VOWELS PRONUNCIATION ASSESSMENT IN THE SPELL SYSTEM

Maria-Gabriella Di Benedetto *, Fabrizio Carraro **, Steven M. Hiller ***, Edmund Rooney ***.

* INFOCOM Dept, Facoltà di Ingegneria, University of Rome 'La Sapienza', Via Eudossiana 18, 00184 Rome, Italy.

** Alcatel Face, Research Center, Via Generale Clark, 21, 84100 Salerno, Italy.

*** CSTR, University of Edinburgh, 80 South Bridge, Edinbutgh, Scotland.

ABSTRACT

The Spell research on the phonetic segmental aspects of speech is focused on the class of vowels. The aim of this aspect of the Spell project is to succeed in indicating to a given speaker speaking a non-native language (Italian, French, or English (American English and British English RP)) how to improve the pronunciation of vowels belonging to the non-native vowel system. In order to automatically provide the information needed to improve the pronunciation, a vowel representation capable of guiding the speaker must be selected. Two different vowel representation strategies, based on two different normalization methods, were implemented. An analysis of the efficiency of these procedures within the Spell system in the case of the Italian vowel system is reported.

I. INTRODUCTION

The Spell research on the phonetic segmental aspects of speech is focused on the segmental class of vowels. The aim of this aspect of the Spell project is to find the way to indicate to a given speaker speaking a non-native language how to improve the pronunciation of vowels belonging to the non-native vowel system.

Vowel pronunciation defects may be classified into two groups. First, the same vowel in two different languages may have different qualities. Secondly, a vowel may not belong to the speaker's native vowel system, and be most often substituted by a different vowel belonging to the speaker's native vowel system. The substituting vowel will be usually selected by the speaker as the one which sounds to the speaker similar, if not in some cases identical, to the right one.

In order to provide the speaker with the information necessary to improve vowel pronunciation, it is necessary to identify the changes in the acoustic waveform which correspond to the pronunciation errors. If the vowels under investigation can be characterized by few acoustic parameters, such as the first and second formant frequencies (F1 and F2), the vowel to be pronounced can be indicated to the speaker as a region in the parameter space. However, some sort of normalization of the data must be applied since the information provided should indicate to the speaker the correct location of the vowel if that speaker (and not a reference one) was pronouncing it correctly .

The present paper is organized as follows: section II consists in a description of the vowel systems of Italian, French, and English (American English and British English RP), which are the languages considered within the Spell project. In section III, a tentative description of the most common vowel pronunciation errors in the Spell languages will be given. Section IV will deal with the problem of vowel representation. On the basis of an analysis on the Italian vowels, two normalization methods will be compared, and discussed in section V.

II. CHARACTERIZATION OF VOWELS IN TERMS OF DISTINCTIVE FEATURES

A classification of vowels belonging to the Italian, French, and English (American English and British English RP) vowel systems will be given in this section. Vowels will be grouped in different classes which can be identified by different features.

Vowels are characterized by at least three features corresponding to specific properties which can be described as follows:

1. High/low: relatively to a cross-section of the vocal tract, it indicates the tongue height during vowel production.
2. Front/back, relatively to the tongue position with respect to an axis which goes from the lips to the epiglottis.
3. Tense/lax: this class characterizes the degree of tension of the tongue pressing against the palate.

By means of these features, vowels in the Italian, English (American English and British English RP), and French vowel systems can be represented as follows ([1], [2], [3]):

Italian vowel system

[i]	[+high]	[+front]	[+tense]	
[e]	[-high]	[-low]	[+front]	[+tense]
[ɛ]	[-high]	[-low]	[+front]	[-tense]
[u]	[+high]	[+back]	[+tense]	
[o]	[-high]	[-low]	[+back]	[+tense]
[ɔ]	[-high]	[-low]	[+back]	[-tense]
[a]	[+low]	[-front]	[-back]	[+tense]

American English vowel system

[i]	[+high]	[+front]	[+tense]	
[I]	[+high]	[+front]	[+lax]	
[ɛ]	[-high]	[-low]	[+front]	[+lax]
[æ]	[+low]	[+front]	[+tense]	
[u]	[+high]	[+back]	[+tense]	
[U]	[+high]	[+back]	[+lax]	
[o]	[-high]	[-low]	[+back]	[+tense]
[ɔ]	[-high]	[-low]	[+back]	[+lax]
[ʌ]	[-high]	[-low]	[+back]	[+lax]
[ɑ]	[+low]	[+back]	[+tense]	

