A CROSS-DIALECT COMPARISON OF PENINSULAR- AND PERUVIAN-SPANISH VOWELS

Geoffrey Stewart Morrison[*] and *Paola Escudero*[†]

[*] Department of Cognitive & Neural Systems, Boston University
gsm2@bu.edu
[†] Institute of Phonetic Sciences, University of Amsterdam
escudero@uva.nl

ABSTRACT

A comparison was made of the acoustic properties of Spanish vowels produced by monolingual Spanish speakers from Spain and Peru. Monophthongs were produced in sentence final position. Peninsular speakers' vowels were shorter, had lower fundamental frequency, and were more likely to be produced with creaky voice. A multivariate test on the whole vowel system did not find a significant cross-dialect difference in formant values. Implications for second-language speech perception and production research are discussed.

Keywords: vowel comparison Spanish dialect

1. INTRODUCTION

Numerous studies of first-language (L1) Spanish speakers' perception or production of second-language (L2) vowels have used data collected from speakers with a mixture of different Spanish dialects, for example [1, 2, 3]. When research is conducted in a non-Spanish-speaking country it is often easier to recruit a large number of L1-Spanish speakers if Spanish dialect is not controlled. The tacit or overt assumption is that there is no or little difference in vowels across Spanish dialects, and that data from speakers of different dialects can therefore legitimately be pooled.

At least one previous study has set out to investigate the issue of cross-dialect vowel differences in Spanish. Godínez [4] found some cross-dialect differences between the monophthongs of Argentinian, Mexican, and Peninsular dialects; however, he pointed out that the results could only be suggestive because of the small number of speakers in each dialect group.

The present study tests whether there are differences in the acoustic properties of monophthongs produced by monolingual Spanish speakers from Madrid, Spain and Lima, Peru.

2. METHODOLOGY

2.1. Data collection

Recordings were made of 20 monolingual speakers of each dialect (10 male and 10 female). The speakers were university students aged 18 to 25. None of the speakers included in the study reported having a knowledge of any language other than Spanish and English, and estimated their speaking ability for English at 2 or below on a range of 0 to 7. Their parents were monolingual Spanish speakers also from Madrid or Lima.

Recordings were made in quiet rooms in Madrid and Lima using an Edirol R-1 recorder and a Sony ECM-MS907 condenser microphone.

Speakers were presented with randomised lists of sentences to read. The sentences were presented on a computer screen. Each sentence had the format: *En* CVCe *y* CVCo *tenemos* V 'In CVCe and CVCo we have V'. In each sentence the V consisted of one of the Spanish monophthongs: orthographic 'i', 'e', 'a', and 'u' which correspond to phonemic /i/, /e/, /a/, /o/, and /u/ (all V within a sentence were the same monophthong). The C varied across sentences; however, the present paper reports only on the sentence-final vowels. In the randomised list of sentences, each vowel phoneme occurred 12 times in sentence-final position.

2.2. Acoustic analysis

Acoustic analysis was performed using software originally developed by T. M. Nearey and adapted by the first author.

The beginning and end of each target vowel were automatically determined in the following manner: Root-mean-square (RMS) amplitude was measured using a Hamming window of width 20 ms for female speakers and 25 ms for male speakers centred at each recording sample (time slice). Measuring away from the sample in the vowel with the maximum amplitude, the beginning and end of the vowel were designated as the first samples in each direction with an amplitude more than 5 dB below the vowel's maximum amplitude.

Vowel duration was extracted as the time between the beginning and end of the vowel as determined above, and fundamental and formant frequency tracks were measured between these markers. Fundamental frequency was measured using the algorithm presented in [5]. The first three formant frequencies (F1, F2, F3) were measured using the procedure outlined in [6]: The number of LPC coefficients was fixed at nine, and formants were tracked using a variant of the algorithm in [7]. The formants were tracked eight times using eight different cutoff values for F3 (range 2500–4000 for male speakers and 3000–4500 for female speakers). Each of the eight formant-track sets were visually displayed overlayed on a spectrogram. On the basis of a set of heuristics, the software indicated the likely best track set. The measured intensity, fundamental frequency, and formant frequencies were also used to synthesise a vowel. The researcher could listen to the original vowel and a synthesised vowel based on any desired track set. On the basis of visual and auditory inspection, the researcher selected the best formant-track set. The researcher also had the option of manually editing formant tracks, and of adjusting parameters for fundamental frequency measurement.

