Acoustic characteristics of American English vowels

James Hillenbrand, Laura A. Getty, Michael J. Clark, and Kimberlee Wheeler
Department of Speech Pathology and Audiology, Western Michigan University, Kalamazoo, Michigan 49008

(Received 10 August 1994; revised 7 November 1994; accepted 17 January 1995)

The purpose of this study was to replicate and extend the classic study of vowel acoustics by Peterson and Barney (PB) [J. Acoust. Soc. Am. **24**, 175–184 (1952)]. Recordings were made of 45 men, 48 women, and 46 children producing the vowels /i,ɪ,e,ɛ,æ,a,ɔ,o,ʊ,u,ʌ,ɝ/ in h–V–d syllables. Formant contours for $F1$–$F4$ were measured from LPC spectra using a custom interactive editing tool. For comparison with the PB data, formant patterns were sampled at a time that was judged by visual inspection to be maximally steady. Analysis of the formant data shows numerous differences between the present data and those of PB, both in terms of average frequencies of $F1$ and $F2$, and the degree of overlap among adjacent vowels. As with the original study, listening tests showed that the signals were nearly always identified as the vowel intended by the talker. Discriminant analysis showed that the vowels were more poorly separated than the PB data based on a static sample of the formant pattern. However, the vowels can be separated with a high degree of accuracy if duration and spectral change information is included.

PACS numbers: 43.70.Fq, 43.71.Es, 43.72.Ar

INTRODUCTION

The most widely cited experiment on the acoustics and perception of vowels is a surprisingly simple study conducted at Bell Telephone Laboratories by Peterson and Barney (1952) shortly after the introduction of the sound spectrograph. Peterson and Barney (PB) recorded two repetitions of ten vowels in /hVd/ context spoken by 33 men, 28 women, and 15 children. Acoustic measurements from narrow-band spectra consisted of formant frequencies ($F1$–$F3$), formant amplitudes, and fundamental frequency ($F0$). The measurements were taken at a single time slice that was judged to be "steady state." The /hVd/ signals were also presented to listeners for identification. The results of the measurement study showed a strong relationship between the intended vowel and the formant frequency pattern. However, there was considerable formant frequency variability from one speaker to the next, and there was a substantial degree of overlap in the formant frequency patterns among adjacent vowels. The listening study showed that the vowels were highly identifiable: The overall error rate was 5.6%, and nearly all of the errors involved confusions between adjacent vowels.

The PB measurements have played a central role in the development and testing of theories of vowel recognition. Acoustic measurements for the signals recorded by PB have been widely distributed to speech research laboratories (e.g., Watrous, 1991) and have been used in numerous studies to evaluate alternative models of vowel recognition (e.g., Nearey, 1978; Nearey *et al.*, 1979; Syrdal, 1985; Syrdal and Gopal, 1986; Nearey, 1992; Lippmann, 1989; Miller, 1989; Hillenbrand and Gayvert, 1993a). Despite the widespread use of the PB measurements, there are several well recognized limitations to the database. Perhaps the most important limitation is that the database consists exclusively of acoustic measurements taken at a single time slice. Duration measurements were not made, and no information is available about the pattern of spectral change over time. There is now a solid body of evidence indicating that dynamic properties such as duration and spectral change play an important role in vowel perception (e.g., Ainsworth, 1972; Bennett, 1968; Di Benedetto, 1989a b; Hillenbrand and Gayvert, 1993b; Jenkins *et al.*, 1983; Nearey, 1989; Nearey and Assmann, 1986; Stevens, 1959; Strange, 1989; Strange *et al.*, 1983; Tiffany, 1953; Whalen, 1989). Other limitations of the PB database include: (1) There is no indication that subjects were screened for dialect, and very little is known about the dialect of either the speakers or the listeners; (2) listening results were not reported separately for men, women, and child talkers; (3) no information is given about the age or gender of the child talkers; (4) measures were made from a relatively small group of children; (5) there is no way to determine the identifiability of individual tokens; (6) measurement reliability was not reported; and (7) since the original signals are no longer available, the database cannot be used to evaluate signal representations other than $F0$ and formant frequencies.

The present study represents an attempt to address these limitations. Recordings were made of /hVd/ utterances spoken by a large group of men, women, and children. Measurements were made of vowel duration, $F0$ contours, and formant frequency contours. The signals were also presented to a panel of listeners for identification. Finally, discriminant analysis was used to classify the signals using various combinations of the acoustic measurements.

I. ACOUSTIC ANALYSIS

A. Methods

1. Talkers

Talkers consisted of 45 men, 48 women, and 46 ten- to 12-year-old children (27 boys, 19 girls). The majority of the speakers (87%) were raised in Michigan's lower peninsula, primarily the southeastern and southwestern parts of the state. The remainder were primarily from other areas of the

 0001-4966/95/97(5)/3099/13/$6.00

upper midwest, such as Illinois, Wisconsin, Minnesota, northern Ohio, and northern Indiana. An extensive screening procedure was used to select these 139 subjects from a larger group. The most important part of the screening procedure was a careful dialect assessment, focusing especially on subjects' production of the /ɑ/-/ɔ/ distinction. The /ɑ/–/ɔ/ distinction is not maintained by many speakers of American English, a fact which we believed (incorrectly, as it turned out) might account for the relatively high confusability reported by PB for this pair of vowels.

The screening procedure began with a 5- to 7-min informal conversation with one of the experimenters. This conversation was tape recorded for later review by an experienced phonetician. Subjects next read a 128-word passage that contained several instances of words with /ɑ/ and /ɔ/. Subjects were eliminated if the phonetician noted any systematic departure from general American English, or if the speaker failed to maintain the /ɑ/–/ɔ/ distinction either in spontaneous speech or in the 128-word passage. Subjects were also required to pass a brief task which tested their ability to discriminate /ɑ/–/ɔ/ minimal pairs. In addition to the dialect assessment, subjects were eliminated if they: (1) were non-native speakers of English; (2) showed any evidence of a speech, language, or voice disorder; (3) showed any evidence of a current respiratory infection; or (4) failed a 20-dB pure-tone screening at 500, 1000, and 2000 Hz.

2. Recordings

Audio recordings were made of subjects reading lists containing 12 vowels: The ten vowels recorded by PB (/i,ɪ,ɛ,æ,ɑ,ɔ,ʊ,u,ʌ,ɝ/) plus /e/ and /o/. Also recorded were four diphthongs in /h–d/ context, and both vowels and diphthongs in isolation. Only results from the 12 /hVd/ utterances will be described in this report. Subjects read from one of 12 different randomizations of a list containing the words "heed," "hid," "hayed," "head," "had," "hod," "hawed," "hoed," "hood," "who'd," "hud," "heard," "hoyed," "hide," "hewed," and "how'd." Subjects were given as much time as needed to practice the task and demonstrate an understanding of the pronunciations that were expected for each key word. Recordings were made of several readings of the list once the experimenter was satisfied that the subject understood the task. Once the recording session began, the experimenter did not audition each stimulus and request additional readings based on the experimenter's judgment of correct pronunciation.[1] An attempt was made to record at least three readings of the list. This was often not possible in the case of the children, who took longer to train than adults and sometimes tired of the task after two readings.

The recordings were made with a digital audio recorder (Sony PCM-F1) and a dynamic microphone (Shure 570-S). One token of each stimulus from each talker was low-pass filtered at 7.2 kHz and digitized at 16 kHz with 12 bits of amplitude resolution on a PDP 11/73 computer. Unless there were problems with recording fidelity or background noise, tokens were taken from the subject's first reading of the list. The gain on an input amplifier was adjusted individually for each token so that the peak amplitude was at least 80% of the ±10-V dynamic range of the A/D, with no peak clipping.

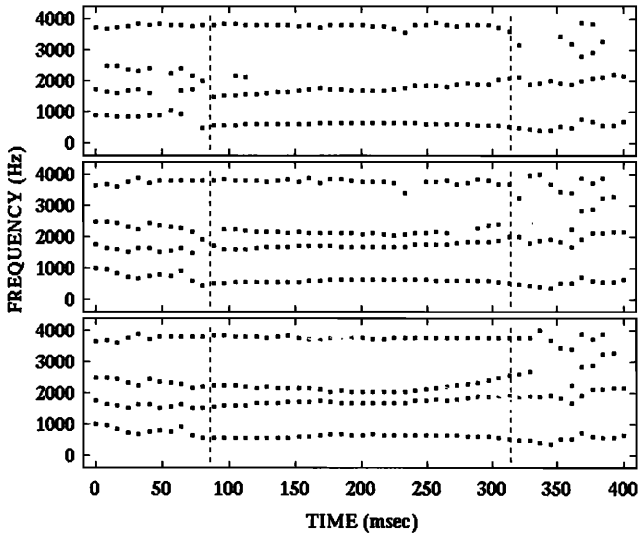

FIG. 1. Spectral peak display of the word "heard" spoken by a child. The dashed vertical lines indicate the beginning and end of the vowel nucleus. The top panel shows the signal after the original 14-pole LPC analysis, the middle panel shows the signal after reanalysis with 18 poles, and the bottom panel shows the signal after hand editing with a custom editing tool.

