Acoustic study of the vowel formant frequencies and F0: a contribution to Catalan forensic phonetics

Jordi Cicres Bosch
University of Girona and University Pompeu Fabra, Spain
E-mail: jordi.cicres@udg.es

ABSTRACT

Fundamental frequency and the frequencies of the F1, F2 and F3 formants of all Catalan vowels have been acoustically measured in a reading-speech corpus of 10 speakers. Several phonological aspects such as stress, context and phrasal intonation have been taken into account. The data was processed using statistical tests (MANOVA) and post-hoc comparisons to evaluate the intra-speaker and inter-speaker variation, and significant differences between speakers, place and manner of articulation. Also, most speaker-specific parameters have been identified. In addition, the value of the results for the forensic phonetics is discussed.

1. INTRODUCTION

Forensic phonetics is the use of phonetic knowledge for legal purposes. In Catalan, phonetic studies with a forensic purpose are rare or inexistent. But general phonetic studies, such as Recasens [1] and [2] are very useful because they present exhaustive analyses of Catalan sounds for a descriptive purpose.

One of the main applications of forensic phonetics is identification of speakers. This consists of providing evidence about a dubitated recorded speech, which is normally compared with undubitated voices, with the aim of matching each voice to a single person e.g., a criminal.

Apart from the acoustic field (environmental noise, reverberation, echo, etc.), acquisition problems (such as microphone, limitations of the transmission channel, etc.), and speaker-related problems (disguise, different speaking styles, etc.), there are limitations in the speaker identification in its phonological context.

In this paper, some data are provided to show the effect of different factors (speaker, place and manner of articulation) with this purpose.

2. THE DATA

The ten speakers were recorded using a Sony Minidisc MZ-R700PC and a Shure 515SD microphone in a quiet room with no other special conditions. The data was then digitalized into the computer using the Creative WaveStudio (version 4.10.3.0) software at the following settings: no compression (PCM), 44.000 Hz sample rate, 16 bits and mono.

The speakers, 5 male and 5 female, were all young (from 17 to 26) and they lived in the same Catalan town. So, interdialectal or age differences were non-existent.

The phonological system of these speakers (as a subdialect of Central Catalan) consists of a set of 6 stressed vowels [i, e, ɛ, a, o, u] and set of 3 unstressed vowels [i, u, ə].

The sound files were labelled by hand using the TextGrid Editor of the Praat software. The label contained information about the vowel, the stress, and the next sound. After that, the next sound was decomposed to place of articulation, manner of articulation and sonority. In total, 3679 vowels were analysed.

3. THE ACOUSTIC ANALYSIS

The software used in the acoustic analysis was Praat (version 4.0.28). An automatic process was used in the analysis of both F0 and formant frequencies for all vowels by means of a script.

This script ran through all the labelled segments of the speech data for all speakers and analysed automatically the given parameters.

For the F0 analysis, the To Pitch… command was used. The *time step* argument (the time between the centres of consecutive analyses frames) was set down at 0.01; the minimum and maximum pitches were set down at 75 Hz and 200 Hz respectively. A post-processing algorithm seeks the cheapest path through the candidates. The duration of the entire vowel was taken into account for this analysis.

For the F1, F2 and F3 analysis, the To Formant (burg)… command was used. The *time step* was set down at 0.01; the *number of formants* argument was set down at 5. The *maximum formant* was set down at 5000 Hz for men, and 5500 Hz for women. The *window length* was 0.025 and the *pre-emphasis from* value (that indicates the frequencies bellow that are not enhanced) was 50 Hz. For each analysis window, Praat applies a Gaussian-like window, and computes the LPC coefficients with the algorithm by Burg. All specific information about the Praat analyses can be found in the Praat manual [3].

ISBN 1-876346-48-5 © 2003 UAB

4. STATISTICAL TESTS

Because more than one dependent variable (F0, F1, F2 and F3) and factor (speaker, place and manner of articulation of the next sound) have been taken into account, the appropriate statistical test is a Multivariate Analysis of Variance (MANOVA).

The speaker variable has ten levels (one for each speaker). The place of articulation has 6 levels: bilabial, dentalveolar –minus [l] and [r]–, palatal, velar, [l] and [r], and none –for the vowels before pause. This classification is based on Recasens [1]. Finally, the manner of articulation has 8 levels: plosive, nasal, fricative, lateral, approximant fricative, trill, tap, and none for the vowels before pause.

