Vowel representation: Some observations on temporal and spectral properties of the first formant frequency

Maria-Gabriella Di Benedetto[a)]
Speech Communication Group, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

(Received 20 May 1987; accepted for publication 4 January 1989)

Acoustic analysis of the vocalic portion of consonant–vowel–consonant (CVC) syllables (where V is one of the five vowels [ɪ,ɛ,æ,ɑ,ʌ] of American English) spoken by three speakers (two males and one female) in the sentence frame "The __ again" is presented. Results of acoustic measurements show that ambiguities between vowels, for each speaker, occur if the vowels are represented by the values of $F1$ and $F2$ sampled at the time where $F1$ reaches its maximum. These ambiguities occur primarily in the $F1$ dimension. Examination of the $F1$ trajectories of the vowels for which confusion occurs shows variations in the way $F1$ reaches its maximum among different vowels. In particular, if two different vowels such as [ɪ] and [ɛ] have the same maximum $F1$, then $F1$ for the lower vowel reaches its maximum value earlier. In addition, results show that the $F1$ onset frequency also might be important in determining vowel height. The implication is that the spectral characteristics at a particular "target," represented by the time at which $F1$ reaches its maximum, are not invariant attributes of the vowel. The results support a hypothesis that time and/or frequency variations of the first formant must be taken into account if an invariant property is to be associated with a vowel.

PACS numbers: 43.71.Es, 43.70.Fq

INTRODUCTION

Vowel sounds have been traditionally classified along several dimensions: height, backness, tenseness, etc. The formant frequencies of vowels have been widely used as acoustic parameters representative of the different dimensions. For example, it is well known that the first formant frequency ($F1$) has been related to vowel height and the second formant frequency ($F2$) to vowel backness. In terms of distinctive feature theory (Chomsky and Halle, 1968), vowels are coded as being [+ vocalic] and [− consonantal]. In the vowel class, different vowels are characterized by different features; for example, the vowel [i] is [+ high] and [− back], while the vowel [u] is [+ high] and [+ back].

According to the theory of acoustic invariance (Stevens and Blumstein, 1981), the search for acoustic correlates of distinctive phonetic features should tend to specify acoustic parameters that are invariant among different speakers, languages, and phonetic contexts, and that are perceptually relevant. The primary aim of the acoustic analysis that will be presented in this paper was to investigate how accurately $F1$ can be used to classify vowels according to vowel height. Stevens and House (1955) related the acoustic event described by a high first formant with the articulatory event characterized by a narrow tongue constriction a few centimeters above the glottis and an unrounded large mouth opening, and low $F1$ values with small and rounded mouth opening or with a narrow tongue constriction near the mouth opening. For vowels characterized by high $F1$ values (for example, the vowel [ɑ]), the tongue body position is low, and for vowels characterized by low $F1$ values (for example, the vowel [i]), the tongue body position is high. For vowels characterized by intermediate values of $F1$ (for example, the vowel [ɛ]), the tongue body position is neither high nor low.

In the acoustic analysis presented in this paper, five vowels of American English were considered: the high vowel [ɪ], two low vowels [ɑ,æ], and two mid vowels [ɛ,ʌ]. In the vowel system of American English, these vowels are characterized by the feature [− round] and by being monophthongal, while the other vowels are all [+ round] or diphthongized. These vowels were considered in the context of voiced and voiceless stops forming CVC syllables and in the hVd and #Vd syllables.

In this analysis, the vowels considered were represented by the first and second formant frequencies ($F1$ and $F2$) in accord with the tradition that has lasted for many years (Stevens and House, 1963; Lisker, 1984). Reducing the vowel space to the plane characterized by $F1$ and $F2$ has been shown to imply a loss of information for vowel identity. Carlson *et al.* (1970) showed that Swedish vowels can be synthesized by two formants $F1$ and $F2'$, where $F2'$ is a combination of $F2$, $F3$, and $F4$. One should note, however, that $F2$ and higher formants are considered to be related to the front–back distinction and not to the high–low dimension. The vowel representation of the present study is motivated by the primary interest of this analysis in vowel height.

Two main problems, which deserve particular attention, may arise when the dimension of vowel height is represented by $F1$. First, when a vowel is pronounced in several consonantal contexts by the same speaker, its acoustic pattern, represented by the $F1$ and $F2$ values, varies, as shown by Stevens and House (1963). Stevens and House noted that the shift in $F1$ values for vowels produced with consonant environments characterized by different places of articula-

[a)] Present address: Department of Information and Communication (INFOCOM), Faculty of Engeineering, University of Rome La Sapienza, Via Eudossiana, 18, 00184 Rome, Italy.

 0001-4966/89/070055-12$00.80

tion is smaller than that found in $F2$ values. However, in order to obtain a nonambiguous specification of vowels in the $F1$ dimension, these shifts should be small enough to allow two different vowel areas, which are contiguous in the $F1$ dimension, not to overlap. Second, markedly different $F1$ and $F2$ values can correspond to the same vowel when this vowel is pronounced by different speakers.

This paper focuses on the problem of nonambiguous specification of vowels in the $F1$ dimension. The results of the acoustical analysis obtained will be compared with previous studies on the acoustic properties of vowels. In particular, we examine the formant displacement of the vowels in all consonantal contexts from the ideal "target" configuration (obtained when V is considered in the hVd or in the #Vd syllables) and attempt to interpret the results obtained either on the basis of acoustic centralization (Delattre, 1969) or on the basis of contextual assimilation (Lindblom, 1963; Stevens and House, 1963). In fact, these two different hypotheses have been proposed by these investigators to account for vowel reduction. An attempt is then made to specify temporal and spectral properties of $F1$, which could be hypothesized to be related to vowel height. The perceptual verification of these hypotheses is the focus of the Di Benedetto (1989) companion paper.

I. EXPERIMENTAL CONDITIONS AND PROCEDURES

A. Speech material

Three of the vowels considered are front vowels [ɪ,ɛ,æ] and two back vowels [ɑ,ʌ]. The vowel [ɪ] is characterized by the feature [+ high], [ɑ,æ] by the feature [+ low], and [ɛ,ʌ] are [− high, − low]. The vowels [ɪ,ɛ,ʌ] are lax vowels, while [ɑ] is tense. The vowel [æ] can be either tense or lax in some dialects, as observed by Halle (1977) and pointed out by Huang (1985). The vowel pairs [ɑ,ʌ] and [æ,ɛ] have been used in studies and perceptual experiments as tense/lax pairs (Huang, 1985).

The vowels under study were considered in the context of voiced and voiceless stop consonants ([b,d,g,p,t,k]), forming CVC syllables, pronounced in the sentence frame "The __ again." All the possible combinations between the five vowels and the six consonants listed above were considered, with the exclusion of nonsymmetrical contexts with respect to voicing. In this way, CVC syllables such as bVg or kVt were included in this analysis, while CVC syllables such as dVp or tVb were excluded. In addition, hVd and #Vd syllables were analyzed. In fact, it is in these contexts that vowels have been assumed in other studies (Peterson and Barney, 1952; Stevens and House, 1963) to be minimally disturbed by coarticulatory effects. It has been possible, then, to compare the acoustic patterns of the vowels under study in stop consonantal contexts with those in the hVd and #Vd syllables.

B. Speakers and recording conditions

Three native speakers of American English, one female and two males, who were phonetically trained, uttered the speech materials described above. The first male speaker (KS) is originally from Canada, but he has been living in Cambridge, MA, for many years. The second male speaker (JP) is from New York and has been living in Cambridge, MA, since 1979. The female speaker (CR) is also from New York. She has been living in Cambridge, MA, since 1983.

The speakers were asked to pronounce the sentences carefully and clearly. If a mistake occurred, the sentence was repeated. The sentences were pronounced in a random order, using the following procedure. The CVC syllables were written in phonetic symbols on cards, one on each card, which were then shuffled. The three speakers knew phonetic symbols and they uttered the sentences reading, from the card, the appropriate CVC syllable. This procedure was repeated three times. Thus three tokens of each vowel in each consonantal context were available. A record of the sentence orders was kept after each repetition.

The speech materials were recorded in a sound-treated room using high-quality equipment. The distance between the microphone and the speaker's mouth was about 20 cm. The recorded materials were then evaluated by a phonetically sophisticated listener; all the syllables were judged to be good samples of the phonemes considered. In fact, there was in all cases a coincidence between what the listener assumed the intended syllable to be and the speaker's intended syllables.

The speech signal was then stored on the MIT-Speech VAX-750. For this purpose, it was first low-pass filtered at 4.8 kHz and then sampled at 10 kHz. The low-pass filter used was a TTE model J97E 5-kΩ passive low-pass antialiasing filter. The A/D conversion was obtained by means of an AD-11k 8-channel (differential) 12-bit, ± 5-V A/D converter.

