Acoustic measures for linguistic features distinguishing the semivowels /w j r l/ in American English

Carol Y. Espy-Wilson
Electrical, Computer and Systems Engineering Department, Boston University, 44 Cummington Street, Boston, Massachusetts 02215 and Research Laboratory of Electronics, Room 36-545, MIT, Cambridge, Massachusetts 02139

(Received 3 December 1990; accepted for publication 23 April 1992)

Acoustic properties related to the linguistic features which characterize the semivowels in American English were quantified and analyzed statistically. The features can be divided into those which separate the semivowels from other sounds and those which distinguish among the semivowels. The features of interest are *sonorant, syllabic, consonantal, high, back, front,* and *retroflex*. Acoustic correlates of these features were investigated in this study of the semivowels. The acoustic correlates, which are based on relative measures, were tested on a corpus of 233 polysyllabic words, each of which was spoken once by two males and two females. For the most part, the appropriate distinctions are made by the chosen acoustic properties for features. However, for each property, there was some overlap in the acoustic correlates of features for the sounds being distinguished. An examination of the sounds in the overlap regions reveals that their surface manifestation varies substantially from the canonical form. In large part, the observed variability can be explained in terms of changes due to feature spreading and lenition.

PACS numbers: 43.70.Fq, 43.72.Ar, 43.70.Hs

INTRODUCTION

An acoustic study of the sounds /w j r l/ was conducted as part of the development of a semivowel recognition system (Espy-Wilson, 1987). Recognition of the semivowels is a challenging task since, of the consonants, the semivowels are most like the vowels and, due to phonotactic constraints, they almost always occur adjacent to a vowel. Thus, acoustic changes between semivowels and vowels are often quite subtle so that there are no clear landmarks to guide the sampling of acoustic properties.

Many studies have examined some of the acoustic and perceptual properties of one or more of the semivowels in English (Lisker, 1957; O'Connor *et al.*, 1957; Lehiste, 1962; Kameny, 1974; Dalston, 1975; Bladon and Al-Bamerni, 1976; Bond, 1976). These studies have primarily focused on the acoustic and perceptual cues that distinguish among the semivowels and the coarticulatory effects between semivowels and adjacent vowels. For the most part, these studies have looked at simple contexts and a limited set of acoustic properties.

While the results of past work were used to guide the present examination of the semivowels, this study differs from previous research in that the acoustic properties investigated were chosen to be closely related to the abstract linguistic features which comprise a phonological description of the semivowels. We examine acoustic properties for features that not only distinguish among the semivowels, but that also separate the semivowels from other sounds. The acoustic properties were analyzed to quantify how the surface manifestation of the semivowels changes with context. The results obtained support previous findings, namely that formant information can be used to distinguish among the semivowels. In addition, there are new findings about the variability of some acoustic properties assumed to be associated with the semivowels. For example, not all prevocalic and intervocalic /l/'s are associated with spectral discontinuities. Finally, the acoustic properties of some of the other sounds also change with context. In particular, some underlyingly voiced obstruents surface as sonorant consonants and, in some cases, they resemble one or more of the semivowels. Some of these variations were also examined. We will argue that the variability observed in the acoustic manifestation of the semivowels and some of the other sounds can be characterized in terms of changes in the phonological features.

I. REVIEW OF THE ACOUSTIC PROPERTIES OF SEMIVOWELS

Spectrograms of the semivowels are shown in Figs. 1 and 2 where they occur in word-initial position before the front vowel /i/ and the back vowel /u/. As can be seen, the semivowels have properties that are similar to both vowels and consonants. Like the vowels, the semivowels are produced orally without complete closure of the vocal tract and without any frication noise. As is also true for the vowels, the degree of constriction needed to produce the semivowels does not inhibit voicing. Thus, as shown in these figures, the semivowels and vowels are both voiced with no evidence of frication noise. In addition, the slower rate of change of the constriction size for the semivowels than other consonants results in slower spectrum changes for these sounds compared to other consonants. For example, the spectrogram of the word "you" in Fig. 1 shows that the formants during the /j/ stay relatively constant for about 130 ms before they move toward the appropriate values for the following vowel. Thus, as in the case of vowels, a voiced steady state is often

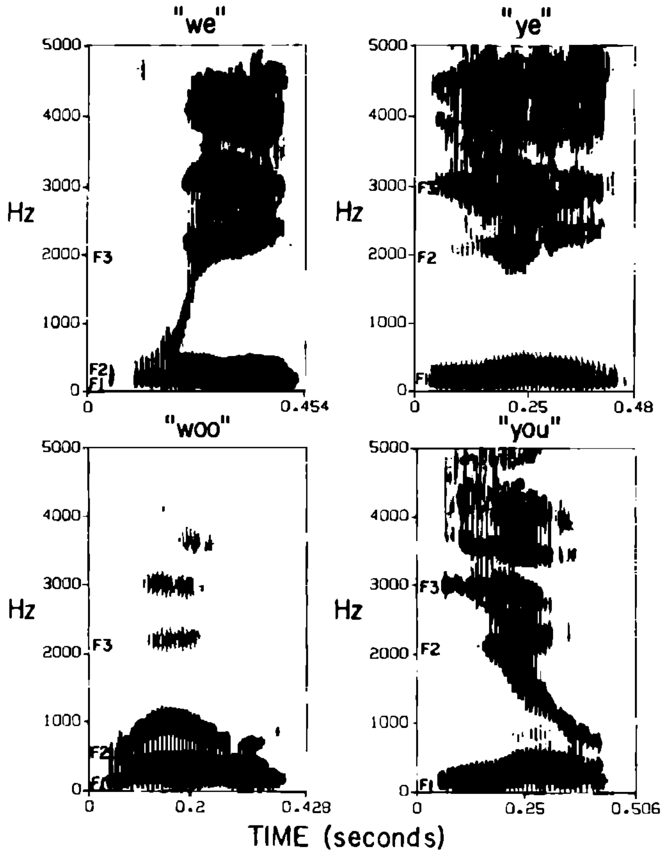

FIG. 1. Wideband spectrograms of the words "we" and "ye" (top), and "woo" and "you" (bottom).

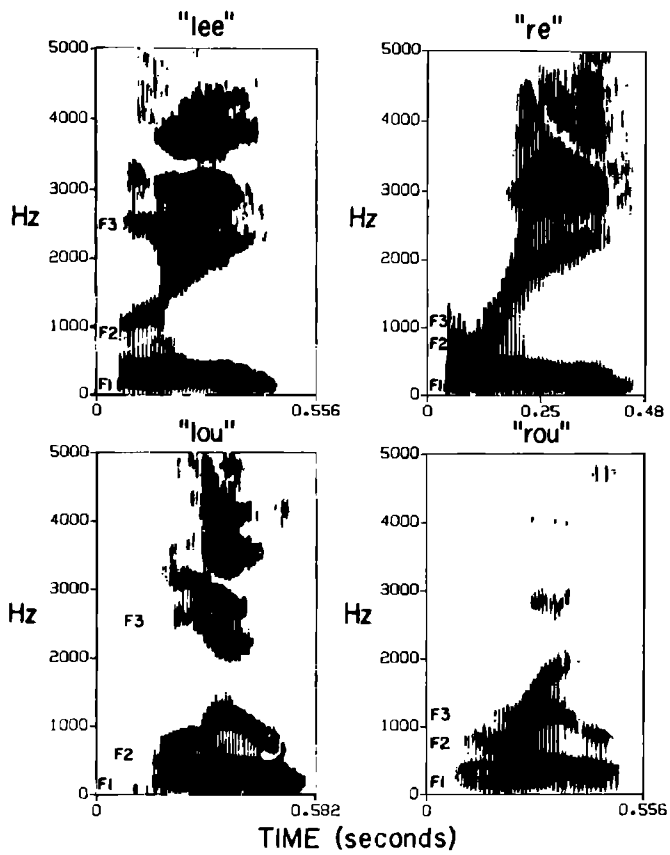

FIG. 2. Wideband spectrograms of the words "lee" and "re" (top), and "rou" and "lou" (bottom).

observed in spectrograms of the semivowels.

Like the other consonants, the semivowels usually occur at syllable margins. That is, they generally do not have or constitute a peak of sonority. (Sonority, in this case, is equated with some measure of acoustic energy.) As shown in Figs. 1 and 2, one or more of the formants during the semivowels is considerably lower in amplitude than it is during the following vowels. In the case of /w/, it is $F3$ and the higher formants which are weaker. In the case of /j/, $F3$ and the higher formants are fairly strong, but $F2$ is not. For /l/, there is less energy in the high-frequency range starting around $F4$ for "lee" and $F3$ for "lou." Finally, $F3$ and the higher formants are lower in amplitude during /r/. The relatively low amplitude of the semivowels as compared to the vowels is probably due to a combination of factors: a low-frequency first formant (Fant, 1960), a large $F1$ bandwidth caused by the narrower constriction (Bickley and Stevens, 1986), or interaction between the vocal folds and the constriction (Bickley and Stevens, 1986). At present, this phenomenon is not well understood.

The semivowels /w/ and /j/ are often referred to as glides or transitional sounds. They are produced with constant motion of the articulators. Consequently, the formants in the transition toward or away from adjacent vowels exhibit a smooth gliding movement. The semivowels /w/ and /j/ are produced with vocal-tract configurations similar to those of the vowels /u/ and /i/, respectively, but with a more extreme constriction. As a result, /w/ has lower $F1$ and $F2$ frequencies than /u/, and /j/ has a lower $F1$ frequency and usually a higher $F2$ or $F3$ frequency than /i/. These differences can be seen in the words "woo" and "ye" of Fig. 1.

The glides occur in prevocalic and intervocalic positions within a word, such as the /j/ in "you" and "yo-yo" and the /w/ in "we" and "away." In addition, they often occur phonetically (even though they are not phonologically specified) as part of the transition between two adjacent vowels. An example of this manifestation of a glide is the intervocalic /j/ sound often observed between /i/ and /ɑ/ in the pronunciation of "radiology" ([redijɑlogi] vs [rediɑlogi]).

The semivowels /l/ and /r/ are often referred to as liquids. Sproat and Fujimura (submitted) found from articulatory and electromyographic data obtained from several speakers that the production of all English /l/'s involves both an apical and a dorsal gestural component. The key articulatory distinction between the two well established variants of /l/, light or clear /l/ (as in "Lee") and dark /l/ (as in "feel"), is that in dark /l/ the tongue body is more retracted than in light /l/, resulting in a much lower $F2$. Sproat and Fujimura argue that this allophonic variation is not categorical, but is the degree to which the apical and dorsal gestures are realized and the timing between the two gestures. Specifically, they found that in addition to a significantly greater retraction of the tongue dorsum for dark /l/ compared to light /l/, the maximum tongue dorsum position for the dark /l/ is achieved well in advance of the maximum tongue tip position. On the other hand, during the gesture for the light /l/, the tongue tip position is reached before the tongue dorsum position is achieved.

In the case of light /l/, the apical gesture usually involves the placement of the center of the tongue tip against the alveolar ridge. The often rapid release of the tongue tip from the roof of the mouth results in a spectral discontinuity between the /l/ and the following vowel (Dalston, 1975). Joos (1948), reported that /l/ is always marked at its beginning and/or end by an abrupt shift in the formant pattern. Along this line, Fant (1960) observed that the identification of an /l/ relies on a sudden shift up of $F1$ from the /l/ into the following vowel. Finally, Dalston (1975) found that this abrupt shift in $F1$ is often accompanied by a transient click in the acoustic spectrum. Some of these properties can be observed at the boundary between the /l/ and the following vowels in Fig. 2.

In the case of dark /l/, Sproat and Fujimura (submitted) report that apical contact is less robust even though it was made during all of the /l/ productions in their study. Giles and Moll (1975) found in an x-ray study of English /l/ that apical contact for dark /l/'s was not always achieved for all speakers and is dependent upon phonetic context and speaking rate. In addition, they found that the mean peak velocity of the tongue apex movement is significantly slower for dark /l/. Furthermore, they found that dark /l/ shows undershoot of articulatory positions with increases in speaking rate. This slower and incomplete apical gesture may help explain why dark /l/ productions are not associated with an abrupt spectral change.

Although the distribution of dark and light /l/ varies across speakers, canonical syllable-final /l/ is dark and, in many dialects, syllable-initial /l/ is light. Sproat and Fujimura (submitted) found in their study of preboundary[1] intervocalic /l/ in the falling stress context /i_ɪ/ that the quality of the /l/ depends upon the phonetic duration of the rime which contains it. (They also show a correlation between the duration of the preboundary rime and the strength of the phonological boundary.) Specifically they found that as the rime becomes longer, the tongue body for /l/ becomes lower and more retracted. Therefore, intervocalic /l/ occurring before a major intonation boundary is dark. On the other hand, they found that the preboundary /l/ preceding the weakest boundaries are as light as initial /l/. These data support previous findings by Lehiste (1962) and Blandon and Al-Bamerni (1976) which show that certain preboundary intervocalic /l/ productions are lighter in quality than preboundary /l/ in prepausal position.

American /r/ may be produced with either a retroflexed or bunched articulation (Delattre and Freeman, 1968). If the upper constriction is at the palate, it is made with either the tongue tip or the tongue blade. If, instead, the upper constriction is further back near the velum, it is made with the tongue body. It is the palatal or palato–velar constriction which lowers $F3$ (Delattre and Freeman, 1968; Stevens, in preparation), whereas the pharyngeal constriction lowers $F2$ and raises $F1$ (Delattre and Freeman, 1968). In terms of perception, Delattre and Freeman found with the use of an electric mouth analog that the palatal constriction is primary in terms of producing the /r/. However, they found that as the pharyngeal constriction is narrowed, the auditory impression of the /r/ is enhanced.

Regardless of whether a bunched or retroflexed /r/ is produced, lip rounding may occur when /r/ is either prevocalic or intervocalic and before a stressed vowel (Delattre and Freeman, 1968). The acoustic consequence of lip rounding is a lowering of all formants. This effect may account for the lower $F1$, $F2$, and $F3$ Lehiste (1962) observed for initial /r/ allophones relative to final /r/ allophones, for which lip rounding does not usually occur.

In summary, many of the acoustic properties which characterize the semivowels have been examined in previous studies. However, this work has largely focused on formant measurements. In this investigation of the semivowels, acoustic properties calculated from formant measurements and energy-based parameters are quantified and analyzed so that we can study to a greater extent some of the variability that occurs in the surface forms of the semivowels. Furthermore, the acoustic properties are related to the linguistic features which provide a framework for understanding the changes that occur.

II. METHOD

A. Stimuli

A data base of 233 polysyllabic words containing semivowels in a variety of phonetic environments was selected from the 20 000-word Merriam-Webster Pocket dictionary. The semivowels occur adjacent to voiced and unvoiced consonants, as well as in word-initial, word-final, and intervocalic positions. (Note that only /l/ and /r/ occur postvocalically.) The semivowels occur adjacent to vowels which are stressed and unstressed, high and low, and front and back. In developing the database, words were chosen that contained several semivowels so that they satisfy more than one category. The distribution of the semivowels in terms of word position and stress is given in Table I. Examples of words in these categories are given in Table II. While most of the contexts contain several words, there are a few for which

TABLE I. Distribution of semivowels in the test words. The number in the parentheses specifies the number of semivowels occurring next to a vowel with primary stress.

