Discriminatory Capacity of the Most Representative Phonemes in Spanish: An Evaluation for Forensic Voice Comparison

Fernanda López-Escobedo[1](✉) and Luis Alberto Pineda Cortés[2]

[1] Licenciatura en Ciencia Forense, Facultad de Medicina, UNAM,
Circuito de la Investigación Científica SN, CU, 04510 Ciudad de Mexico, Mexico
flopeze@unam.mx

[2] Departamento de Ciencias de la Computación, IIMAS,
UNAM, Circuito Escolar SN, CU, 04510 Ciudad de Mexico, Mexico
www.fernandalopez-escobedo.com

Abstract. In this paper, a study of the discriminatory capacity of the most representative segments for forensic speaker comparison in Mexican Spanish is presented. The study is based on two corpora in order to assess the discriminatory capacity of the fundamental frequency and the three first vocalic formants acoustic parameters for reading and semi-spontaneous speech. We found that the context /sa/ has 73% of discriminatory capacity to classify speakers using the three first formants of the vowel /a/ with a dynamic analysis. We used several statistical techniques and found that the best methodology for the recognition of patterns consists of using linear regression with a quadratic fitting to reduce the number of predictors to a manageable level and apply discriminant analysis on the reduced set. This result is consistent with previous research data despite the methodology for Mexican Spanish had never been used.

Keywords: Pattern recognition · Forensic speech recognition · Linear Discriminant Analysis · Principal Component Analysis · Linear Regression

1 Introduction

Forensic linguistics is a relative recent development area that has the aim to use the linguistic knowledge in order to support legal cases. One of its applications is the forensic voice comparison that consists of comparing an unknown voice sample (criminal voice) to a set of known voice samples (voices of suspects) and select those which share the greater number of acoustics characteristics.

The development of different tools and methodologies to facilitate and optimize forensic voice comparison has become a fundamental aim in forensic phonetic research. This article explores a methodology to improve the discrimination between speaker voices and examines whether the analysis of formant dynamics can be applied to classify a vocalic segment of Mexican Spanish.

G. Sidorov and O. Herrera-Alcántara (Eds.): MICAI 2016, Part I, LNAI 10061, pp. 127–140, 2017.
DOI: 10.1007/978-3-319-62434-1_11

In the forensic field, it is very common to rely on short recordings for speaker comparison; therefore, it is very helpful to study sounds that occur often in a short period of time. Hence, the selection of the sound used in this study is based on the frequency of occurrence of the Spanish phonemes. Several papers [1–7] documented that the vocalic phoneme /a/ is one of the most frequent phonemes in this language. Other sounds, like diphthongs, have a much lower probability of occurrence. Consequently, the sound /a/ was selected for the present study. In addition, this sound was studied in the most frequent syllabic context [8], which is formed by a consonant followed by a vowel (CV). The consonant phonemes preceding /a/ in this syllabic context are also the most frequent consonants in Spanish (i.e., /s/, /t/, /p/, /l/, /n/). If it turns out that these contexts improve the discriminatory potential of /a/ and present stability across different types of speech then they have a high value in Spanish forensic phonetic studies.

The empirical resource used in the present research consisted of two corpora: one in semi-spontaneous and the other in read speech. Speakers in the two corpora were different. All instances of the vocalic phoneme /a/ were segmented and analyzed. For this, two acoustic parameters were used: (1) the fundamental frequency (F0); and, (2) the three first vocalic formants (F1, F2 and F3). Both parameters were studied with two methods that we named static and dynamic. The first is based on the value of the acoustic parameter at the midpoint of the vowel, and the second uses several values of the parameter along the total duration of the vowel. The dynamic method, which takes into account the three first vocalic formants in nine equidistant points, was used to analyze the Australian English diphthong /a/ by McDougall [9]. The same method was used to study intervocalic /r/ in the sequence /ərV/, where the vowels considered were /iː,æː,ɑː,ɔː,uː/ in British English [10]. In 2008, Eriksson and Sullivan [11] examined whether this methodology could be applied to another segment of the same size constituted by /jœː/ in Swedish. In addition, this latter study considered the value of the fourth vocalic formant. In the present investigation, the parameters considered were the fundamental frequency and the three first formants, and each parameter was analyzed with both the static and the dynamic methods. For the dynamic method, the value of both parameters at nine equidistant points along the total duration of /a/ was used.