C. Measurement procedures

The speech materials were analyzed using a software program KLSPEC developed by Dennis Klatt on the Speech-VAX and described extensively by Klatt (1984). This program allows visualization of a 512-point DFT transform of slices of the signal (predifferenced and premultiplied by a Hamming window) and the corresponding time waveform on the screen of a VT 125 terminal. The duration of the Hamming window was 30 ms at the sampling rate considered. In addition, fundamental frequency was determined and displayed in those cases in which local spectral maxima occurred with regularity. The estimation of fundamental frequency was obtained by collecting frequencies of local maxima occurring below 3000 Hz and judging the $F0$ to be that frequency which accounted for most peaks as harmonics.

The program KLSPEC also calculates and displays a smoothed wideband spectrum. This pseudospectrum is obtained by windowing a slice of signal (for example, 256 samples, which correspond to 25.6 ms at a sampling rate of 10 kHz) and computing a 256-point DFT. An approximation to the filter set used in a broadband spectrogram display is then obtained by forming a weighted sum of adjacent DFT sample energies for each of the 128 spectrogramlike filters.

The use of the pseudospectrum is of interest in the estimation of formant frequency positions. In fact, local maxima in this spectrum are most often indicative of the frequency positions of the formants. An interpolation algorithm im-

proves the accuracy over the 40-Hz resolution implied by a 128-sample spectrum over 5 kHz. The program provides a display of the location of the prominent spectral peaks. Figure 1(a) shows the DFT spectrum of the vowel [ɑ] (speaker KS), considering a portion of signal located in the central part of this vowel. Figure 1(b) shows the display of the pseudospectrum of the same speech segment considered in Fig. 1(a).

The possibility of computing the linear prediction coding (LPC) spectrum was also available. Figure 1(c) shows the LPC spectrum display when the same speech segment considered in Fig. 1(a) and 1(b) was analyzed (the number of LPC coefficients was 14).

The smoothed wideband spectrum was used for the estimation of the formant frequencies of the vowels under analysis. In some cases, in which formant tracking gave results that were particularly doubtful and where this algorithm was not successful, the formant frequencies were manually extracted. This happened mainly in the analysis of low vowels, especially of the female speaker, when the two first formants hide under a single peak in the pseudospectrum. DFT spectrum slices sampled every 5 ms were plotted, and the frequency positions of the formants were evaluated by visual examination of the evolution of the locations of the DFT spectrum peaks in time.

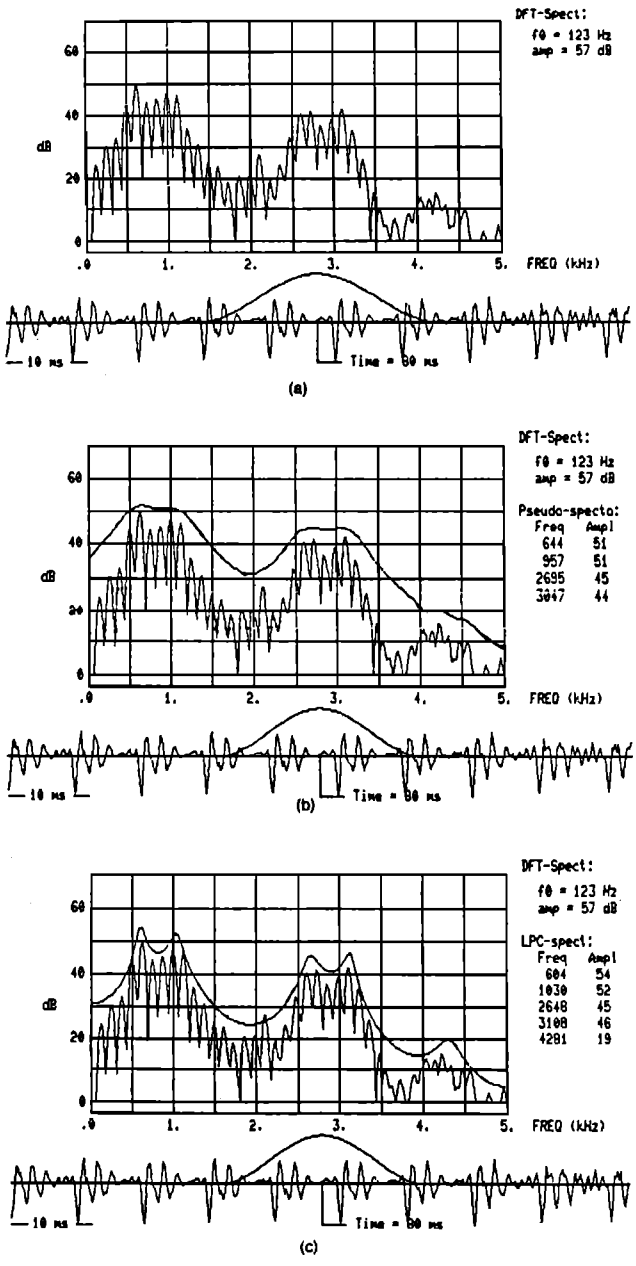

FIG. 1. (a) DFT, (b) spectrogramlike, and (c) LPC magnitude spectrum of the vowel [ɑ] (speaker KS, obtained by using the program KLSPEC (Klatt, 1984).

The LPC spectrum method was not adopted in this analysis, due to its numerous limitations. However, for the sake of completeness and due to the extensive use of this analysis method, the $F1$ and $F2$ values obtained with the pseudospectrum and with the LPC spectrum were systematically compared. The $F1$ values, obtained with the LPC spectrum method, were usually lower than those observed with the pseudospectrum method, while the $F2$ values, obtained with the two methods, were similar. Quantitatively, a typical example (the vowel [æ], speaker KS) showed that the average difference in $F1$ values between the two methods (obtained by averaging the differences in $F1$ values obtained with the two methods for the vowel in each consonantal context) was 74.6 Hz and in $F2$ values was 4.2 Hz, and that the standard deviation for $F1$ values was 7.4 Hz and for $F2$ values was 16.8 Hz. In addition, the [æ] areas displayed in the $F1$ vs $F2$ space, obtained with the two methods, had similar shape and similar "orientation," indicating a similar degree of correlation between the $F1$ and $F2$ parameters. In the example in Fig. 1, it can be noticed that a difference of 40 Hz was obtained for the $F1$ value using these two analysis methods. In this case, a noticeable difference of about 90 Hz was also found in the $F2$ value. This example shows how the location of the peaks in the pseudospectrum is different from that in the LPC spectrum, due to the different way of considering the harmonics in the spectrum.

In addition, to ascertain the reliability of the manual extraction of the formants and of the pseudospectrum, formant frequencies obtained with the two methods for all vowels and speakers were compared. The average difference of $F1$ values, obtained with the pseudospectrum and the manual extraction, was 14 Hz, and of $F2$ values was 10 Hz (higher for the pseudospectrum values). The standard deviation for $F1$ values was 26 Hz and for $F2$ values was 13 Hz. For example, in the case of [æ] for speaker JP, the average difference of $F1$ values, obtained with the pseudospectrum and by manual extraction, was 8 Hz (higher for the pseudospectrum values) and of $F2$ values was 0 Hz. The standard deviation for $F1$ values was 22 Hz and for $F2$ values was 7.7 Hz. The [æ] areas obtained with the two techniques were similar in shape and had similar orientation.

D. Temporal sampling point of the $F1$ and $F2$ trajectories

In previous work on acoustic analysis of American English vowels, vowels were characterized by the values of $F1$ and $F2$ sampled at one or more instants of time in the vocalic portion. Different ways of choosing these sampling points were employed in the various studies, and there is a lack of uniformity in the procedures adopted. For example, in a

study by Lisker (1984), the sampling point corresponded to the time at which $F1$ reached its maximum. In a study by Stevens and House (1963), vowels were represented by the $F1$ and $F2$ values sampled in three contiguous points, selected midway between the beginning and the end of the vowel. In a study of Swedish vowels by Lindblom (1963), vowels were represented by the values of the first three formants at the time at which the first derivative of the corresponding formant curves was equal to zero.

In the present study, the trajectories were sampled at the time at which $F1$ is maximum. A motivation for this choice was the shape of the $F1$ trajectory. As pointed out by Stevens *et al.* (1966), the shape of the first formant curve for the labial, alveolar, and velar consonantal contexts should be characterized by relatively low values for the initial and final $F1$ frequencies and a maximum at some point between these two boundaries. This trajectory shape is consistent with predictions of acoustical theory. In the present study, the predictions of the acoustical theory on the $F1$ trajectory shape were verified. In almost all cases, the $F1$ values at the beginning and at the end of the vowel were lower than in any other point in the vowel; the maximum of $F1$ was reached, varying from case to case, more toward the beginning or more toward the end of the vowel.