British English vowel system

[i]	[+high]	[+front]	[+tense]		
[I]	[+high]	[+front]	[+lax]		
[ɛ]	[-high]	[-low]	[+front]	[+lax]	
[æ]	[+low]	[+front]	[+tense]		
[u]	[+high]	[+back]	[+tense]		
[U]	[+high]	[+back]	[+lax]		
[ɔ]	[-high]	[-low]	[+back]		
[ɒ]	[+low]	[+back]	[-tense]		
[ʌ]	[-high]	[-low]	[+back]	[+lax]	
[ɑ]	[+low]	[+back]	[+tense]		
[ɜ]	[-high]	[-low]	[-front]	[-back]	[+tense]
[ə]	[-high]	[-low]	[-front]	[-back]	[-tense]

The Spell project is supported by the European Community's ESPRIT program, under contract no.5192.

French vowel system

The vowels [œ] and [ɑ] were not considered in the French vowel system, since they tended to be substituted by [e] and [a] respectively, and disappeared in most pronunciations.

[i]	[+high]	[+front]	[+tense]	
[e]	[-high]	[-low]	[+front]	[+tense]
[ɛ]	[-high]	[-low]	[+front]	[+lax]
[y]	[+high]	[+front]	[+tense]	
[a]	[+low]	[-front]	[-back]	[+tense]
[ø]	[-high]	[-low]	[+front]	[+tense]
[œ]	[-high]	[-low]	[+front]	[+lax]
[u]	[+high]	[+back]	[+tense]	
[o]	[-high]	[-low]	[+back]	[+tense]
[ɔ]	[-high]	[-low]	[+back]	[+lax]
[ɛ̃]	[+high]	[+front]		
[ã]	[+low]	[-front]	[-back]	
[õ]	[-high]	[-low]	[+back]	

According to this description, the Italian vowels are univocally described, while the American English and French vowels are not; two additional features, associated with rounding and nasalisation, must be introduced. In American English, [o] is distinguished from [ʌ] since [o] is [+rounded] and [ʌ] is [-rounded]. Similarly, in British English, [ɑ] is distinguished from [o] because [ɑ] is [-rounded] and [o] is [+rounded] (note that these two vowels are also distinguished by tenseness). In French, [e,ø], [ɛ,œ] are distinguished since the first vowel in the pair is [-rounded] and the second vowel is [+rounded]. In the pairs [e,ɛ̃], [a,ã], [ɔ,ɔ̃], the first vowel is [-nasal] and the second vowel is [+nasal].

III.IDENTIFICATION OF ERRORS IN VOWEL PRONUNCIATION

A tentative description of the possible errors in vowel pronunciation can be derived from the classification reported in the previous section. Pronunciation errors correspond to changes in either the quality or the identity of a vowel. The first case may occur for vowels which are known by the speaker because belonging to the native vowel system, but having in the non-native language a different quality. Differences in the acoustic realization and associated perceived quality of vowels which are assigned the same phonetic symbol have been reported [4]. The second case occurs when a vowel is not known by the speaker because it does not belong to the speaker's native vowel system. In this latter case, the speaker often tends to substitute the unknown vowel by a known vowel perceived as 'close' to the 'right' one. This correspondence is often made by selecting the native vowel which differs from the non-native one by the least number of features. An analysis of this type of errors will be presented in this paragraph for English, Italian, and French.

American English and British English

From the previous section, it is noted that Italian and French speakers do not know the English vowels [I,æ,ʌ, ɑ,U]. Consequently, these speakers have no experience in pronouncing the tense/lax high vowel pairs [i,I] and [u,U] and will in general pronounce the lax vowel in the pair as the tense vowel in the pair ([I] will be substituted by [i] and [U] will be substituted by [u]). In the pairs [i,I] and [u,U] the distinction seems to be based directly on duration and less on average formant frequency [5]. The mispronunciation of the pairs [i,I] and [u,U] often leads to misunderstanding of words, since there are several minimal pairs based on this distinction (for example: beet-bit). Other tense/lax pairs to be analyzed are [ɛ,æ], [ɑ,ʌ]. There is some controversary on whether these are 'real' tense/lax pairs. Huang [5] showed that for these pairs, the distinction may be based directly on average formant frequency, and indirectly on duration.