Category	w	j	r	l
Prevocalic	65	41	63	52
word-initial (prestressed)	11(9)	8(5)	10(5)	6(3)
stop cluster (prestressed)	18(8)	11(4)	18(10)	10(4)
fricative cluster (prestressed)	19(13)	7(3)	18(9)	18(6)
stop fricative cluster (prestressed)	10(7)	8(3)	10(7)	10(5)
adjacent to sonorant consonant	7	7	7	9
Intervocalic	9	6	29	32
poststressed	2	1	13	16
prestressed	5	4	10	8
unstressed	2	1	6	8
Postvocalic			25	26
word-final (poststressed)			13(10)	19(14)
obstruent cluster			6	4
adjacent to sonorant consonant			6	3

TABLE II. Examples of test words.

Category	w	j	r	l
Prevocalic				
word-initial: prestressed	walnut	yell	requiem	leapfrog
other	walloon	eurologist	rhinoceros	linguistics
stop cluster: prestressed	aquarius	pule	brilliant	bless
other	quadruplet	bucolic	fibroid	chlorination
fricative cluster: prestressed	swollen	view	frivolous	flourish
other	Swahili	behavior	anthrax	grizzly
stop fricative cluster: prestressed	disqualify	spurious	astrology	exclaim
other	misquotation	promiscuously	widespread	exploitation
adjacent to sonorant consonant	carwash	brilliant	walrus	harlequin
Intervocalic				
poststressed	forward	Ghanaian	caloric	astrology
prestressed	unaware	reunion	fluorescence	unilateral
unstressed	unctuous	diuretic	correlation	fraudulent
Postvocalic				
work-final: poststressed			clear	dwell
other			memoir	whippoorwill
obstruent cluster			cartwheel	oneself
adjacent to sonorant consonant			forewarn	walnut

only a small number of words were available in the dictionary. For example, only the word "Ghanaian" had a poststressed intervocalic /j/.

B. Speakers and recordings

For recording, the words were embedded in the carrier phrase "________pa." The final "pa" was added in order to avoid glottalization and other types of utterance-final variability. Each word was spoken once by two males and two females. Given that there are several words in most categories (see Table I), one repetition of each word by each speaker provides at least 24 to 260 productions of each semivowel in each major category, e.g., prevocalic /w/. In fact, as explained in Sec. II C, some categories may contain more instances of a semivowel than what is indicated in Table I since speakers often insert semivowels between adjacent vowels.

The speakers were students and employees at the Massachusetts Institute of Technology. The female speakers were from the northeast and the male speakers were from the midwest. All were native speakers of English and reported having normal hearing. The speakers were recorded in a quiet room with a pressure-gradient close-talking noise canceling microphone (part of Sennheiser HMD 224X microphone/headphone combination). They were instructed to say the utterances at a natural pace.

C. Initial processing

The utterances were digitized using a 6.4-kHz low-pass filter and a 16-kHz sampling rate. The speech signals were also pre-emphasized to compensate for the relatively weak spectral energy at high frequencies (a particular issue for sonorants). Finally, the test words were excised and hand transcribed. This process resulted in 2378 vowels, 1689 semivowels, 479 nasals, and 1894 obstruents (stops, fricatives, and affricates). Specific characteristics of measures and measurement procedures will be indicated in Sec. III.

Segmentation and labeling of the waveforms was performed by the author with the help of playback and displays of several attributes including LPC and wide-band spectra, the speech signal and various bandlimited energy waveforms (Cyphers, 1985; Shipman, 1982; Zue *et al.*, 1986). The Merriam-Webster Pocket dictionary provided a baseline phonemic transcription of the words. However, modifications of some of the labels were made based on the speakers' pronunciations. In addition, when transcribing the database, we did not normally consider the /w/ and /j/ offglides of diphthongs as being separate from the vowel. In some instances, however, the offglide of a diphthong which was followed by another vowel was articulated with a narrow enough constriction that a semivowel label was inserted. On the other hand, some underlying postvocalic liquids, particularly /l/, in words like "almost" were not always clearly heard. In these instances, the liquid was often omitted from the transcription.

D. Feature analysis

Distinctive feature theory was used to provide a guide to understanding what acoustic properties we should look for to characterize the semivowels. In addition, distinctive feature theory provides a basis for understanding how the acoustic properties of the semivowels may change as a function of context. A feature specification of the semivowels is given in Tables III and IV. Table III contains features that separate the semivowels as a class from other sounds and Table IV contains features that distinguish among the semivowels.

The features listed are modifications of ones proposed by Jakobson *et al.* (1952) and later by Chomsky and Halle (1968). For example, in Table IV, we list both the features *back* and *front*. For practical reasons, we chose to use both features and classify /r/ and prevocalic /l/ as *-back* and

TABLE III. Features which characterize various classes of consonants. A " + " indicates that the designated feature is present in the representation of the sound, and a " − " indicates the absence of the feature.

	Sonorant	Syllabic	Nasal
Fricatives, stops, affricates	−	−	−
Semivowels	+	−	−
Nasals	+	−	+
Vowels	+	+	−

TABLE IV. Features which discriminate among the semivowels. A " + " indicates that the designated feature is present in the represention of the sound, and a " − " indicates the absence of the feature.

	Consonantal	High	Back	Front	Retroflex
/w/	−	+	+	−	−
/j/	−	+	−	+	−
/r/	−	−	−	−	+
prevocalic /l/	+	−	−	−	−
postvocalic /l/	−	−	+	−	−

-front.[2] Their $F2$ values clearly lie between those of the back and rounded semivowel /w/ and the front semivowel /j/.

We also found it necessary to distinguish between initial and final /l/ allophones on the basis of the features *consonantal* and *back.* As stated earlier, several researchers have observed a sharp spectral discontinuity between a prevocalic /l/ and a following vowel due to the rapid release of the tongue tip from the alveolar ridge, as we would expect with a change in the feature *consonantal.* On the other hand, in the production of postvocalic /l/, alveolar contact is often not realized or is realized only gradually, so that the spectral change between it and a preceding vowel is usually gradual. In addition, a final /l/ is more velarized than an initial /l/. Thus, $F2$ is much lower and close in value to that of the back and rounded /w/.

Finally, the feature *retroflex* is used to distinguish /r/ from all other sounds. Although the term "retroflex" is used, this feature relates to the acoustic consequence of either a bunched or retroflexed tongue shape.

Table V shows the acoustic properties for the features in Table III and IV (with the exception of the feature *nasal*) and the parameters used for their extraction. To make them insensitive to variations in speaker, speaking rate, and speaking level, all of the properties are based on relative measures instead of absolute thresholds. That is, a measure either examines an attribute in one speech frame[3] in relation to another frame, or, within a given frame, examines one part of the spectrum in relation to another nearby part of the spectrum.

In the following sections, we will examine how well the properties listed in Table V distinguish among the semivowels and separate them from other sounds. The results also indicate how the characteristics of the semivowels and other sounds change as a function of context.

III. RESULTS AND DISCUSSION

A. Sonorant measure

Like the vowels, the semivowels are sonorant sounds. That is, the main source of excitation is at the glottis, so that all of the natural frequencies of the vocal tract are excited. Thus, unlike the obstruents, where the main source of excitation is further forward in the vocal tract, there is significant energy at low frequencies. The only other consonants that share these properties are the nasals.

The parameter used to extract the acoustic correlate of the *sonorant* feature is the bandlimited energy computed from 100 to 400 Hz. More specifically, the value of the parameter in each frame is the difference (in dB) between the maximum energy within the word and the energy in each frame. An example of this parameter is shown in the lower part of Fig. 3 for the word "chlorination." The energy difference is small in the sonorant regions (vowels, semivowels, and nasals), and is large in the obstruent regions (stops, fricatives, and affricates).

Figure 4 shows how all of the sounds differ in sonority, as determined with this measure. For each sound, the minimum energy difference occurring within the hand-transcribed region is used. There is considerable overlap between the distributions of the vowels, semivowels, and nasals. If we set a threshold of −20 dB to divide sonorant and nonsonor-

TABLE V. Mapping of features into acoustic properties. $B0$, $B1$, $B2$, $B3$, and $B4$ are the bark transformations of $F0$, $F1$, $F2$, $F3$, and $F4$, respectively.

Feature	Acoustic correlate	Parameter	Property
Sonorant	No significant decrease in energy at low frequencies	Energy 100–400 Hz	high[a]
Nonsyllabic	Dip in midfrequency energy	Energy 640–2800 Hz	low[a]
		Energy 2000–3000 Hz	low[a]
Consonantal	Abrupt amplitude change	First difference of adjacent spectra	high
High	Low $F1$ frequency	$B1$–$B0$	low
Back	Low $F2$ frequency	$B2$–$B1$	low
Front	High $F2$ frequency	$B3$–$B2$	low
		$B4$–$B3$	low
Retroflex	Low $F3$ frequency &	$B4$–$B3$	high
	Close $F2$ and $F3$	$B3$–$B2$	low

[a] Relative to a maximum value within the utterance.

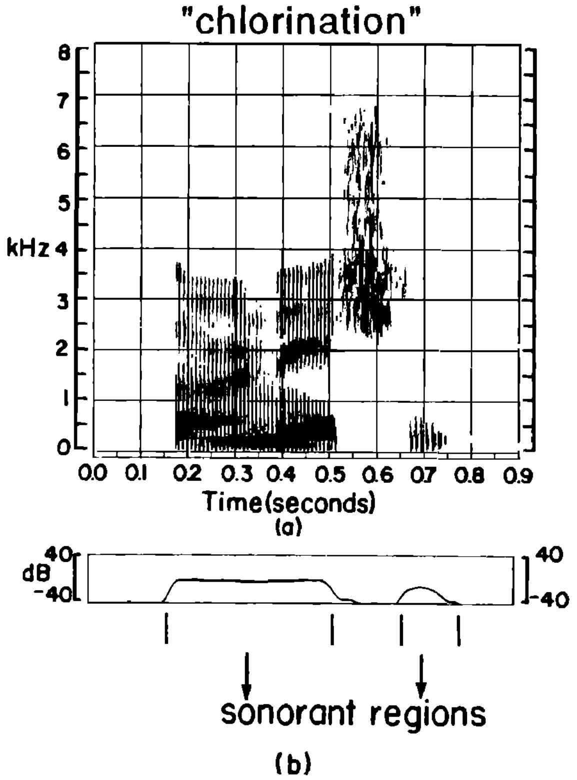

FIG. 3. An illustration of the parameter used to capture the feature *sonorant.* (a) Wideband spectrogram of the word "chlorination." (b) The difference between the maximum value of the low-frequency energy (computed from 100 to 400 Hz) in the word and the value in each frame.

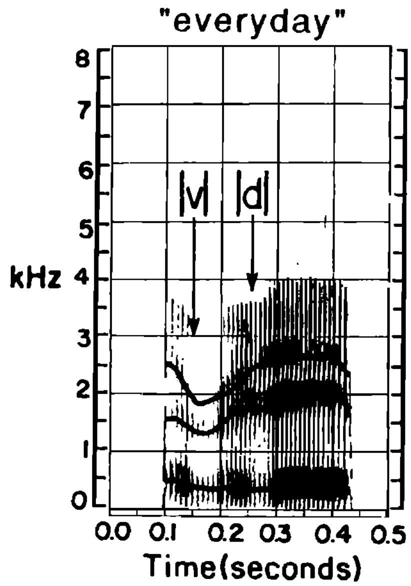

FIG. 5. A spectrogram of the word "everyday" which contains the two obstruents /v/ and /d/ that have undergone lenition.

ant sounds, then only about 12.5% of the typically nonsonorant consonants overlap with the sonorant sounds. Of these consonants, 72% are produced with a weakened constriction (this process referred to as *lenition* is discussed in Catford, 1977) so that they are realized as sonorants. Two examples are shown in Fig. 5, which contains a spectrogram of the word "everyday." Both the /v/ and /d/ surface as sonorant consonants.

The remaining segments which overlap with the sonorants are the closed portions of voiced stops. The low-frequency energy they exhibit is presumably caused by vibrations of the vocal cords which are transmitted through the tissues around the neck.

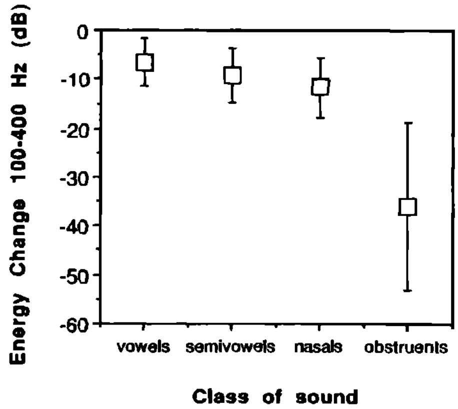

FIG. 4. Averages and standard deviations of the change (in dB) in the low-frequency energy computed from 100 to 400 Hz within sonorant and nonsonorant sounds with respect to the maximum energy within the word.

B. Consonantal measure

Consonantal sounds are produced with a narrow constriction at some point along the midline of the vocal tract. Due to the narrow constriction, the release of the consonantal sound into the following vowel involves rapid movement of some of the formants. The result of this formant movement is an abrupt change in the spectrum over at least some part of the frequency range (Stevens and Keyser, 1989).

The parameter used to capture the rate of spectral change between consonants and vowels is based on the outputs of a bank of 40 linear critical band filters to which some nonlinearities (designed to model the hair-cell/synapse transduction process in the inner ear) are applied to enhance onsets and offsets (Seneff, 1986). An example is shown in part (b) of Fig. 6 for the word "correlation." The waveforms that are spaced about a half bark apart show sharp onsets and offsets between /l/ and the surrounding vowels in the frequency region between 800 and 1200 Hz and between 1800 and 2400 Hz.

