial F_o in V_2 is greater than a threshold frequency (155 Hz for ASH, 157 for GWH), C_1 is unvoiced; otherwise it is voiced".

Figure 2 also clearly shows that high vowels (near the right and left extremes of the plots) generally yield higher F_o values than low vowels (near the centers of the plots).

Consonant C_2 showed no substantial effects on $PF_o(V2)$, $F_o(i)$, or other critical F_o values defined in Figure 1.

Effects of phonetic sequences on transitions in F_o values from phoneme to phoneme were also determined. With unvoiced consonantal contexts, F_o increased approximately 40 Hz (about 30%) in transition from the unstressed [ə] to the beginning of the stressed vowel (and then steadily fell throughout the remainder of the utterance). For voiced stops and fricatives, F_o fell about 12 Hz (about 10%) in transition from the [ə] into the consonant, then rose an average of about 25 Hz (20%) to the initial value $F_o(i)$ in the stressed vowel, and continued to rise to a peak F_o about 10 or 15 Hz higher than $F_o(i)$. No substantial dip occurs in medial nonvowel sonorants, for which F_o generally rises steadily from $F_o(V1)$, to $F_o(C1)$, to $F_o(i)$, and to the peak value $PF_o(V2)$ in the stressed vowel.

Table I illustrates the reliability of detecting voicing/unvoicing from the values of either ΔF parameter. For example, the simple hypothesis that "a rise in F_o at vowel onset marks a preceding voiced consonant, and a fall marks an unvoiced consonant" would be correct for 95% of all consonants for ASH, and 90% for GWH, if "a rise" meant $\Delta_5 F > 0$ and "a fall" that $\Delta_5 F \leq 0$. Separation between unvoiced and voiced consonants by the size of any ΔF is more reliable for $\Delta_{10} F$ then for $\Delta_5 F$.

Thus, while the rise or fall of F_o after vowel V_2 onset in [həC$\acute{V}_2$C] utterances is a good cue to the state of voicing of the prevocalic consonant, it is not perfectly reliable.

The results showed a slight dependency of ΔF parameters on the manner and place of consonant articulation, but considerably more data, for other talkers, would be needed to substantiate the slight trends indicated (Lea, 1972).

Contrary to a conjecture set forth by Haggard, Ambler, and Callow (1970, p. 616), voiced strident fricatives did cause substantial rises in F_o following onset of the following vowel, just as other voiced consonants did. Haggard, Ambler, and Callow had contended that, since the glottis is open partially for the strident fricatives, cessation of glottal vibration will occur with them, and the F_o jump and following fall of an unvoiced fricative may then be observed. While it is true that glottal vibration, as indicated by computer-determined F_o values, ceased for some portion of the closure time for 17 (71%) of the 24 medial strident fricatives [z], spoken by the two talkers, it also temporarily ceased for 25 (52%) of the 48 nonstrident medial fricatives [v] and [ð]. What is more, the average F_o rises (that is, ΔF parameters) for strident fricatives (see Figure 4) are comparable to those of their nonstrident [v] and [ð] counterparts, and the number of $\Delta_5 F$ falls after strident [z] fricatives was 1 (4%), compared to 4 (8%) for (twice as many) nonstrident fricatives.

The most dominant effect of vowel identity on the contour parameters $\Delta_5 F$ and $\Delta_{10} F$ was due to the vowel features high/low. Low vowels tended to yield somewhat flatter contours at vowel onset, for both unvoiced and voiced preceding consonants. Place of consonant articulation had little effect on ΔF parameters. Manner of consonant articulation had some effect on ΔF parameters (at least for talker ASH), with voiced fricatives and sonorants yielding slightly larger ΔF values than voiced stops did. For talker ASH, unvoiced fricatives were accompanied by more prominant falls for F_o after vowel onset than unvoiced stops were.

3. Effects of Word-Stress Contrasts on F_o Contours

All [həC_1VC_2] words exhibit stress on the second vowel, so that studying F_o contours in the C_1V section of such words demonstrated effects when the consonant precedes a stress peak. Studying the VC_2 sections indicated the lack of any significant effects of a word-final consonant following a stressed vowel. To study effects of word stress on F_o contour parameters, sixty bisyllabic words with contrasting stress patterns, (e.g., permit/permit; mystic/mystique) were recorded by the same two talkers. A general tendency toward higher F_o values in stressed syllables was clearly evident. However, it was also evident that the individual talker's mode of list reading (most specifically, his rising or falling contour) can affect relative F_o values in stressed and unstressed syllables. In all 120 words, talker ASH produced higher peak F_o in the stressed syllable than in the unstressed syllable of the word. No such simple distinction was made by talker GWH. This is evident in the scatter plots of Figure 3, where peak F_o in the second vowel is plotted versus peak F_o in the first vowel. Pairs of peak F_o values cluster into two completely isolatable groups for ASH, corresponding to whether the first or the second vowel is stressed. Any straight line between those two clusters would separate $\acute{V}_1 - V_2$ from $V_1 - \acute{V}_2$ stress patterns with 100% accuracy. For talker GWH, the two groups overlapped some.