In addition, note that Italian and French speakers only know the [-back,-front] [a] vowel as a low vowel. On the contrary, the English vowel system includes two vowels [ɑ,ac] which are [+low] ([ɑ] is [+back] and [æ] is [+front]). Italian and French speakers tend to substitute [ɑ] by [a], and [æ] by [ɛ].

French

The English and Italian speakers do not know the nasal vowels [ɛ̃,ã,õ]. These speakers tend to add the extra nasal [n] after the vowel and to substitute the nasal vowel by the corresponding non-nasal vowel.

The English and Italian speakers also do not know the rounded vowels in the pairs [e,ø], [e,œ], [i,y]. In general, [y] will be substituted by [u]. In the case of [ø,œ] the substitution strategy is less systematic.

In the case of English speakers, the problem of diphthongization must also be considered. These speakers tend to markedly diphthongize most vowels (for example [e] pronounced as [e^y]).

Italian

All the vowels of the Italian vowel system are included in the French vowel system. French speakers will therefore make errors mostly of the first type (same vowel but with different vowel quality).

In the case of English speakers, problems arise with [e,a,o] which tend to be diphtonguized. The vowel [a] tends to be substituted by [ɑ].

IV. VOWEL REPRESENTATION

In order to provide the speaker with the information necessary to improve vowel pronunciation, it is necessary to code vowels in terms of acoustic parameters by which each vowel corresponds to a different acoustic pattern. Since vowels are characterized univocally by the features described in section II, it would be ideal if the acoustic parameters were acoustic correlates of the distinctive phonetic features previously introduced. Commonly, the acoustic parameters considered correspond to formants (for example vowel height is related to F1 values, and vowel backness to F2 values). However, two problems arise. First, two different vowels pronounced in different contexts by the same speaker may have similar acoustic patterns (problem of vowel coarticulation). Secondly, two different vowels pronounced by two different speakers (for example one male and one female speaker) may have similar acoustic patterns (problem of vowel normalization).

Vowel coarticulation provokes a change in the formant values. Consequently a given vowel spoken by the same speaker is characterized by several formants frequencies values. How to represent the vowel by a reference acoustic pattern? A possible way of bypassing this problem, is to consider and teach the pronunciation of vowels included in a predetermined phonetic context in order to have comparable vowels under examination. For example, as regards the tense/lax distinction, the training could be made on minimal pairs (for example in American English: beet-bit, fool-full, pop-pup). A similar training could teach the rounded-non rounded or the nasal-non nasal distinction (for example in French passer-penser).

Another possibility is to teach vowels in minimally disturbed consonantal contexts and suppose that the speaker will be able to extrapolate what learned and correctly pronounce the same vowels in any context. In American-English, the h-d context can be taken into consideration since hVd syllables form meaningful words and were extensively investigated in the past [6]. In Italian and French, most isolated vowels can be taken into consideration since they form meaningful words (with the exception of [ɔ] and [œ] in French).

In this paragraph, the analysis will be focused on isolated Italian vowels. The correspondence between Italian vowels and meaningful words is the following:

vowel	word	meaning	
[i]		i	the (plural)
[e]		e	and
[ɛ]		è	is
[a]		a	to, in, at
[o]		o	or
[ɔ]		ho	have
[u]		uh	exclamation particle

Two normalization methods will be compared. The first method is based on the use of the Bark-transformed formant differences F1-F0 and F2-F1. Acoustic analyses showed that Italian vowels could be well represented by these parameters [7]. This method does not need any a-priori knowledge on the vowels of the speaker under examination. The second method is the isomorphic transformation proposed in [8]. This method is based on a 'vowel mapping' transformation of the vowels of a specific speaker onto an 'average' vowel space of a given language. In this method, a-priori knowledge on the first and second formant values of three vowels of the speaker under investigation, in the native language is needed. The area used by the speaker in absolute terms is identified by acquiring information on the representation of (for example) the cardinal vowels [i,a,u] of the specific speaker.