The purpose of the MANOVA is to determine whether the means of the dependent variables are significantly different from each other for each subgroup formed by the levels of the factors. Individual factors and the interaction between them are analysed.

Also, the use of MANOVA tests provides a multivariate analysis of effects by taking into account the correlations between dependent variables.

MANOVA is a parametric statistical procedure. So, some assumptions of the data are necessary. These assumptions are (a) independence of observations; (b) multivariate normality; and (c) homogeneity of the covariance matrices. Nevertheless, only the independence of observations is strictly necessary. "The violation of the multivariate normality assumption has a small effect on the actual alpha level with which the researcher is working" (Stevens [4]). And "In practice, MANOVAs tend to be performed on data, regardless of whether the data violate this assumption, because the general consensus is that the MANOVA is a robust procedure." (Weinfurt [5]). The only handicap is the reduction of the power of the analysis in extremely platykurtic distributions (Olson [6]). On the other hand, "Stevens (1986) concluded that violation of the homogeneity of covariances matrices assumption when the number of subjects in each experimental group is approximately equal will lead to a slight reduction in statistical power" (Weinfurt [5]).

The output of a MANOVA test is multiple. First, the Pillai's Trace value indicates if a significant multivariate effect has been obtained at an alpha level; Second, the eta-squared is the value that indicates the proportion of variance explained by the independent variable's effect; finally, if the MANOVA yields significance, multiple ANOVAs are calculated individually with each dependent variable. The use of the Pillai's Trace instead of the Wilks's Lambda value is better when the covariances matrices assumption is violated, because it is more robust.

All statistic tests have been made using the SPSS for Windows package, version 11.0.1.

5. HYPOTHESIS

The main null hypothesis tested in the MANOVA is that the vector of dependent variables means (F0, F1, F2, and F3) is equal for each effect (each factor or interaction of factors). The alternative hypothesis postulates that there are differences.

So, in this study, it is expected that the speaker is a significant factor. This is essential for the forensic task in speaker recognition. Second, it postulates that place and manner of articulation are also significant (because of articulatory phenomenons, as is shown in the literature, i.e. [1, 2]). The interaction of place and manner of articulation factors should also be significant for the same reason. Nevertheless, no significances should be shown in the interaction between the speaker and the place or manner of articulation, nor between the interaction of all three factors, because articulatory phenomena should affect all speakers in the same way.

Regarding the eta-squared values, it is expected that the speaker factor will explain the largest portion of the variance in the dependent measures. Univariate ANOVAs for each of the dependent variables (and the interactions between them) are expected to show that F1 and F2 are more related with place and manner of articulation, whereas F0 and F3 are more related with the speaker.

Finally, post-hoc comparisons are expected to show significant differences between speakers (above all in F0 and F3 parameters) and between the place and manner of articulation factors (above all in F1 and F2). As the measures of the formant values were done at the middle of the vowel, coarticulation phenomena should have a low effect in the formant values.

6. RESULTS

The output of a MANOVA test is multiple. First, the Pillai's Trace value indicates if a significant multivariate effect has been obtained at an alpha level; second, the Eta-squared is the proportion of the total variability in the dependent variable accounted for by the variation in the independent variable; finally, if the MANOVA yields significance, then multiple ANOVAs are calculated with each factor.

The assumptions of multivariate normality and homogeneity of the covariance matrices were violated. Nevertheless, see [4, 5, 6].

Table 1 shows the significance and the eta-squared values for each vowel and each effect (including the interaction between the factors). The shading cells indicate significance at the 0.05 level. The bold values indicate those which are significant at the 0.01 level.

The values of p in Table 1 indicate that speaker, place of articulation and manner of articulation are significant ($p < 0.05$) in all vowels. On the one hand, the interaction between speaker*place of articulation, speaker*manner of articulation, and speaker*place of articulation*manner of articulation is not significant in any vowel except for [i] –and for [ə] in the interaction speaker*manner of articulation–. On the other hand, the results indicate that the interaction of place of articulation*manner of articulation is significant in all vowels except for [e] and [ɛ].