The shape of the $F1$ trajectory, which, as noticed above, in the case of the present analysis, was verified to be concave upward, resulted in a maximum of $F1$ in all vowels and consonantal contexts. In addition, this event specifies a time that is not dependent, as it would be if the sampling time were the middle point of the vowel, on the determination of the onset and offset of the vowel, and consequently on measurements of its duration.

It is of interest to point out that, in several cases, the maxima of $F1$, $F2$, and $F3$ occurred at different instants of time. This made it impossible to consider as a sampling point the time at which the first derivative of the formant curves was equal to zero, as proposed by Lindblom (1963).

II. RESULTS OF ACOUSTIC MEASUREMENTS

A. Formant frequencies for each speaker individually

The results of the analysis for speakers KS, JP, and CR are presented in Figs. 2, 3, and 4, respectively. These figures show the location of the vowel areas, for all tokens, in the $F1$ vs $F2$ space. These areas represent the regular and convex polyhedra, which include all the F patterns of each vowel. This way of representing the vowel areas is somewhat unorthodox, as it is more common to represent vowel areas by the ellipses of equiprobability of each vowel. The reason for preferring such a representation is justified by the fact that, in describing the vowel areas in terms of statistical parameters, some important details may be obscured. The regular and convex polyhedra should then be considered in this view as a convenient schematic way of representing the data while preserving critical details. Detailed results on the location of the vowel areas for the vowels [ɪ,ɛ,æ,ɑ,ʌ], for each speaker, can be found in Di Benedetto (1987).

Figures 2–4 show that overlapping occurred in the $F1$ dimension between vowel areas that were contiguous in the

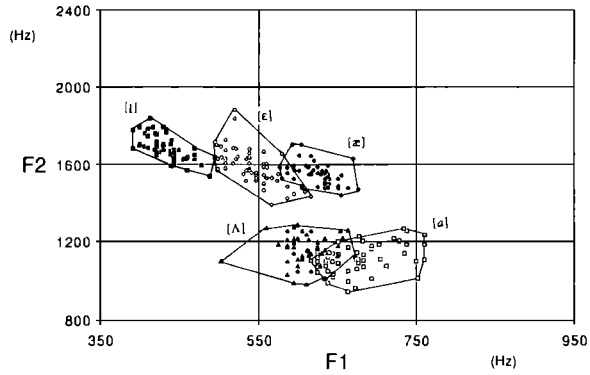

FIG. 2. Results of the analysis in the $F1$ vs $F2$ space of the vowels [ɪ,ɛ,æ,ɑ,ʌ] (speaker KS) for the three tokens. Each vowel is considered in 20 different consonantal contexts. The vowel areas, represented by the regular and convex polyhedron, which included all the F patterns of that particular vowel, are also shown: ■ = [ɪ], ◇ = [ɛ], ◆ = [æ], □ = [ɑ], ▲ = [ʌ].

$F1$ dimension. One can observe, for speaker KS, the little overlap between [ɪ] and [ɛ], and the more relevant overlap between [ɛ] and [æ], and [ɑ] and [ʌ]. For speaker JP, problems of overlapping occurred between [ɪ] and [ɛ], [ɛ] and [æ], and [ɑ] and [ʌ]. In the case of the female speaker CR, problems occurred between [ɑ] and [ʌ]. The three front vowels [ɪ,ɛ,æ] are well separated. It should be noted that, for the three speakers, no overlapping occurred between vowel areas of front and back vowels in the $F2$ dimension.

Figure 5 shows on the same plot the location of the vowel areas of the three speakers. The comparison of the results obtained for the three speakers shows that a large overlap was found in the $F1$ dimension between different vowel areas of different speakers. The $F1$ values of the front vowel [ɪ] were similar for the three speakers, while the $F1$ difference for [ɛ] and [æ] among the speakers was related to the degree of vowel height; the difference was greater for [æ] than for [ɛ]. A similar but less significant effect was found for the two back vowels [ɑ,ʌ]; there was a greater difference between the $F1$ values of the speakers for [ɑ] than for [ʌ]. The

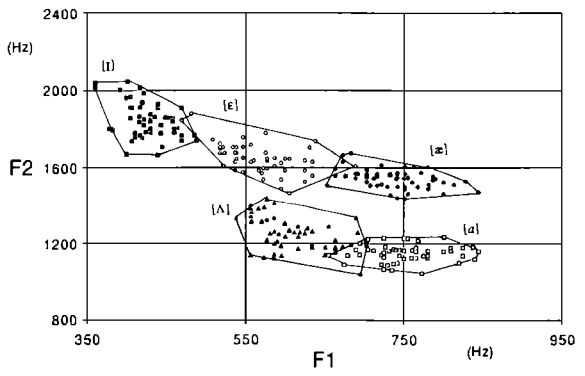

FIG. 3. Results of the analysis in the $F1$ vs $F2$ space of the vowels [ɪ,ɛ,æ,ɑ,ʌ] (speaker JP) for the three tokens. Each vowel is considered in 20 different consonantal contexts. The vowel areas, represented by the regular and convex polyhedron, which included all the F patterns of that particular vowel, are also shown: ■ = [ɪ], ◇ = [ɛ], ◆ = [æ], □ = [ɑ], ▲ = [ʌ].

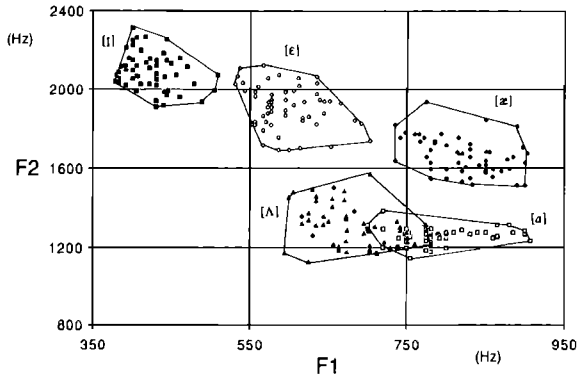

FIG. 4. Results of the analysis in the $F1$ vs $F2$ space of the vowels [ɪ,ɛ,æ,ɑ,ʌ] (speaker CR) for the three tokens. Each vowel is considered in 20 different consonantal contexts. The vowel areas, represented by the regular and convex polyhedron, which included all the F patterns of that particular vowel, are also shown: ■ = [ɪ], ◇ = [ɛ], ◆ = [æ], □ = [ɑ], ▲ = [ʌ].

$F2$ values were higher for the front vowels of the female speaker CR than for the front vowels of the two male speakers KS and JP, while the $F2$ values of the back vowels were similar for the three speakers. This was most probably due to the fact that, for front vowels, $F2$ is primarily affiliated with the back cavity while, for back vowels, $F2$ is affiliated with the front cavity, and the biggest vocal tract difference between men and women is the pharyngeal cavity. Mainly, no overlap was found between the vowel areas in the $F2$ dimension, except for the small overlap between the back vowel [ʌ] of CR, the front vowel [ɛ] of JP, and the front vowel [æ] of KS, and for the small overlap between [ʌ] of JP and [ɛ] of KS.

In order to quantify the amount of overlapping between vowels contiguous in the $F1$ dimension, for each speaker, a linear discriminant analysis was carried out on each vowel of the pairs [ɪ]–[ɛ], [ɛ]–[æ], and [ɑ]–[ʌ]. This analysis allows one to determine the straight lines, in the $F1$ vs $F2$ plane, which best separate the sets represented by the $F1$–$F2$ values of [ɪ]–[ɛ], [ɛ]–[æ], and [ɑ]–[ʌ], under the hypotheses that the statistical distribution of the measurements is Gaussian and that the covariance matrix is similar for the measurements of the vowels in each pair. The Mahalanobis distance was also computed. The Mahalanobis distance is a generalized Euclidean distance that can give an indication of the distance between two sets by considering the distance between the mean values and the spreading of the two sets. The Mahalanobis distance is obtained by dividing the distance between the mean values by the amount of spreading in each set, and corresponds to a dimensionless parameter; for similar values of the distance between the means, when the spreading of the areas increases, the Mahalanobis distance decreases. The results of our analysis showed that the hypotheses were well verified. The classification rates and Mahalanobis distance values obtained, presented in Table I, give an indication on the degree of accuracy obtained by the $F1$ vs $F2$ representation and are in agreement with the previous observations made on the vowel areas of Figs. 2–4.

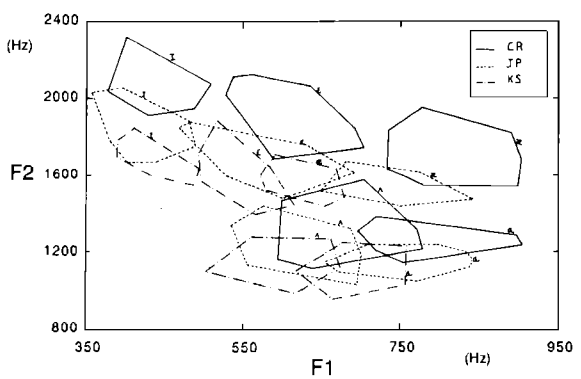

FIG. 5. Vowel areas of the three speakers in the $F1$ vs $F2$ space. Each vowel area is represented by the regular and convex polyhedron, which included all the F patterns of that particular vowel. As indicated, different line patterns correspond to different speakers.