3. RESULTS

Data from 17 Peruvians (8 males and 9 females) and from 17 Spaniards (7 males and 10 females) were subjected to statistical analysis. Data were excluded from two Peruvian males and one Peruvian female because of poor recording quality, and from three Peninsular males because all or almost all their sentence-final vowels were produced with creaky voice. Occasional tokens from other speakers were excluded for the same reasons.

3.1. Duration and fundamental frequency

Geometric means for vowel duration and fundamental frequency (*f0*) are given in Figures 1 and 2. Means were calculated giving equal weight to male and female groups of speakers. Mean fundamental frequency had very little change from 25% to 75% of the duration of the vowel (magnitude 1 Hz or less).

Figure 1: Mean vowel duration.

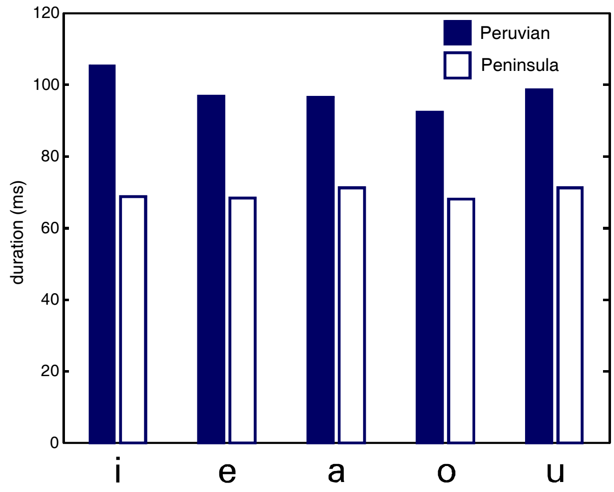


Figure 2: Mean fundamental frequency.

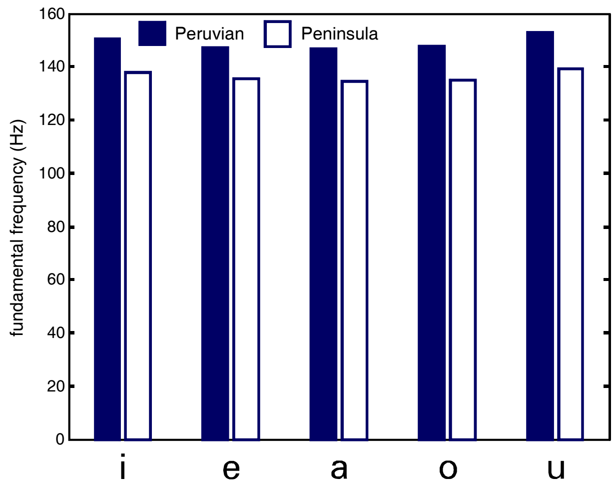
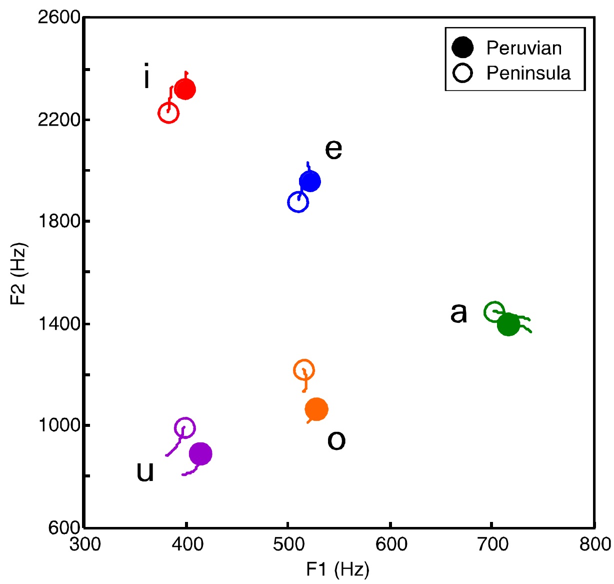


Figure 3: Mean normalised formant values plotted from 25% of the duration of the vowel (comet head) to 75% of the duration of the vowel (comet tail).