In order to assess the discriminatory capacity of /a/ in the referred contexts, three statistic techniques were used: Linear Discriminant Analysis (LDA), Principal Components Analysis (PCA) and Linear Regression. The Linear Discriminant Analysis is a statistical technique whose main objective is to identify the characteristics that differentiate between two or more groups of subjects and most of the time is useful to classify new observations as belonging to one group or another [12]. PCA is a statistical technique that reduces the number of original variables through their linear combination, where the reduced set of variables is called 'components'. Finally, the Linear Regression technique consists of an optimal linear model that represents the distribution of the data [13].

McDougall [9,10] and Eriksson and Sullivan [11] used LDA to determine the discriminatory capacity of the sounds analyzed in their corresponding studies.

In this study, LDA was used to assess the discriminatory capacity of /a/ in all the different contexts of the two corpora, considering the two acoustic parameters and with both the static and the dynamic methods. The higher percentages of classification were obtained with the dynamic method, using the three first formants in the semi-spontaneous speech corpora. In particular, the contexts that best discriminate among speakers were /sa/ and /ta/.

However, LDA yields unstable results when the number of tokens is not above the number of predictors at least by two. For this reason, McDougall [9] proposed an approach to achieve higher percentages of classification with a less number of predictors using different combinations of variables. Later on she proposed a new methodological approach in which the three first formant contours were fitted with a quadratic and a cubic polynomial equation using linear regression, and the parameters of the polynomial fittings were used as predictors in LDA [10]. In this study we also tested McDougall's methodologies, but in addition, we used PCA to reduce the number of variables. In particular, the contexts /sa/ and /ta/ in semi-spontaneous speech corpora were analyzed with all of these methodologies, as these are the best classified contexts.

The overall conclusion of this study is that the contexts /sa/ and /ta/ analyzed with the dynamic method using the three first vocalic formants are frequent enough and have the best discriminatory capacity to classify speakers in semi-spontaneous speech in Mexican Spanish.

2 Empirical Resource

As aforementioned in the introduction, the vocalic segment /a/ was chosen because of its high frequency of occurrence in Spanish. The effect of neighboring sounds was considered so that the /a/ segment was analyzed within a syllabic context formed by CV preceded by the five most frequent consonants in Spanish as shown in Table 1.

Table 1. Occurrence percentage of the five most frequent consonants in Spanish [6]

Non vocalic phonemes	Frequency of ocurrence (%)
/s/	20.99
/t/	10.94
/p/	10.81
/l/	10.39
/n/	10.21

In order to determine whether dynamic measures of vocalic formants and the fundamental frequency can be used as idiosyncratic properties of the speakers' voice independently of the type of speech, the vowel /a/ has been evaluated by making use of two types of corpora: a read speech (RS) corpus and a semi-spontaneous speech (SS) corpus.

The read speech corpus consists of the recorded speech of five male and five female native speakers of Mexican Spanish, 16–36 age range. All ten speakers were born in Mexico City and have completed secondary school. Corpus RS is a subcorpus of corpus DIMEx100 [7] 'designed and collected to support the development of language technologies, especially speech recognition, and also to provide an empirical base for phonetic studies of Mexican Spanish'. It included recordings from 100 Mexican Spanish speakers. The corpus DIMEx100 was collected from the Web. A very large Spanish sentences were extracted and its content was measured according to the perplexity value. The 5,010 sentences with the largest value were selected. Hence, the corpus is complete and phonetically balanced. All 10 speakers in corpus RS read 60 five-to-fifteen word sentences. The list of sentences contains 50 individual sentences for each speaker and 10 equal sentences for all of them. So, corpus RS consists of 510 different sentences. Typical sentences are for instance[1]:

> Ofrecemos la mejor calidad y el mejor precio en tinas de hidromasaje
> Si acaso, de vez en cuando, pasan como tormentas de verano por mis asquerosos pensamientos
> La cuenta la pagaría, por tanto, el último que quedara sentado a la mesa
> Para el mes de enero, las fases de la luna son las siguientes
> En ágora, estamos convencidos de que la participación ciudadana debe ser libre y voluntaria

The recordings were made in a sound studio at CCADET[2] UNAM[3] with an Audigy Platinum ex (24 bit/96khz/100db SNR) sound blaster. The WaveLab 4.0 program with a sampling format of mono at 16 bits and sampling rate of 44.1 Hz was used. Each speaker was recorded with a single diaphragm studio condenser microphone Behringe B-1.

The semi-spontaneous speech corpus consisted of recordings from three native male and two native female speakers of Mexican Spanish. All speakers participating in SS were different from the ones in read speech corpus. The speakers were between 18 and 30 years old, held a higher education diploma and were born in Mexico City. The recordings consisted of sociolinguistic interviews, based on the criteria and techniques proposed by the Labovian tradition of sociolinguistic variation [14,15]. Recordings were made in a silent room with a MiniDisc Sony MZ-R900 and a lapel microphone. Each speaker's interview lasted from 35 to 45 min.

3 Dynamic Method

In the last two decades, forensic phonetic research has shown an interest in the dynamic characteristics of the acoustic signal [9–11,16–19]. McDougall [9] points

[1] The full corpus is available on request to the main author.

[2] Centro de Ciencias Aplicadas y Desarrollo Tecnológico.

[3] Universidad Nacional Autónoma de México.

out that the study of the change undergone by acoustic properties throughout time provides more information about the individual characteristics of a speaker's voice than the measurement of these acoustic features in a single point in time. Previous studies that have considered the dynamic acoustic properties of the signal have been structured around two criteria: length of the analyzed segment and type of alignment. For example, Greisbach, Esser and Weinstock in 1995 [16] compared the discriminatory potential of first and second vocalic formants, both measured in a single point of the segment and in five intervals of equal duration. The analyzed segments were three German vowels ([a:], [e:], [o:]) and three German diphthongs ([ae:], [ao:], [oe:]). Rose in 1999 [17] undertook a study of the word hello produced by six Australian English speakers, by extracting the four first vocalic formants from seven intervals and defining their length in terms of the middle of the first vowel, the middle of the /l/ phoneme and five intervals of equal duration in the /oʊ/ diphthong. It is the criterion of how alignments are made that establishes the main difference between both studies. According with Greisbach, Esser and Weinstock, the alignment is normalized, since the segment is divided into five intervals of equal duration and five frequency measures of each interval are extracted. By contrast, Rose delimits the duration of the interval based on each phoneme's duration; therefore, the alignment is not normalized. In addition, the consideration of the length of the segment establishes another difference between these two studies. Thus, while Greisbach, Esser and Weinstock studies a short segment constituted by a vowel, Rose analyzes a longer stretch of speech. In the latter case, the analyses can be based on a word [17] or on the whole recording [19].

McDougall [9,10] and Eriksson and Sullivan [11] analyzed a short speech segment with a time-normalized alignment. The methodology proposed by McDougall was used to analyze the Australian English diphthong /ai/ and to extract the three first vocalic formants at nine equidistant points. Each of these points was employed as a predictor variable in a Linear Discriminant Analysis (LDA). This analysis allows the classification of each speaker's tokens, drawing from the linear functions based on the information contained in the set of predictor variables. Furthermore, the number of predictors was reduced, since LDA yields unstable results when the number of tokens is not above the number of predictors at least by two. The criterion followed by McDougall was to disregard the variables that display the smallest F-ratio in an ANOVA and to test different combinations of predictor variables in order to determine which one discriminates the speakers better. The best result showed 95% of correct classification versus the 68% obtained with the predictor variables corresponding to the midpoint of the diphthong.