The formant displacement of the vowels in all consonantal contexts from the formant pattern of the vowel in the [#–d] and in the [h–d] contexts was then examined. Two different hypotheses have been proposed to account for the displacement of the vowels from the ideal target configuration (vowel reduction): centralization and contextual assimilation. The first hypothesis describes vowel reduction acoustically in terms of centralization (Delattre, 1969). According to this hypothesis, regardless of contextual assimilation, a vowel, when not pronounced in isolation, tends to degenerate into a neutral vowel (schwa). The second hypothesis assumes that contextual assimilation could account for vowel reduction (Lindblom, 1963; Stevens and House, 1963). It was hypothesized that to each vowel corresponded an ideal articulatory configuration represented acoustically by an ideal formant pattern, which could be satisfactorily represented by the acoustic pattern of the vowel in isolation or in the [h–d] consonantal context. When, on the contrary, a vowel was pronounced in a different consonantal context, the maneuver implied (from a consonantal configuration to a vowel configuration to a consonantal configuration) might cause a displacement from the ideal target configuration, and undershoot might occur.

The results of the present study are summarized in Fig. 6. This figure shows the value of $F1$ and $F2$, for each vowel, averaged over all the contexts under study and tokens. These values (labeled with the letter C on the figure) can be compared with the ideal target patterns, represented by the $F1$

TABLE I. Classification rates and Mahalanobis distance values, for the three vowel pairs [ɪ]–[ɛ], [ɛ]–[æ], and [ɑ]–[ʌ], obtained with the $F1$ and $F2$ values, sampled at the time where $F1$ reaches its maximum, for the three speakers.

		Vowel pairs					
	Speaker	[ɪ]	[ɛ]	[ɛ]	[æ]	[ɑ]	[ʌ]
KS	Classification rate %	98	100	94	96	92	68
	Mahalanobis distance	20		14		3	
JP	Classification rate %	100	96	94	98	92	98
	Mahalanobis distance	13		14		10	
CR	Classification rate %	100	100	100	100	79	90
	Mahalanobis distance	24		27		5	

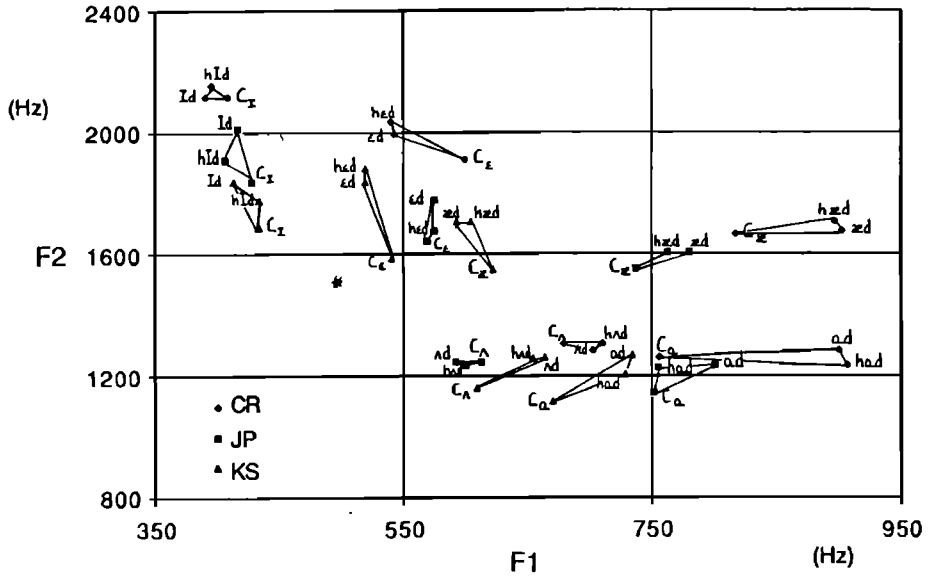

FIG. 6. Results of the analysis in the $F1$ vs $F2$ space showing the formant displacement for vowels in different consonantal contexts from the ideal target configuration for KS, JP, and CR. As indicated, different dots patterns correspond to different speakers. The dots labeled C represent the $F1$ and $F2$ values for the vowel averaged over all consonantal contexts (except [h–d] and [#–d]) and all versions.

and $F2$ values in the [h–d] and [#–d] contexts. [Some systematic regularities, due to the postvocalic consonant place of articulation, were observed in Di Benedetto (1987); for all vowels, the labial contexts caused the $F1$ values to be higher than in the other contexts considered, while the $F1$ values in alveolar and velar contexts were almost unchanged].

Figure 6 shows that the results of the present study cannot be accounted for by the first hypothesis of acoustic centering. In fact, for example, the vowels [ʌ] and [ɑ], for speaker KS, had lower $F2$ values, when pronounced in a consonantal context (which was not the [h–d] or the [#–d] context) than in the minimally distrubed context (target). As another example, the $F1$ and $F2$ values of [æ] for KS and [ɛ] for CR tended to be far from those characterizing the center of the vowel chart ($F1 = 500$ Hz, $F2 = 1500$ Hz) in terms of $F1$. The divergencies found from our expectation, on the basis of Delattre's study, may have several justifications; the vowels studied by Delattre were considered in stressed and unstressed form, they belonged to meaningful words, and they appeared in only one consonantal context.

The undershoot hypothesis could account for the results obtained in the present study. First, one should note that the [ɪ] area for the three speakers extended to the right of the hypothetical target. This event was in accord with an undershoot hypothesis; the $F1$ loci for the consonants considered are at lower frequencies than the $F1$ values of the vowels under analysis, and contextual assimilation would always imply a downward shift of the $F1$ values of vowels from the target values to the $F1$ loci of the consonants. However, except in the case of velar contexts, discounting the undershoot in the articulators that form the constriction (the undershoot provoked by the fast movements of the tongue tip and the lips can be neglected here), if the tongue body position was too low, $F1$ could assume higher values than the hypothetical $F1$ target. One should note, in addition, that the $F1$ values for the vowel [ɛ] (speaker CR) and the vowels [ɛ,æ] (speaker KS) were higher than the target values. It could be hypothesized that the tongue body position was too low, and that this effect could be introduced in order to make the consonant clearer.

In Stevens and House's study, the nonhigh vowels [ɛ,æ] had the highest $F1$ and $F2$ values in the hypothetical target. However, the consonantal contexts considered by Stevens and House included the ones used in the present analysis, but formed a larger set. In addition, only symmetrical syllables were considered, and the target values were represented by an average between the $F1$ and $F2$ values of the vowel in the [h–d] context and of the vowel pronounced in isolation. This could lead to divergencies, especially in the case of lax vowels, which are difficult to pronounce in isloation. In fact, we have observed that, for Stevens and House's data, if the vowel [ɛ] were considered only in stop consonantal context, and the target value were represented by $F1$ and $F2$ for [ɛ] pronounced in the [h–d] context, then the [ɛ] area for one of the speakers would extend to a region characterized by higher values of $F1$ than the target.

B. Temporal and spectral properties of *F*1

In the preceding paragraph, it has been shown how the location of the vowel areas, with respect to a hypothetical ideal target, can be justified on the basis of an undershoot hypothesis, and it has been argued that the undershoot hypothesis could account for the results obtained also in the case of vowel areas that extend to higher values than the target values.

However, this hypothesis, as $F1$ loci for consonant configurations are always at low frequencies, does not explain why the articulatory system should pass through a configuration characterized by a value of $F1$ typical of the hypothetical target to reach another value of $F1$ that is higher than the ideal one. Even though this effect can be explained on the basis of an undershoot hypothesis, in some cases there does not seem to be any particular evidence in the behavior of the system for a tendency to try to reach some hypothetical ideal target configuration.

The concept of an ideal target configuration that one would tend to approach, at least in terms of $F1$, may be misleading. Instead of considering the vowel as *minimally* or *maximally* disturbed by coarticulatory effects, one may want to consider the vowel as affected *differently* in different context. In fact, when a displacement in $F1$ values, with respect to an hypothetical ideal target, is observed, there may be some different or additional cues to a single value of $F1$ that could supply the listener with the information necessary to identify the vowel.