Both normalization methods were applied to an Italian data-base. Vowels pronounced by 25 male and 11 female speakers were analyzed. These vowels were extracted from a data-base created by Ferrero (1968). Vowel measurements were available from a previous study [7]. For each vowel, the fundamental frequency F0 was computed using an algorithm based on the cepstrum of the signal, and the formants were found by manual comparison of the local maxima in the Fourier transform (FFT) of the signal and the maxima of the autoregressive analysis (AR) of the spectrum. The comparison between the two spectra was necessary since frequently the peaks in the AR spectrum under 2000 Hz were slightly lower than those found from the examination of the FFT, when F0 is high (this happens for most of the female speakers).

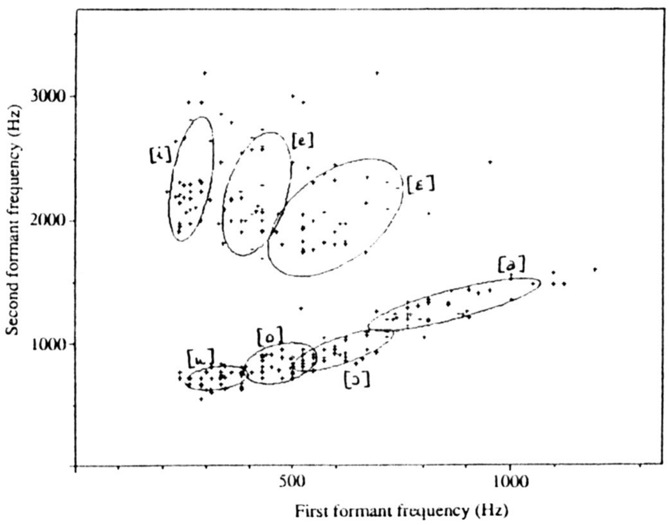

Figure 1. Representation of the Italian vowels and of the ellipses of equiprobability corresponding to P=0.7 in the F_1 vs F_2 plane.

Presentation of the data in the unnormalized F1 vs F2 space is given in Fig.1. Figure 1 also shows, according to a method adopted in [8], the ellipses of equiprobability corresponding to P=0.7. The data transformed by the isomorphic method are presented in Fig.2. The same data in the Bark-transformed formant differences plane (F1-F0) vs (F2-F1) are presented in Fig.3. The end-correction of the Bark scale proposed by Traunmüller [9] was applied.

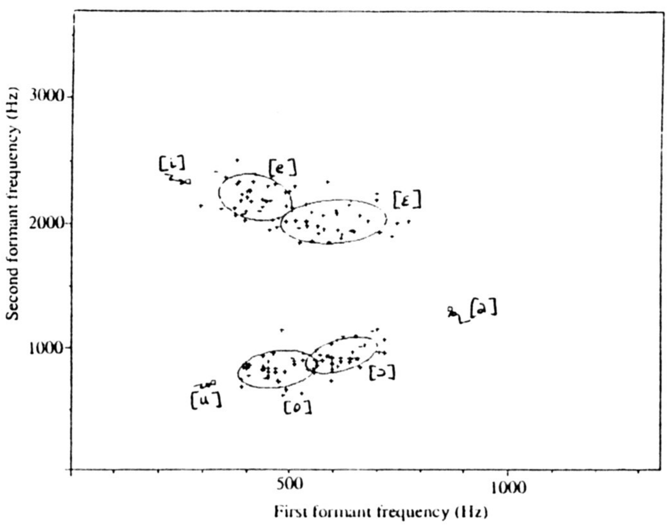

Figure 2. Representation of the Italian vowels and of the ellipses of equiprobability corresponding to P=0.7 in the F_1-normalized vs F_2-normalized plane.

In order to quantify the distances between vowel areas in the three above spaces, the Mahalanobis distance was computed. This distance indicates the actual euclidean distance between the mean vectors of two sets, normalized with respect to the variances and correlation coefficients of the two sets. The Mahalanobis distances between vowel areas in the F1 vs F2 plane, in the F1-normalized vs F2-normalized plane, and in the (F1-F0) vs (F2-F1) plane are reported in Tables I,II, and III, respectively. Tables I, II, and III show that by using the isomorphic transformation as well as the Bark-transformed formant differences, the distances between vowel areas were increased.