Based on the first differences in time of waveforms like the ones shown in part (b) of Fig. 6, we computed global onset and offset waveforms for each consonant. The onset waveform is computed by summing, in each frame, all the negative first differences in time. Similarly, the offset waveform is obtained by summing, in each frame, all the positive first differences in time of the channel outputs. The resulting onset and offset waveforms for the word "correlation" are shown in parts (c) and (d) of Fig. 6 where the sharp amplitude changes between the /l/ and the surrounding vowels show up as a valley and a peak, respectively. Note that since a 25-ms time window is used, there is a limit to the maximum

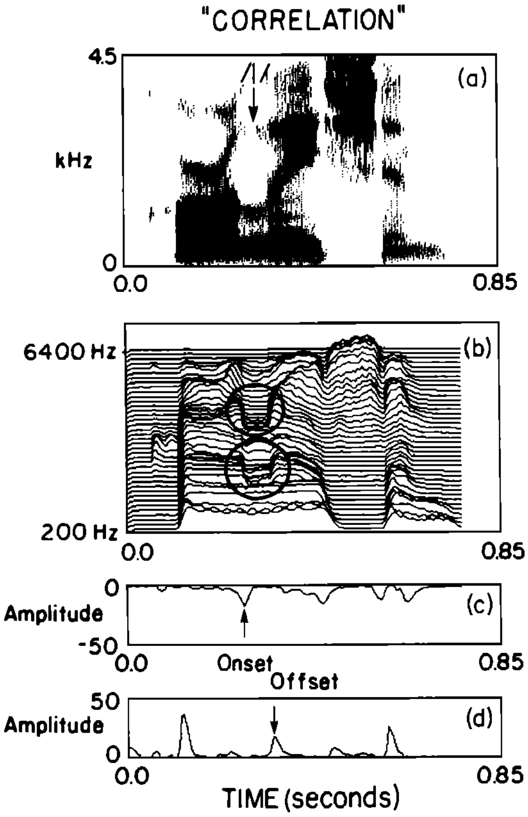

FIG. 6. An illustration of parameters which capture abrupt amplitude changes. (a) Wideband spectrogram of "correlation." (b) Channel outputs of an auditory model which show abrupt spectral changes in two frequency regions between the /l/ and adjacent vowels. (c) Onset waveform (computed from the sum of the negative first differences of the channel outputs) which shows a sharp valley at the onset of the /l/. (d) Offset waveform (computed from the sum of the positive first differences of the channel outputs) which shows a sharp peak at the offset of the /l/.

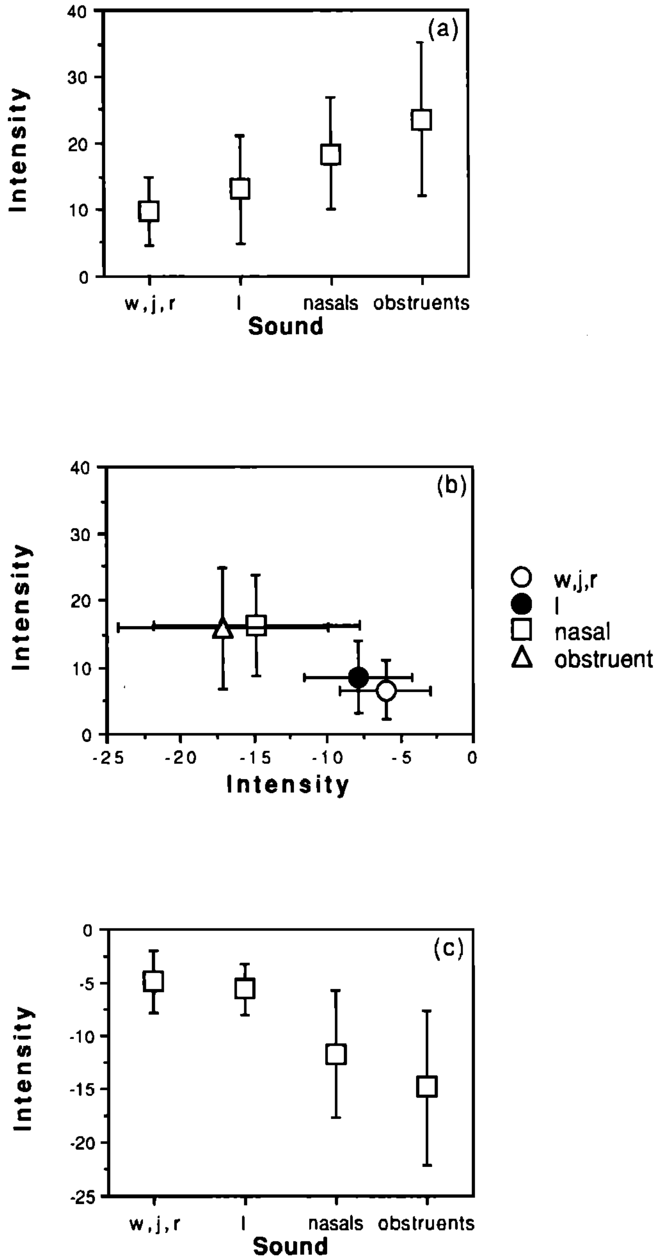

FIG. 7. Averages and standard deviations of (a) the offsets between prevocalic consonants and following vowels, (b) the onsets and offsets between intervocalic consonants and adjacent vowels, and (c) the onset between postvocalic consonants and preceding vowels.

rate of change that can be captured by this measure.

The onset and offset waveforms were examined during the time interval between each consonant and its neighboring vowel(s). We defined the onset of the consonant to be the maximum absolute value of the onset waveform occurring between the preceding vowel and the consonant. Likewise, we defined the offset of the consonant to be the maximum value of the offset waveform occurring between the consonant and the following vowel. The time at which these values occur are indicated by arrows in parts (c) and (d) of Fig. 6.

Figure 7 shows the data on the onsets and offsets across all words and all speakers. The units of the onset and offset values are like dB since the channel outputs after nonlinearities have been applied are approximately linear with amplitude at low signal levels and logarithmic at higher signal levels (Seneff, 1986, p. 88). The data for /l/ are separated from /w j r/ since, of the semivowels, /l/ is most associated with spectral discontinuities. Several observations can be made from the data. First, in general, the spectral changes between obstruent consonants and adjacent vowels are more rapid than the spectral changes between semivowels and adjacent vowels. Second, the spectral change between /l/ and adjacent vowels tends to be more abrupt than the spectral change between the other semivowels and adjacent vowels. However, as can be seen from the standard deviations, there is often a wide spread in the distribution of onset and offset values.

We also observed a strong relationship between the stress pattern of the words and the rate of spectral change between the consonants and adjacent vowels. That is, onsets and offsets associated with consonants that precede stressed vowels are significantly stronger than those associated with consonants that precede unstressed vowels, presumably because the constriction is tighter and the release is more rapid. For example, compare the rate of spectral change between the prevocalic /l/ and adjacent vowels in the words "blurt" and "linguistics," and between the intervocalic /l/ and surrounding vowels in "walloon" and "swollen" shown in Fig. 8. The offset associated with the /l/ in "blurt" (at about 130

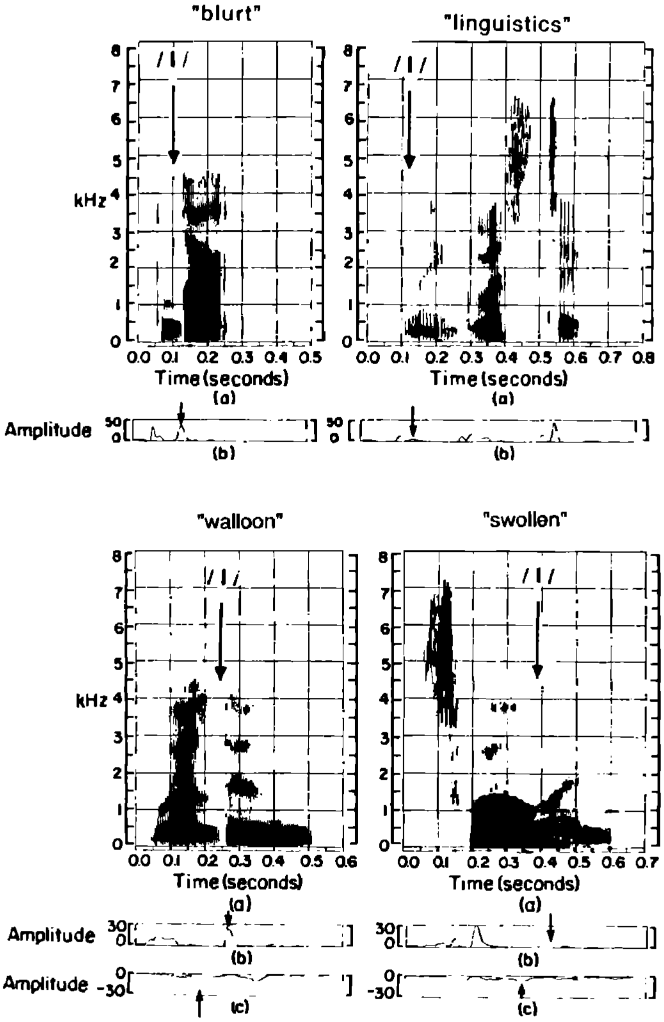

FIG. 8. An illustration of the rate of spectral change associated with the /l/'s in "blurt," "linguistics," "walloon," and "swollen." (a) Wideband spectrograms. (b) Offset waveform. (c) Onset waveform.

ms) is much more abrupt than the one associated with the /l/ in "linguistics" (at about 145 ms). Similarly, the onset and offset associated with the intervocalic /l/ in "walloon" (at 190 and 260 ms, respectively) are much more abrupt than those associated with the intervocalic /l/ in "swollen" (at 350 and 410 ms, respectively).

C. Syllabic measure

Because they are more constricted and hence have a relatively low $F1$, the semivowels usually have considerably less energy in the low- to midfrequency range than the vowels. Like other consonants, the semivowels usually occur as nonsyllabic sounds adjacent to syllable nuclei at a syllable boundary. That is, they generally do not have or constitute a peak of sonority, where we are equating sonority in this case with a mid-frequency acoustic energy measure. An acoustic manifestation of a syllable boundary appears to be a significant dip within some bandlimited energy contour.

To access the difference in energy between semivowels and vowels, and, more generally, between consonants and vowels, we used to bandlimited energies in the frequency ranges 640–2800 Hz and 2000–3000 Hz. We chose the frequency range 640–2800 Hz because, relative to the vowels, the lower $F1$ for the semivowels is expected to cause a decrease in the amplitudes of the formants in this region. However, we found that several intervocalic /r/'s have energy levels in this range which do not differ from those found on surrounding vowels, presumably because of the proximity of $F2$ and $F3$. To avoid this problem, we also examined the bandlimited energy from 2000 to 3000 Hz. Since $F3$ is normally between 2000 and 3000 Hz for vowels, but falls near or below 2000 Hz for /r/, /r/ will usually be considerably weaker in the 2000- to 3000-Hz range than an adjacent vowel(s).

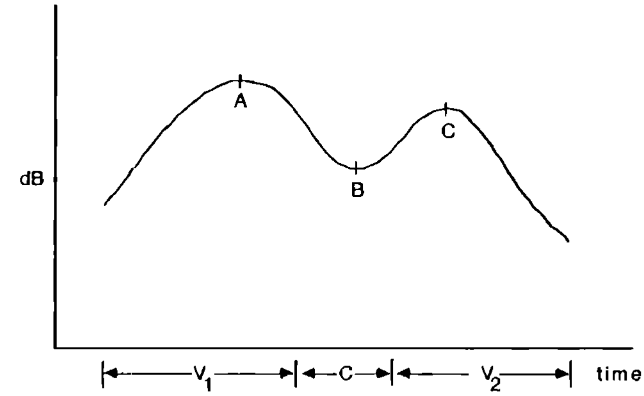

FIG. 9. A schematic of an energy waveform for a vowel–consonant–vowel (VCV) sequence. The extrema within the waveform are used to compute the energy difference between consonants and vowels. In the case of prevocalic consonants (V_1 does not exist), points C and B are used. In the case of postvocalic consonants (V_2 does not exist), points A and B are used. Finally, in the case of intervocalic consonants, the smaller of the difference between points C and B and between points A and B is taken as a measure of the energy dip.

Measurements of the midfrequency energy of semivowels are based on energy contours like the one in Fig. 9. All measures are relative to energy in an adjacent vowel. The depth of the energy dip is considered to be the difference (in dB) between the minimum energy within the consonant, point B, and the maximum energy within the adjacent vowel(s), point A and/or point C.

In the case of syllables with prevocalic consonants, the difference in energy between the prevocalic consonant (point B) and the following vowel (point C) was computed. For syllables with postvocalic consonants, the difference in energy between the postvocalic consonant (point B) and the preceding vowel (point A) was computed. Finally, for intervocalic consonants, both the differences in energy at points C and B and points A and B were computed. The depth of the energy dip was taken to be the smaller of the two differences.

As a basis for comparison with the semivowels, the depths of several types of intravowel energy dips were computed as well. An illustration of this procedure is shown in Fig. 10, which shows a schematic representation of an energy contour of a vowel. First, an estimate of the natural rise in energy within word-initial vowels was computed by calculating the energy difference at points W and T. This vowel energy onset is compared with the energy difference between

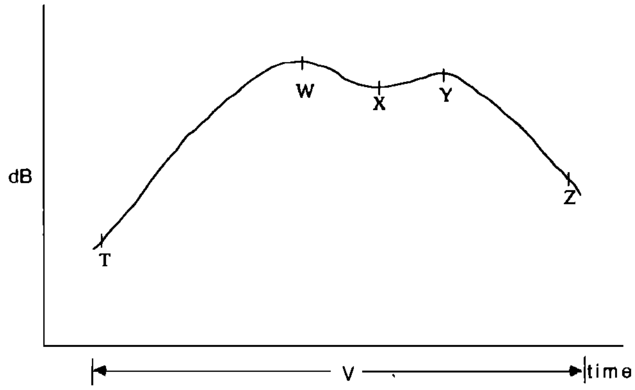

FIG. 10. A schematic of an energy waveform for a vowel. The energy difference between points W and T is used to determine the energy rise within word-initial vowels. The energy difference between points W and Z is used to determine the energy taper within word-final vowels. The smaller of the energy differences between points W, X and between points X and Y is used in all vowels with the appropriate energy waveform to determine within vowel energy dips.