The peak value of fundamental frequency <u>in the first vowel alone</u> separates $\acute{V}_1 - V_2$ and $V_1 - \acute{V}_2$ stress patterns of both talkers very well. A simple hypothesis (shown by the dashed lines in Figure 3) that "peak value of F_o in vowel V_1 is greater than 127 Hz if and only if vowel V_1 is stressed" yields 100% correct stress classification for ASH, and 96% correct for GWH. The use of any absolute frequency threshold (like 127 Hz) for distinguishing between stress patterns would undoubtedly be futile for talkers with different physiological structures, and is certainly not advocated here. However, the success of such an hypothesis here does illustrate that peak F_o values in initial syllables are more reliably related to stress patterns than are values in the utterance-final syllables. We would expect initial syllables to be less affected by confusing influences such as the individual's mode of list reading (his rising and falling terminal contours) and the sloppiness, reduced amplitudes, and falling sub-glottal pressure that prevail at the ends of utterances.

The <u>initial</u> values of F_o in both vowels were also studied. As with peak F_o values, a simple hypothesis that "initial F_o value in vowel V_1 is greater than 120 Hz if and only if vowel V_1 is stressed" yielded fairly accurate stress classification (98% for ASH and 82% for GWH). These results are obviously not as good as those with peak F_o values, and are influenced by the identities of previous consonants.

The above results agree with previous studies showing that F_o tends to be higher in stressed syllables. However, the effects of stress on the <u>transitions</u> in F_o at consonant-vowel and vowel-consonant boundaries are also of interest. Various hypotheses are listed in Table 2 which relate falling or rising F_o contours at vowel onset to the stress condition of the syllable, the voicing/unvoicing of preceding consonants, and the exceptional condition where vowel V_1 is word-initial (as in <u>as</u>ter, as<u>tir</u>; etc.).

Published works that suggest that "pitch is a voicing cue" (Haggard, Ambler and Callow, 1970; Chistovich, 1969), and the results with [həCV́C] utterances, would suggest that voiced consonants are followed by rising F_o contours, while falling F_o contours follow unvoiced consonants. The simple "consonantal hypotheses," listed as hypotheses 1 and 2 in Table 4, state these claims. However, the data show that only about half of the falling [rising] contours result from unvoiced [voiced] consonants. The one exception was with rising contours in second syllables. These were regularly associated with previous voiced consonants (in 100% of the syllables of ASH, and 84% of those of GWH).

Other hypotheses that weaken the implications of a rising F_o contour, (number 3) or that consider stress effects alone (numbers 4 and 5) are given in Table 4, but hypotheses based on the combination of stress and voiced environments yielded the most accurate predictions of stress and phonetic context. As hypothesis 6 shows, a falling F_o contour may be evidence either of a preceding unvoiced consonant, or of an unstressed syllable (with previous unvoiced or voiced consonant). A rising F_o contour at vowel onset may indicate (by hypothesis 8) either that the vowel is word-initial or preceded by a voiced consonant, or that the syllable is stressed (with a prevocalic voiced or unvoiced consonant). These reliable hypotheses are, however, predictively quite weak. All that they positively eliminate from consideration are (1) the possibility of a stressed vowel with preceding voiced consonant yielding a falling contour; and (2) the possibility of an unstressed vowel with previous unvoiced consonant yielding a rising contour.

It is evident that falling or rising F_o contours at vowel onset do not simply mark either stress or state of voicing, but are functions of a complex combination of stress and phonetic context.

ACKNOWLEDGEMENTS

Research reported herein was supported in part by the Air Force Cambridge Research Laboratory, under contracts F 19628-69-C-0157 and F 19628-72-C-0154, and was presented in more detail in a doctoral desertation prepared at Purdue University. The advise and data of George W. Hughes and Arthur S. House are gratefully acknowledged, as are the programming aids of K.-P. Li and Thomas Snow.

REFERENCES

CHISTOVICH, L. A. (1969) Variations of the Fundamental Voice Pitch as a Discriminatory Cue for Consonants. Soviet Physics-Acoustics, 14: 372-8.