3. Acoustic measurements

a. Vowel duration and "steady-state" times. The starting and ending times of vocalic nuclei were measured by hand from high-resolution gray-scale digital spectrograms using standard measurement criteria (Peterson and Lehiste, 1960). In an attempt to produce a data set comparable to PB, two experimenters, working independently, made a judgment of steady-state time for each signal. The measures were made while viewing a spectral peak display (Fig. 1) and a gray-scale spectrogram. PB provide a very brief description of how steady-state times were located, indicating only that the spectrum was sampled, "... following the influence of the /h/ and preceding the influence of the /d/, during which a practically steady state is reached" (Peterson and Barney, 1952, p. 177). The two experimenters worked from this brief description, and from the ten examples shown in Fig. 2 of PB.

In addition to the hand measurements of steady-state times, we experimented with several methods of determining steady-state times automatically through an analysis of edited formant contours. Of the several methods that were tried, the technique that seemed to produce the best results defined steady state as the center of the sequence of seven analysis frames (56 ms) with the minimum slope in log $F2$–log $F1$ space (Miller, 1989).

b. Formant contours. Formant-frequency analysis began with the calculation of 14-pole, 128-point linear-predictive coding (LPC) spectra every 8 ms over 16 ms (256 point) hamming windowed segments. The frequencies of the first seven spectral peaks were then extracted from the LPC spectrum files. The frequencies of spectral peaks were estimated with a three-point parabolic interpolation, yielding a finer resolution than the 61.5-Hz frequency quantization. Files containing the LPC peak data served as the input to a custom interactive editor. The editor allows the experimenter

TABLE I. Percentage of utterances showing a formant merger anywhere in the vowel nucleus. Shown in parentheses are the percentage of utterances showing a formant merger at "steady state."

Vowel	$F1-F2$		$F2-F3$	
/i/	0.0	(0.0)	10.0	(8.0)
/ɪ/	0.0	(0.0)	0.0	(0.0)
/e/	0.0	(0.0)	9.3	(4.7)
/ɛ/	0.0	(0.0)	0.0	(0.0)
/æ/	0.0	(0.0)	3.3	(3.3)
/ɑ/	2.0	(2.0)	0.7	(0.0)
/ɔ/	4.0	(2.0)	0.7	(0.7)
/o/	2.7	(1.3)	0.0	(0.0)
/ʊ/	0.0	(0.0)	0.0	(0.0)
/u/	1.3	(0.7)	0.7	(0.7)
/ʌ/	0.7	(0.7)	0.0	(0.0)
/ɝ/	0.0	(0.0)	15.3	(11.3)

to reanalyze the signal with different LPC analysis parameters and to hand edit the formant tracks.

Editing and analysis decisions were based on an examination of the LPC peak display overlaid on a gray-scale spectrogram and, in some cases, on an examination of individual LPC or Fourier spectral slices. General knowledge of acoustic phonetics also played a role in the editing process. For example, editing decisions were frequently influenced by the experimenter's knowledge of the close proximity of $F2$ and $F3$ for vowels such as /i/ and /ɝ/, the close proximity of $F1$ and $F2$ for vowels such as /ɔ/ and /u/, and so on (see Ladefoged, 1967, for an excellent discussion of the inherent circularity in this method of estimating vowel formants, and for other insightful comments on the formant analysis). Considerations such as these often led the experimenter to conclude that a formant merger occurred. In these cases, the LPC spectra were recomputed with a larger number of poles until the merged formants separated.

Once the experimenter was satisfied with the analysis, editing commands could be used to hand edit any formant tracking errors that remained. Figure 1 shows an example of the utterance "heard" spoken by a ten-year-old boy: (a) after the original 14-pole analysis, (b) after reanalysis with 18 poles, and (c) after hand-editing. (For simplicity, the gray-scale spectrogram underlying the peak display is not shown.) The vertical lines indicate the beginning and end of the vowel nucleus. Two commands are available for hand editing the formant contours. One command allows the experimenter to use the mouse to delete a spurious peak, and a second command allows the experimenter to use the mouse to interpolate through "holes" in the formant contour. For example, in the center panel of Fig. 1, there is a gap in the $F3$ contour toward the end of the vowel. Clicking the mouse on either side of this gap causes the program to linearly interpolate formant frequencies through this gap.

It was not uncommon for utterances to show formant mergers throughout all or part of the vocalic nucleus that could not be resolved using these methods. In these cases, zeros were written into the higher of the two formant slots showing the merger (e.g., $F3$ was zeroed out in the case of an $F2-F3$ merger). Table I shows the frequency of occurrence of formant mergers for each of the 12 vowels.

TABLE II. Average absolute difference between formant frequencies sampled at "steady-state" times determined by two judges. Figures in parentheses are differences as a percent of average formant frequency.

	Men	Women	Children	Overall
$F1$	7.5 (1.3%)	9.2 (1.5%)	10.7 (1.8%)	9.2 (1.5%)
$F2$	14.2 (0.8%)	20.0 (1.1%)	18.5 (1.1%)	17.6 (1.0%)
$F3$	18.6 (0.7)	21.2 (0.7%)	27.6 (1.0%)	22.5 (0.8%)
$F4$	20.6 (0.5%)	31.5 (0.8%)	36.4 (0.9%)	29.5 (0.7%)

For the present study, formants were edited only between the starting and ending times of the vowel. Contours for $F1-F3$ were measured for all signals, except in cases of unresolvable formant mergers. The fourth formant was measured only when a well-defined $F4$ contour was clearly visible both on the LPC peak display and the gray-scale spectrogram. The fourth formant was judged to be unmeasurable for 15.6% of the utterances.

c. Fundamental frequency contours. $F0$ contours were extracted with an autocorrelation pitch tracker (Hillenbrand, 1988), followed by hand editing using the tool described above. Gross tracking errors such as pitch halving and pitch doubling were corrected by reanalyzing the signal with an option that imposes an upper or lower limit on the search for the autocorrelation peak. Any errors that remained were corrected using the editing commands that were described above.

B. Results

1. Measurement reliability

a. Vowel duration. Vowel durations for 10% of the utterances were remeasured independently by a second experimenter. The utterances chosen for remeasurement were drawn at random from the total of 1668 signals, but with approximately equal numbers of men, women, and children. The averaged absolute difference between the original and remeasured durations was 6.9 ms. This result is in line with reliability data for vowel duration reported by Allen (1978) and Smith *et al.* (1986).

b. Steady-state times. Steady-state times were measured by two experimenters for all 1668 utterances. The average absolute difference between the two measurements was 21.1 ms, or 7.7% of average vowel duration. However, more important than the time difference between these two measurements is the difference in the formant frequency pattern at these two sample points. These results, shown in Table II, indicate that formant frequencies at the two sample points typically differed by roughly 1% of average formant frequency.

c. Formant frequencies. Two methods were used to estimate the reliability of the formant frequency measurements. The first method involved a simple reanalysis of 10% of the utterances using the LPC-based signal processing and editing techniques described previously. The second method involved a reanalysis of 10% of the utterances using the same peak picking and editing techniques but with 128-point cepstrally smoothed spectra instead of LPC spectra. The primary motivation for this comparison was Di Benedetto's (1989a)

TABLE III. Measurement–remeasurement reliability for formant frequencies obtained from a randomly selected 10% of the signals. Values are given as average absolute differences and average signed differences.

	Men		Women		Children		Overall	
	Abs	Signed	Abs	Signed	Abs	Signed	Abs	Signed
*F*1	8.1	−1.8	12.2	−3.3	14.2	−2.5	11.7	−2.6
*F*2	20.7	2.8	26.4	2.8	27.4	2.4	25.2	3.4
*F*3	23.1	1.2	28.2	1.2	34.0	8.9	28.7	2.5
*F*4	56.2	−3.1	50.7	−3.1	69.1	15.4	59.0	4.0

report that LPC produced comparable estimates of *F*2 and *F*3 but estimates of *F*1 that were low when compared with smoothed wideband Fourier spectra. The analysis carried out in the present study consisted of calculating Fourier spectra over 16 ms (256 point) hamming windowed segments every 8 ms followed by cepstral smoothing. Cepstral smoothing was implemented with the "smooft" algorithm from Press *et al.* (1988). The size of the smoothing window was adjusted individually for each utterance to minimize spurious peaks or eliminate formant mergers. In this sense, the degree-of-smoothing parameter performed a role in the cepstrum analysis comparable to the number of poles in the LPC analysis. The peak picking and editing procedures described previously were used to extract formant frequencies from the cepstrally smoothed spectra.