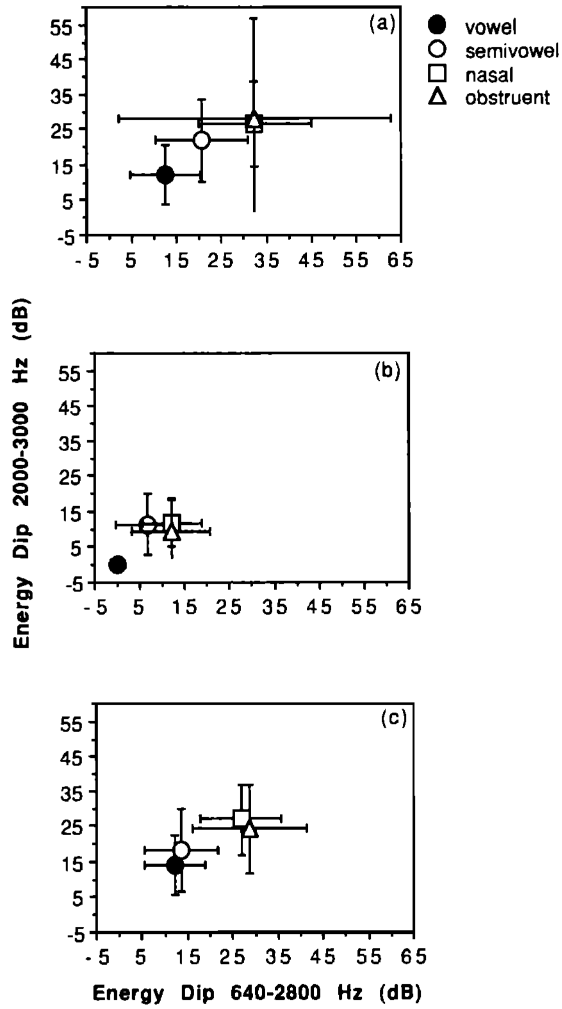

FIG. 11. Averages and standard deviations of the midfrequency energy changes (in dB) between consonants and adjacent vowels and within vowels. Data are shown separately for the (a) prevocalic, (b) intervocalic, and (c) postvocalic consonants.

prevocalic consonants and following vowels. Second, an estimate of the natural energy taper within word-final vowels was computed by calculating the energy difference at points W and Z. This vowel energy offset is compared with the energy difference between postvocalic consonants and preceding vowels. Finally, in cases where there was an intravocalic dip, X, we computed the difference between the energy at points W and X and between points Y and X. In this case, the smaller of the two differences was recorded. Of course, not all vowels will have this type of energy waveform shape so that there will not always be a point X and a point Y. In these cases, the intravowel energy dip is simply 0 dB. This energy measure is compared with the energy difference between intervocalic consonants and surrounding vowels.

The results of these measurement procedures are plotted separately in Fig. 11 for prevocalic, intervocalic, and postvocalic consonants. In each plot, the consonants are divided into obstruents, nasals, and semivowels. Also included in the figure are the data for the energy changes within vowels. The data show that the difference in midfrequency energy between the consonants and vowels is, on average, much greater than the energy change within vowels. Of the energy changes between consonants and adjacent vowels, the energy change associated with the semivowels is almost always smallest. In addition, as the standard deviations show, there is sometimes considerable overlap between the distributions of the energy changes within vowels and the energy changes between semivowels and adjacent vowels. On closer examination of these data, patterns in their distribution emerge across consonantal contexts.

1. Prevocalic consonants

In general, the difference in midfrequency energy between the prevocalic semivowels and following vowels is greater than the midfrequency energy change within the beginning portions of word-initial vowels. However, word position has a strong effect on the phonetic realization of the semivowels. There is a more significant energy change between semivowels and following vowels if the semivowel is not in a cluster with another consonant, but is word-initial. Furthermore, if the semivowel is in a cluster with another consonant, there is a greater energy change between it and the following vowel if the preceding consonant is voiced, ensuring that the semivowel is also completely voiced. In addition to the contextual influence of preceding consonants, the degree of stress of the following vowel also matters. There is a more pronounced energy change between the semivowel and vowel if the vowel is stressed.

2. Intervocalic consonants

In intervocalic positions, most of the semivowels showed substantial differences in energy compared to neighboring vowels. The energy dip computed for these VCV segments was greater than 2 dB for 90% of the semivowels. Of the other semivowels which did not show a substantial difference in energy relative to adjacent vowels, 33% were /j/'s, 14% were /r/'s and 5% were /l/'s. Most of these semivowels follow a stressed vowel and precede an unstressed vowel, such as the /l/ in "astrology" and the /r/ in "guarantee." This lack of an energy dip for semivowels in this environment may be a case of phonetic lenition.

While the majority of vowels do not normally have such energy dips, there were several instances of vowels with energy dips comparable to those between intervocalic consonants and adjacent vowels. An examination of such vowels showed that, in general, those with such significant energy dips were either and /ɚ/, such as the one in "plurality" where an intervocalic /r/ was not included in the transcription, or a diphthong, such as the /iʲ/ in "queer" and the /ɑʷ/ in "flour."

3. Postvocalic consonants

The patterns in energy change within word-final vowels and between vowels and following semivowels (including only the liquids /r/ and /l/) are very similar. Two factors contribute to the overlap. First, the /j/ or /w/ offglides of word-final diphthongs often result in energy changes that are comparable to the changes observed between a word-final liquid and the preceding vowel. Such a large energy taper can be seen in Fig. 12 for the word "view" which has a substantial energy change in the frequency range 640–2800 Hz. Second, postvocalic liquids that are followed by another consonant are often as strong as the preceding vowels, as can be seen by comparing the amplitudes of the formants in the /ɑr/ region (0.1 to 0.2 s) in the word "cartwheel" shown in Fig. 13. In many such cases, there is significant assimilation between the postvocalic consonant and the preceding vowel. The fairly constant formant amplitudes and the steady $F3$ frequency during the /ɑr/ region of "cartwheel" suggest

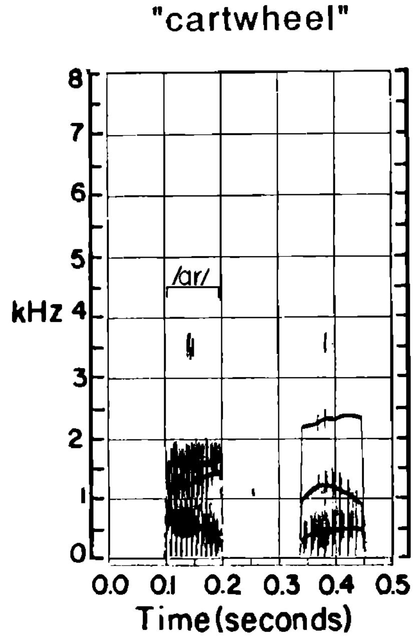

FIG. 13. A spectrogram with automatically extracted formant tracks overlaid on the word "cartwheel."

that the /ɑ/ and /r/ are coarticulated so that they are realized acoustically as one segment.

This energy continuity between vowels and following liquids also occurs when the postvocalic liquids are followed by another sonorant consonant which is not in the same syllable, such as the /l/ in "bellwether." In words like this, there is a nonsyllabic region between the vowel preceding the postvocalic liquid and the vowel after the second sonorant consonant (in this case, the /ɛ/ before the /l/ and the /ɛ/ after the /w/). However, there is little energy change between the postvocalic liquid and the preceding vowel. Instead, the energy offset (referred to as consonant onset in Sec. III B) between the nonsyllabic region and the preceding vowel occurs after the liquid and before the following sonorant consonant. On the other hand, when nasals occupy this postvocalic position, there is substantial energy change between them and the preceding vowel so that the energy offset occurs before the postvocalic nasal consonant.

This difference in where the energy offset occurs is illustrated in Fig. 14 which contains information relating to the intersonorant sequences /rm/ and /nr/ in the words "harmonize" and "unreality," respectively. In the case of "harmonize" (shown on the left), the nonsyllabic dip occurs during the /m/ and, as indicted by the arrows, the energy offset between the /rm/ cluster and the previous vowel occurs after the /r/, at the beginning of the /m/. In contrast, the energy offset between the /nr/ cluster and the preceding vowel in "unreality" (shown on the right) occurs at the point of implosion for the /n/ at about 175 ms as indicated by the arrow.

To capture this difference in the temporal properties of the energy offset for nasal-sonorant consonant sequences and liquid-sonorant consonants sequences, we computed the duration of the intersonorant nonsyllabic region. The duration of this energy dip region was taken to be the difference in

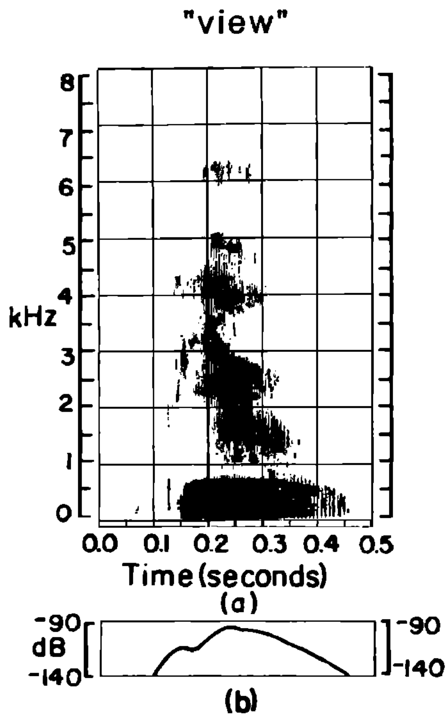

FIG. 12. Illustration of a large energy taper in word-final diphthongs. The energy towards the end of the vowel is 30 dB or more less than the maximum value within the vowel. (a) Wideband spectrogram of the word "view." (b) Energy 640 to 2800 Hz.

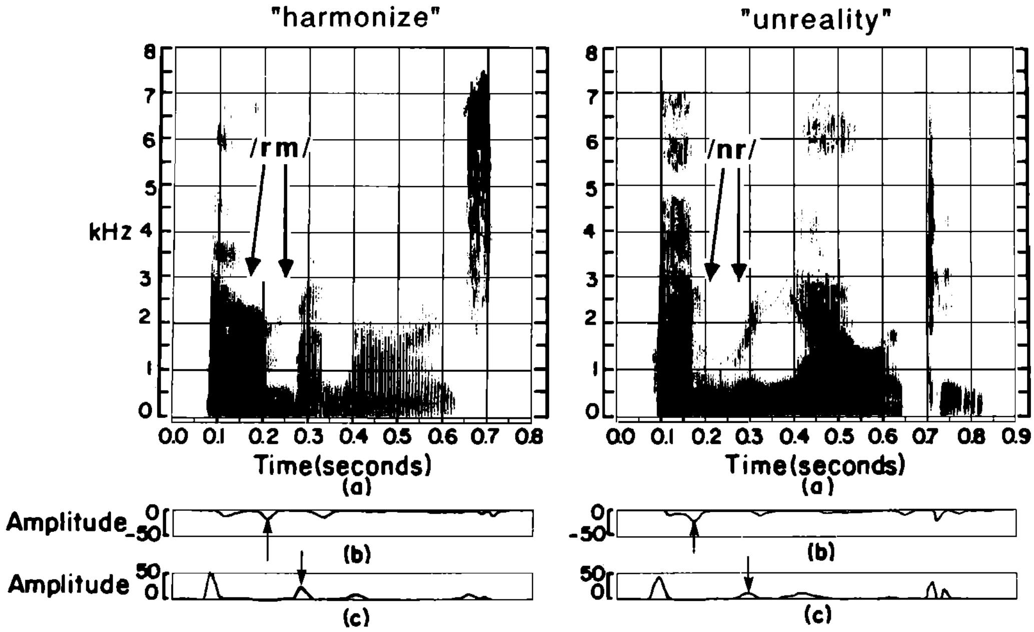

FIG. 14. (a) Wideband spectrograms of the words "harmonize" and "unreality." (b) Onset waveforms that show a valley at the energy offset indicating the beginning of the nonsyllabic region. (c) Offset waveforms which show a peak at the energy onset indicting the end of the nonsyllabic region.

time between the energy offset and the energy onset immediately surrounding the energy dip. In the case of "harmonize," the energy onset occurs after the/r/ and before the following vowel at about 295 ms. Thus the energy dip region includes only the/m/ and is 75 ms in duration. In the case of "unreality," the energy onset also occurs after the second sonorant consonant and before the following vowel. However, in this case, the energy dip region includes both consonants and is 120 ms in duration.

Data across all words containing intersonorant clusters are shown in Fig. 15. For comparison, we also included the duration of the energy dip regions when there is only one sonorant consonant occurring between two vowels, an intervocalic nasal or semivowel. In this case, the energy offset and energy onset will correspond to the consonant onset and offset, respectively. Although there is no normalization for variability in speaking rates, the results in Fig. 15 show a distinct pattern. The distributions of the duration of energy dip regions associated with only one sonorant consonant and those associated with two sonorant consonants where the first consonant is a liquid are essentially the same. However, the average duration of the energy dip regions associated with two sonorant consonants where the first is a nasal is considerably longer than those of the other cases. We can infer from this pattern that the energy offset in the cluster where the first member is a postvocalic liquid occurs after the postvocalic liquid so that only one of the sonorant consonants is contained in the energy dip region. On the other hand, the energy offset in the cluster where the first member is a postvocalic nasal occurs before the postvocalic nasal so that both sonorant consonants are part of the energy dip region.

These results show that postvocalic liquids which are followed by another sonorant consonant are not a part of the energy dip region. Instead, they appear to be a part of the syllable nucleus. Thus it may be more appropriate to think of the liquid and preceding vowel as a diphthong where the liquid, like the glides in this context, is considered to be a part of the vowel.

The assertion that postvocalic /l/ acts as the second element of a diphthong has also been made by Giles and Moll (1975). Based on x-ray data of prevocalic and postvocalic /l/, they found that postvocalic /l/ shows relatively slow movement characteristics and undershoot of articulatory position. On the other hand, prevocalic /l/ had a relatively high rate of articulatory movement and no undershoot char-

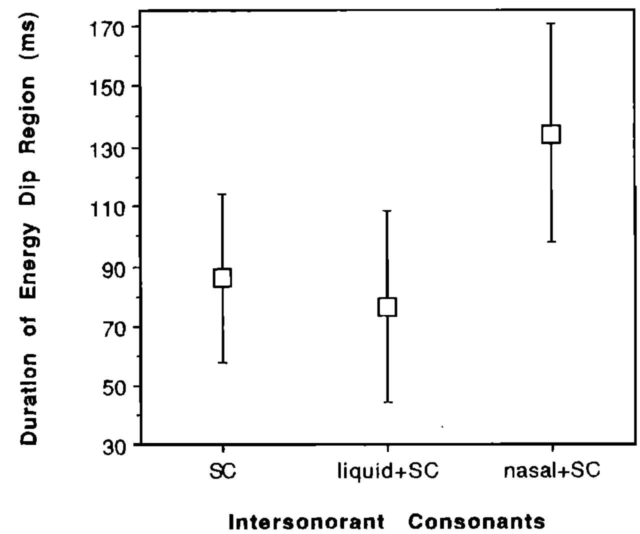

FIG. 15. A comparison of the average durations of the nonsyllabic regions of words containing one intervocalic sonorant consonant (SC), words containing an intervocalic liquid-sonorant consonant sequence (liquid + SC) and words containing an intervocalic nasal-sonorant consonant sequence (nasal + SC).

acteristics. Thus they conclude the prevocalic /l/ functions as a consonant while postvocalic /l/ is vocalic in nature. Along this line, Sproat and Fujimura (submitted) postulate that a gesture involving a nonperipheral articulator (such as tongue dorsum retraction) is attracted to the syllable nucleus whereas a gesture involving a peripheral articulator (such as the tongue tip) is attracted to syllable margins. With this assumption, they too conclude that postvocalic /l/, which has a more significant tongue dorsal retraction than prevocalic /l/, should be considered more vocalic.