HAGGARD, M., AMBLER, S., and CALLOW, M. (1970) Pitch as a Voicing Cue, J. Acoust. Soc. Amer., 47: 613-7.

LEA, W. A. (1972). Intonational Cues to the Constituent Structure and Phonemics of Spoken English, Ph.D. Dissertation, Purdue University, Lafayette, Ind.

LEHISTE, I. (1970) Suprasegmentals. Cambridge: M.I.T. Press.

SNOW, T. B. and HUGHES, G. W. (1969). Fundamental Frequency Estimation by Harmonic Identification, J. Acoust. Soc. Amer., 45: 316(A).

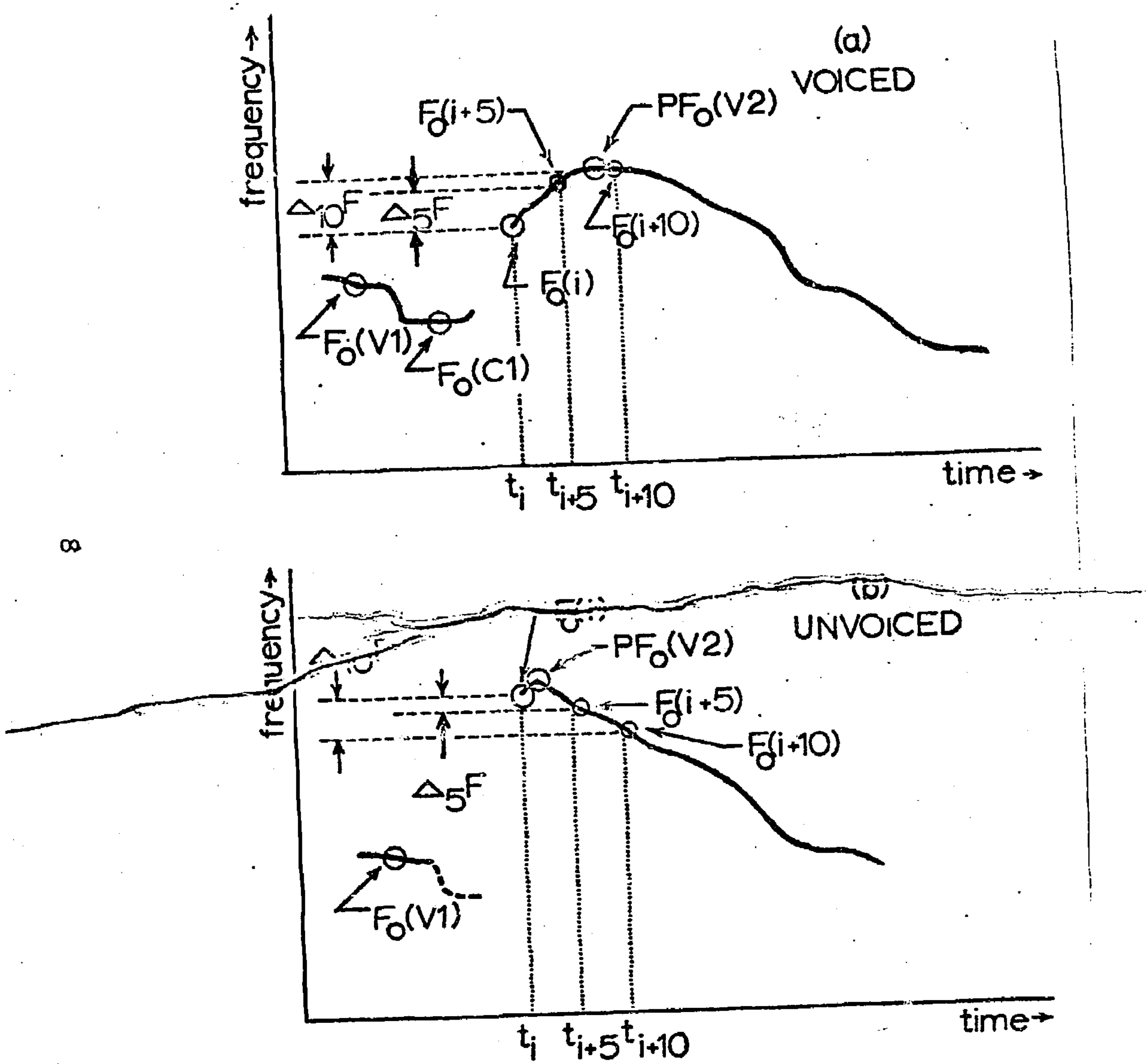

Figure 1. Contours Illustrating F_o Parameter Definitions for (a) Voiced and (b) Unvoiced Consonants in [həCVC] Utterances.