In particular, it may be possible that this additional information is contained in properties associated with temporal and/or spectral variations of $F1$. In order to verify this hypothesis, an analysis of the $F1$ trajectories of the vowels under study was carried out. Particular attention was given to two points. The first regarded the study of temporal properties of $F1$; the correlation between the $F1$ maximum values and a parameter that represents the $F1$ onglide duration in relation to the total vowel duration was examined. This parameter will be called $F1$ timing. The second point was related to the spectral changes of $F1$; the correlation between the $F1$ maximum and the $F1$ onset values (of a particular vowel for a particular speaker) was examined. It should be noted, however, that the two effects observed regarding temporal versus spectral changes of $F1$ may not be independent. In fact, although the relation between the $F1$ maximum and $F1$ onset values is specifically connected to spectral changes of $F1$, it may be also indirectly related to $F1$ temporal properties and vice versa. For example, a larger difference between the $F1$ onset and the $F1$ maximum value could imply a larger $F1$ onglide duration and, consequently, a larger $F1$ timing value. This possible relation between $F1$ temporal and spectral properties will be considered in Sec. II B 3.

1. Temporal properties of F1

Before proceeding to the description of the results obtained, one should note that two vowel pairs analyzed in this study ([ɛ,æ] and [ɑ,ʌ]) are tense/lax pairs (House, 1961, considered [æ] as tense, but mentioned that this classification was open to question); the vowels of these pairs are considered to differ in the dimension representing vowel height, but they also differ in duration, due to the tense/lax distinction. From both an acoustic and a perceptual point of view, properties based on measurements of the time-varying spectrum have been shown to be related to the tense feature (Lehiste and Peterson, 1961; Huang, 1985). In the case of the two tense/lax pairs of this study, it may then be difficult to disambiguate the time-varying properties related to the tense feature and those related to the low feature. The pair [ɪ,ɛ] is the only one, in the present study, consisting of two lax vowels that differ only by the high feature. The results will show whether the same behavior is observed in the tense/lax pairs and in the lax/lax pair analyzed.

As a first observation on $F1$ temporal properties, consider, for example, Fig. 7, which shows the details of the results of the analysis in the $F1$ vs $F2$ space for [ɪ] and [ɛ] of speaker KS. In this figure, each dot corresponds to a label representing the consonantal context in which the vowel having that particular $F1$–$F2$ pattern was spoken. As shown, the contexts in which [ɪ] assumed the highest $F1$ values and [ɛ] the lowest $F1$ values (i.e., the cases in which the [ɪ] area overlaps with the [ɛ] area) were the [b–b], [t–p], and [k–p] contexts for [ɪ] and the [d–d], [g–g], [b–g], and [d–g] contexts for [ɛ]. Figures 8 and 9 present the $F1$ trajectories for these particular cases. The examination of the $F1$ trajec-

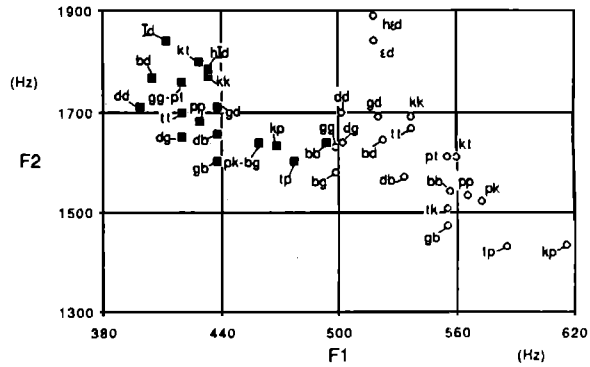

FIG. 7. Results of the analysis in the $F1$ vs $F2$ space of the vowels [ɪ] and [ɛ] (speaker KS) for the second token. The vowels are considered in 20 different consonantal contexts: ■ = [ɪ], ◇ = [ɛ].

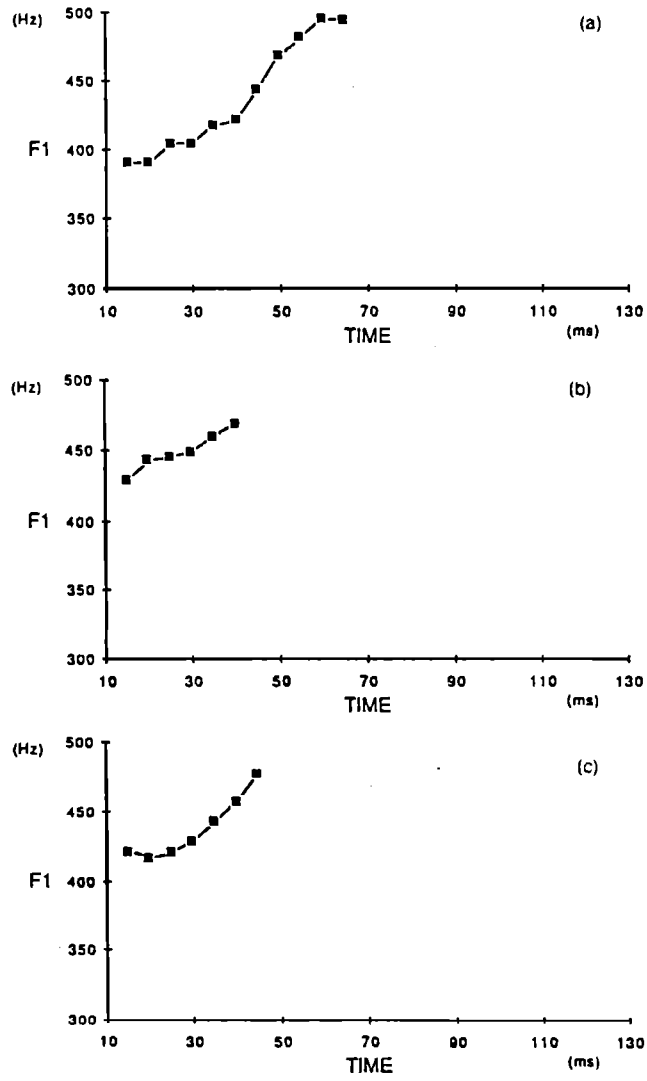

FIG. 8. The $F1$ trajectories of the vowel [ɪ] in the syllable (a) bIb, (b) kIp, and (c) tIp, for speaker KS for the second token.

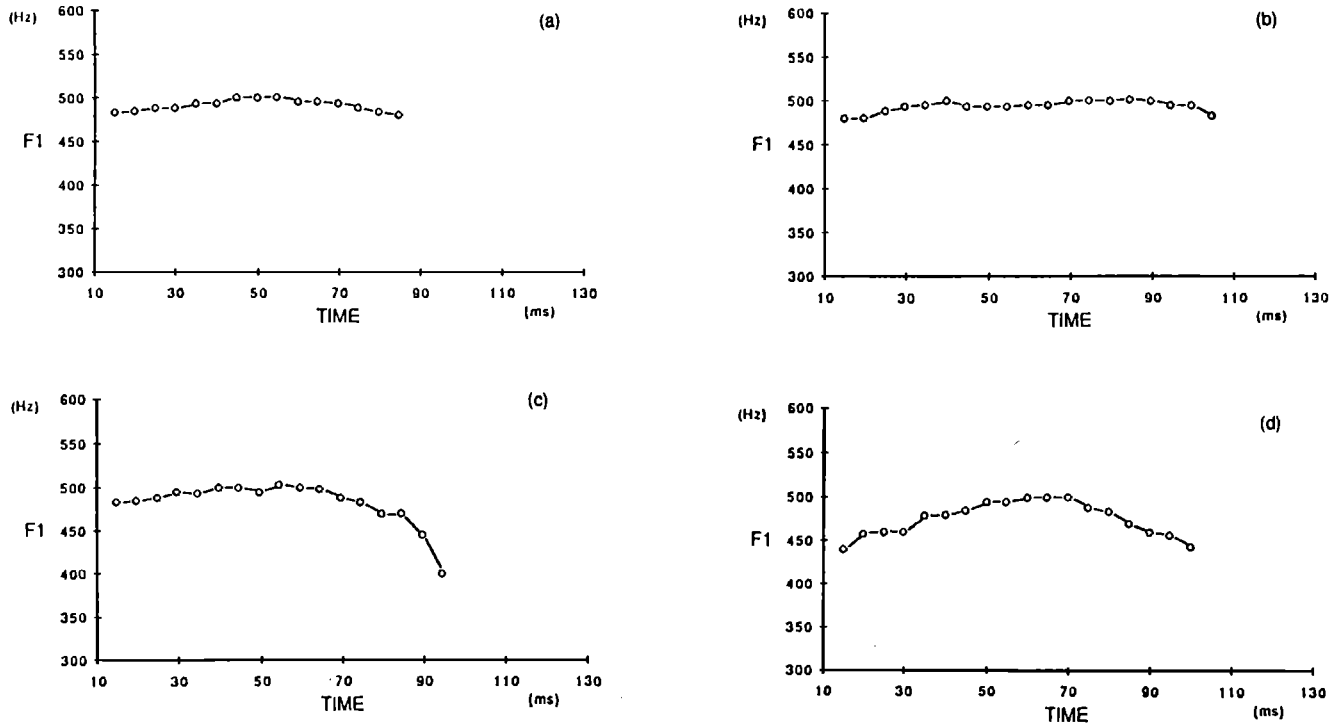

FIG. 9. The $F1$ trajectories of the vowel [ɛ] in the syllable (a) bɛg, (b) dɛd, (c) dɛg, and (d) gɛg, for speaker KS for the second token.

tories of these vowels for which overlap occurred shows that, in the case of the high vowel [ɪ], the maximum of $F1$ was reached towards the end of the vocalic portion, whereas, in the case of the lower vowel [ɛ], the maximum of $F1$ was reached earlier. In particular, for the vowel [ɛ], $F1$ reached its maximum at the beginning of the vocalic portion within approximately 40–50 ms after the vowel onset (except for cases with a very small additional increase after 40–50 ms). Note that, in these examples, the vowel [ɪ] occurred in voiceless consonantal contexts, except for the [b–b] context, and, in these cases, had a shorter duration than [ɛ].