D. Formant frequency measures

Important information for distinguishing among the semivowels are the frequencies of the first three formants ($F1$, $F2$, and $F3$). Given minimal-pair words, it has been shown (Lisker, 1957; O'Connor *et al.*, 1957) that $F1$ separates the glides /w/ and /j/ from the liquids /l/ and /r/, $F2$ separates /w/ from /l r/ from /j/, and $F3$ separates the liquids /l/ and /r/. The data in this study concur with these observations.

A formant tracker (Espy-Wilson, 1987) was used to automatically extract the first four formants during the sonorant regions of the words in the database. The frequencies of $F1$, $F2$, $F3$, and $F4$ were estimated by averaging the value at the time of a minimum or maximum in a particular formant track and the samples in the preceding and following frames within the hand-transcribed semivowel region. In the case of /w/ and /l/, the values of the formants were averaged

TABLE VI. Formant frequencies (in Hertz) and formant differences (in Hertz and in bark) of semivowels averaged across all speakers.

Prevocalic

	F1 (Hz)	F2 (Hz)	F3 (Hz)	F4 (Hz)
w	381	848	2320	3525
l	399	1074	2553	3767
r	419	1285	1779	3350
j	317	2142	2827	3661

	F1–F0 (Hz)	F2–F1 (Hz)	F3–F2 (Hz)	F4–F3 (Hz)	F4–F2 (Hz)	B1–B0 (bark)	B2–B1 (bark)	B3–B2 (bark)	B4–B3 (bark)	B4–B2 (bark)
w	241	467	1472	1204	2676	2.4	3.6	6.4	2.5	8.9
l	258	675	1479	1214	2693	2.6	4.9	5.5	2.3	7.8
r	242	866	493	1571	2064	2.8	6.0	2.1	3.9	5.9
j	174	1825	684	834	1518	1.7	10.2	1.7	1.6	3.2

Intervocalic

	F1 (Hz)	F2 (Hz)	F3 (Hz)	F4 (Hz)
w	349	771	2340	3508
l	445	1060	2640	3762
r	460	1240	1720	3433
j	361	2270	2920	3824

	F1–F0 (Hz)	F2–F1 (Hz)	F3–F2 (Hz)	F4–F3 (Hz)	F4–F2 (Hz)	B1–B0 (bark)	B2–B1 (bark)	B3–B2 (bark)	B4–B3 (bark)	B4–B2 (bark)
w	211	422	1570	1169	2737	2.1	3.4	7.0	2.4	9.4
l	305	610	1580	1123	2707	3.0	4.5	5.8	2.1	7.9
r	317	783	473	1717	2190	3.1	5.4	2.1	4.2	6.2
j	213	1910	648	906	1554	2.1	10.1	1.5	1.6	3.1

Postvocalic

	F1 (Hz)	F2 (Hz)	F3 (Hz)	F4 (Hz)
l	465	898	2630	3650
r	503	1300	1830	3391

	F1–F0 (Hz)	F2–F1 (Hz)	F3–F2 (Hz)	F4–F3 (Hz)	F4–F2 (Hz)	B1–B0 (bark)	B2–B1 (bark)	B3–B2 (bark)	B4–B3 (bark)	B4–B2 (bark)
l	323	433	1740	1015	2752	3.2	3.2	6.9	1.9	8.8
r	363	799	531	1554	2088	3.5	5.4	2.1	3.7	5.8

around the time of the $F2$ minimum. For /j/, the formant values were averaged around the time of the $F2$ maximum, and for /r/ the formant values were averaged around the time of the $F3$ minimum. Thus the formants were measured during the time when the vocal tract could be expected to be most constricted.

We normalized the formants by computing bark differences to reduce the acoustic variability due to contextual effects and speaker differences and to better capture some of the acoustic properties. Chistovich and Lublinskaya (1979) have postulated that when two formants are within a critical distance of 3.0 to 3.5 bark of each other, they are interpreted by the auditory system as one spectral peak whose frequency is at the center of gravity of the prominence. Syrdal and Gopal (1986), in an acoustic study using the Peterson and Barney (1952) vowel data, investigated whether this con-

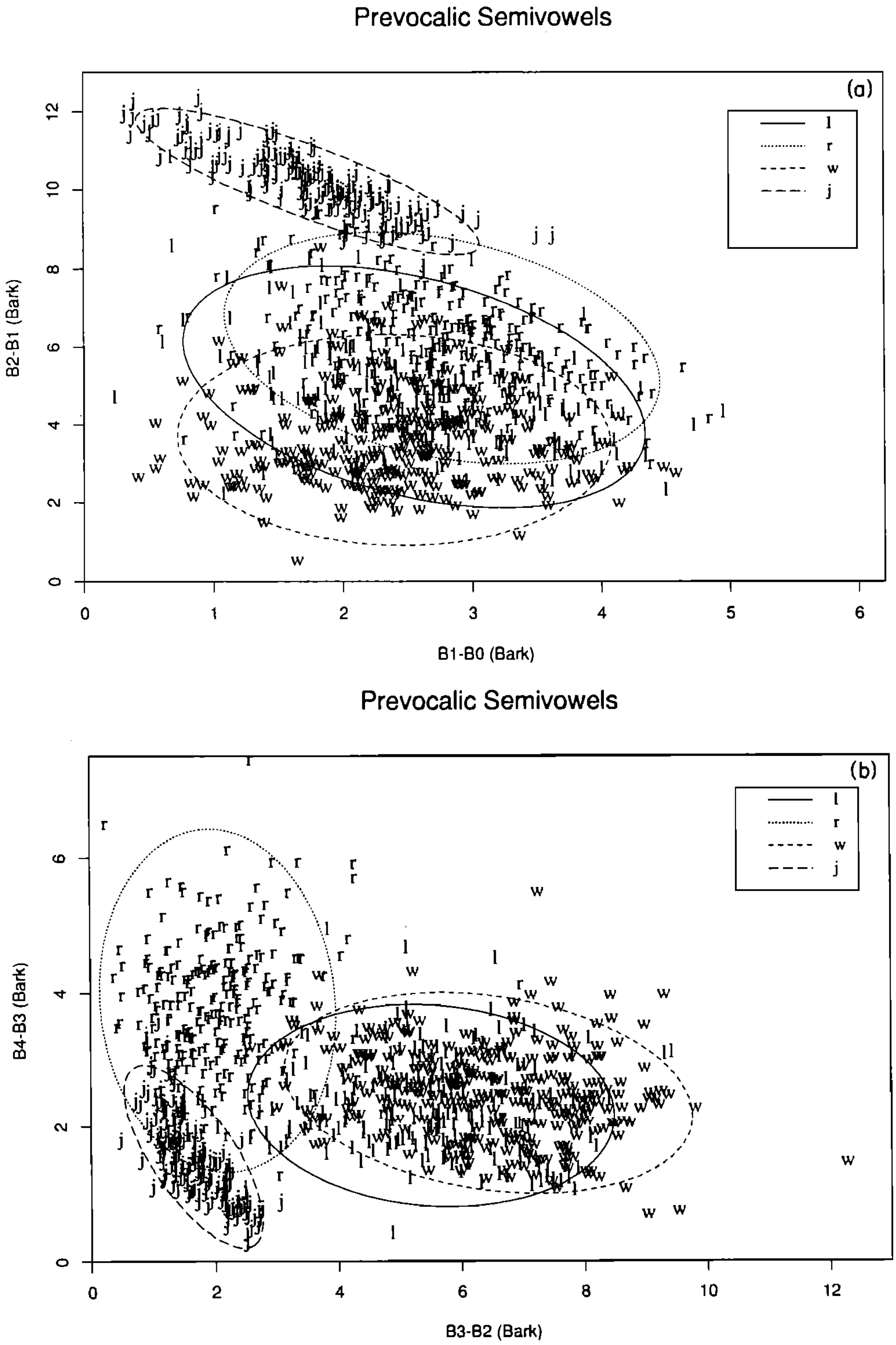

FIG. 16. Scatter plots of the prevocalic semivowels spoken by two males and two females according to the bark transformed (a) $F2$–$F1$ vs $F1$–$F0$ and (b) $F4$–$F3$ vs $F3$–$F2$. A two-dimensional 90% confidence region is drawn around the data for each semivowel.

stant in auditory units held between several formants and between the first formant and the fundamental frequency ($F0$). They found that with a critical distance of 3 bark, the difference between $F1$ and $F0$ provided a reasonable representation of the high–nonhigh vowel distinctions independent of speaker, and the difference between $F3$ and $F2$ represented the front–back vowel distinctions. In addition, they found that the bark difference transformations reduced greatly the acoustic variability between vowels spoken by different talkers.

The formant frequencies obtained in this study are in agreement with previously reported data. The results across speakers are shown in Table VI for prevocalic, intervocalic, and postvocalic semivowels. Also included in the table are normalized formant values ($F1$–$F0$, $F2$–$F1$, $F3$–$F2$, and $F4$–$F3$) and bark differences ($B1$–$B0$, $B2$–$B1$, $B3$–$B2$, and

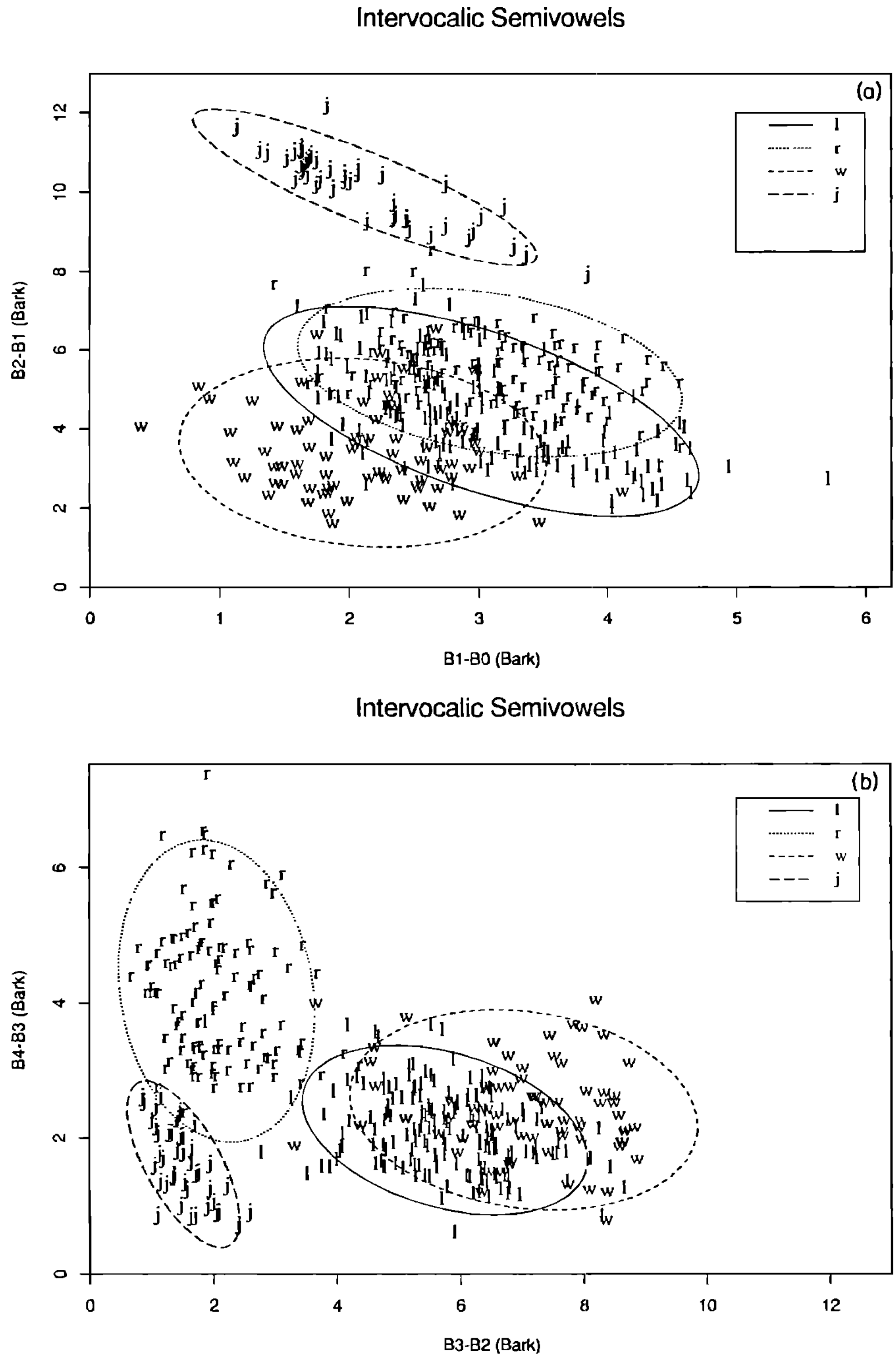

FIG. 17. Scatter plots of the intervocalic semivowels spoken by two males and two females according to the bark transformed (a) $F2$–$F1$ vs $F1$–$F0$ and (b) $F4$–$F3$ vs $F3$–$F2$. A two-dimensional 90% confidence region is drawn around the data for each semivowel.