Results for the LPC remeasurement are shown in Table III. The results are based on a frame-by-frame comparison of the signals, excluding from consideration any frame in which either signal showed a merger in the formant slot being compared. Results are given as average absolute differences and as signed differences. Overall, the absolute differences ranged from about 12 to 60 Hz, or between 1.0% and 2.0% of average formant frequency.

Table IV compares formant measurements obtained from LPC and cepstrally smoothed spectra. Positive numbers in the signed-difference columns indicate that the LPC-derived formants were higher in frequency than those derived from cepstrally smoothed spectra. In light of Di Benedetto's (1989a) findings, the signed differences are of particular interest. Consistent with Di Benedetto's results, the signed differences are quite small for formants above *F*1, especially as a percent of formant frequency. However, unlike Di Benedetto's findings, our results showed slightly higher first formants from LPC spectra. This discrepancy might be due to differences between the cepstral smoothing method used in the present study and the "pseudospectrum" method used by Di Benedetto. However, it should be noted that Di Benedetto's findings were based on analyses of utterances spoken by just two men and one woman.

d. Fundamental frequency. Remeasurement of fundamental frequency contours for 10% of the utterances showed a frame-by-frame average absolute difference of 1.7 Hz and an average signed difference of 0.6 Hz.

2. Measurement results

Acoustic measurements for the /hVd/ signals are shown in Table V. Fundamental frequency and formant values were sampled at the steady-state times determined by one of the two experimenters who made these judgments. The averages shown in the table, and the data displayed in the subsequent figures, are based on measurements from individual tokens that were well identified in the listening study, to be described in the next section. Specifically, for the purposes of these calculations, measurements were not included from individual tokens that produced an identification error rate of 15% or greater, where "error" simply means any instance in which a signal was identified as a vowel other than that intended by the talker. Using this criterion, the averages in this table are based on measurements from 88.5% of the signals. This allows an analysis of measurements for signals for which the talkers and listeners are in good agreement about the vowel that was spoken. In general, the removal of the more ambiguous signals had very little effect on the averages, with the important exception of /ɔ/. As will be discussed in the next section, there were several instances of attempts at /ɔ/ that were poorly identified and, in some cases, consistently identified as /ɑ/.

a. Vowel duration. The pattern of durational differences among the vowels is very similar to that observed in connected speech. Our vowel durations from /hVd/ syllables are two-thirds longer than those measured in connected speech by Black (1949), but correlate strongly ($r=0.91$) with the connected speech data. There were significant differences in vowel duration across the three talker groups ($F[2,33]=9.04$, $p<0.001$). Newman-Keuls *post-hoc* analyses showed significantly shorter durations for the men when compared to either the women or the children. Longer durations for the children were expected based on numerous developmental studies (e.g., Smith, 1978; Kent and Forner, 1980) but the differences between the men and the women were not expected. We do not have an explanation for this finding and do not know if these male–female duration differences would also be seen in conversational speech samples.

b. Fundamental frequency. Figure 2 compares our average values of fundamental frequency with those of PB for

TABLE IV. Comparison of formant frequencies derived from LPC analysis and cepstral smoothing. Results are given as average absolute differences and as average signed differences (LPC-cepstrum). Figures in parentheses are differences as a percent of average formant frequency.

	Absolute	Signed
*F*1	53.5 (8.9%)	41.5 (6.9%)
*F*2	60.4 (3.5%)	12.7 (0.7%)
*F*3	74.0 (2.6%)	10.9 (0.4%)
*F*4	91.1 (2.3%)	7.4 (0.2%)

TABLE V. Average durations, fundamental frequencies, and formant frequencies of vowels produced by 45 men, 48 women, and 46 children. Averages are based on a subset of the tokens that were well identified by listeners (see text for details). The duration measurements are in ms; all others are in Hz.

		/i/	/ɪ/	/e/	/ɛ/	/æ/	/ɑ/	/ɔ/	/o/	/ʊ/	/u/	/ʌ/	/ɝ/
Dur	M	243	192	267	189	278	267	283	265	192	237	188	263
	W	306	237	320	254	332	323	353	326	249	303	226	321
	C	297	248	314	235	322	311	319	310	247	278	234	307
*F*0	M	138	135	129	127	123	123	121	129	133	143	133	130
	W	227	224	219	214	215	215	210	217	230	235	218	217
	C	246	241	237	230	228	229	225	236	243	249	236	237
*F*1	M	342	427	476	580	588	768	652	497	469	378	623	474
	W	437	483	536	731	669	936	781	555	519	459	753	523
	C	452	511	564	749	717	1002	803	597	568	494	749	586
*F*2	M	2322	2034	2089	1799	1952	1333	997	910	1122	997	1200	1379
	W	2761	2365	2530	2058	2349	1551	1136	1035	1225	1105	1426	1588
	C	3081	2552	2656	2267	2501	1688	1210	1137	1490	1345	1546	1719
*F*3	M	3000	2684	2691	2605	2601	2522	2538	2459	2434	2343	2550	1710
	W	3372	3053	3047	2979	2972	2815	2824	2828	2827	2735	2933	1929
	C	3702	3403	3323	3310	3289	2950	2982	2987	3072	2988	3145	2143
*F*4	M	3657	3618	3649	3677	3624	3687	3486	3384	3400	3357	3557	3334
	W	4352	4334	4319	4294	4290	4299	3923	3927	4052	4115	4092	3914
	C	4572	4575	4422	4671	4409	4307	3919	4167	4328	4276	4320	3788

each of the ten vowels common to the two studies. Data points that lie on the solid line in the scatter plot indicate identical values, while data points above the line indicate higher *F*0 values for PB. Average *F*0 values for the men and the women typically differed by only a few Hz when compared to the corresponding vowels recorded by PB. *F*0 values for our children averaged 28 Hz lower than the PB data.

c. Formant frequencies. Figure 3 shows the average frequencies for *F*1 and *F*2 for the three talker groups, along with ellipses fit to each vowel category. Figure 4 shows the individual data points. To improve the clarity of the display, data from /e/ and /o/ have been omitted, and the database has been thinned of redundant data points, resulting in the display of approximately half of the individual points. While there are clearly gross similarities to the PB data, there are numerous differences as well. The degree of crowding among adjacent vowel categories appears much greater than in the PB data, and many of the vowels are in different locations in *F*1–*F*2 space than in PB.

Figures 5–7 show acoustic vowel diagrams based on average formant frequencies from the present study and from PB for men, women, and children, respectively. It is difficult to arrive at a simple summary of the differences that are seen in these figures. However, to the extent that conventional articulatory interpretations of formant data are valid, a few general observations can be made. The formant data seem to imply a general tendency toward lower tongue positions in

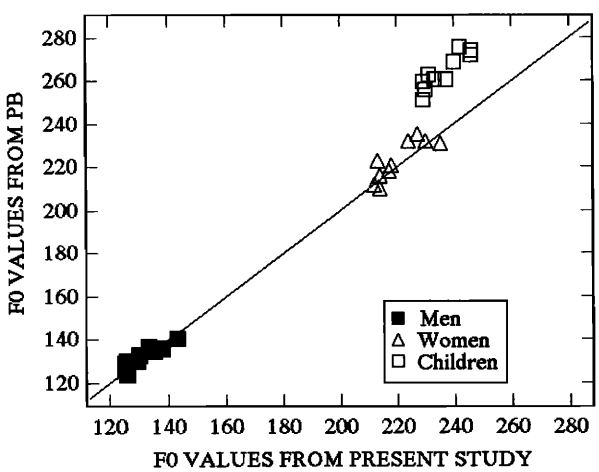

FIG. 2. Scatter plot of *F*0 values from the present study and from Peterson and Barney (1952) for each of the ten vowels common to the two studies. Data points above the solid line indicate higher *F*0 values for the Peterson and Barney data.