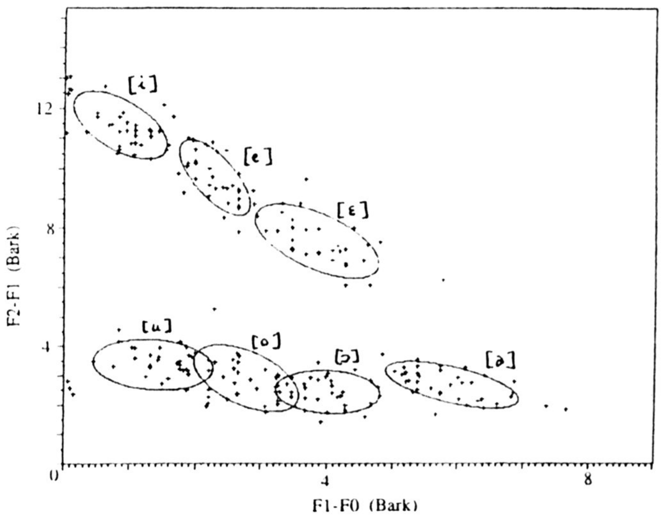

Figure 3.Representation of the Italian vowels and of the ellipses of equiprobability corresponding to P=0.7 in the (F1-F0) vs (F2-F1) plane (F0, F1, and F2 are expressed in Bark, with the end-correction proposed in [9]).

Table I - Mahalanobis distances between vowels in F1 vs F2 plane.

	[e]	[ε]	[a]	[o]	[ɔ]	[u]
[i]	16	31	98	119	85	62
[e]		9	62	64	43	42
[ε]			32	32	27	35
[a]				9	19	37
[o]					5	23
[ɔ]						8

Table II - Mahalanobis distances between vowels in F1-normalized vs F2-normalized plane. The normalization was obtained with respect to the three vowels [i,*a*,u]. Consequently, the distances for these three vowels are not significative and are not reported.

	[e]	[ε]	[a]	[o]	[ɔ]	[u]
[i]	*	*	*	*	*	*
[e]		10	*	162	160	*
[ε]			*	116	126	*
[a]				*	*	*
[o]					7	*
[ɔ]						*

Table III Mahalanobis distances between vowels in (F1-F0) vs (F2-F1) plane. The Bark values were end corrected according to Traunmuller [9]

	[e]	[ε]	[a]	[o]	[ɔ]	[u]
[i]	13	36	195	217	146	170
[e]		11	108	119	100	115
[ε]			53	71	77	90
[a]				14	37	57
[o]					6	25
[ɔ]						7

A linear discriminant analysis was carried out in order to quantify the differences between the representations described above in terms of misclassification rates on three measurement sets: F1-F2 measurements, F1 normalized-F2 normalized measurements (refering to the data after isomorphic transformation), (F1-F0) (F2-F1) measurements.

The results of this analysis are given in terms of misclassification scores. Table IV shows the results in the case of the F1-F2 measurements, Table V shows the results in the case of the F1 normalized-F2 normalized measurements, and Table VI shows the results in the case of (F1-F0) (F2-F1) data.

Averaging the misclassification scores over all vowel pairs (excluding in all cases the values obtained for [i,a,u] since these vowels were used to apply the isomorphic transformation) led to the following scores: 3.2% for the unnormalized values, 2.3% for the isomorphic normalization method, and 2.7% for the Bark-transformed formant differences. Note that each vowel set was composed of 36 tokens and the discriminant analysis is applied to 6 different vowel pairs, leading to 216 classifications.

Table IV - Misclassification scores between vowels in the F1 vs F2 space

	[e]	[ε]	[a]	[o]	[ɔ]	[u]
[i]	0.0	0.0	0.0	0.0	0.0	0.0
[e]		6.9	0.0	0.0	0.0	0.0
[ε]			0.0	0.0	0.0	0.0
[a]				5.6	0.0	0.0
[o]					12.5	0.0
[ɔ]						6.9

Table V - Misclassification scores between vowels in the F1-normalized vs F2-normalized space. The normalization was obtained with respect to the three vowels [i,*a*,u]. Consequently, the scores for these three vowels are not significative and are not reported.