The $F1$ timing values were systematically obtained for each speaker by dividing, for each vowel in each consonantal context, the $F1$ onglide duration by the total vowel duration. The linear fit of the $F1$ timing values and the correlation coefficient R between the $F1$ maximum and the $F1$ timing values, were found. The correlation coefficient R can acquire values that fall between 0 and 1. Here, $R = 1$ means that the points in the $F1$ maximum versus $F1$ timing representation were well fitted by a straight line and that the $F1$ timing increased with the $F1$ maximum; this is equivalent to saying that, when $F1$ was in the high-frequency range, the $F1$ maximum was reached later in the vocalic portion than when it was in the low-frequency range. Also, $R = 0$ means that the points could not be well fit by a straight line and that it was not possible to assert that the $F1$ timing increased with the $F1$ maximum. Intermediate R values represent situations that fall in between these two extreme cases. Table II shows the R values obtained for each speaker and each vowel. For all speakers, R was significantly higher for [ɪ] and [ɛ] (minimum value = 0.53, maximum value = 0.71) than for [æ], [ɑ], and [ʌ] (minimum value = 0.03, maximum value = 0.27). Figure 10 shows the plots in the $F1$ maximum versus $F1$ timing space for speaker KS for the three vowel pairs [ɪ,ɛ], [ɛ,æ], and [ɑ,ʌ] [Fig. 10(a), (b), and (c), respectively]. These results are representative of the ones obtained for the three speakers. It should be noted that, confirming the R values, when $F1$ was in the high-frequency range, the $F1$ maximum for [ɪ] and [ɛ] was reached later in the vocalic portion than in the low-frequency range. For the other vowels, such a systematic effect was not revealed. It should be noted, however, that, in these cases, when $F1$ was in the high-frequency range (for [æ] in the last third of its frequency range, for [ʌ] in the last fifth, and for [ɑ] in the two last thirds), the $F1$ timing parameter never assumed low values ($F1$ timing values > 0.4, approximately). On the contrary, when $F1$ was in the low-frequency range (complementary sets with respect to the ones mentioned above), no systematic $F1$ timing behavior was present.

The analysis described above showed, for [ɪ] and [ɛ], the existence of a direct relation (well represented by a straight line) between the $F1$ maximum values and the $F1$ trajectory differences ($F1$ timing values). In the cases of [ɪ]

TABLE II. Values of the correlation coefficient R, obtained for the three speakers by finding a linear fit to the $F1$ maximum versus $F1$ timing values for [ɪ], [ɛ], [æ], [ɑ], and [ʌ].

Speaker	R value [ɪ]	[ɛ]	[æ]	[ɑ]	[ʌ]
KS	0.62	0.53	0.11	0.03	0.27
JP	0.71	0.7	0.2	0.1	0.22
CR	0.56	0.53	0.22	0.11	0.15

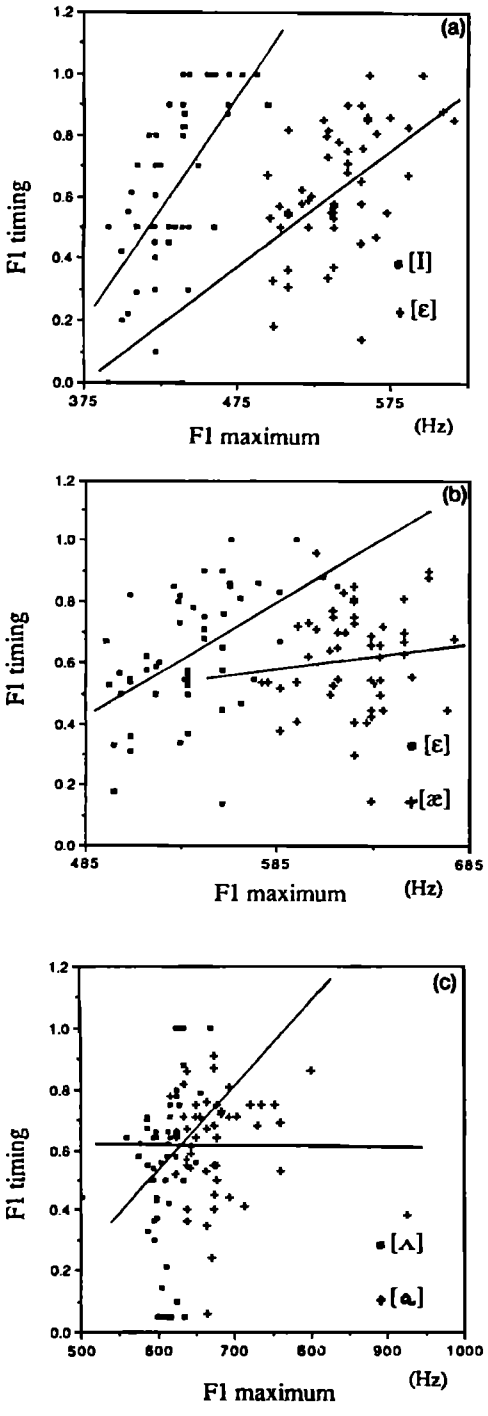

FIG. 10. The $F1$ maximum versus $F1$ timing values for the three vowel pairs [ɪ]–[ɛ], [ɛ]–[æ], and [ɑ]–[ʌ] [parts (a)–(c), respectively] for speaker KS.

and [ɛ], an investigation of whether the trajectory differences varied systematically, with postvocalic consonant place of articulation, was carried out. In the case of [ɛ], a very systematic effect was found. In fact for the three speakers, an increase of the $F1$ maximum corresponded to a different postvocalic consonant place of articulation, according to the following order: voiced velar [g], voiced alveolar [d], voiced labial [b], voiceless alveolar [t], voiceless velar [k], voiceless labial [p]. In the case of [ɪ], for the three speakers, the highest $F1$ maximum values corresponded to postvocalic labials and the lowest values to voiced velars. For intermediate $F1$ maximum values, the effect was not as systematic as for [ɛ]. These results indicate the existence of a relation between $F1$ maximum values, the postvocalic place of articulation, and the $F1$ timing values, suggesting anticipatory coarticulation effects on vowels dependent on postovcalic consonant place of articulation.

Finally, it was interesting to examine whether, in the $F1$ maximum versus $F1$ timing space, the vowels [ɪ] and [ɛ] could be well discriminated. If this were the case, one could not only assert that the $F1$ timing increased with the $F1$ maximum, but also that, at the [ɪ]–[ɛ] boundary, the $F1$ timing values were different. For this purpose, a linear discriminant analysis was carried out. The results of this analysis showed that the two sets corresponding to [ɪ] and [ɛ] could be very well separated by a straight line. In fact, for the three speakers, there was no ambiguity between the two vowels analyzed (100% of correct classification). The Mahalanobis distance was also computed. Results of this analysis led to the following values: 23 for KS, 20 for JP, and 31 for CR. The comparison of these classification rates and Mahalanobis values with the ones obtained in the analysis of the $F1$–$F2$ values (Table I) shows that [ɪ] and [ɛ] could be better discriminated in the $F1$ maximum versus $F1$ timing plane than in the $F1$ vs $F2$ plane.

2. Spectral properties of F1

The relation between the $F1$ maximum and the $F1$ onset values of each vowel, for each speaker, was analyzed. For this purpose, first the $F1$ maximum and the $F1$ onset values were found, and then a linear fit of the values obtained in the $F1$ maximum versus $F1$ onset values space was obtained. As before, the correlation coefficient R gave a valuable indication of whether the linear fit was a satisfactory approximation to the results. The R values found for each vowel and speaker are listed in Table III. These values show that, for the three speakers, for each vowel, the $F1$ onset increased with the $F1$ maximum value. Figure 11 shows the $F1$ maximum versus the $F1$ onset values for speaker KS for the pair [ɪ]–[ɛ] [Fig. 11(a)], [ɛ]–[æ] [Fig. 11(b)], and [ɑ]–[ʌ] [Fig. 11(c)]. These results were representative of what was obtained for the three speakers. It can be seen in Fig. 11 that, in the $F1$ maximum versus $F1$ onset plane, the two sets of measurements were well distinct for [ɪ] and [ɛ], while there was some overlapping between [ɛ] and [æ] and [ɑ] and [ʌ]. The same effect was found for the two other speakers, showing that, in the region in which the vowels of the pairs assumed the same $F1$ maximum, the $F1$ onset value was dis-

TABLE III. Values of the correlation coefficient R, obtained for the three speakers by finding a linear fit to the $F1$ maximum versus $F1$ onset values for [ɪ], [ɛ], [æ], [ɑ], and [ʌ].