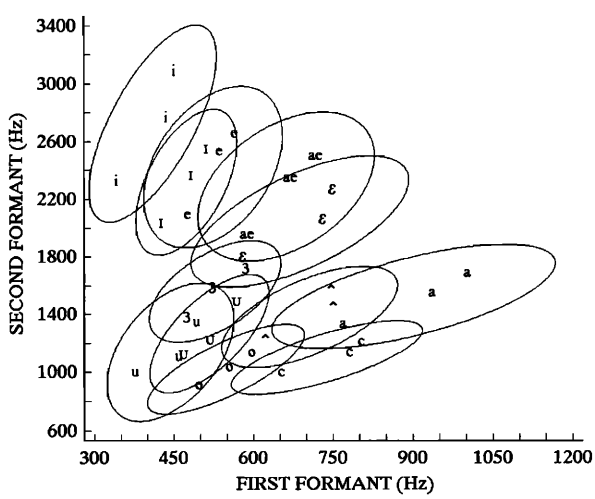

FIG. 3. Average values of *F*1 and *F*2 for men, women, and child talkers for 12 vowels with ellipses fit to the data ("ae"=/æ/, "a"=/ɑ/, "c"=/ɔ/, "ʌ"=/ʌ/, "ɜ"=/ɝ/).

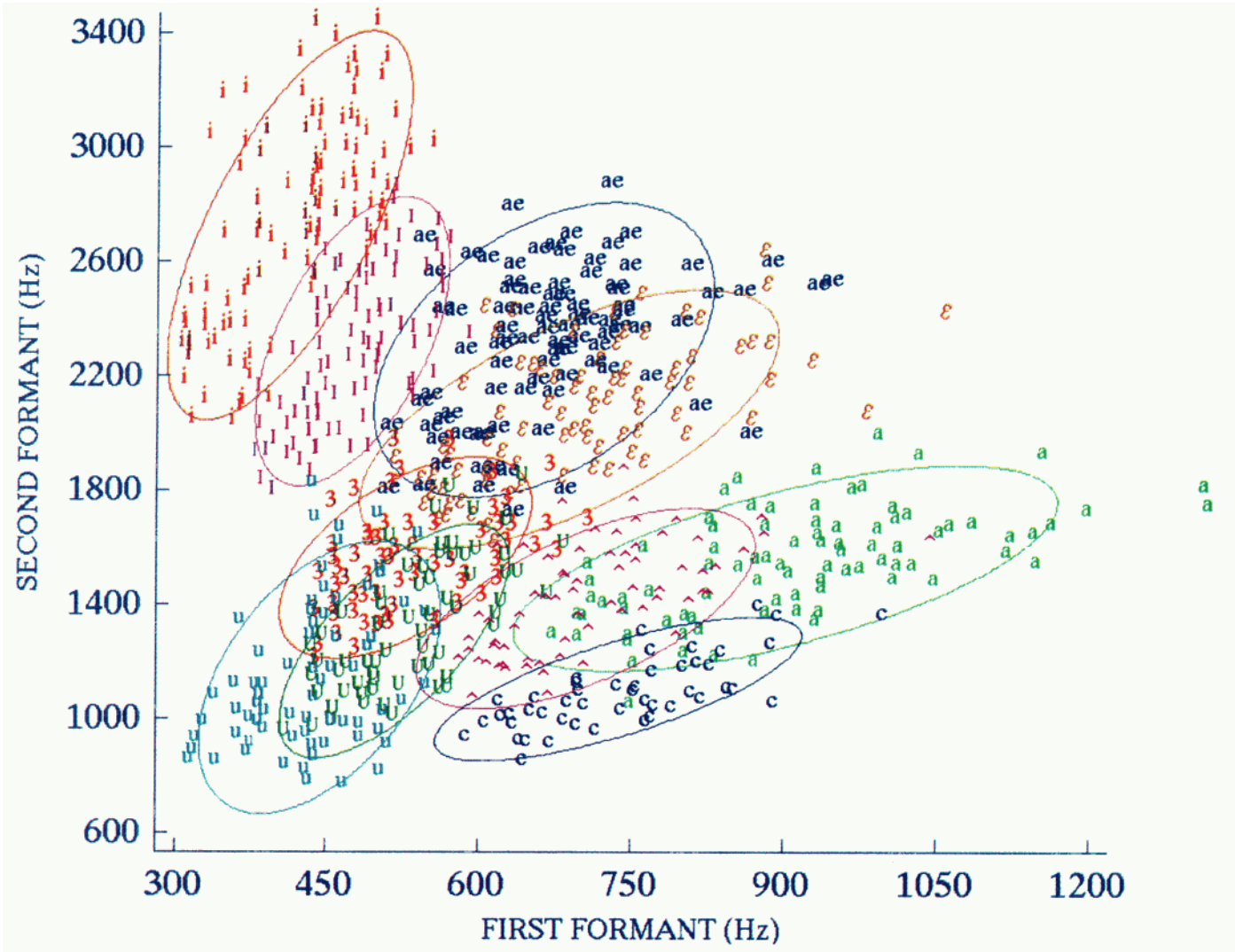

FIG. 4. Values of $F1$ and $F2$ for 46 men, 48 women, and 46 children for 10 vowels with ellipses fit to the data ("ae"=/æ/, "a"=/ɑ/, "c"=/ɔ/, "ʌ"=/ʌ/, "ɜ"=/ɝ/). Measurements for /e/ and /o/ have been omitted, and the data have been thinned of redundant data points.

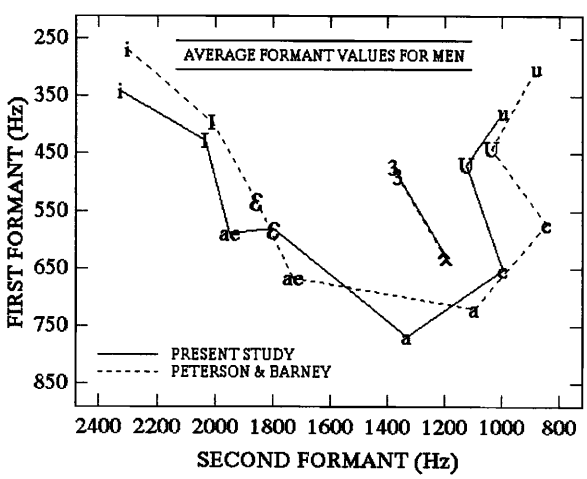

FIG. 5. Acoustic vowel diagrams showing average formant frequencies for men from the present study and from Peterson and Barney ("ae"=/æ/, "a"=/ɑ/, "c"=/ɔ/, "ʌ"=/ʌ/, "ɜ"=/ɝ/).

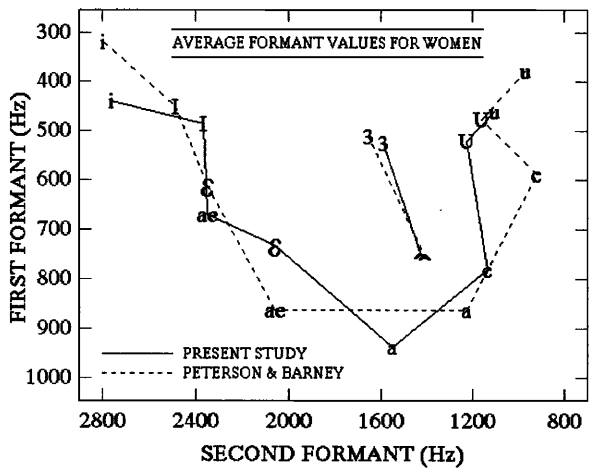

FIG. 6. Acoustic vowel diagrams showing average formant frequencies for women from the present study and from Peterson and Barney ("ae"=/æ/, "a"=/ɑ/, "c"=/ɔ/, "ʌ"=/ʌ/, "ɜ"=/ɝ/).

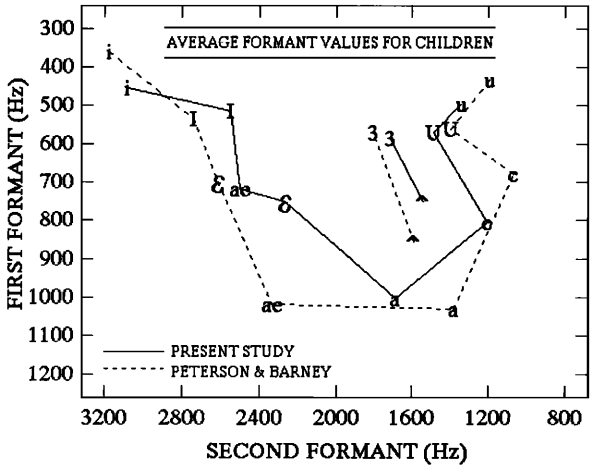

FIG. 7. Acoustic vowel diagrams showing average formant frequencies for children from the present study and from Peterson and Barney ("ae"=/æ/, "a"=/ɑ/, "c"=/ɔ/, "ʌ"=/ʌ/, "ɜ"=/ɝ/).