The high rates obtained by McDougall encouraged Eriksson and Sullivan to examine whether the methodology could be applied to Swedish and to another segment of the same size constituted by /jœː/. In addition, these researchers included the fourth vocalic formant measurement. The nonparametric Kruskal-Wallis test was used instead of ANOVA, as a criterion to reduce the number of predictor variables. Those with the smallest values in the k statistic were excluded from the analysis. Also, the combination of predictors achieved from the nonparametric test was assayed along with the different combinations proposed by McDougall. The best result yielded a classification rate of 88%.

In 2006, McDougall presented another study applying the same methodology for the analysis of a British English segment, formed by intervocalic /r/ in the sequence /ərV/, where the vowels considered were /iː,æː,ɑː,ɔː,uː/, which carried the nuclear stress of the sentence. In this new study, a classification rate of 88% was reached. Additionally, McDougall [10] proposed a new methodological approach in which the three first formant contours of /ai/ diphthong were fitted with a quadratic and a cubic polynomial equation by using the linear regression technique. The three first formant contours of intervocalic /ai/ were fitted with cubic, quartic and quintic polynomial equations thus reducing the number of variables required to describe the dynamics of the formant contours. The parameters of polynomial fittings were used as predictor variables in LDA. The rate yielded with the polynomial fittings was equivalent to the rate yielded with direct measurements used in the first methodological approach. However, the number of predictors used with the polynomial fitting is lower than that of direct measurements. Thus, it opens the possibility of including additional acoustic information from each segment in the analysis.

The results of all these studies show that formant transitions discriminate well between the speakers' voices and can contribute to robust forensic voice comparison. In this study, the vocalic segment /a/ in Mexican Spanish is analyzed by employing the methodology proposed by McDougall in 2004, and the new one reported by this researcher in 2006. Moreover, fundamental frequency transitions are also analyzed in our research work following procedures of several studies [20–24] which agree on their discriminatory potential to undertake forensic voice comparison.

4 Statistical Analysis

The three statistical techniques employed in this study were Linear Discriminant Analysis (LDA), Principal Components Analysis (PCA) and Linear Regression (LR)[4]. LDA is a statistical technique whose main aims are: (a) discrimination, in order to determine the characteristics that differentiate between groups, as well as to find the optimal representation where the observation projection distinguishes between groups; and, (b) classification, which is used to assign a new observation to one of the existing groups, and is based on the original variables [12].

The discriminant functions generated by LDA out of the data are used for the classification process. In a first approach, the whole dataset is used for generating discriminant functions and testing classification ability. However, this approach overestimates the classification since an observation used to generate

[4] All statistical analyses were performed with R program. Each statistical analysis was programmed in order to reproduce the methodology proposed in this work.

the discriminant functions is also employed to test their capacity of classification. A better approach is to partition the data in two sets, one for generating the discriminant functions and the other for testing. In this study, the latter was employed using different partitions in a leave-one-out cross-validation method.

However, as aforementioned, LDA can only be used with reliability if the number of tokens exceeds the number of predictor variables by two, which was not always the case. Hence, we used PCA to reduce the number of variables. PCA is a technique that maps the original set of variables into an orthogonal space in which it is possible to model the data as a linear combination in the reduced set of dimensions called components. The net effect of the model is to filter out the natural dependency among the variables in the original set [13].

In this study, PCA was carried out with contexts in which the amount of tokens did not exceed the number of predictor variables in at least by two. The number of components that explained 80% of the total variance was used as a predictor variable in LDA. Figure 1 shows the two analysis carried out depending on the number of tokens obtained.