	R value				
Speaker	[ɪ]	[ɛ]	[æ]	[ɑ]	[ʌ]
KS	0.25	0.69	0.55	0.48	0.64
JP	0.44	0.77	0.68	0.79	0.6
CR	0.28	0.7	0.65	0.54	0.7

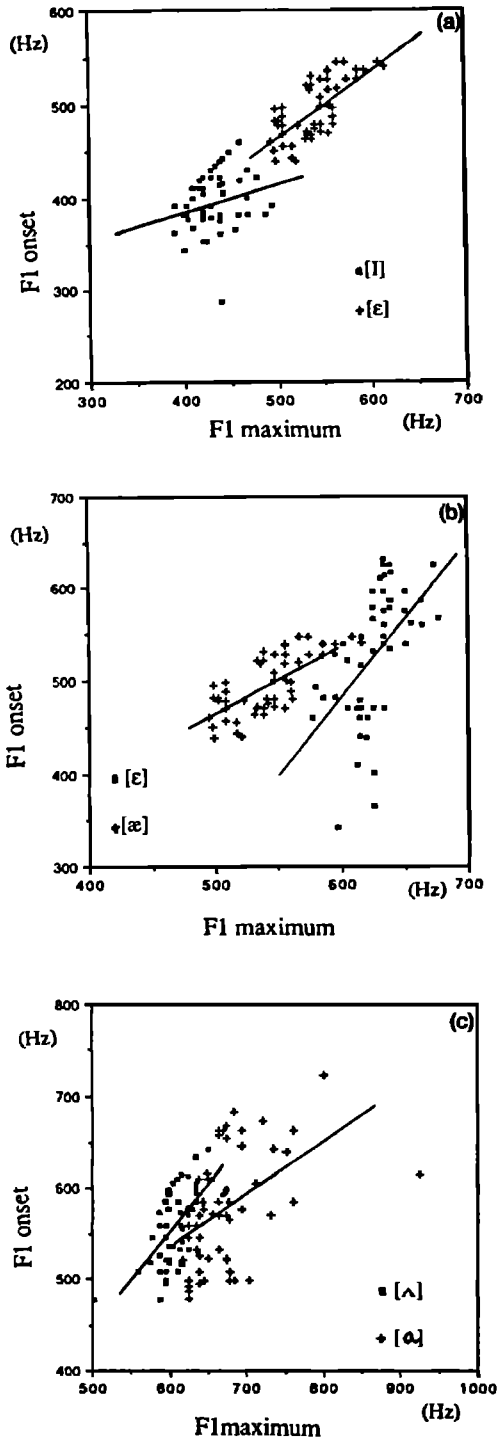

FIG. 11. The $F1$ maximum versus $F1$ onset values for the three vowels pairs [ɪ]–[ɛ], [ɛ]–[æ], and [ɑ]–[ʌ] [parts (a)–(c), respectively] for speaker KS.

tinctive at the boundary between [ɪ] and [ɛ], but not at the boundary between the vowels of the other two pairs. A linear discriminant analysis was carried out to support quantitatively this observation. The classification rates and the Mahalanobis distance values, for all vowels and speakers, are presented in Table IV. It should be noted that the $F1$ onset value could disambiguate [ɪ] and [ɛ] in all cases, while the [ɛ]–[æ] boundary was not as sharp, as is also shown by the decrease in the Mahalanobis distance values. The classification rate and the Mahalanobis distance values were not satisfactory for the [ɑ,ʌ] pair. The comparison of these classification rates and the Mahalanobis values with the ones obtained in the analysis of the $F1$–$F2$ values (Table I) shows that [ɪ] and [ɛ] could be perfectly discriminated in the $F1$ maximum versus $F1$ onset plane, while this was not the case in the $F1$ vs $F2$ plane. For the [ɑ]–[ʌ] pair, the results were similar in the two analyses both in terms of classification rates and in terms of Mahalanobis values. For the [ɛ]–[æ] pair, the results of the $F1$–$F2$ analysis were more satisfactory, except for a slight improvement in the opposite direction of the Mahalanobis distance for speaker JP.

TABLE IV. Classification rates and Mahalanobis distance values, for the three vowel pairs [ɪ]–[ɛ], [ɛ]–[æ], and [ɑ]–[ʌ], obtained by using as variables the $F1$ maximum and $F1$ onset values, for the three speakers.

		Vowel pairs					
	Speaker	[ɪ]	[ɛ]	[ɛ]	[æ]	[ɑ]	[ʌ]
KS	Classification rate %	100	100	90	98	96	77
	Mahalanobis distance	23		11		3	
JP	Classification rate %	100	100	94	100	98	87
	Mahalanobis distance	20		16		9	
CR	Classification rate %	100	100	95	98	95	80
	Mahalanobis distance	24		15		6	

3. Relation between temporal and spectral properties of F1

The results presented in the two preceding paragraphs have shown that both spectral and temporal properties of $F1$ could disambiguate [ɪ] and [ɛ]. In this paragraph, the relation between the temporal and the spectral properties observed will be considered.

These two factors, as already mentioned, may not be independent. It was observed that, for [ɪ] and [ɛ], an increase in the $F1$ maximum values corresponded systematically to both an increase in the $F1$ onset and the $F1$ timing values. Let us find in which proportion: Let us consider the ratio between the $F1$ onglide duration and the difference between the $F1$ maximum and the $F1$ onset values. This ratio, which will be called the $F1$ speed, represents the slope of the $F1$ onglide portion of the trajectory. If the $F1$ speed increased for increasing $F1$ maxima, for both [ɪ] and [ɛ], and if the $F1$ speed were different for these two vowels at their boundary in the $F1$ dimension, it would then be possible to hypothesize that the temporal and spectral properties observed contributed to the variation of the same parameter, i.e., the $F1$ speed; the analysis could then be interpreted on the basis of the same phenomenon, mainly connected to the $F1$ time-varying properties.

The analysis consisted of determining the $F1$ speed values for all vowels and speakers and in finding the correlation coefficient R between the $F1$ speed and the $F1$ maximum values. Our results showed that R was very low in all cases (minimum value = 0.02, maximum value = 0.28). In addition, it was observed that, at the boundary between [ɪ] and [ɛ], the $F1$ speed values were not distinctive for the two

vowels. This effect was verified on the plots in Fig. 12, which shows the $F1$ speed values (on a logarithmic scale) versus $F1$ maximum values for the three speakers [Fig. 12(a) for KS, (b) for JP, and (c) for CR]. The linear fits of the values are not shown in Fig. 12, since the R values were low in all cases. It was concluded that the temporal and spectral properties observed did not contribute to the variation of a single parameter, which is represented by the $F1$ speed.

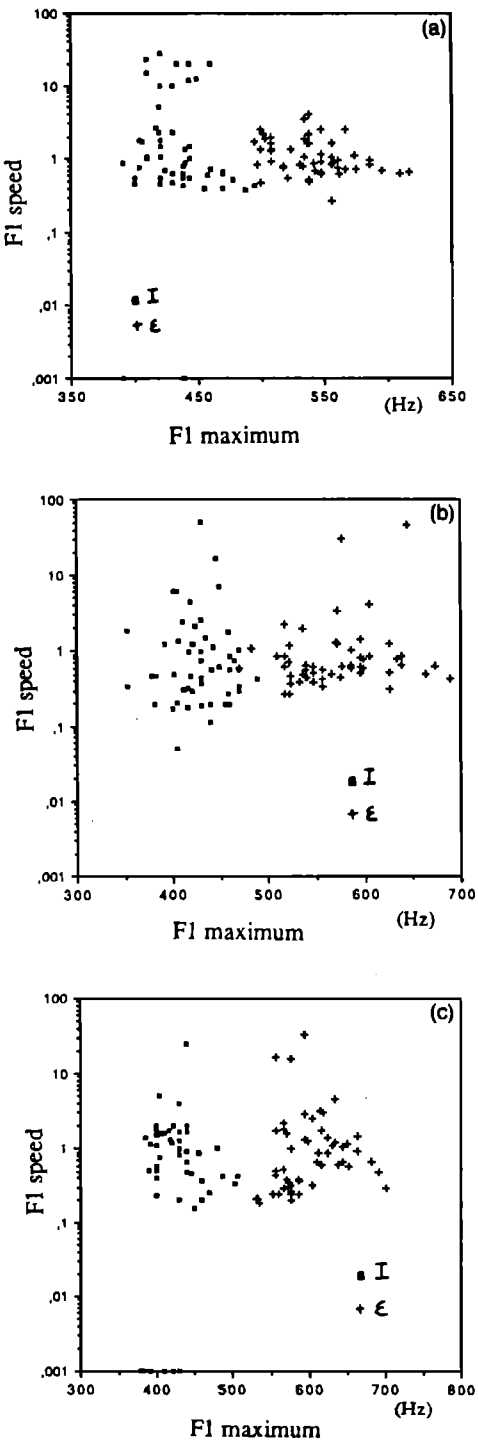

FIG. 12. The $F1$ maximum value versus $F1$ speed values (on a logarithmic scale), in the case of vowel pair [ɪ]–[ɛ] for (a) KS, (b) JP, and (c) CR.