$B4–B3$). ($F0$ was obtained automatically with the pitch detector described in Gold and Rabiner, 1969.) The distributions of the bark differences are shown in Figs. 16–18 for the prevocalic, intervocalic, and postvocalic semivowels, respectively. A two-dimensional 90% confidence region is drawn around the data for each semivowel. Finally, Table VII summarizes the classification of the semivowels according to a 3.5-bark critical distance criterion and the five bark-difference dimensions. A + indicates that a majority of the semivowels are within 3.5 bark in the bark-difference dimension. Conversely, a − indicates that a majority of the semivowels exceeds the 3.5 bark in the bark-difference di-

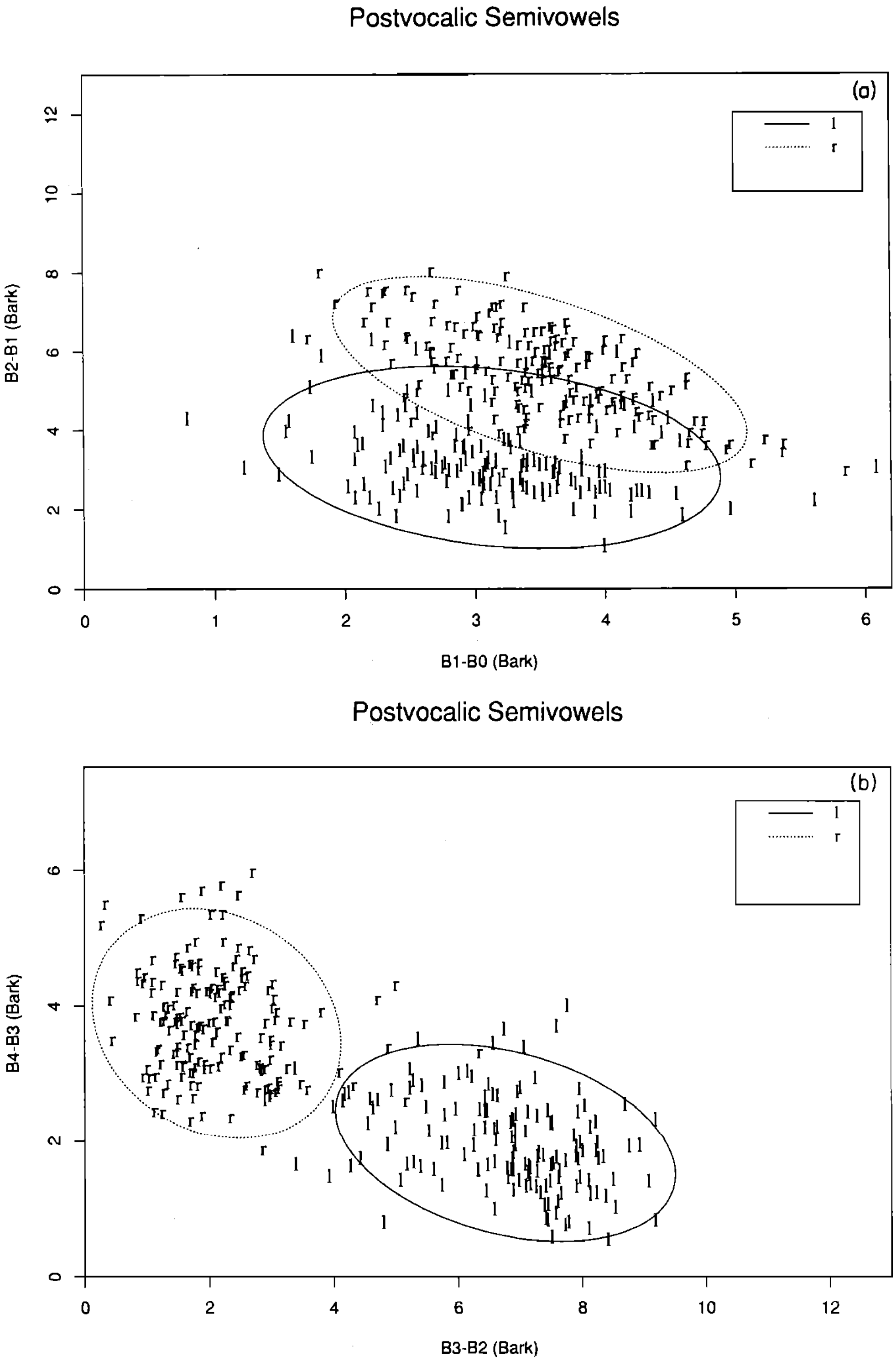

FIG. 18. Scatter plots of the postvocalic semivowels spoken by two males and two females according to the bark transformed (a) $F2–F1$ vs $F1–F0$ and (b) $F4–F3$ vs $F3–F2$. A two-dimensional 90% confidence region is drawn around the data for each semivowel.

TABLE VII. Semivowel classification based on critical distance features in five bark-difference dimensions.

	Dimensions				
Semivowels	$B1-B0$ <3.5 bark	$B2-B1$ <3.5 bark	$B3-B2$ <3.5 bark	$B4-B3$ <3.5 bark	$B4-B2$ < bark
Prevocalic					
w	+	+	−	+	−
l	+	−	−	+	−
r	+	−	+	−	−
j	+	−	+	+	+
Intervocalic					
w	+	+	−	+	−
l	+	−	−	+	−
r	+	−	+	−	−
j	+	−	+	+	+
Postvocalic					
l	−	+	−	+	−
r	−	−	+	−	−

mension. Below, we discuss separately the bark-difference dimensions used to quantify the acoustic property for the features *high*, *back*, *front*, and *retroflex*.

1. High measure

Segments designated as [+ high] are produced with the tongue body raised above the level it holds in the neutral position. This configuration of the tongue body results in a lowered $F1$. The measure used to capture the acoustic property for the feature *high* is the difference $B1-B0$. Syrdal and Gopal (1986) found that the $B1-B0$ dimension separated the high vowels /i ɪ u ʊ/ from the nonhigh vowels.

The data in Table VI show that the difference $B1-B0$ is, on average, less than 3.5 bark for the prevocalic and intervocalic semivowels and, in this sense, this measure puts them in the same class as the high vowels. In addition, $B1-B0$ is smaller for the glides /w/ and /j/ than for the liquids /l/ and /r/. Finally, postvocalic liquids have the highest mean difference. This is not surprising since they tend to be less constricted than the prevocalic allophones.

2. Back–front measures

In sounds that are [+ back], the body of the tongue is retracted from the neutral position, resulting in a lowered $F2$ that is closer to $F1$ than $F3$. Of the semivowels, this lowering of $F2$ is especially salient for /w/ since $F2$ is lowered further by rounding of the lips (introduction of the feature *round*) and by a greater narrowing of the lip opening (introduction of the feature *labial*). For front sounds, on the other hand, the tongue body is displaced forward in the mouth relative to the neutral position. Consequently, $F2$ is raised so that it is closer to $F3$ than to $F1$. Of the semivowels, this $F2$ raising is especially marked for /j/. In addition to fronting of the tongue, the production of /j/ involves a raising of the tongue blade toward the roof of the mouth (introduction of the feature *coronal*), creating a narrow channel between the front of the tongue and the hard palate. This configuration results in a further reduction in the distance between $F2$ and $F3$ and it produces an $F3$ which is close to $F4$. The "spectral center of gravity" of the broad prominence formed by $F2$, $F3$ and $F4$ is in the region of $F3$ or higher (Carson *et al.*, 1970).

Syrdal and Gopal (1986) found that the $B3-B2$ dimension distinguished between front and back vowels. As shown in Table VII, the $B3-B2$ dimension classifies /r/ and /j/ as *front*, and /w/ and /l/ as *back*. In the case of /w/, the large distance between $F3$ and $F2$ implies a close spacing between $F1$ and $F2$. As stated above, the feature *back* is strengthened in /w/ by the features *round* and *labial* (Stevens *et al.*, 1986). This enhancement is evidenced by the $B2-B1$ difference in Table VI. If we exclude those semivowels that are in clusters with unvoiced consonants where they are likely to be at least partially devoiced, the mean $B2-B1$ difference for the prevocalic /w/ is reduced from 3.6 to 3.1 bark and it remains substantially higher than 3.5 bark for the other prevocalic semivowels. In terms of their distribution, 69% of /w/ productions have a difference $B2-B1$ less than 3.5 bark, whereas only 13% of /l/ and 4% of /r/ productions fall into this category. Similarly, in the case of the intervocalic semivowels, 61% of /w/'s have a $B2-B1$ difference less than 3.5 bark whereas only 26% of /l/'s and 3% of /r/'s fall into this category. Thus, the spacing between $F2$ and $F1$ for most /w/'s is close enough that they may be perceived by the auditory system as one formant according to the conclusions of Chistovich and Lublinskaya (1979). Most of the /w/'s with a larger spacing between $F1$ and $F2$ are either adjacent to a nonback vowel(s) and/or they are in an unstressed context such as those in the words "withhold" and "periwig."

As compared to the /l/ in prevocalic or intervocalic positions, the postvocalic /l/ has a much closer spacing between $F1$ and $F2$. In fact, the difference $B2-B1$ for a postvocalic /l/ is comparable to the values obtained for the /w/'s. That is, 84% of postvocalic /l/'s have a difference $B2-B1$ less than 3.5 bark (less than 8% of the postvocalic /r/'s have such a close spacing between $F1$ and $F2$). This difference in the values of $B2-B1$ for a postvocalic /l/ compared to the prevocalic and intervocalic /l/ supports previous findings with regard to its allophonic variation. That is, a postvocalic /l/ is more velarized, with less of a constriction formed by the tongue blade, resulting in a much lower $F2$, a higher $F1$ and, therefore, a smaller $F2-F1$ difference. Those postvocalic /l/'s that fall outside the critical range are not word-final, but they are adjacent to the semivowel /j/, a fronted sonorant, in words like "brilliant" and "cellular." Thus the back articulation of the /l/ is influenced by the front articulation of the following /j/.

The classification of /r/ as a front sonorant may seem a bit unusual. [Syrdal and Gopal (1986) also found that the $B3-B2$ dimension classified the retroflexed vowel /ɚ/ as being front.] However, Peterson and Barney (1952) found in a perceptual experiment using real speech that 80% of the /ɚ/'s that were not heard as such were confused with the front vowels /ɛ/ and /æ/. As can be seen from the distributions in Figs. 16–18, there are a few /r/'s in each plot with a $B3-B2$ difference greater than 3.5 bark. In the case of the prevocalic /r/, this large difference may be due to "additional" rounding which would lower all of the formants. That is, as Delattre and Freeman (1968) found, lip rounding accompanies all prestressed prevocalic American /r/'s. However, from listening to these tokens and judging from their formant frequencies, it appears that speakers, two in particular, produced a sound that is a cross between /w/ and /r/. For example, in one repetition of the word "rule," $F2$ within the /r/ is as low as 620 Hz, which is in the expected $F2$ range of /w/. However, $F3$ is 1240 Hz which is too low for a /w/, but in the $F3$ range of /r/. In this case, $B3$–$B2$ is 4.3 bark.

The postvocalic /r/'s with a $B3$–$B2$ difference greater than 3.5 bark are not word-final, but occur in words like "forewarn" and "harlequin," where they are followed by a /w/ or an /l/. In this environment, two effects can occur. First, $F2$ in the /r/ is usually reduced by the velarization or rounding occurring within these back sonorants. Second, the /r/ is often merged with the preceding vowel so that the surface manifestation of the vowel and following /r/ is an /r/-colored vowel. An example where both phenomena occur is shown in Fig. 19, which is a spectrogram of the word "Norwegian." Both the word-initial /n/ and the vowel /ɔ/ are retroflexed. The visible portion of $F3$ at the beginning and end of the first sonorant region (60 to 300 ms) as well as the automatically extracted $F3$ track show that the lowest point of $F3$ occurs at the beginning of the /n/ around 60 ms. $F2$ during the region that is transcribed as /r/ becomes as low as 612 Hz due to the /w/.

Finally, the situation for the intervocalic /r/ for which $B3$–$B2$ falls outside the 3.5-bark range is similar to that of the postvocalic allophones. That is, either the /r/ is merged with a preceding vowel or it occurs in words like "already" and "bulrush" where an underlying and preceding /l/ was not transcribed, but the preceding sonorant region is produced with a back articulation which results in a lowered $F2$ within the /r/.

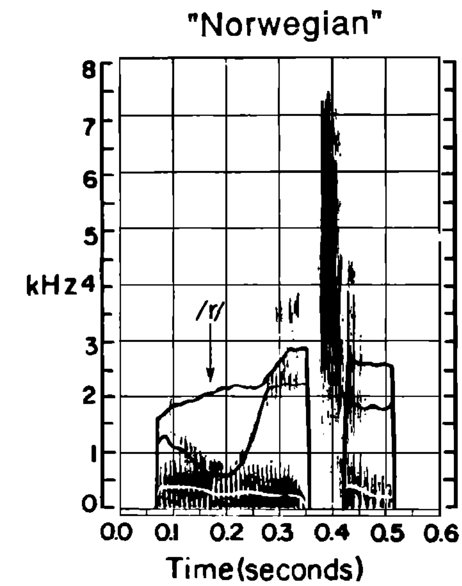

FIG. 19. Shown is a wideband spectrogram of the word "Norwegian" with automatically extracted formant tracks overlaid. The phonetically transcribed /r/ shows a large separation between $F3$ and $F2$ which results in a difference greater than 3.5 bark.

While /r/ and /j/ both have a close spacing between $F3$ and $F2$, the frequency range of this spectral prominence is quite different. For /r/, this prominence occurs in a midfrequency range between 1000 and 2000 Hz. On the other hand, for /j/, this spectral prominence occurs in a high-frequency range between 2000 and 3000 Hz. Thus, as stated above, the feature *front* for /j/ is enhanced by the feature *coronal* (Stevens *et al.*, 1986), resulting in a broad spectral prominence which includes $F2$, $F3$, and $F4$. Figures 16 and 17 show that /j/ always has a $B3$–$B2$ difference less than 3 bark. In fact, as can be seen from Table VI, the average spacing between $F2$ and $F4$ is also within the critical separation of 3.0 to 3.5 bark.

3. Retroflex measure

The major acoustic consequence of the feature *retroflex* appears to be a low third formant. For a typical male speaker, the frequency of $F3$ for an /r/ is usually at or below 2000 Hz. As a result, $F3$ and $F4$ are usually well separated whereas the difference between $F3$ and $F2$ is small. The data in Table VI show that /r/ can normally be separated from the other semivowels based on the difference $B4$–$B3$, which is usually greater than 3.5 bark for /r/ and less than 3.5 bark for the other semivowels. However, there are some /r/'s with a $B4$–$B3$ difference that is less than 3 bark.

The exceptions occur for several reasons. First, in addition to lowering $F3$ within /r/, some speakers lower $F4$ as well. Second, the articulation of /r/ is sometimes considerably affected by the articulation of a following sound so that $F3$ is higher than normal. Finally, for some word-final /r/'s $F3$ does not get very low, suggesting that the degree of palatal constriction may be lessened.

E. Formant transition measures

The wide spread in the distribution of average formant values given in Figs. 16–18 shows that the formant frequencies of the semivowels are affected by adjacent sounds. Because of this contextual effect, there is some overlap in the distributions of the bark differences for the semivowels. For example, there is substantial overlap between the formant distributions of /w/ and /l/ in Figs. 16 and 17. However, the figures also show that the influence of adjacent sounds could lead to confusions between /w/ and /r/ based on the bark differences alone. Most of the /w/'s and /r/'s which have similar formant values are in clusters with unvoiced consonants and they are partially devoiced, so that the information which distinguishes between them is often outside of the sonorant regions. In the few cases where voiced /w/ and /r/ productions are confused on the basis of their formant values, they are distinguishable if the formant transitions are taken into account. Thus, in addition to the spacing between the formants, the formant transitions are sometimes needed to help make certain distinctions.

To determine the direction and extent of these formant movements, the average semivowel formant values were subtracted from the average formant values of the adjacent vowel(s). The target vowel formant frequencies were computed from the value at the time of the maximum $F1$ frequency within the hand-transcribed vowel region and the values in the previous and following frames.

1. F1 transitions

The average and standard deviation of the $F1$ differences between vowels and adjacent semivowels are plotted in Fig. 20. The data show that $F1$ normally increases from a prevocalic semivowel into the following vowel. $F1$ is also consistently lower in an intervocalic /w/ and /j/ relative to its value in the adjacent vowels. This is also the case for many intervocalic /r/'s and /l/'s. However, when the liquids are adjacent to both a high vowel and a low vowel, their $F1$ values will sometimes lie somewhere between the $F1$ frequencies of the neighboring sounds. In the few cases where a postvocalic liquid had a higher $F1$ than that of the preceding vowel, the vowel is characterized as *high* and, therefore, it normally has quite a low $F1$ frequency. Examples of such words are "cartwheel" and "clear."