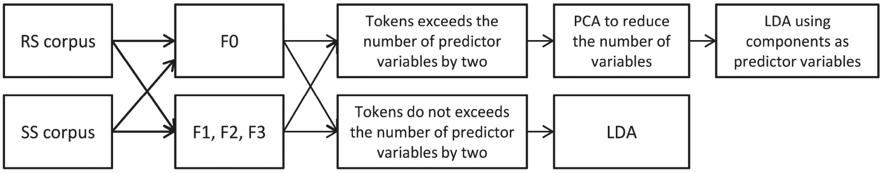
RS corpus
SS corpus
F0
F1, F2, F3
Tokens exceeds the number of predictor variables by two
Tokens do not exceeds the number of predictor variables by two
PCA to reduce the number of variables
LDA
LDA using components as predictor variables

Fig. 1. Type of analysis carried out in each corpus

In addition, we also used the LR method as suggested by McDougall [10]. The LR model specifies that for any value of the independent variable x, the population mean of the dependent or response variable, y, is described by a straight line [25]. This model is written as:

$$y = \beta_0 + \beta_1 x + \varepsilon \tag{1}$$

However, it is not always possible to fit the data into a straight line and often a more general kind of figure is needed. Hence, a polynomial model is frequently used. This is the case of formant contours, where the nine measures made along the total duration of vowel /a/ in contexts /sa/ and /ta/ in semi-spontaneous corpora can be fitted into a polynomial model as follows:

$$y = \beta_o + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + ... + \beta_m x^m + \varepsilon \tag{2}$$

This model is called an mth-order polynomial in which the m corresponds to the degree of polynomial. The greater the value of m, the better the fit. However, models with fewer parameters are preferred. In this study, the F1, F2 and F3 contour of each token of /a/ was fitted with a quadratic and a cubic polynomial. For the quadratic model, three parameters were used as predictor variables in LDA and four in the case of the cubic model.

5 Analysis of the Data

All recordings containing /a/ vowels in the contexts defined in section one were analyzed. The beginning and end of each vowel were labeled with Praat [26]. Additionally, a script was created in order to segment the /a/ in ten equal intervals. Finally, the fundamental frequency and the three first vocalic formants in all intervals were extracted, as shown in Fig. 2.

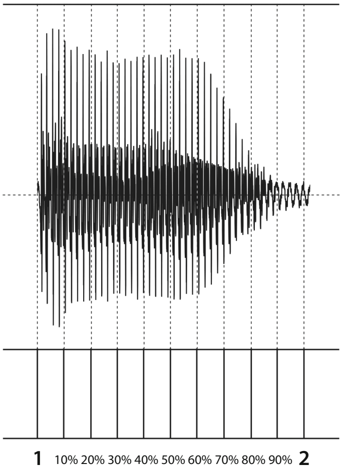

Fig. 2. /a/ vowel divided into ten time intervals of equal duration

The script displays frequency measures for each of the eleven points shown in Fig. 2; however, only nine measurements – represented by the nine percentiles in Fig. 2 – were used; this is, the first and the last measurements were eliminated in order to avoid possible errors at the vowel borders. Consequently, nine predictor variables were obtained for the fundamental frequency and nine more for each of the three first formants. Table 2 shows the predictor variables used in the statistical analysis.

Table 2. Predictor variables obtained for the statistical analysis

	Predictor variables								
	10%	20%	30%	40%	50%	60%	70%	80%	90%
F0	F01	F02	F03	F04	F05	F06	F07	F08	F09
F1	F11	F12	F13	F14	F15	F16	F17	F18	F19
F2	F21	F22	F23	F24	F25	F26	F27	F28	F29
F3	F31	F32	F33	F34	F35	F36	F37	F38	F39

Although the number of tokens per context is the same for all speakers, the number of tokens for each context varies as illustrated in Table 3. In the case of context /sa/, it was not possible to obtain at least five tokens from any of the speakers in RS corpus. However, in SS corpus, the number of tokens analyzed for this context was 24 in all five speakers.

Table 3. Number of occurrences per speaker in each corpus

Context	Number of tokens per speaker in RS	Number of tokens per speaker in SS
/sa/	-	24
/ta/	6	32
/pa/	7	34
/la/	18	32
/na/	7	33

Two Linear Discriminant Analyses were performed for each parameter. LDA analyses were carried out for F0: (1) using the nine predictors obtained from a dynamic measurement, and (2) midpoint measurement that corresponds to the F05 predictor. Additionally, two LDA were performed for the acoustic parameter of the three first formants. One analysis used 27 predictors, while the other, the three central ones (i.e., F15, F25 and F35). A Principal Component Analysis was carried out for both parameters in the dynamic case when the amount of predictor variables did not surpass the number of tokens by at least a factor of two. Then, the components that explain 80% of the total variance were used as predictor variables in LDA as shown in Fig. 1. For instance, the number of tokens in the context /na/ in the RS corpora for F0 was seven (see Table 3) and the number of predictor variables was nine; therefore, PCA was performed before LDA.