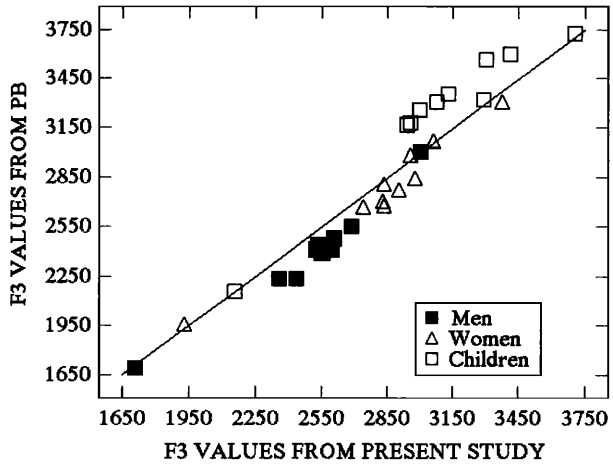

FIG. 8. Scatter plot of $F3$ values from the present study and from Peterson and Barney (1952). Data points above the solid line indicate higher $F3$ values for the Peterson and Barney data.

our data and, for the back vowels, more anterior tongue positions. It should be noted, of course, that these differences in the formant patterns could be explained by differences in lip posture instead of—or perhaps in addition to—differences in tongue position. The children in our study seem to show a general tendency toward centralization when compared to the PB data. The central vowels /ʌ/ and /ɝ/ produced by the adult talkers are alone in occupying nearly identical positions in the two sets of formant data.

The vowels occupy similar relative positions in the two sets of data, with the notable exception of /ɛ/ and /æ/. Our data indicate higher $F2$ values for /æ/ as compared with /ɛ/, and slightly lower $F1$ values for /æ/ than /ɛ/, although the difference in $F1$ is not consistent across talker groups. Analysis of data for individual talkers showed that the $F2$ differences between these two vowels were highly consistent: 91% of the talkers produced an /æ/ with a higher $F2$ than their /ɛ/. The $F1$ differences were less consistent, with 68% of the talkers producing an /æ/ with a lower $F1$ value than their /ɛ/. Our findings for these two vowels contrast not only with PB, but also with Di Benedetto's (1989a) results from three adult talkers, and with a large Texas Instruments database described by Syrdal (1985). As can be seen in Figs. 3 and 4, these two vowels show a very high degree of overlap in $F1-F2$ space. As will be seen below, these vowels are well identified by listeners, and can be separated well based on acoustic measurements only if spectral change is taken into account.

Figure 8 compares our average values of $F3$ with those of PB. Overall, the $F3$ values from the two studies are quite similar, with our measures averaging 113 Hz (4.7%) higher for the men, 47 Hz (1.7%) higher for the women, and 174 Hz (5.5%) lower for the children. The slightly higher $F3$ values for PB's children are consistent with the $F0$ differences described previously.

d. Spectral change patterns. Although the primary purpose of this study was to compare our static formant measurements with those of PB, a preliminary analysis was conducted of the patterns of formant frequency change associated with these utterances. Figure 9 is based on the formant pattern sampled at 20% and 80% of vowel duration, averaged across vowels produced by the three groups of talkers. $F0$ was sampled just once at steady state. The values have been converted to a mel scale using the technical approximation from Fant (1973) and are represented as $F1-F0$ vs $F3-F2$. (Note that the $F3-F2$ axis has been inverted to produce a display that more closely resembles a conventional $F1/F2$ plot.) This representation is similar to models proposed by Miller (1989) and Syrdal (1985), except that a mel scale is used in place of Miller's log scale and Syrdal's Bark scale.[2] The symbol identifying the vowel is plotted at the location of the second sample of the formant pattern, and a line connects this point to the first sample. As the figure

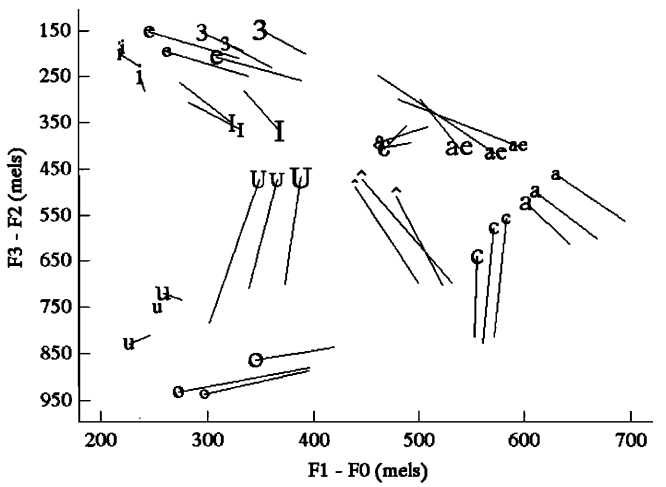

FIG. 9. Spectral change patterns associated with the 12 vowels ("ae" =/æ/, "a"=/ɑ/, "c"=/ɔ/, "ʌ"=/ʌ/, "ɜ"=/ɝ/). The abscissa is the difference between mel-transformed values of $F1$ and $F0$, and the ordinate is the difference between mel-transformed values of $F2$ and $F3$. The $F3-F2$ axis has been inverted to produce a display that more closely resembles a conventional $F1/F2$ plot. The symbol identifying the vowel is plotted at the location of the second sample of the formant pattern, and a line connects this point to the first sample. The largest symbols are used for the men and the smallest symbols are used for the children.

indicates, nearly all of the vowels show a good deal of formant frequency change. Further, the formants are moving in such a way as to enhance the contrast between vowels with similar static positions in formant space. For example, the /æ/–/ɛ/ pair shows a high degree of overlap when the formants are sampled at steady state. As the figure shows, these two vowels appear to exhibit distinct spectral change patterns. Likewise, /u/ and /ʊ/ show a high degree of overlap in static $F1/F2$ space but appear to show distinct patterns of spectral change. The influence of spectral change patterns on the separability of vowel categories will be examined in a more systematic way in the discriminant analysis studies described in Sec. III.

II. VOWEL IDENTIFICATION

A. Methods

1. Listeners

Listeners consisted of 20 undergraduate and graduate students in the Speech and Pathology and Audiology Department at Western Michigan University, none of whom had participated as talkers. The choice of phonetically trained listeners was motivated by the findings of Assmann *et al.* (1982) indicating that a relatively large proportion of identification errors produced by untrained listeners are due primarily to the listeners' uncertainty about how to map perceived vowel quality onto orthographic symbols. All of the listeners had taken an undergraduate course in phonetics, although as a group they would not be considered experienced phoneticians. The dialect screening procedure and subject selection criteria described for the subjects who served as talkers were used for the listeners as well.

2. Procedures

Listeners were tested individually in a quiet room in two sessions lasting approximately 1 h each. Signals were low-pass filtered at 7.2 kHz at the output of a D/A converter (Tucker and Davis DD1), amplified, and delivered to a single loudspeaker (Boston Acoustics A60) at an average intensity of 77 dBA at the listener's head (approximately 70 cm from the loudspeaker). Over the course of the two sessions listeners identified one presentation of each of the 1668 /hVd/ signals. The signals were presented in fully random order (i.e., not blocked by talker), and the randomization was changed daily. This randomization method differs from PB, who tested subjects in blocks of trials which presented listeners with randomly ordered tokens from ten talkers.[3]

Subjects responded by pressing one of 12 keys on a computer keyboard that had been labeled both with the phonetic symbols and the corresponding key words (e.g., "heed," "hid," "head," etc.). Each listening test was preceded by a brief practice session to ensure that subjects understood how the key labels were to be interpreted.

3. Results

Table VI presents a summary of identification rates for each vowel category, along with comparable data from PB. The full confusion matrix is shown in Table VII. The results are generally quite similar to PB. The overall identification rates from the two studies are quite similar, as are the rates for most of the individual vowels. Our results show slightly poorer identification of the /ɑ/–/ɔ/ pair, but somewhat better identification of /ɪ/ and of the /æ/–/ɛ/ pair. The relatively high identifiability of /æ/ and /ɛ/ is interesting in light of the poor separation of these vowels based on static measures of $F1$ and $F2$.