III. CONCLUSION

Results of the acoustic analysis in the $F1$ vs $F2$ space of the five vowels of American English [ɪ,ɛ,æ,ʌ,ɑ] have shown that overlapping occurred between vowel areas in the $F1$ dimension when the formants were sampled at the time where $F1$ reached its maximum, even when measurements on vowels pronounced by a single speaker were considered. For a single speaker, the front vowels [ɪ,ɛ,æ] were well separated from the back vowels [ɑ,ʌ]. These results showed that $F1$ could not accurately classify vowels along a dimension height if, over the whole $F1$ trajectory, one considered only the $F1$ value at the time where $F1$ reaches its maximum. The location of the vowel areas, with respect to an hypothetical ideal target, could be justified on the basis of an undershoot hypothesis, as proposed by Stevens and House (1963). The results showed, however, that in the behavior of the system there did not seem to be, in some cases, any particular evidence for a tendency to try to reach some particular target configuration.

The hypothesis was made that, since different consonantal contexts affect vowels differently due to coarticulatory effects, in the search for invariant attributes of vowels one should take into account properties associated with temporal and spectral variations of $F1$. The study of time-varying properties of $F1$ showed that the way in which $F1$ reaches its maximum may be relevant for vowel identification. In particular, if two different vowels, such as [ɪ] and [ɛ], were characterized by the same $F1$ maximum, $F1$ reached its maximum earlier in [ɛ] than in [ɪ]. It was observed that, for [ɪ] and [ɛ], the onglide duration, relative to the total vowel duration, increased with increasing $F1$ maximum values. This effect was shown to be related to systematic changes in the postvocalic consonant place of articulation; the trajectory differences observed could be explained by general articulatory phenomena and, in particular, by anticipatory coarticulatory effects on vowels depending on the postvocalic place of articulation. Finally, it was noticed that [ɪ] and [ɛ] could be very well discriminated (better than by the $F1$ and $F2$ values) by the relative onglide duration parameter and the $F1$ maximum for the three speakers. The same effects were not found for the [æ]–[ɛ] and the [ɑ]–[ʌ] pairs. A possible explanation of this finding was that the vowels of these pairs differ not only by the degree of height, but also by tense/lax characteristics, such that acoustic properties of the $F1$ trajectory related with the tense/lax distinction and properties related with vowel height may be confounded in these pairs.

The examination of the $F1$ trajectories showed, in addition, that vowels contiguous in the $F1$ dimension could be also discriminated by the $F1$ onset frequency. Systematic analysis of the relation between $F1$ maxima and $F1$ onset values was carried out. Results led to the conclusion that the $F1$ onset frequency could perfectly discriminate [ɪ] and [ɛ]. Results of a linear discriminant analysis did not indicate such a satisfactory classification for the vowels in the [ɛ]–[æ] and in the [ʌ]–[ɑ] pairs. For the [ɑ]–[ʌ] pair, the classification rates and the Mahalanobis distances were similar to the ones obtained with the $F1$–$F2$ values, and for the [æ]–[ɛ] pair these parameters indicated a better discrimi-

nation by the $F1$ and $F2$ values. Finally, an additional investigation showed that it was not possible to determine whether the temporal and spectral properties observed contributed to the variation of a parameter directly associated with time-varying properties of $F1$. The interpretration was that properties associated with temporal and spectral variations may both be important in discriminating vowels along a dimension of height.

The perceputal relevance of the properties observed is investigated in Di Benedetto (1989), in which it is shown that perceptual experiments with listeners of different native languages lead to the hypothesis that these properties may have either an articulatory or an auditory basis.

The properties observed in the present study appear in agreement with the results of previous investigations on the perception of coarticulated vowels (Strange *et al.*, 1976; Strange and Gottfried, 1980; Gottfried and Strange, 1980; Strange *et al.*, 1983; Rakerd *et al.*, 1984; Verbugge and Rakerd, 1986). These investigations have focused on a dynamic specification theory of vowel perception, which could account for the perceptual constancy over variations of the acoustic patterns due to coarticulatory effects. The results of these studies showed that vowel perception may be aided by consonantal context (vowels were considered in the context of stop consonants). The agreement of the results of the present analysis with the results of the studies mentioned above, in the light of the perceptual data, is highlighted in Di Benedetto (1989).

ACKNOWLEDGMENTS

I am grateful to Paolo Mandarini for his generous support and advice, and I wish to express my deep gratitude to Ken Stevens for his unique help, criticism, and guide throughout this study.

Carlson, R., Granström, B., and Fant, C. G. M. (**1970**). " Some studies concerning perception of isolated vowels," R. Inst. Technol. Stockholm, Q. Prog. Stat. Rep. 2–3.

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

Delattre, P. (**1969**). "The general phonetic characteristics of languages," final report, University of California (Santa Barbara), chapter entitled "An acoustic and articulatory study of vowel reduction in four languages," 33–80.

Di Benedetto, M. G. (**1987**), "An acoustical and perceptual study on vowel height," Ph.D. thesis, University of Rome, Italy.

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

Gottfried, T. L., and Strange, W. (**1980**). "Identification of coarticulated vowels," J. Acoust. Soc. Am. **68**, 1626–1635.

Halle, M. (**1977**). "Tenseness, vowel shift and the phonology of the back vowel in modern English," Ling. Inq. **8**(4), 611–625.

House, A. A. (**1961**). "On vowel duration in English," J. Acoust. Soc. Am. **33**, 1174–1178.

Huang, C. B. (**1985**). "Perceptual correlates of the tense/lax distinction in general American English," Master's thesis, MIT, Cambridge, MA.

Klatt, D. H. (**1984**). "M.I.T. Speech VAX user's guide," preliminary version.

Lehiste, I., and Peterson, G. E. (**1961**). "Transitions, glides, and diphtongs," J. Acoust. Soc. Am. **67**, 268–277.

Lindblom, B. (**1963**). " On vowel reduction," R. Inst. Technol. Stockholm, Speech Transmission Lab. Rep. No. 29.

Lisker, L. (**1984**). "On reconciling monophtongal vowel percepts and continuously varying F patterns," Haskins Lab., Stat. Rep. Speech Res. **SR-79/80**.

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

Rakerd, B., Verbrugge, R. R., and Shankweiler, D. P. (**1984**). "Monitoring for vowels in isolation and in consonantal context," J. Acoust. Soc. Am. **76**, 27–31.

Stevens, K. N., and Blumstein, S. R. (**1981**). "The search for invariant acoustic correlates of phonetic features," in *Perspectives on the Study of Speech*, edited by P. Eimas and J. Miller (Erlbaum, Hillsdale, NJ).

Stevens, K. N., and House, A. S. (**1955**). "Development of a quantitative description of a vowel articulation," J. Acoust. Soc. Am. **27**, 484–493.

Stevens, K. N., and House, A. S. (**1963**). "Perturbation of vowel articulations by consonantal context: An acoustical study," J. Speech Hear. Res. **6**(2), 111–128.

Stevens, K. N., House, A. S., and Paul, A. P. (**1966**). "Acoustical description of syllabic nuclei: an interpretation in terms of a dynamic model of articulation," J. Speech Hear. Res. **40**(1), 123–132.

Strange, W., Verbrugge, R. R., Shankweiler, D. P., and Edman, T. R. (**1976**). "Consonant environment specifies vowel identity," J. Acoust. Soc. Am. **60**, 213–222.

Strange, W., and Gottfried, T. L. (**1980**). "Task variables in the study of vowel perception," J. Acoust. Soc. Am. **68**, 1622–1625.

Strange, W., Jenkins, J. J., and Johnson, T. L. (**1983**). "Dynamic specification of coarticulated vowels," J. Acoust. Soc. Am. **74**, 695–705.

Verbrugge, R. R., and Rakerd, B. (**1986**). "Evidence of talker-independent information for vowels," Lang. Speech **29**, Part 1, 39–57.