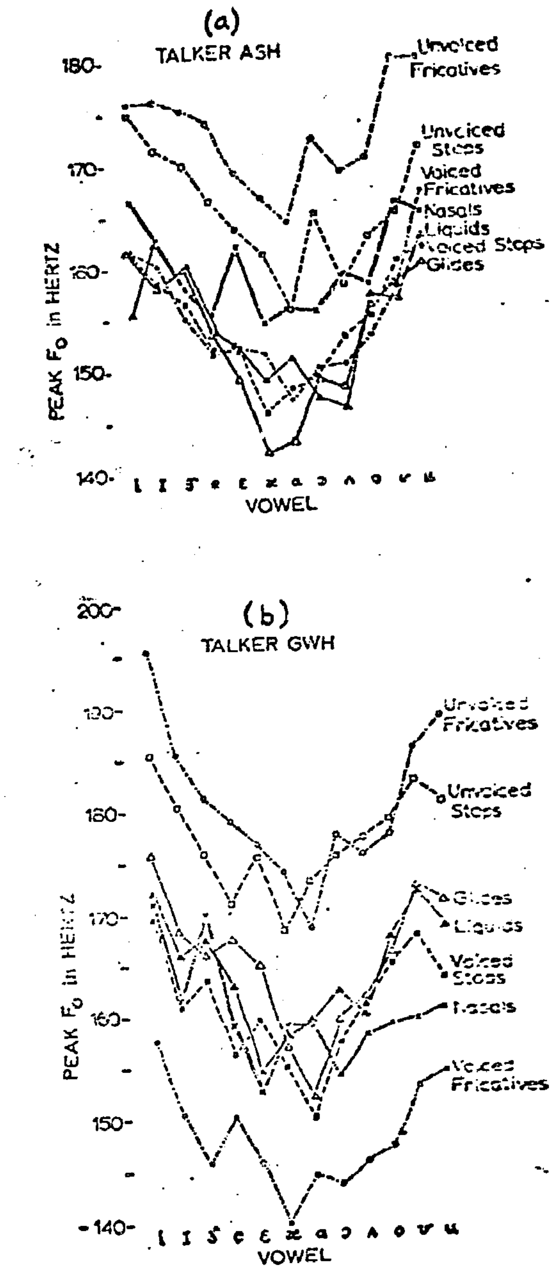

Figure 2. Effects of Vowel Identity on Pea[k] F_o in V_2 of [həCVC] Utterances, as Spoken (a) by talker ASH and (b) by talker GWH. Cons[o]nants Pooled by Categories.