Figure 10 shows a histogram of identification rates for the individual signals. The majority of the signals (65%) were identified unanimously by the listeners, and 89% were identified at rates of 90% or greater. For 33 signals (2%) the majority vote of the listeners was a vowel other than that intended by the talker. Most of these (55%) were intended as /ɔ/ but heard as /ɑ/, but there were also instances such as /æ/ heard as /ɛ/, /ɛ/ heard as /æ/, /ɑ/ heard as /ɔ/, and /ɪ/ heard as /ɛ/. It seems logical to interpret signals consistently heard as a vowel other than that intended by the talker as production errors. While the number of tokens involved is a small proportion of the total, by this criterion nearly 14% of the attempts at /ɔ/ were production errors. It is clear, then, that the dialect screening procedures were ineffective in some cases; that is, there were some speakers who passed the /ɑ/–/ɔ/ dialect screening but, for reasons that are not clear, did not produce a convincing /ɔ/ when the /hVd/ syllables were recorded.

Overall identification rates were highest for the women talkers and lowest for the children. A repeated-measures analysis of variance for talker gender using arcsine transformed overall identification rates for the 20 listeners was significant [$F(2,57)=14.7$, $p<0.001$]. *Post-hoc* analysis showed significant differences among all three talker groups. It is important to note, however, that the magnitude of the effect is quite small, with only 1.9% separating the highest and lowest identification rates. It is interesting that there is no evidence at all that signals with higher fundamental frequencies are more poorly identified, despite the fact that formant peaks are more poorly defined in signals with wide harmonic spacing.

III. DISCRIMINANT ANALYSIS

The purpose of the discriminant analyses was to determine how well the vowels could be separated based on various combinations of the acoustic measurements and, where appropriate, to compare these results with similar analyses of the PB database. A quadratic discriminant analysis technique (Johnson and Winchern, 1982) was used for classification. In all cases, the "jackknife" technique was used in which statistics for an individual token are removed from the training data prior to the attempt to classify the token (see also Syrdal and Gopal, 1986). Tokens showing a merger in a formant slot that was included in the parameter list were not included in the analyses.

Table VIII shows classification results for static parameter sets, i.e., measurements sampled once at steady state. To facilitate comparisons with PB, /e/ and /o/ were not included in these tests. As might have been expected based on inspection of Figs. 3 and 4, our vowels do not separate as well as the PB data based on static measures of $F1$ and $F2$. The

differences in category separability become smaller as more parameters are added, but in all cases the PB vowels are classified with greater accuracy than ours.

Table IX demonstrates the effects of including vowel duration and spectral change information on classification accuracy. Only data from the present study are included, again omitting /e/ and /o/. The one-sample results are based on a single sample of the formant pattern at steady state, the two-sample results are based on samples taken at 20% and 80% of vowel duration, and the three-sample results are based on samples taken at 20%, 50%, and 80% of vowel duration. For the two- and three-sample parameter sets including $F0$, a single sample of $F0$ at steady state was used. Results are shown for all tokens in the database, and for a subset of the data that excluded individual tokens that showed identification error rates of 15% or greater (11.5% of the tokens).

It can be seen that including vowel duration in the parameter set results in a consistent improvement in performance, especially for the simplest parameter sets such as single-sample $F1-F2$. However, the most dramatic effect is seen when comparing a single sample of the formant pattern with two samples. The improvement in classification accuracy averages 11.2%, and is especially large for the parameter sets involving $F1$ and $F2$ alone. Adding a third sample of the formant pattern produces little or no improvement in classification accuracy. This would seem to suggest that only a very coarse representation of the spectral change pattern is needed for classification. It can also be seen that omitting tokens with relatively high identification error rates produces a consistent improvement in classification accuracy. This finding indicates that the errors produced by the pattern classifier tend to occur more often for tokens that are poorly identified by listeners.

Although not shown in the table, the same classification tests were conducted using the full set of 12 vowels. The overall pattern of results was quite similar to that shown in Table IX, except that the improvement in classification accuracy with two samples of the formant pattern was somewhat larger. This is a logical result given that the two vowels that were added, /e/ and /o/, are nearly always diphthongized.

IV. DISCUSSION

The original intent of this study was to collect a database of acoustic measurements for /hVd/ utterances comparable to PB, but with additional measures of duration and spectral change that could be used to study the role of dynamic properties in vowel recognition. The differences that were observed between our static measurements of formant patterns and those of PB were not anticipated. One possible explanation of these differences has to do with our use of LPC as opposed to the more direct spectrum analysis method used by PB. This possibility cannot be eliminated entirely, particularly since we attempted no direct comparisons of our LPC measurements with measures obtained using PB's spectrographic technique. However, the differences between our data and those of PB strike us as both too numerous and too diverse to be explained by differences in spectrum analysis methods. Although not entirely conclusive, the comparisons that were made between formants measured from LPC and cepstrally smoothed Fourier spectra also make it seem unlikely that the differences can be attributed to spectrum analysis methods. The close similarity in $F3$ values and the nearly identical formant values for the central vowels produced by the adult talkers from the two studies would also seem to argue against this interpretation.

It also seems unlikely that the discrepancies can be attributed to differences in the times at which the formant patterns were sampled. Our data showed that steady-state times could be located with a surprisingly high degree of reliability. In addition, in data not reported here, we determined for three vowels (/ɛ/, /æ/, and /u/) that even the most procrustean method of locating steady state times could not bring the two sets of formant data into convergence.[4]

If we assume, then, that the acoustic measurements were made with roughly equal precision in the two studies, then it must be the case that many of these vowels were simply produced in different ways by the two groups of talkers. As was indicated previously, little is known about the dialect of the PB talkers except: (1) Most of the women were raised in the mid-Atlantic region; (2) the men represented "... a broad regional sampling of the United States ..." (Peterson and Barney, 1952, p. 177); (3) a "few" of the talkers learned English as a second language; and (4) "most" of the talkers spoke General American. Perhaps more important than potential differences in regional dialect is the passage of some 40 years in the times at which the two sets of recordings were made. It is well known that significant changes in speech production can occur over a period of several decades. For example, Bauer (1985) showed clear evidence of significant diachronic vowel shifts when comparing recordings of British RP speakers made in 1949 with similar recordings made in 1966. A second comparison of comparable recordings made in 1982 showed a continuation of the same vowel shifts. There are almost certainly some differences between our data and those of PB that can be attributed, at least in part, to diachronic change. For example, the "raising" of /æ/, with the consequent reduction in contrast in the static formant positions for /æ/ and /ɛ/, has been well documented in several dialects of American English (e.g., Labov *et al.*, 1972).

It should also be noted that the Texas Instruments database described by Syrdal (1985) shows some rather large differences from PB. The TI values for the front vowels are quite similar to PB, but rather large differences are seen for the back vowels. The differences from PB are in the same direction as in our data (i.e., implying lower and more anterior tongue positions compared to PB), but the discrepancies are even larger. As with the PB study, little is known about the dialect of the speakers in the TI study.

There is one final point worth noting about the discrepancies in formant frequencies across these three studies. There has been a tendency to view the PB database as a benchmark of sorts, establishing *the* set of formant frequencies for American English vowels. For example, the PB measurements are frequently used as control parameters in speech synthesis studies, and often serve to define prototypes for vowel categories. The PB measurements have also been

heavily used to evaluate vowel normalization algorithms and are frequently used in cross-language comparisons and comparisons between normal and disordered speech. The present results, along with those of Syrdal (1985) and Bauer (1985), serve as a reminder that a study of this kind can only hope to establish a set of formant frequencies that are typical of a specific dialect at a specific time in the history of that dialect.

In contrast to the numerous differences in acoustic measurements between our study and PB, the two listening studies produced very similar results. The overall identification rates from the two studies are quite similar, as are the rates for individual vowels. Although a detailed analysis of the relationships between the acoustic and perceptual data from the present study will have to await further study, it seems quite clear that the frequencies of $F1$ and $F2$ at steady state are not good predictors of the identification results. The clearest example is the /æ/–/ɛ/ pair, which was identified quite well by listeners despite very poor separation in static $F1-F2$ space.

Another indication that static measures of $F1$ and $F2$ are poor predictors of vowel identification is the general finding that the significantly increased crowding of vowels in static $F1-F2$ space relative to PB was not accompanied by an increase in perceptual confusions among vowels. Although it can only be guessed at, one possibility that might be considered is that our talkers produced more heavily diphthongized vowels than PB's talkers. According to this view, the greater degree of crowding in the static formant space of our talkers might be offset by an increase in spectral change, resulting in a set of vowels that are as distinct as the PB vowels. However, this speculation cannot be confirmed and, in the absence of the original PB recordings, it is difficult to go beyond the general suggestion of dialect differences between the two groups of talkers.