2. F2 transitions

The average and standard deviation of the $F2$ differences between vowels and adjacent semivowels are plotted in Fig. 21. The data show that $F2$ generally increases between a /w/ or /l/ and an adjacent vowel, and it usually decreases between a /j/ and an adjacent vowel. However, between an /r/ and surrounding vowels, $F2$ may increase or decrease. In

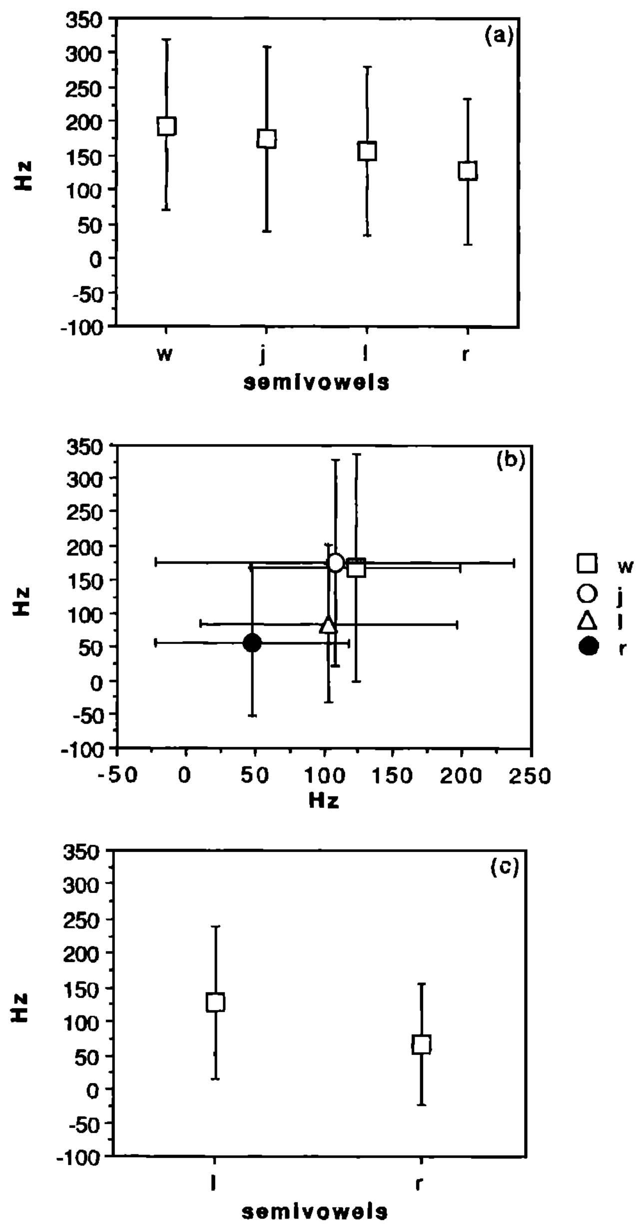

FIG. 20. Averages and standard deviations of the differences (in Hz) between the average $F1$ frequency of (a) prevocalic semivowels and following vowels, (b) intervocalic semivowels and surrounding vowels (the change relative to the preceding vowel is on the horizontal axis and the change relative to the following vowel is on the vertical axis), and (c) postvocalic semivowels and preceding vowels.

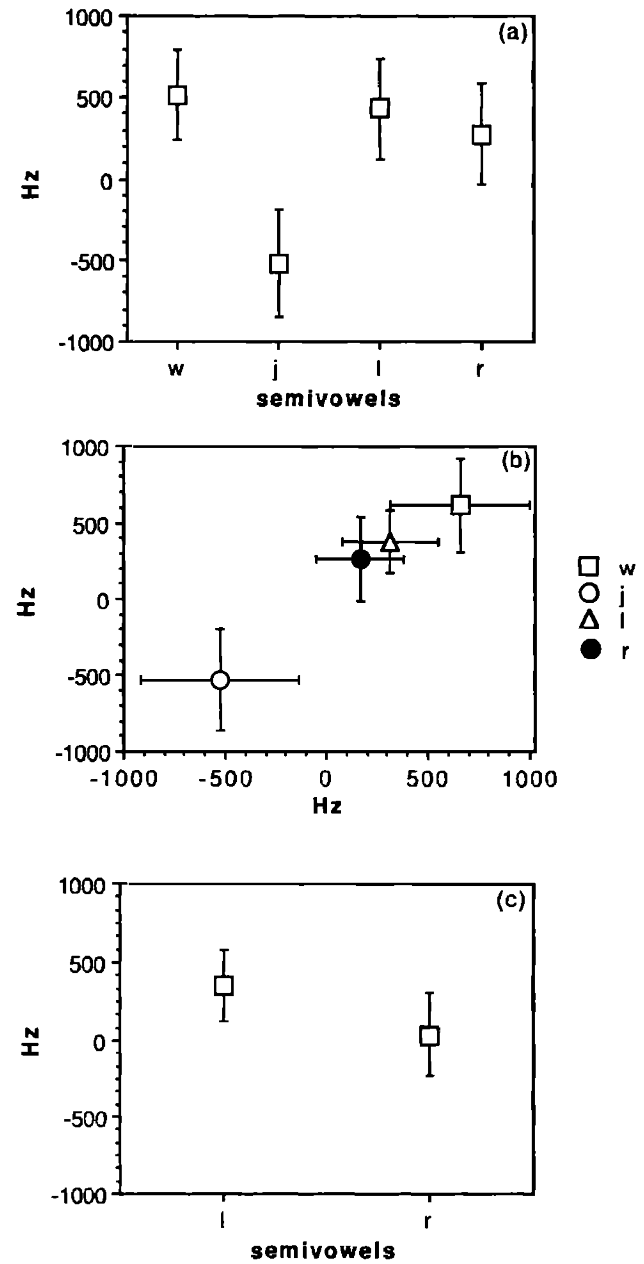

FIG. 21. Averages and standard deviations of the differences (in Hz) between the average $F2$ frequency of (a) prevocalic semivowels and following vowels, (b) intervocalic semivowels and surrounding vowels (the change relative to the preceding vowel is on the horizontal axis and the change relative to the following vowel is on the vertical axis), and (c) postvocalic semivowels and preceding vowels.

the case of prevocalic /r/, a decrease in $F2$ from /r/ into the following vowel mainly occurred when /r/ was in a cluster with a preceding coronal consonant such as the /d/ in "withdraw." In addition, there were a few cases where the $F2$ differences between the vowel /u/ and the preceding word-initial /r/ in the words "rule" and "roulette" were also negative. However, in all but one case, there was an initial rise in $F2$ from the /r/ before it fell into its lower value for the /u/. This type of $F2$ trajectory was also noted by Lehiste (1962).

As for intervocalic /r/, most have a lower $F2$ value than that of adjacent vowels. However, if an intervocalic /r/ is preceded by a back vowel and followed by a front vowel, as in "chlorination" ([klɔrɪne^{y}šən]), then there may be a rise in $F2$ from the back vowel through the /r/ and into the front vowel. Likewise, if the /r/ is preceded by a front vowel and followed by a back vowel, as in "heroin" ([hɛro^{w}ɪn]), then $F2$ may fall steadily from the front vowel through the /r/ and into the back vowel.

Finally, in the case of postvocalic /r/ and preceding vowels, $F2$ may increase or decrease, depending upon whether the vowel is front or back. That is, if the vowel is back, $F2$ may rise while $F3$ falls, narrowing the difference between $F3$ and $F2$. However, if the vowel is front, both $F2$ and $F3$ will fall into the appropriate values for an /r/.

3. F3 transitions

The average and standard deviation of the $F3$ differences between vowels and adjacent semivowels are plotted in Fig. 22. The data show that $F3$ is almost always substantially higher in /j/ than it is in an adjacent vowel. In the few cases when this is not true, another coronal consonant was nearby, such as the /l/ in "uvula" ([jUvjUlə]). In words like this $F3$ steadily rose from its value in the /j/ to a somewhat higher value in the nearby consonant.

As expected, Fig. 22 shows that $F3$ is almost always substantially lower in /r/ than it is in an adjacent vowel. However, there are several instances where $F3$ for /r/ is comparable to or higher than that of the preceding vowel. With the exception of the intervocalic /r/ in one pronunciation of the word "guarani," these /r/'s occur in postvocalic, but not word-final, position. That is, they are always followed by another consonant, such as the /r/'s in "cartwheel," "harlequin," and "Norwegian." An example of this type of $F3$ trajectory is shown in the spectrogram of the word "cartwheel" in Fig. 13. As can be seen, the lowest point of $F3$ within the /ɑr/ region (0.1 to 0.2 s) occurs near the beginning of the /ɑ/. Acoustically, the vowel and /r/ appear to be completely assimilated in that no discernible acoustic cue points to separate /ɑ/ and /r/ segments. Thus it appears as if the /ɑ/ and following /r/ are merged into an r-colored vowel.

The $F3$ transition for /w/ and an adjacent vowel can be positive or negative. A negative $F3$ transition from /w/ into an adjacent vowel may seem surprising since /w/ is produced with a labial constriction. However, we found this to be the case mainly when /w/ is adjacent to a retroflexed vowel. The average change in $F3$ between prevocalic /w/ and following retroflexed vowels is about -215 Hz. In the case of intervocalic /w/, the average increase in $F3$ from a

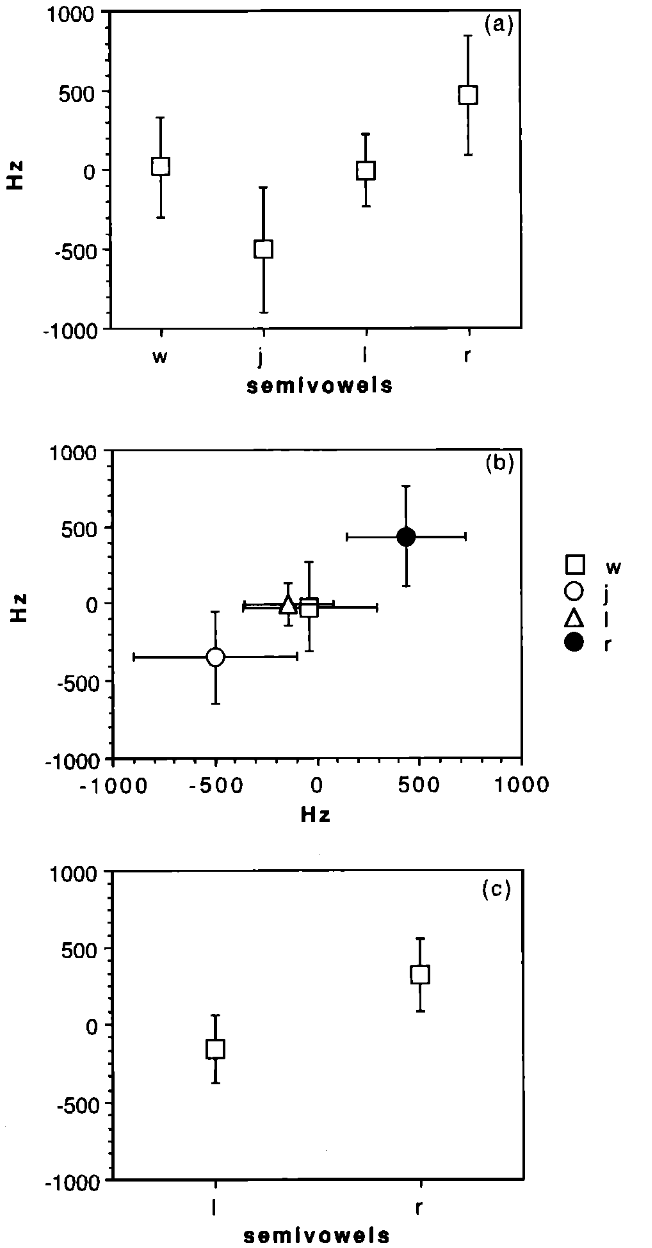

FIG. 22. Averages and standard deviations of the differences (in Hz) between the average $F3$ frequency of (a) prevocalic semivowels and following vowels, (b) intervocalic semivowels and surrounding vowels (the change relative to the preceding vowel is on the horizontal axis and the change relative to the following vowel is on the vertical axis), and (c) postvocalic semivowels and preceding vowels.

preceding retroflexed vowel is about 300 Hz, and the average decrease in $F3$ into a following retroflexed vowel is about 200 Hz. An example of this phenomenon can be seen in the spectrogram and formant tracks of the word "froward," which is displayed in Fig. 23. Although $F3$, due to its low amplitude, is not always visible within the /w/ the direction of the $F3$ movement can be inferred from the visible transitions in the adjacent vowels, and it is apparent in the accompanying formant tracks.

If we exclude those /w/'s which are either adjacent to a retroflexed sound or one segment removed from a retroflexed sound (e.g., "guarani"), the average increase in $F3$ from a prevocalic /w/ into a following vowel is 105 Hz. For

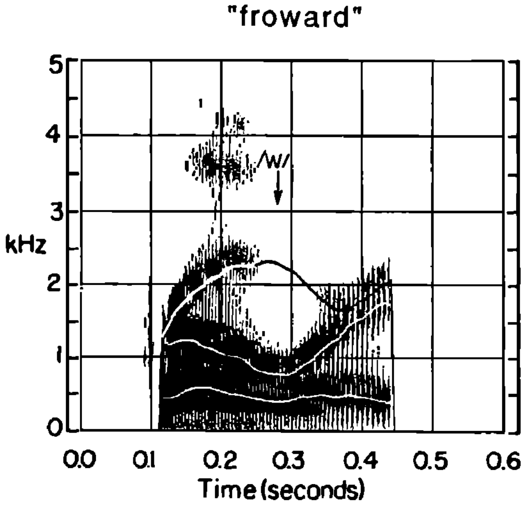

FIG. 23. An illustration of $F3$ movement between /w/ and nearby retroflexed sounds in "froward." Shown is a wideband spectrogram with automatically extracted formant tracks overlaid.

intervocalic /w/'s, the average increase in $F3$ into the following vowel is 70 Hz and the average increase into the preceding vowel is 164 Hz.

Finally, the $F3$ differences involving /l/ show that $F3$ is almost always substantially higher in /l/ relative to a preceding vowel, and that there is usually little change in $F3$ between /l/ and a following vowel. These data support previous findings (Lehiste, 1962) which show that $F3$ for /l/ tends to be equal to or higher than that of adjacent vowels. However, as can be inferred from the standard deviations, there are several instances where a prevocalic and intervocalic /l/ had a substantially lower $F3$ frequency than that of the adjacent vowel. This phenomenon, which usually occurs when /l/ is adjacent to a front vowel, was observed in words such as "leapfrog" and "Swahili," where $F3$ is already at a high frequency, close to $F4$.