With regard to eta-squared values, Cohen [7] suggested that eta-squared values around 0.01 were small, 0.09 were medium and 0.025 or more were large. So, viewing the results, we can conclude that all factors and interaction of factors reveal a considerable proportion of variance. But the speaker factor is, by far, the factor that explains more variance.

	Vowel	p-value	η^2		Vowel	p-value	η^2
speaker	[i]	**0**	.261	**place of articulation**	[i]	**0**	.045
	[e]	**0**	.311		[e]	**0**	.074
	[ɛ]	**0**	.328		[ɛ]	.020	.029
	[a]	**0**	.268		[a]	**0**	.054
	[o]	**0**	.288		[o]	**0**	.070
	[u]	**0**	.283		[u]	**0**	.053
	[ə]	**0**	.226		[ə]	**0**	.019
manner of articulation	[i]	**0**	.040	**speaker*place of art.**	[i]	**.004**	.058
	[e]	**.001**	.063		[e]	.959	.068
	[ɛ]	**0**	.064		[ɛ]	1	.046
	[a]	**0**	.067		[a]	.986	.044
	[o]	**0**	.094		[o]	.699	.042
	[u]	**0**	.049		[u]	.868	.073
	[ə]	**0**	.056		[ə]	.588	.033
speaker*manner	[i]	**0**	.088	**place*manner of art.**	[i]	**0**	.021
	[e]	.970	.152		[e]	.081	.044
	[ɛ]	.984	.087		[ɛ]	.843	.008
	[a]	.999	.070		[a]	**0**	.027
	[o]	.297	.111		[o]	**0**	.059
	[u]	.629	.105		[u]	**0**	.043
	[ə]	.039	.080		[ə]	**.001**	.015
speaker*place * manner	[i]	.025	.070		[o]	.409	.055
	[e]	.981	.032		[u]	.854	.068
	[ɛ]	1	.009		[ə]	.643	.035
	[a]	.993	.017				

Table 1: Summary of the p-values and eta-squared values using the Pillai's Trace test in the multivariate tests.

One of the possible ways to analyse a significant effect in MANOVA is to perform univariate ANOVAs for each of the dependent variables (and the interactions between them). Nevertheless, this technique does not take into account the effect of the other dependent variables.

Tables 2 to 5 summarise the univariate ANOVA results in the globally significant MANOVA effects. The significance of the p-value and eta-squared have the same interpretation as in the multivariate analysis above.

vowel	F0	η^2	F1	η^2	F2	η^2	F3	η^2
[i]	**0**	.479	**0**	153	**0**	317	**0**	435
[e]	**0**	.585	**0**	.303	**0**	.360	**0**	.624
[ɛ]	**0**	.643	**0**	.437	**0**	.493	**0**	.400
[a]	**0**	.407	**0**	.153	**0**	.317	**0**	.435
[o]	**0**	.522	**0**	.478	**0**	.334	**0**	.333
[u]	**0**	.702	**0**	.279	**0**	.163	**0**	.377
[ə]	**0**	.498	**0**	.224	**0**	.247	**0**	.283

Table 2: P-values and eta-squared values of all vowels of univariate ANOVA for factor speaker.

vowel	F0	η^2	F1	η^2	F2	η^2	F3	η^2
[i]	**0**	.094	**.003**	.023	.150	.009	.070	.012
[e]	.629	.011	.024	.029	**0**	.108	**.002**	.093
[ɛ]	.638	.006	.032	.031	**.009**	.041	.310	.013
[a]	.275	.012	.156	.015	**0**	.145	**0**	.050
[o]	**.002**	.038	.179	.013	**0**	.094	**0**	.075
[u]	.120	.021	.014	.037	**0**	.091	.086	.023
[ə]	.591	.003	.022	.015	**0**	.044	.277	.006

Table 3: P-values and eta-squared values of all vowels of univariate ANOVA for factor place of articulation.

vowel	F0	η^2	F1	η^2	F2	η^2	F3	η^2
[i]	**0**	.051	**0**	.045	**0**	.085	**0**	.040
[e]	**.008**	.085	.012	.080	.090	.050	**.004**	.095
[ɛ]	**.009**	.048	**0**	.170	.080	.030	.235	.020
[a]	.073	.026	**0**	.075	**0**	.123	**.002**	.047
[o]	.359	.015	**0**	.114	**0**	.175	**0**	.113
[u]	**.001**	.066	433	.014	.012	.045	**.001**	.064
[ə]	.086	.015	**.005**	.027	**0**	.045	**0**	.106

Table 4: P-values and eta-squared values of all vowels of univariate ANOVA for factor manner of articulation.