It is important to note that listeners were asked to identify utterances by choosing a label from a closed set of broad phonemic categories. It should not be concluded that all utterances that were assigned the same phonemic label are phonetically equivalent (see Ladefoged, 1967, for a discussion). Even a casual listening by an experienced phonetician shows clearly that there is a range of phonetic qualities within the vowel categories, even when considering only tokens that were well identified by listeners. It might be useful to reanalyze a subset of the utterances that are judged by experienced phoneticians to be good examples of the various vowel qualities. It seems likely that within-vowel-category variability in the acoustic measures for a subset of this kind would be substantially reduced.

TABLE VI. Overall percent correct identification by vowel category. for the present study (HGCW) and for Peterson and Barney (PB).

	HGCW	PB
/i/	99.6	99.9
/ɪ/	98.8	92.9
/e/	98.3	a
/ɛ/	95.1	87.7
/æ/	94.1	96.5
/ɑ/	92.3	87.0
/ɔ/	82.0	92.8
/o/	99.2	a
/ʊ/	97.5	96.5
/u/	97.2	99.2
/ʌ/	90.8	92.2
/ɝ/	99.5	99.7
Total:	95.4	94.4
Men:	94.6	b
Women:	95.6	b
Children	93.7	b

[a]These vowels were not recorded by Peterson and Barney.
[b]Peterson and Barney did not report results separately for men, women, and child talkers.

One other aspect of the listening test deserves comment. The large number of speakers and full randomization of both talkers and vowels make it very unlikely that listeners could have made use of a vocal tract normalization process of the kind described by Ladefoged and Broadbent (1957) and others. While there is some evidence that listeners can make use of speaker-specific normalizing information in certain kinds of tasks, the relatively high identification rates obtained in the present study would seem to indicate that accurate vowel identification does not require calibration to individual speakers.

The discriminant analysis results showed that our vowels could not be separated well based on a single sample of

TABLE VII. Confusion matrix for /hVd/ utterances produced by 45 men, 48 women, and 46 children.

		Vowel identified by listener											
		/i/	/ɪ/	/e/	/ɛ/	/æ/	/ɑ/	/ɔ/	/o/	/ʊ/	/u/	/ʌ/	/ɝ/
	/i/	**99.6**	0.1	0.1									0.1
	/ɪ/		**98.8**	0.2	0.9								
	/e/	0.6	0.3	**98.3**	0.3		0.2	0.1		0.1			
Vowel	/ɛ/		0.5		**95.1**	3.7	0.2	0.1		0.1		0.2	0.1
intended	/æ/				5.6	**94.1**	0.2						
by	/ɑ/		0.1		0.1	0.3	**92.3**	3.5	0.1	0.2	0.1	3.3	
talker	/ɔ/						13.8	**82.0**		0.1	0.1	3.8	0.2
	/o/							0.1	**99.2**	0.5			0.2
	/ʊ/				0.1					**97.5**	1.3	1.0	
	/u/		0.3						0.4	1.9	**97.2**		0.1
	/ʌ/						3.7	1.8	0.3	3.2	0.2	**90.8**	0.2
	/ɝ/	0.1				0.2				0.2			**99.5**

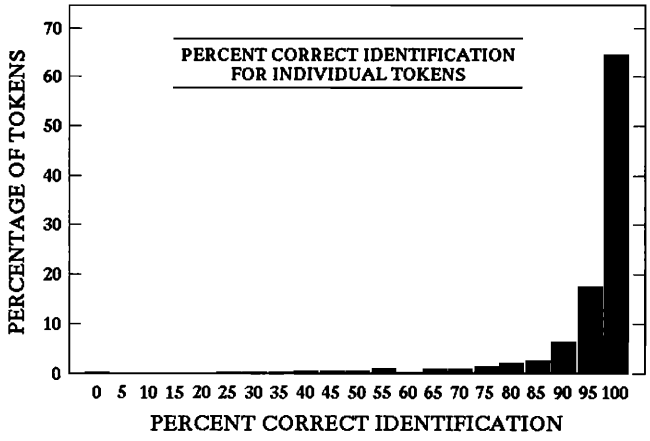

FIG. 10. Histogram of percent correct identification rates for individual tokens, where "correct" means that the listener identified the vowel as the one intended by the talker.

the formant pattern, especially $F1$ and $F2$ alone. The same is true of the PB data, but to a lesser degree. Adding vowel duration measures resulted in consistent but fairly modest improvements in classification accuracy, and including two samples of the formant pattern produced large improvements in category separability. These findings are consistent with Zahorian and Jagharghi (1993), who showed much better classification rates for dynamic versus static representations of both formants and overall spectral shape. Zahorian and Jagharghi are also in agreement in showing much larger improvements in classification accuracy with the addition of spectral change as compared to vowel duration. The present results, along with those of Zahorian and Jagharghi, are consistent with many recent findings suggesting that the vowels of American English are more appropriately viewed not as points in phonetic space but rather as trajectories through

TABLE VIII. Quadratic discrimination results for the present data (HGCW) and for the PB data set based on a single sample of the formant pattern. The table shows overall classification accuracy using the "jackknife" method in which measurements for individual tokens are removed from the training statistics prior to classification.

Parameter set	HGCW	PB
$F1,F2$	68.2	74.9
$F1,F2,F3$	81.0	83.6
$F0,F1,F2$	78.2	85.9
$F0,F1,F2,F3$	84.7	86.6

phonetic space. For example, several studies have shown very high identification rates for "silent center" stimuli consisting of onglides and offglides only (e.g., Jenkins *et al.*, 1983; Nearey, 1989; Nearey and Assmann, 1986). Complementing these results are Hillenbrand and Gayvert's (1993b) findings showing that steady-state vowels synthesized from the PB measurements are not well identified by listeners (see also Fairbanks and Grubb, 1961). Taken together, the silent center and steady-state resynthesis results suggest that static spectral targets are neither necessary nor sufficient for accurate vowel recognition.

It is interesting in this regard that the importance of dynamic information in vowel identification was recognized by PB, who commented, "It is the present belief that the complex acoustical patterns represented by the words are not adequately represented by a single section, but require a more complex portrayal" (Peterson and Barney, 1952, p. 184). The precise nature of this "more complex portrayal" remains unclear, however, since we still do not know how listeners map spectral change patterns onto perceived vowel quality. Our discriminant analysis findings indicated that a fairly coarse, two-sample representation of the formant pattern is all that is necessary for accurate vowel classification.

TABLE IX. Quadratic discrimination results for the present data showing the effect of including duration and spectral change information on classification accuracy. The table shows overall classification accuracy using the "jackknife" method in which measurements for individual tokens are removed from the training statistics prior to classification. The one-sample results are based on a single sample of the formant pattern at "steady state;" the two-sample results are based on samples taken at 20% and 80% of vowel duration; the three-sample results are based on samples taken at 20%, 50%, and 80% of vowel duration. ("NoDur"=vowel duration not included; "Dur"=vowel duration included.) Entries under the heading of "All tokens" used the full database; entries under the heading "well identified tokens only" are based on a data set that did not include tokens with error rates of 15% or greater (11.5% of the tokens).

	All tokens					
	One sample		Two samples		Three samples	
Parameter set	NoDur	Dur	NoDur	Dur	NoDur	Dur
$F1,F2$	68.2	76.1	87.9	90.3	87.7	90.4
$F1,F2,F3$	81.0	84.6	91.6	92.7	91.8	93.1
$F0,F1,F2$	78.2	82.0	91.6	92.5	91.0	92.6
$F0,F1,F2,F3$	84.7	87.8	93.6	94.1	92.8	94.8
	Well identified tokens only					
	One sample		Two samples		Three samples	
Parameter set	NoDur	Dur	NoDur	Dur	NoDur	Dur
$F1,F2$	71.4	80.0	90.8	93.6	90.7	93.3
$F1,F2,F3$	85.3	89.1	95.4	96.2	95.3	95.8
$F0,F1,F2$	82.3	86.3	95.5	96.3	94.8	96.0
$F0,F1,F2,F3$	88.7	91.6	97.3	97.8	96.6	97.3

However, that does not imply that the details of the formant change pattern are unimportant to the listener. Additional studies using synthesis methods are needed to learn more about the specific mapping relations that are involved in vowel recognition. Also needed are studies of spectral change patterns in more complex phonetic environments than the /hVd/ utterances examined here (e.g., Stevens and House, 1963). While it seems certain that the associations which we observed between vowel categories and spectral change patterns will be less straightforward in more complex phonetic environments, the extent of the oversimplification in our data is as yet unknown.