IV. GENERAL DISCUSSION

The results of this investigation have shown that the acoustic measures used in this study for the features *sonorant, syllabic, consonantal, high, back, front,* and *retroflex* are, for the most part, extracting relevant information from the speech signal. That is, the quantified acoustic properties for these features generally separate the semivowels from other sounds and distinguish among the semivowels. The acoustic measures are summarized in Table V. The acoustic property for *− syllabic* separates the semivowels from vowels and the acoustic property for *+ sonorant* separates the semivowels from most consonants (the nasals are also sonorant consonants). In addition, as has been shown in past studies, the acoustic property for *+ retroflex* separates /r/ from /w j l/ and the acoustic property for *+ front* separates /r j/ from /w l/. However, no one acoustic measure investigated in this study provides a clear distinction between /w/ and /l/. While the acoustic property for *+ consonantal* separates prestressed /l/ from the other semivowels and the acoustic property for *+ back* separates many /w/'s from /j r l/, considerable overlap remains in the distributions for /w/ and /l/. Subsequent work (Espy-Wilson, 1989) suggests that other features such as *coronal* and *labial* maybe useful in distinguishing between these two sounds. Other information such as phonotactic constraints (/l/ can occur in postvocalic position while /w/ cannot) and the direction of the $F3$ transition between the semivowel and adjacent vowel are also helpful in making the /w/–/l/ distinction.

In addition to the overlap between the distributions for /w/ and /l/, overlap between the distributions for the semivowels and other classes of sounds show that a recognition system based only on the acoustic properties for features investigated in this study will not always be able to recognize correctly the semivowels.

In light of our general goal of a semivowel recognition system, it is important to understand why this overlap occurred. From our analysis, we attribute the exceptions to several reasons. First, the results of this study suggest refinements in the acoustic properties. For example, the feature *high* groups all of the prevocalic semivowels together even though a division should be made between the liquids and glides. This lack of discrimination suggests that other considerations may need to be taken into account in the development of the appropriate acoustic properties for features. In particular, it may be the case that the acoustic correlates of some or all of the features differ depending upon whether the sound is vocalic and thus the vocal tract is relatively open, or nonvocalic and thus the vocal tract is more constricted.

Second, there is the segmentation problem. As the $F3$ transition data for /r/ of Sec. III E show, speech sounds often overlap, at least to some extent, so that some of the strongest acoustic evidence for a feature that is distinctive for a particular sound may occur outside of the region transcribed for that sound.

Finally, labeling of the segments is an issue. The labeling of some sounds is inherently subjective even with a phonemic transcription available as a reference. In some pronunciations of words like "flower," a clear /w/ will be heard and in others the presence of a /w/ will be questionable. In addition, there are feature changes which occur in some contexts, but they are not anticipated in the transcription. For instance, even though the underlying /v/ in "everyday" shown in Fig. 5 is realized as a sonorant consonant as opposed to a fricative consonant, the label assigned to this portion of the speech signal does not reflect this large acoustic change. Thus, mismatches between assigned labels and the expected acoustic properties are to be expected. By relating the measured acoustic properties to the articulartory correlates for features, we are able to understand these feature modifications as changes in articulation.

As the results of Secs. III A and C show, the semivowels can in general be separated from other sounds (except nasals) on the basis of the acoustic properties for the features *+ sonorant* and *− syllabic*. However, blurring of this distinction between semivowels and most other sounds does sometimes occur (e.g., the example of "everyday" mentioned above). This apparent neutralization is caused by a reduction in the degree of the constriction normally made in the production of consonants—usually poststressed conson-

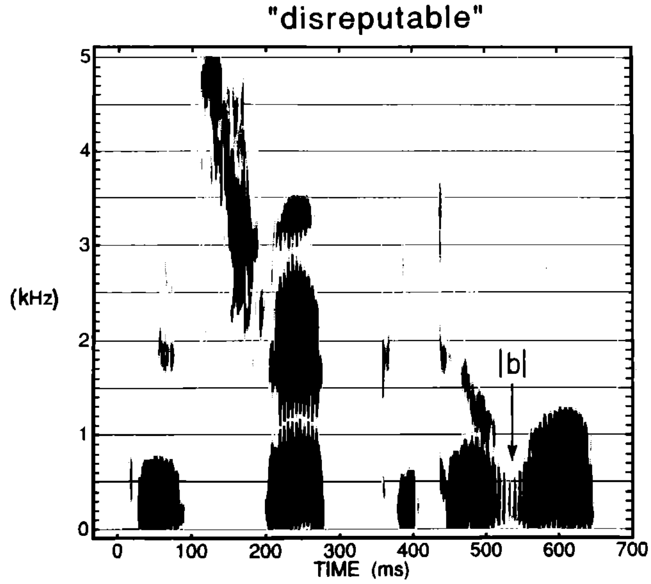

FIG. 24. A wideband spectrogram of the word "disreputable" which contains a lenited /b/ that resembles a /w/ and an /l/.

ants. The term often used to refer to this type of variability is *lenition* (cf. Catford, 1977). Data from Sec. III A show that underlying intersonorant voiced stops and voiced fricatives are sometimes sufficiently weakened that they surface as voiced approximants which exhibit the property of sonorancy. Most of the lenited consonants observed in this study are preceded by a vowel with a higher degree of stress than the vowel occurring either after them or after the following sonorant consonant. In some cases, these weakened consonants bear a close resemblance to one or more semivowels. For example, the /b/ in "disreputable" shown in Fig. 24 is acoustically similar to a /w/ and an /l/. In addition to being sonorant, $F1$, $F2$, $F3$, and $F4$ at the lowest point of $F2$ during the /b/ (around 540 ms) are 348, 820, 2714, and 3672 Hz, respectively. These values are close to the average formant frequencies listed in Table VI for intervocalic /w/ and /l/. Furthermore, since /b/ is a labial consonant, it has formant transitions similar to those of /w/. Data from Secs. III C and C 3 show that poststressed intervocalic and postvocalic semivowels are also sometimes weakened so that they are not nonsyllabic. Instead, they appear to form a part of the syllable nucleus.

Another distinction which is sometimes obscured by lenition is the separation of /l/ from the other semivowels on the basis of the acoustic property for the feature *consonantal*. The results of Sec. III B show that not all prevocalic /l/'s are associated with abrupt spectral changes. As stated in Sec. I, several researchers (Joos, 1948; Fant, 1960; Dalston, 1975) have noticed some type of sharp spectral discontinuity between /l/ and a following vowel. In agreement with this observation, the data in Sec. III B show abrupt spectral changes between /l/ and following stressed vowels. However, the results also show that the spectral change between prevocalic /l/ (as well as other consonants) and following unstressed vowels can be quite gradual. In this case, the prevocalic /l/ does not appear to be consonantal.

Another category of variability observed often in this study is *feature spreading* (cf. Henke, 1966; Moll and Daniloff, 1971; Kent *et al.*, 1974). The $F3$ transition data of Sec. III E 3 point to the merging of postvocalic /r/ with preceding vowels, resulting in r-colored vowels. Such assimilation is evident in Fig. 13 between the /ɑ/ and /r/ in "cartwheel" and in Fig. 19 between the /n/, /ɔ/ and /r/ in "Norwegian." Espy-Wilson (1991) found that this anticipation of a postvocalic /r/ depends on several factors including speaking rate, consonantal context, and speaker differences.

In addition, we have observed the backward spreading of retroflexion from a prevocalic /r/, across a labial consonant, to a preceding vowel. An example of this occurrence can be seen in the word "everyday" of Fig. 5. The lowest point of $F3$ occurs during the /v/ which not only is retroflexed, but also is lenited. From a speech production viewpoint, retroflex spreading in this context is not surprising since the /v/ is a labial consonant and, therefore, does not require a particular placement of the tongue. Thus the tongue configuration for the /r/ can be anticipated during the consonant and possibly earlier.

V. SUMMARY

In conclusion, we have studied several acoustic properties of the semivowels which generally separate them as a class from other speech sounds, and which distinguish among them. In addition, we have observed how the acoustic properties of the semivowels and of other sounds can change as a function of context. The observed variability can be understood in terms of production and, at the more abstract level of features, in terms of lenition and assimilation. These feature altering processes can occur independently and simultaneously. Understanding variability in terms of how and when features may change and which features are invariant has implications for the representation of lexical items. Such knowledge should not only help us understand the human speech system, but it may also contribute to the development of feature-based systems for speech recognition and to the synthesis of natural sounding speech.

ACKNOWLEDGMENTS

This work was supported in part by a Xerox Fellowship and NSF Grant BNS-8920470. It is based in part on a Ph.D. thesis by the author submitted to the Department of Electrical Engineering and Computer Science at MIT. The author gratefully acknowledges the encouragement and advice of Professor Kenneth N. Stevens. I also want to acknowledge the assistance of Marie Huffman, Suzanne Boyce, Sharon Manuel, and Corine Bickley, whose comments greatly improved the quality and clarity of this paper. The editorial comments and suggestions of Jesse G. Kennedy III, John H. Saxman, and two anonymous reviewers are also very much appreciated. Special thanks to Jeffrey Marcus and David Goodine for their technical assistance in generating the formant plots.

[1]The preboundary /l/'s occurred before phonological boundaries which varied in strength so that in some cases (e.g., before a major intonation break) the /l/ is syllable final and in other cases (e.g., between two vowels) it is unclear whether the /l/ is syllable final or syllable initial.

[2]A change in the phonological representation of these sounds is not being argued for. However, for the kinds of phonetic characterizations being made, it is appropriate to separate /w/ from /l r/ from /j/ in the front–back continuum.

[3]Frames occur at 5-ms intervals and they are obtained by windowing the speech signal with a 25.6-ms Hamming window.

Bickley, C., and Stevens, K. N. (1986). "Effects of a Vocal Tract Constriction on the Glottal Source: Experimental and Modeling Studies," J. Phon. **14**, 373–382.

Bladon, R. A. W., and Al-Bamerni, A. (1976). "Coarticulation Resistance in English /l/," J. Phon. **4**, 137–150.

Bond, Z. S. (1976), "Identification of Vowels Excerpted from /l/ and /r/ contexts," J. Acoust. Soc. Am. **60**, 906–910.

Carlson, R., Granström, B., and Fant, G. (1970). "Some Studies Concerning Perception of Isolated Vowels," Speech Transmiss. Lab.: Q. Prog. Stat. Rep. **2–3**, 19–35.

Catford, J. C. (1977). *Fundamental Problems in Phonetics* (Indiana U. P., Bloomington, IN).

Chistovich, L. A., and Lublinskaya, V. V. (1979). "The 'Center of Gravity' Effect in Vowel Spectra and Critical Distance between the Formants: Psychoacoustical Study of the Perception of Vowel-like Stimuli," Hear. Res. **1**, 185–195.

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

Cyphers, D. S. (1985), "Spire: A Research Tool," S. M. thesis, MIT, Cambridge, MA.

Dalston, R. M. (1975). "Acoustic Characteristics of English /w,r,l/ Spoken Correctly by Young Children and Adults," J. Acoust. Soc. Am. **57**, 462–469.

Delattre, P., and Freeman, D. (1968). "A Dialect Study of American *R*'s by X-ray Motion Picture," Language **44**, 29–68.

Espy-Wilson, C. Y. (1987). "An Acoustic-Phonetic Approach to Speech Recognition: Application to the Semivowels," Technical Report No. 531, Research Laboratory of Electronics, MIT.

Espy-Wilson, C. Y. (1989). "The Influence of Selected Acoustic Cues on the Perception of /l/ and /w/," J. Acoust. Soc. Am. Suppl. 1 **86**, S49.

Espy-Wilson, C. Y. (1991). "Consistency in /r/ Trajectories in American English," Proc. 12th Int. Cong. Phon. Sci. **4**, 370–373.

Fant, G. (1960). *Acoustic Theory of Speech Production* (Mouton, The Netherlands).

Gold, B., and Rabiner, L. (1969). "Parallel Processing Techniques for Estimating Pitch Periods of Speech in the Time Domain," J. Acoust. Soc. Am. **2**, 442–448.

Giles, S. B., and Moll, K. L. (1975). "Cinefluorographic Study of Selected Allophones of English /l/," Phonetica **31**, 206–227.

Henke, W. (1966). "Dynamic Articulatory Model of Speech Production Using Computer Simulation," Ph.D. dissertation, MIT, Cambridge, MA.

Jakobson, R., Fant, G., and Halle, M. (1952). "Preliminaries to Speech Analysis," MIT Acoustics Lab. Tech. Rep. No. 13.

Joos, M. (1948). "Acoustic Phonetics," Lang. Suppl. **24**, 1–136.

Kameny, I. (1974). "Comparison of the Formant Spaces of Retroflexed and Non-retroflexed Vowels," IEEE Symp. Speech Recog. **80-T3–84-T3**.

Kent, R., Carney, P., and Severeid, L. (1974). "Velar Movement and Timing: Evaluation of a Model for Binary Control," J. Speech Hear. Res. **17**, 470–488.

Lehiste, I. (1962). "Acoustical Characteristics of Selected English Consonants," Report No. 9, University of Michigan, Communication Sciences Laboratory, Ann Arbor, MI.

Lisker, L. (1957). "Minimal Cues for Separating /w,j,r,l/ in Intervocalic Position," Word **13**, 256–267.

Moll, K., and Daniloff, R. (1971). "Investigation of the Timing of Velar Movements During Speech," J. Acoust. Soc. Am. **50**, 678–694.

O'Connor, J. D., Gertsman, L. J., Liberman, A. M., Delattre, P. C., and Cooper, F. S. (1957). "Acoustic Cues for the Perception of Initial /w,j,r,l/ in English," Word **13**, 24–43.

Peterson, G. E., and Barney, H. (1952). "Control Methods Used in a Study of the Vowels," J. Acoust. Soc. Am. **24**, 175–184.

Seneff, S. (1986). "A Computational Model for the Peripheral Auditory System: Application to Speech Recognition Research," IEEE Int. Conf. Acoust. Speech Sig. Process. **3**, 1938–1986.

Shipman, D. W. (1982). "Development of speech research software on the MIT lisp machine," J. Acoust. Soc. Am. Suppl. 1 **71**, S103.

Sproat, R., and Fujimura, O. (submitted). "On the Nature of Allophonic Variation and the Phonology/Phonetics Mapping: the Case of /l/ in English," J. Phon.

Stevens, K. N., Keyser, S. J., and Kawasaki, H. (1986). "Toward a Phonetic and Phonological Theory of Redundant Features," *Invariance and Variability in Speech Processes*, edited by J. Perkell and D. Klatt (Erlbaum, Hillsdale, NJ), pp. 426–449.

Stevens, K. N., and Keyser, S. J. (1989). "Primary Features and their Enhancement in Consonants," Language **65**, 81–106.

Stevens, K. N., "Acoustic Phonetics" (in preparation).

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.

Zue, V., Cyphers, D., Kassel, R., Kaufman, D., Leung, H., Randolph, M., Seneff, S., Unverferth, J., and Wilson, T. (1986). "The Development of the MIT LISP-Machine Based Speech Research Workstation," IEEE Int. Conf. Acoust. Speech Process., 7.6.1–7.6.4.