ISBN 1-876346-48-5 © 2003 UAB

vowel	F0	η²	F1	η²	F2	η²	F3	η²
[i]	.438	.008	**.008**	.026	**0**	.043	**.002**	.031
[e]	.879	.002	.017	.051	.191	.021	.191	.021
[ɛ]	.294	.009	.647	.003	.616	.004	.616	.004
[a]	**.001**	.035	**0**	.040	.156	.012	.156	.012
[o]	.541	.006	**.007**	.032	**0**	.115	**0**	.115
[u]	**0**	.078	.746	.007	**.001**	.066	**.001**	.066
[ə]	.620	.004	**.003**	.025	**.001**	.028	**.001**	.028

Table 5: P-values and eta-squared values of all vowels of univariate ANOVA for the interaction between factors place and manner of articulation.

Table 2 is clear: all dependent variables are significant for each vowel. Nevertheless, in tables 3, 4 and 5, variation is shown between dependent variables and vowels. In the univariate ANOVA for the factor place of articulation, F0 and F3 present less significance. In the univariate ANOVA for the manner of articulation factor, however, all dependent variables present enough significance. Table 5 presents also enough significances in F1, F2 and F3, but less in F0.

The eta-squared values indicate that the highest proportion of the variance explained by the independent variable's effect is in the F0 and F3. With respect to the place and manner of articulation, F1 and F2 explains more variance than F0 and F3.

Post-hoc comparisons (using the Dunnet T3 analysis) showed, on the one hand, that significant pairwise comparisons between speakers was very high in F0 and F3. On the other hand, most significant pairwise comparisons between different places and manners of articulation were in F1 and F2.

7. CONCLUSIONS

The results of the experiment reported above have shown that the MANOVA test is significant for some effects. So, the null hypothesis can be rejected. The most significant effects are, in order of importance, the speaker, the manner of articulation, the place of articulation of the nex sound and the interaction between the place and manner of articulation.

Also, it has been seen in the multivariate ANOVAs and post-hoc tests that F0 and F3 carry important information about the identity of the speaker, and F1 and F2 carry more information about the place and the manner of articulation. Nevertheless, F1 and F2 are also useful for speaker identification purpose. The criteria to determine the importance of the dependent variable was the number of significant pairwise comparisons.

In conclusion, the analysis of all analyzed variables (F0 to F3) are significant to the speaker identification task with forensic purposes –with the limitations described in the introduction–, but above all F0 and F3, which carrie the most speaker-specific information.

ACKNOWLEDGMENTS

This paper was made possible by a grant from the Spanish Ministry of Education, Culture, and Sports. Also, I am very grateful to Lluís de Yzaguirre, Salvador Oliva and Angela Buxton for their generous help.

REFERENCES

[1] D. Recasens, *Fonètica descriptiva del català*, Barcelona: Institut d'Estudis Catalans (2nd edition), 1996.

[2] D. Recasens, *Estudis de fonètica experimental del català central*, Barcelona: Publicacions de l'Abadia de Montserrat, 1986.

[3] Boersma, P. & D. Weenink, *Praat, a system for doing phonetics by computer*, version 4.0.28, Institute Of Phonetic Sciences University of Amsterdam (up-to-date version of the manual at http://www.fon.hum.uva.nl/praat/).

[4] J. Stevens, *Applied multivariate statistics for the social sciences*, Hillsdale, NJ: Erlbaum, 1986.

[5] K. P. Weinfurt, "Multivariate Analysis of Variance", in *Reading and Understanding Multivariate Statistics*, L. G. Grimm and P. R. Yarnold, Ed., pp. 245-276. Washington: American Psychological Association, 1995.

[6] C. L. Olson, "Comparative robustness of six tests in multivariate analysis of variance", *Journal of American Statistical Association*, vol. 69, pp. 894-908.

[7] J. Cohen, *Statistical power analysis for the behavioral sciences*, San Diego, CA: Academic Press, 1977.