3.2. Formants

Formant values were normalised using a variant of the cross-dialect vowel normalisation procedure from [8]: In log hertz, each speaker's mean formant value $\overline{G_s}$ was calculated across all formants, tokens, and vowel categories (means were calculated for the middle 50% of each formant track for each token, then the mean over F1, F2, and F3, then the mean over all tokens within a vowel category, then the mean over all vowel categories). The mean of the $\overline{G_s}$ within each dialect, $\overline{G}_{Peru}$ and $\overline{G}_{Spain}$, were also calculated (equal weight was given to male and female groups of speakers). Each speaker's mean formant value $\overline{G_s}$ was subtracted from all of that speaker's formant measurements, then the dialect mean for that speaker's dialect, $\overline{G}_{Peru}$ or $\overline{G}_{Spain}$ as appropriate, was added back in. Mean normalised formant values for F1 and F2, converted back to hertz, are plotted in Figure 3.

3.3. Univariate analyses of variance

In order to determine whether there were significant differences between the acoustic properties of Peruvian and Peninsular speakers' vowels, a series of univariate analyses of variance (ANOVA) were conducted. An ANOVA was conducted on each of vowel duration and fundamental frequency. Each analysis was conducted on the non-normalised log values of the dependent variable, with dialect, gender, and vowel phoneme treated as fixed factors and speaker as a random factor nested within dialect and gender.

There was a significant dialect main effect for duration [$F(1, 30) = 41.14, p < .001$], and none of the interactions involving dialect were significant. The Peninsular speakers' mean vowel duration was 69 ms, which was 28 ms (33.9%) shorter than the Peruvian speakers' mean of 98 ms.

There was a significant dialect main effect for fundamental frequency [$F(1, 30) = 4.62, p = .040$], and none of the interactions involving dialect were significant. The Peninsular speakers' *f0* was 136 Hz, which was 13 Hz (8.8%) less than the Peruvian speakers' mean of 149 Hz.

For formants the dependent variables considered were the non-normalised log-hertz formant values measured at 25% of the duration of the vowel (*F1*, *F2*, *F3*), and the change in non-normalised log-hertz formant values from 25% to 75% of the duration of the vowel ($\Delta F1$, $\Delta F2$, $\Delta F3$). A series of ANOVA were conducted on each of the dependent variables separately for each vowel phoneme with dialect and gender as fixed factors and speaker as a random factor nested within dialect and gender.

Main effects for dialect which were significant at a nominal α level of .05 are given in Table 1. *F2* in /o/ was the only effect which was significant after a Bonferroni correction of one-sixth was applied to account for the six univariate tests run on each vowel category. No dialect by gender interactions were significant.

The effect size in Table 1 is the amount by which the Peninsular speakers' mean value is greater than that of the Peruvian speakers (calculated giving equal weight to male and female groups of speakers). Effect size is expressed in hertz and as a percentage of the grand mean of the specified formant for the specified vowel category calculated at 25% of the duration of the vowel.

Table 1: Significant dialect main effects for formants.

vowel	var.	*df*	*F*	*p*	effect size (Hz)	effect size (%)
/i/	$\Delta F2$	1, 30	5.54	.025	+36	+1.5
/i/	*F3*	1, 30	5.61	.025	–177	–6.0
/o/	*F2*	1, 30	8.65	.006	+119	+10.8
/o/	$\Delta F2$	1, 30	4.47	.043	–38	–3.5

3.4. Multivariate discriminant analyses and randomisation test

The following procedure was adopted in order to quantify the difference between the two dialects in terms of the whole set of vowels and all formant values together.

A discriminant analysis model was trained on the Peninsular speakers' normalised formant values, (*F1*, $\Delta F1$, *F2*, $\Delta F2$, *F3*, $\Delta F3$) and used to predict vowel category membership for all the vowel tokens from both dialects. A second model was likewise trained on the Peruvian speakers' data and also used to predict vowel category membership for all the vowel tokens. The correlation between these two models was calculated on the basis of the predicted a posteriori probabilities of vowel category membership for each vowel token. The correlation, calculated using the formula given in [8, 9], was .9905.