ACKNOWLEDGMENTS

We are very grateful to Michelle Malta, who donated many hours to this project and set the tone for dedication and attention to detail. The able assistance of Emily Weil and Matt Phillips is also gratefully acknowledged. We would also like to thank Terry Nearey for helpful advice provided at various points along the way, and Jim Flege for comments on a previous draft. This work was supported by a research grant from the National Institutes of Health (NIDCD 1-R01-DC01661) and by the Air Force Systems Command, Rome Air Development Center, Griffiss Air Force Base, and the Air Force Office of Scientific Research (Contract No. F30602-85-C-0008).

[1]The issue of whether to screen talkers' productions based on the experimenter's phonetic judgments generated a great deal of discussion as the project was being planned. We ultimately settled on the method described above, in which we trained the subjects as carefully as possible but did not intervene once the tape recorder was started, except in cases of dysfluency or obvious reading errors. The reason that we chose this approach is that, along with PB, we wanted to determine how accurately listeners identified naturally produced vowels, where "accurate" is defined as agreement between the listener and the intent of the speaker. If stimuli are screened based on the experimenter's judgments of correct production, the question changes from a two-way agreement between the talker and the listener to a three-way agreement among the talker, the listener, and the experimenter. It is not known whether stimuli were screened for correct pronunciation when the PB recordings were made.

[2]The mel scale was chosen over the Bark and log transforms because this transformation did a better job of grouping vowels produced by men, women, and children. This representation is similar to a normalization scheme proposed by Peterson (1951).

[3]Four additional listeners were tested both on the 12-vowel, full randomization task described here and on a ten-vowel, blocked-by-speaker task comparable to PB. Results for the ten vowels common to the two tasks were very similar.

[4]A graduate student was given a copy of the PB data and asked to choose a steady-state location in such a way that the values of $F1$ and $F2$ would be closest to the averages from PB, even if that location made no sense in relation to the objective of finding the steadiest portion of the vowel. Differences between our data and those of PB remained even when this procedure was used.

Ainsworth, W. A. (**1972**). "Duration as a cue in the recognition of synthetic vowels," J. Acoust. Soc. Am. **51**, 648–651.

Allen, G. D. (**1978**). "Vowel duration measurement: A reliability study," J. Acoust. Soc. Am. **63**, 1176–1185.

Assmann, P., Nearey, T. E., and Hogan, J. (**1982**). "Vowel identification: Orthographic, perceptual, and acoustic factors," J. Acoust. Soc. Am. **71**, 975–989.

Bauer, L. (**1985**). "Tracing phonetic change in the received pronunciation of British English," J. Phon. **13**, 61–81.

Bennett, D. C. (**1968**). "Spectral form and duration as cues in the recognition of English and German vowels," Lang. Speech **11**, 65–85.

Black, J. W. (**1949**). "Natural frequency, duration, and intensity of vowels in reading," J. Speech Hear. Disord. **14**, 216–221.

Di Benedetto, M-G. (**1989a**). "Vowel representation: Some observations on temporal and spectral properties of the first formant frequency," J. Acoust. Soc. Am. **86**, 55–66.

Di Benedetto, M-G. (**1989b**). "Frequency and time variations of the first formant: Properties relevant to the perception of vowel height," J. Acoust. Soc. Am. **86**, 67–77.

Fairbanks, G., and Grubb, P. (**1961**). "A psychophysical investigation of vowel formants," J. Speech Hear. Res. **4**, 203–219.

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

Hillenbrand, J. (**1988**). "MPITCH: An autocorrelation fundamental-frequency tracker," [Computer Program], Western Michigan University, Kalamazoo, MI.

Hillenbrand, J., and Gayvert, R. T. (**1993a**). "Vowel classification based on fundamental frequency and formant frequencies," J. Speech Hear. Res. **36**, 647–700.

Hillenbrand, J., and Gayvert, R. T. (**1993b**). "Identification of steady-state vowels synthesized from the Peterson and Barney measurements," J. Acoust. Soc. Am. **94**, 668–674.

Jenkins, J. J., Strange, W., and Edman, T. R. (**1983**). "Identification of vowels in 'vowelless' syllables," Percept. Psychophys. **34**, 441–450.

Johnson, R. A., and Winchern, D. W. (**1982**). *Applied Multivariate Statistical Analysis* (Prentice-Hall, Englewood Cliffs, NJ).

Kent, R. D., and Forner, L. L. (**1980**). "Speech segment durations in sentence recitations by children and adults," J. Phon. **8**, 157–168.

Labov, W., Yaeger, M., and Steiner, R. (**1972**). "A quantitative study of sound change in progress," Report on National Science Foundation Contract NSF-FS-3287.

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

Ladefoged, P., and Broadbent, D. E. (**1957**). "Information conveyed by vowels," J. Acoust. Soc. Am. **29**, 98–104.

Lippmann, R. P. (**1989**). "Review of neural networks for speech recognition," Neural Computation **1**, 1–38.

Miller, J. D. (**1989**). "Auditory-perceptual interpretation of the vowel," J. Acoust. Soc. Am. **85**, 2114–2134.

Nearey, T. M. (**1978**). *Phonetic Feature Systems for Vowels* (Indiana University Linguistics Club, Bloomington, IN).

Nearey, T. M. (**1989**). "Static, dynamic, and relational properties in vowel perception," J. Acoust. Soc. Am. **85**, 2088–2113.

Nearey, T. M. (**1992**). "Applications of generalized linear modeling to vowel data," in *Proceedings ICSLP 92*, edited by J. Ohala, T. Nearey, B. Derwing, M. Hodge, and G. Wiebe (University of Alberta, Edmonton, AB), pp. 583–586.

Nearey, T. M., and Assman, P. (**1986**). "Modeling the role of vowel inherent spectral change in vowel identification," J. Acoust. Soc. Am. **80**, 1297–1308.

Nearey, T. M., Hogan, J., and Rozsypal, A. (**1979**). "Speech signals, cues and features," in *Perspectives in Experimental Linguistics*, edited by G. Prideaux (Benjamin, Amsterdam).

Peterson, G. E. (**1951**). "The phonetic value of vowels," Language **27**, 541–553.

Peterson, G. E., and Barney, H. L. (**1952**). "Control methods used in a study of the vowels," J. Acoust. Soc. Am. **24**, 175–184.

Peterson, G. E., and Lehiste, I. (**1960**). "Duration of syllable nuclei in English," J. Acoust. Soc. Am. **32**,693–703.

Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, M. W. T. (**1988**). "Numerical Recipes in C," (Cambridge U.P., Cambridge, MA).

Smith, B. L. (**1978**). "Temporal aspects of English speech production: A developmental perspective," J. Phon. **6**, 37–67.

Smith, B. L., Hillenbrand, J., and Ingrisano, D. R. (**1986**). "Comparison of temporal measures of speech using spectrograms and digital oscillograms," J. Speech Hear. Res. **29**, 270–274.

Stevens, K. N. (**1959**). "The role of duration in vowel identification," Q. Progr. Rep. 52, Research Laboratory of Electronics, MIT.

Stevens, K. N., and House, A. S. (**1963**). "Perturbation of vowel articulations by consonantal context: An acoustical study," J. Speech Hear. Res. **6**, 111–128.

Strange, W. (**1989**). "Dynamic specification of coarticulated vowels spoken in sentence context," J. Acoust. Soc. Am. **85**, 2135–2153.

Strange, W., Jenkins, J. J., and Johnson, T. L. (**1983**). "Dynamic specifica-

tion of coarticulated vowels," J. Acoust. Soc. Am. **74**, 695–705.
Syrdal, A. K. (**1985**). "Aspects of a model of the auditory representation of American English vowels," Speech Commun. **4**, 121–135.
Syrdal, A. K., and Gopal, H. S. (**1986**). "A perceptual model of vowel recognition based on the auditory representation of American English vowels," J. Acoust. Soc. Am. **79**, 1086–1100.
Tiffany W. (**1953**). "Vowel recognition as a function of duration, frequency modulation and phonetic context," J. Speech Hear. Disord. **18**, 289–301.
Watrous, R. L. (**1991**). "Current status of the Peterson–Barney vowel formant data," J. Acoust. Soc. Am. **89**, 2459–2460.
Whalen, D. H. (**1989**). "Vowel and consonant judgments are not independent when cued by the same information," Percept. Psychophys. **46**, 284–292.
Zahorian, S. A., and Jagharghi, A. J. (**1993**). "Spectral shape versus formants as acoustic correlates for vowels," J. Acoust. Soc. Am. **94**, 1966–1982.