In addition, LDA for a number of combinations of predictor variables for the most discriminating contexts /sa/ and /ta/ in semi-spontaneous speech corpus were carried out, along the lines suggested by McDougall [9], as shown in Table 4.

Contexts /sa/ and /ta/ in semi-spontaneous speech corpora were also chosen to test the methodology proposed by McDougall [10]. The F1, F2 and F3 contours of each token of /a/ were fitted with a quadratic and a cubic polynomial. The parameters of these fittings were used as predictor variables in LDA.

Table 4. Different combinations of predictor variables

P	F1	F2	F3	F1+F2	F2+F3	F1+F3	F1+F2+F3	Optimal* F3+F2
1	F15	F25	F35					
2				F15,25	F25,35	F15,35		
3	F11,15,19	F21,25,29	F31,35,39				F15,25,35	
5	F11,13,15, 17,19	F21,23,25, 27,29	F31,33,35, 37,39					
6				F11,15,19, 21,25,29	F21,25,29, 31,35,39	F11,15,19, 31,35,39		F32,33,34, 35,36,26
9	F11,12,13, 14,15,16, 17,18,19	F21,22,23, 24,25,26, 27,28,29	F31,32,33, 34,35,36, 37,38,39					
10				F11,13,15, 17,19,21, 23,25,27, 29	F21,23,25, 27,29,31, 33,35,37, 39	F11,13,15, 17,19,31, 33,35,37, 39		
15							F11,13,15, 17,19,21, 23,25,27, 29,31,33, 35,37,39	
18				F11,12,13, 14,15,16, 17,18,19, 21,22,23, 24,25,26, 27,28,29	F21,22,23, 24,25,26, 27,28,29, 31,32,33, 34,35,36, 37,38,39	F11,12,13, 14,15,16, 17,18,19, 31,32,33, 34,35,36, 37,38,39		
27							F11,12,13, 14,15,16, 17,18,19, 21,22,23, 24,25,26, 27,28,29, 31,32,33, 34,35,36, 37,38,39	

P = Predictors included in each analysis,
*The optimal combination mentioned by Eriksson and Sullivan [11]

6 Results and Conclusions

The results obtained with dynamic and midpoint measurements for each type of genre speech (read and semi-spontaneous) are shown in Table 5.

The F0 parameter in both tables shows that the classification performance of the dynamic and static methods is the same for both corpora. Hence, the methodology applied to the F0 parameter is not useful for forensic voice comparison in Mexican Spanish. However, the dynamic method is better when the three first formants are used.

Table 5. Average of percentage of classification

Parameter	RS corpus (10 speakers)		SS corpus (5 speakers)	
	Dynamic	Static	Dynamic	Static
F0	24	24	43	43
F1, F2, F3	48	42	63	61

Although a direct comparison between the classification rates between the two corpora is not strictly granted, especially because the number of speakers is not the same, it can be appreciated that the classification obtained with the dynamic method is 20% higher in the semi-spontaneous speech corpus.

The results for each particular phonetic context of F1, F2 and F3 with the dynamic method for the semi-spontaneous speech case are shown in Table 6.

Table 6. Percentages of correct classification with dynamic measurements in semi-spontaneous speech corpus

Context	Correct classification (%)
/sa/	66
/ta/	64
/na/	62
/pa/	62
/la/	59

The contexts /sa/ and /ta/ show the best classification performances, and thus were chosen to run other Linear Discriminant Analyses using the different combinations of predictor variables proposed by McDougall [9]. Additionally, the optimal combination proposed by Eriksson and Sullivan [11], and the combination based on the parameters of a polynomial equation proposed by McDougall [10] were tested. The results are shown in Table 7.