If there is no difference between the Peninsular and Peruvian groups' formant values, then the correlation between the predictions of the two models' above should on average be the same as the correlation between the predictions of pairs of models based on two groups of speakers drawn randomly from the whole set of speakers without regard for the dialect spoken. On the other hand, if there is a difference between the two original groups' formant values, then the correlation between the Peninsular versus Peruvian models predictions' will on average be less than the correlation between predictions from models based on randomly selected groups of speakers.

Randomisation tests were conducted (the ratios of male to female speakers were matched to the ratios in the original groups). The correlation between the models trained on the original Peninsular and Peruvian groups was less than the correlation between pairs of models trained on randomised groups for 755 of 1000 randomisations. As a test of the difference between the two dialect groups, the randomisation test therefore yielded a p value of .245.

4. DISCUSSION & CONCLUSION

The Peninsular speakers produced vowels which were substantially (33.9%) shorter than those produced by Peruvian speakers. In the Spanish-speaking world, Spaniards have a reputation for speaking quickly. A speaking-rate difference may account for the vowel-duration difference observed.

Compared to the Peruvian speakers, the Peninsular speakers had a substantially (8.8%) lower fundamental frequency. It should also be noted that data were excluded from three Peninsular males because they consistently produced creaky voice in the sentence-final vowels. It therefore appears that there are differences between speakers of the two dialects in terms of laryngeal activity or physiology.

In univariate analyses of individual vowels, the largest formant difference was for $F2$ in /o/. The Peninsular speakers' mean $F2$ was 10.8% greater than that of the Peruvian speakers. This was the only univariate test which was significant after a Bonferroni correction. The remainder of the dialect effects for formants which were significant at a nominal α level of .05 had effect sizes of 6.0% or less, and are therefore unlikely to be important.

In the multivariate analysis of the whole vowel system, the inter-dialect difference in formant values did not reach the customary α level of .05. Thus we do not have a compelling reason to reject the null hypothesis that there is no difference between vowel formant values in the two dialects.

In conclusion, from an analysis of formant values in Spanish monophthongs in sentence final position, it appears unlikely that mixing results from speakers of Madrid and Lima dialects will adversely affect the results of L2 vowel perception and production studies involving L1-Spanish speakers. However, if the observed vowel duration difference between the two dialects is intrinsic, rather than simply a speaking rate effect, then this could have an effect on L2 speech studies where vowel duration is a relevant factor. Whether these results will generalise to other phonetic contexts or to other sets of Spanish dialects remains to be tested.

5. REFERENCES

[1] Flege, J. E., Bohn, O.-S., Jang, S. 1997. Effects of experience on non-native speakers' production and perception of English vowels. *J. Phonetics* 25, 437–470.

[2] Escudero, P., Boersma, P. 2004. Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition* 26, 551–585.

[3] Morrison, G. S. 2006. *L1 & L2 Production and Perception of English and Spanish vowels: A Statistical Modelling Approach*. PhD diss., University of Alberta.

[4] Godínez, M. Jr. 1978. A comparative study of some romance vowels. *UCLA Working Papers in Phonetics* 41, 3–19.

[5] Boersma, P. 1993. Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. *Proceedings of the Institute of Phonetic Sciences, University of Amsterdam*. 17, 97–110.

[6] Nearey, T. M., Assmann, P. F., Hillenbrand, J. M. 2002. Evaluation of a strategy for automatic formant tracking. *J. Acoust. Soc. Am.* 112, 2323.

[7] Markel, J. D., Gray, A. H. 1976. *Linear Prediction of Speech*. Berlin: Springer-Verlag.

[8] Morrison, G. S., Nearey, T. M. 2006. A cross-language vowel normalisation procedure. *Canadian Acoustics* 34(3), 94–95.

[9] Nearey, T. M., Assmann, P. F. 1986. Modeling the role of vowel inherent spectral change in vowel identity. *J. Acoust. Soc. Am.* 80, 1297–1308.

ACKNOWLEDGMENTS

Thanks to Dr. Francisco Aliaga García, Dr. Javier Ferreiros López, and Dr. Jorge Iván Pérez Silva for help with recruiting participants and for providing locations in which to record. Thanks to Dr. Paul Boersma for comments on an earlier version of this paper. This research was supported by the Social Sciences and Humanities Research Council of Canada (SSHRC) and by the Netherlands Organization for Scientific Research (NWO).