	[e]	[ε]	[a]	[o]	[ɔ]	[u]
[i]	*	*	*	*	*	*
[e]		2.8	*	0.0	0.0	*
[ε]			*	0.0	0.0	*
[a]				*	*	*
[o]					11.1	*
[ɔ]						*

Table VI - Misclassification scores between vowels in the (F1-F0) vs (F2-F1) space.

	[e]	[ε]	[a]	[o]	[ɔ]	[u]
[i]	1.4	0.0	0.0	0.0	0.0	0.0
[e]		5.6	0.0	0.0	0.0	0.0
[ε]			0.0	0.0	0.0	0.0
[a]				1.4	0.0	0.0
[o]					11.1	0.0
[ɔ]						11.1

V. DISCUSSION

In this paper, the aspects of the Spell project which deal with the phonetic segmental aspects of speech have been presented.

In particular, a description of the vowel systems of the Spell languages ((Italian, French, or English (American English and British English RP)) in terms of distinctive features has been proposed. A tentative description of the possible errors in vowel pronunciation in terms of distinctive feature variations has been presented.

The need for normalizing the acoustic parameters characterizing vowels has been emphasized. Two normalization methods were applied to an Italian vowel data-base. The first method was based on the Bark transformed formant differences (F1-F0) and (F2-F1). This method did not need any a-priori knowledge on speaker's characteristics and can be classified as an intrinsic normalization method. The second method was based on the 'vowel mapping' criterion (isomorphic transformation) described in [8]. Results of the analyses showed that by the use of both methods an improvement was obtained in terms of reduction of vowel scattering and increase in the distance between vowel areas. However, as been thouroughly illustrated by Disner [4], when evaluating normalization procedures, particular care should be given to the liguistic validity of the methods. As emphasized by Disner, normalization techniques might be usually effective and appropriate for analyzing a single language, but usually also prove to be totally ineffective for comparing different languages. This aspect rises the question on whether the techniques investigated in the present paper are capable of not destroying the linguistic differences between vowels of two languages. Both normalization methods seem to satisfy this property. As regards the Bark-transformed formant differences method, this procedure has the privilege of avoiding the need for a-priori information on the speaker. However, although effective, this method provides the speaker with the information on average values of vowel patterns. The isomorphic transformation requires a-priori input data on three vowels of the speaker under test. Nevertheless, this method presents the appealing property of finding the rule which governs the location of vowels of a given speaker with respect to a reference vowel system. Consequently, when comparing vowels of two different languages, the isomorphic transformation can first find the rule governing the location of a given speaker with respect to a reference vowel system in the speaker's native language. Secondly, the isomorphic transformation can find the rule governing the location of the reference native vowel system with respect to a reference non-native vowel system. Consequently, the isomorphic transformation is capable, given that speaker S is located with reference to system S1 (native vowel system) according to a formalized rule, of indicating how speaker S is located with reference to system S2 (non-native vowel system). As example, suppose that Mario (an Italian speaker) was to learn American-English. Given the [i,a,u] Italian vowels of Mario, the isomorphic transformation would indicate to Mario where would his American-English vowel [æ] be located if he correctly pronounces it.

REFERENCES

[1] Z. Muljacic. *Fonologia della lingua italiana*, Società editrice il Mulino, Bologna, 1972.

[2] K.N. Stevens *Acoustic Phonetics*, (in press).

[3] J.S. Liénard. personal communication, 1991.

[4] S.F. Disner. "Evaluation of vowel normalization procedures", *J.Ac.Soc.Am.* 67(1), pp.253-261, 1980.

[5] C.B. Huang. "Perceptual correlates of the tense/lax distinction in general American English". Master's thesis, Massachusetts Institute of Technology, 1985.

[6] G.E. Peterson and H.L. Barney. "Control methods used in a study of the vowels", *J.Ac.Soc.Am.* 24(2), pp.175-184,1952.

[7] M.G Di Benedetto and G. Flammia. "Vowel distinction along auditory dimensions: a comparison between a statistical and a neural classifier", Verba 90, Rome, 1990.

[8] J.S. Liénard and M.G. Di Benedetto. "Extrinsic normalization of vowel formant values based on cardinal vowels mapping", ICSLP, Banff, Alberta, Canada, 1992.

[9] H. Traunmüller. "Perceptual dimension of vowel openness in vowels", *J.Ac.Soc.Am.* 68(5), pp.1465-1475, 1981.