The best classification percentage (73%) refers to the /sa/ context using nine predictors of F1, F2 and F3 followed by 72% fitting a quadratic polynomial model to the vocalic formants. For context /ta/, using the same predictors, the best classification percentage was 66%.

A comparison of the best percentage obtained in this study for the /sa/ context (73%) with McDougall's [9] and Eriksson and Sullivan's [11] shows that this percentage is 4% and 22% below the one obtained by Eriksson and Sullivan's [11] and McDougall [9], respectively. However, the method used by McDougall to test the discriminatory capacity of the linear functions overestimates the classification rate as she included the test data within the training data instead of the cross validation method used in the present study and also in that of Eriksson and Sullivan [11] in which test and training data are always kept apart.

Table 7. Correct classification rates obtained in contexts /sa/ and /ta/ using different combinations of predictors in LDA

Combination of formants	Predictors number	Correct classification SS corpus (%)	
		/sa/	/ta/
F1	1	59	53
	3	67	65
	5	65	64
	9	68	62
F2	1	48	35
	3	53	43
	5	48	44
	9	45	44
F3	1	39	27
	3	37	35
	5	37	38
	9	34	36
F1 + F2	2	60	57
	6	63	62
	10	66	64
	18	68	60
F2 + F3	2	53	38
	6	53	52
	10	50	48
	18	49	48
F1 + F3	2	65	47
	6	68	63
	10	62	63
	18	59	60
F1 + F2 + F3	3	64	58
	6	60	56
	9	**73**	**66**
	15	68	65
	27	-	64
Optimal F3 + F2	6	45	41
Quadratic	9	**72**	**66**
Cubic	12	66	65

In this study, one of the best results obtained by fitting the three first formant contours with a polynomial equation using linear regression was reached with the quadratic equation as well as McDougall [10]. A comparison of the best percentages obtained with this methodology shows a difference of 17% between the result reached by McDougall [10] (89%) and the corresponding to this study (72%). However, these results cannot be compared directly as the studies employed different segments in different languages.

The overall conclusion is that the dynamic analysis using only F0, is not informative enough to classify reliably speakers' voices. On the other hand, the dynamic analysis using F1, F2 and F3 improves the classification results over the static method that uses only the midpoint frequency. These results are in accordance with McDougall's [9,10] and Eriksson and Sullivan's [11] findings, which indicate that the dynamic method is better than the static one.

The highest percentages of classification was obtained with two methods: using nine predictors for the three first formants (F11, F15, F19, F21, F25, F29, F31, F35, F39) and fitting a quadratic equation (better than the cubic) using linear regression (i.e., down to nine variables), and applying the discriminant functions (i.e., discriminant analysis) on this reduced set, as suggested by McDougall [10], who obtained up to 89% of classification performance. Both methodologies are better to discriminate between speakers in the RS corpus. It is possible to obtain up to 73% of discrimination rate for the /sa/ context and, therefore, it can be used as evidence for forensic speaker comparison in Mexican Spanish.

Acknowledgements. Authors want to thank Mrs. Josefina Bolado, Head of the Scientific Paper Translation Department, from División de Investigación at Facultad de Medicina, UNAM, for editing the English-language version of this manuscript.

References

1. Zipf, G.K., Rogers, F.M.: Phonemes and Variphones in four present-day Romance Languages and Classical Latin from the viewpoint of dynamic Philology. Archives Néerlandaises de Phonétique Experimentale **15**, 111–147 (1939)
2. Navarro, T.: Estudios de Fonología Española. Syracuse University Press, New York (1946)
3. Alarcos, T.: Estudios de Fonología Española. Gredos, Madrid (1991)
4. Guirao, M., García, J.: Estudio Estadístico del Español. Consejo Nacional de Investigaciones Científicas y Técnicas, Buenos Aires (1993)
5. Pérez, E.: Frecuencia de fonemas. E-rthabla, Revista de Tecnologías del habla **1**, 1–7 (2003). http://gth-www.die.upm.es/numeros/N1/N1_A4.pdf
6. Cuétara, J.: Fonética y fonología del habla espontánea de la ciudad de México. Su aplicación en las tecnologías del habla [disertation]. Universidad Nacional Autónoma de México, México (2004)
7. Pineda, L., Castellanos, H., Cuétara, J., Galescu, L., Juárez, J., Llisterri, J., Pérez, P., Villaseñor, L.: The Corpus DIMEx100: transcription and evaluation. Lang. Resourc. Eval. **44**, 347–370 (2010)