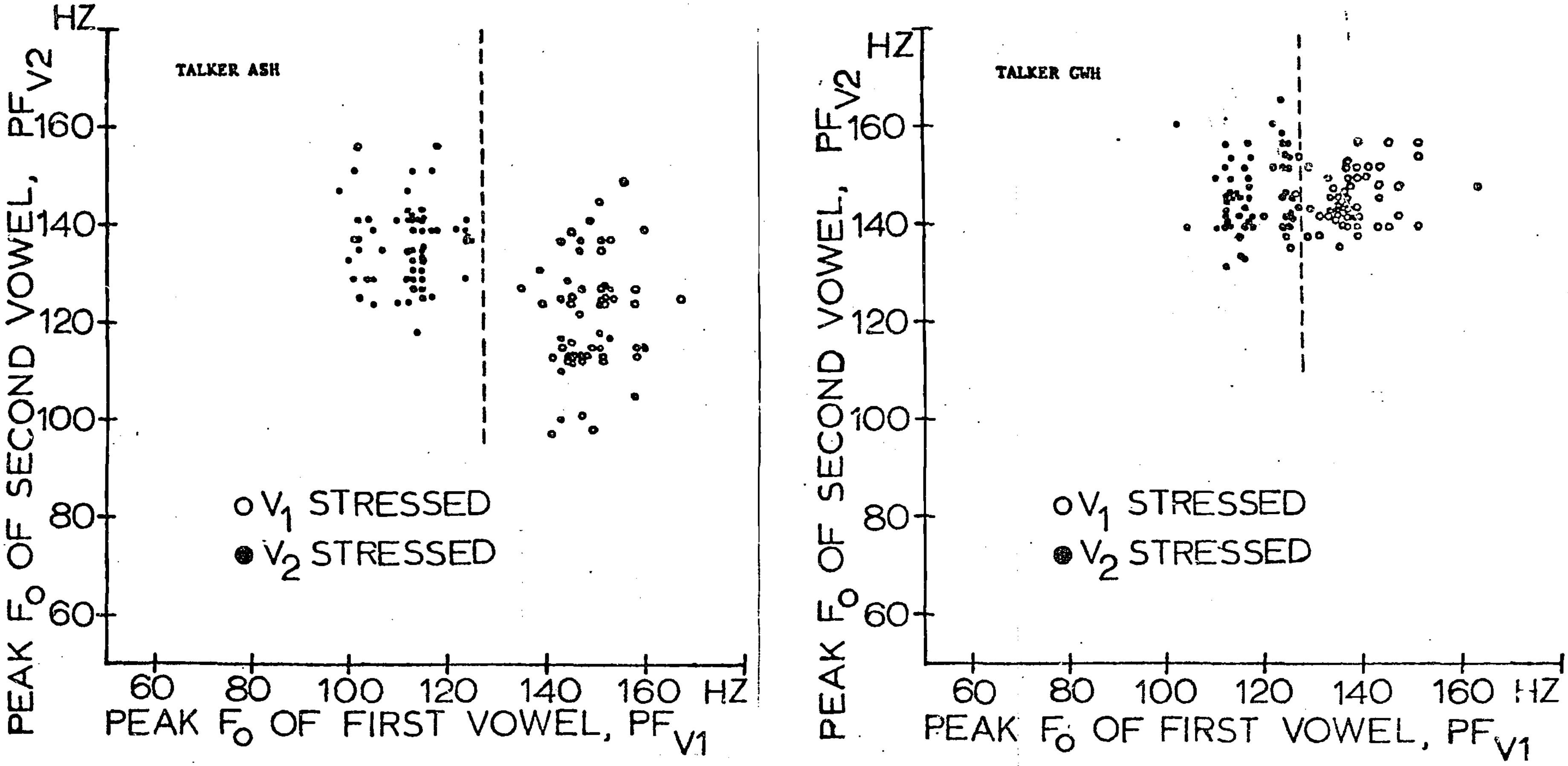

Figure 3. Peak F_o Values in the First and Second Vowels of Bisyllabic words, Showing Clustering Due to Stress Contrasts.

Table I. Detection of Voiced/Unvoiced State of Consonant C_1 From ΔF Parameters of 240 Symmetric həCVC Utterances. All Values in Percent.

% Correct Classification of Voicing Feature ⇓ / ΔF Parameter Hypothesis ⇒		$\Delta_5 F > 0$ ⇓ Voiced	$\Delta_{10} F > 0$ ⇓ Voiced
	Unvoiced Stops	72	94
	Voiced Stops	94	100
TALKER ASH	Fricatives	100	100
	Sonorants	100	100
	TOTAL ALL CONSONANTS	95	99
	Unvoiced Stops	92	100
	Voiced Stops	86	92
TALKER GWH	Fricatives	91	96
	Sonorants	90	83
	TOTAL ALL CONSONANTS	90	92

Table II. Hypotheses for Relating F_o Contours to Stress Patterns and Prevocalic Environments.

	HYPOTHESIS	TALKER	DETECTION SCORES (%) FIRST SYLLABLE	SECOND SYLLABLE	BOTH SYLLABLES
	CONSONANTAL HYPOTHESES				
1.	FALLING CONTOUR → PREVIOUS UNVOICED CONSONANT	ASH	53	49	50
		GWH	46	55	54
2.	RISING CONTOUR → PREVIOUS VOICED CONSONANT	ASH	39	100	60
		GWH	33	84	60
3.	RISING CONTOUR → PREVIOUS VOICED CONSONANT OR VOWEL V_1 WORD-INITIAL	ASH	80	100	87
		GWH	89	84	87
	SIMPLE STRESS HYPOTHESES				
4.	FALLING CONTOUR → UNSTRESSED VOWEL	ASH	75	67	70
		GWH	63	62	62
5.	RISING CONTOUR → STRESSED VOWEL	ASH	78	100	85
		GWH	72	65	68
	STRESS AND ENVIRONMENT HYPOTHESES				
6.	FALLING CONTOUR → UNSTRESSED VOWEL OR PREVIOUS UNVOICED CONSONANT	ASH	94	91	92
		GWH	87	87	87
7.	RISING CONTOUR → STRESSED VOWEL OR PREVIOUS VOICED CONSONANT	ASH	88	100	92
		GWH	77	94	86
8.	RISING CONTOUR → STRESSED OR PREVIOUS VOICED CONSONANT OR VOWEL V_1 WORD-INITIAL	ASH	98	100	99
		GWH	98	94	96