8. Guerra, R.: Estudio estadístico de la sílaba en español. In: Esgueva, M., Cantanero, M. (eds.) Estudios de Fonética, vol. I, pp. 9–112. Consejo Superior de Investigaciones Científicas, Madrid (1983)
9. McDougall, K.: Speaker-specific formant dynamics: an experiment on Australian English /aI/. Int. J. Speech Lang. Law **11**(1), 103–130 (2004)
10. McDougall, K.: Dynamic features of speech and the characterization of speakers: toward a new approach using formant frequencies. Int. J. Speech Lang. Law **13**(1), 89–126 (2006)
11. Eriksson, E., Sullivan, K.: An investigation of the effectiveness of a Swedish glide+vowel segment for speaker discrimination. Forensic Linguist. **5**(1), 51–66 (2008)
12. Pardo, A., Ruiz, M.A.: SPSS 11. Guía Para el Análisis de Datos. McGraw-Hill, Madrid (2002)
13. Rencher, A.: Methods of Multivariate Analysis. John Wiley & Sons Inc., Publication, USA (2002)
14. Labov, W.: Field methods used by the project on linguistic change and variation. In: Baugh, J., Sherzer, J. (eds.) Language in use: Readings in Sociolinguistics. Prentice Hall, New Jersey (1984)
15. Turell, M.: La base teórica y metodológica de la variación lingüística. In: Turell, M. (ed.) La Sociolingüística de la Variación. Promociones y Publicaciones Universitarias, Barcelona (1995)
16. Greisbach, R., Esser, O., Weinstock, C.: Speaker identification by formant contours. In: Braun, A., Köster, J. (eds.) Studies in Forensic Phonetics, pp. 49–55. Wissenschaftlicher Verlag, Trier (1995)
17. Rose, P.: Long- and short-term within-speaker differences in the formants of Australian hello. In: Braun, A., Köster, J. (eds.) Studies in Forensic Phonetics, pp. 49–55. Wissenschaftlicher Verlag, Trier (1995)
18. McDougall, K.: Nolan, F: Discrimination of speakers using the formant dynamics of /u:/ in british English. In: Trouvain, J., Barry, W.J. (eds.) Proceeding of the 16th International Congress on Phonetic Sciences, pp. 1825–1828. Universität des Saarlandes, Saarbrücken (2007)
19. Kinoshita, Y., Ishihara, S., Rose, P.: Exploring the discriminatory potential of F0 distribution parameters in traditional forensic speaker recognition. Int. J. Speech Lang. Law **16**(1), 91–111 (2009)
20. Nolan, F.: The Phonetic Bases of Speech Recognition. Cambridge University Press, Cambridge (1983)
21. Hollien, H.: The Acoustics of Crime. The New Science of Forensic Phonetics. Plenum Press, New York (1990)
22. Jiang, M.: Fundamental frequency vector for a speaker identification system. Forensic Linguist. **3**(1), 95–106 (1996)
23. Jessen, M.: Speaker-specific information in voice quality parameters. Forensic Linguist. **4**(1), 84–103 (1997)
24. Foulkes, P., Barron, A.: Telephone speaker recognition amongst members of a close social network. Forensic Linguist. **7**(2), 180–198 (2000)
25. Freund, R., Wilson, W., Sa, P.: Regression Analysis. Statistical Modeling of a Response Variable. Elsevier, Amsterdam (2006)
26. Boersma, P., Weenink, D.: Praat: doing phonetics by computer [Computer program]. 5.1.25 Version. University of Amsterdam (2010)