LOCUS EQUATIONS AS A METRICS FOR PLACE OF ARTICULATION IN AUTOMATIC SPEECH RECOGNITION

Eugenio M. Celdrán and Xavier Villalba
Laboratorio de Fonética, Universitat de Barcelona

ABSTRACT

In this communication it is showed that locus equations are a powerful metrics for classifying Spanish stops regarding place of articulation. 10 subjects (5 male and 5 female) produced a series of labial [p]-[b], dental [t]-[d], and velar [k]-[g] tokens for 5 vowels. The resultant three locus equations nicely characterized the three places. Moreover, a discriminant analysis using both slopes and *y* intercepts and slopes alone yielded a 100% correct classification.

INTRODUCTION

After having failed in the search of reliable invariant cues for place of articulation of Spanish stops, following Blumstein and Stevens's [1] steps, we decided to look at Sussman's [2] and Sussman et alii's [3], [4] new proposal based on locus equations. This concept was originally conceived by B. Lindblom [5], who sensed that, although variations caused by coarticulation were noticeable, there seemed to exist a close relationship between the onset F2 values —which would roughly correspond to the vowel transition— and the F2 values of the midvowel nucleus. In order to prove it, he calculated the correlation between these F2 values. Thereby he obtained a series of lineal functions such as the following:

$$F2_{onset} = k * F2_{vowel} + c$$

where the constants *k* and *c* stand for the slope and the *y* intercept, respectively. Lindblom found that the values of the constants were clearly different for each place of articulation. In other words, the slope of the regression line varied depending on the place of articulation. Therefore, the various lineal functions obtained represented different places of articulation, more precisely, different locus equations.

[3] presented a particular proposal about the basis of voiced English stops. The results they achieved were enormously encouraging: using locus equations as metrics, they obtained 93% classification rates. That is why we found it necessary to carry out a study about the usefulness of locus equations in the discrimination of place of articulation for Spanish stops. This would serve us to see the universality of this method and we would be able to study in a quantitative manner, the effects of coarticulation of vowels on stop consonants. The degree of success of this method has direct consequences on the studies about automatic speech recognition: it provides us with a metrics to classify place of articulation starting from a specific sound stimulus.

METHOD

Material of study

The study was based on the speech of five male and five female subjects of ages 20 to 30. The subjects produced the sequence [kan'CVna], where C={[p], [b], [t], [d], [k], [g]} and V={[i], [e], [a], [o], [u]}; they repeated this sequence five times for each vowel. We obtained 150 stimuli for each subject (6 stops x 5 vowels x 5 productions = 150). The informants' productions were recorded in a soundproof booth using a "Shure SM58" microphone and a cassette recorder (Marantz, model CP430).

Measurements

The stimuli were reproduced and analyzed with the Kay CSL 4300 B. The formant measurements were based on measurements via cursor on a wide-band spectrogram, with the additional LPC derivation values of each formant.

Points of analysis

Considering that Spanish stops are non-aspirated, we measured the first glottal pulse after the burst (F2 onset) and the midvowel nucleus (F2 vowel). When the vowel formant trajectory was either ascending or descending, we took the value of F2 at a middle position in the formant. In the cases in which the formant trajectory was ascending-descending or descending-ascending, we took either the maximum or the minimum value, respectively.

RESULTS

We generated thirty locus equations (3 places of articulation × 10 speakers). The results are presented in Table 1.

Table 1. Slope and y intercept values for all speakers and place of articulation

	/p/-/b/		/t/-/d/		/k/-/g/	
Subjects	Slope	Intercept	Slope	Intercept	Slope	Intercept
M1	0.74	224.42	0.59	606.83	0.93	65.62
M2	0.79	151.06	0.69	516.53	0.95	60.7
M3	0.9	-1	0.52	803.27	1.19	-242.74
M4	0.88	45.04	0.65	545.86	1.12	-167.64
M5	0.81	195.56	0.51	829.94	0.97	74.43
F6	0.87	71.62	0.64	706.28	0.95	149.37
F7	0.83	134.33	0.63	662.41	0.99	5.66
F8	0.82	94.96	0.53	920.51	0.99	79.54
F9	0.80	241.86	0.53	966.08	1.01	42.13
F10	0.84	102.99	0.50	899.88	0.92	227.21
mean	**0.83**	**126.08**	**0.58**	**745.76**	**1**	**29.43**

The mean labial slope was 0.83 (s.d. 0.05) and the labial *y* intercept mean was 126.08. The dental mean slope was 0.58 (s.d. 0.07) with a mean *y* intercept of 745.76. Finally, the velar mean slope was 1 (s.d. 0.09) with a mean *y* intercept of 29.43. In all the cases

Slope variability

To test the variability of slope values, two ANOVAs were performed. The first one made a comparison with respect to gender and showed no significant difference: $F(1,28) = 0.138$, $p<0.7172$. The other ANOVA made a comparison with respect to stop place and significant difference was found: $F(2,27) = 93.013$, $p<0$. Further comparisons between place pairs also showed significant difference: labial vs. dental yielded $F(1,18) = 88.22$, $p<0.01$; labial vs. velar yielded $F(1,18) = 30.92$, $p<0.01$; and finally dental vs. velar yielded $F(1,18) = 145.01$, $p<0.01$. To sum up, the variability of slopes was not significantly affected by gender but by place of articulation.

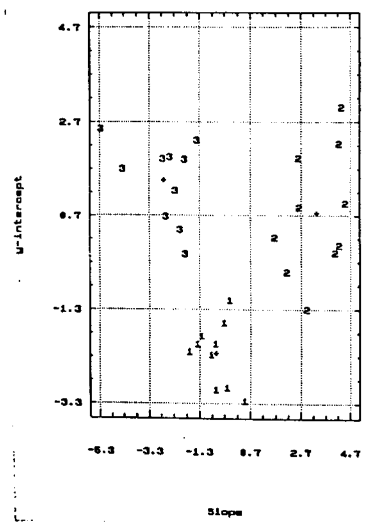
Y intercept variability

Again, in spite of the great variability of *y* intercept values, ANOVAs analysis showed non significant effects due to gender ($F(1,28) = 0.699$, $p<0.419$), but very significant effects due to place of articulation ($F(2,27) = 88.024$, $p<0$).

Discriminant analysis

Following [3], we decided to set up a discriminant analysis taking place as the classificator and the values of all locus equations (3 places × 10 speakers = 30 locus equations). Using only *y* intercepts the correct classification rates were 70% for labials, 100% for dentals, and 70% for velars (an overall mean of 80%). Nevertheless, using either slopes alone or both slopes and *y* intercepts the correct classification rates were 100% for all three place categories. Figure 1 shows the means plot for both slope and *y* intercept values.

Figure 1. Means plot of the discriminant analysis. 1 corresponds to labial stops, 2 to dental stops, and 3 to velar stops. Points show group centroids.



DISCUSSION

The experiment has shown that three clearly distinguished locus equations describe place of articulation of Spanish stops. Note that, as Table 2 shows, Spanish stops yield no slope overlapping, even though the minimum slope value of velar stops and the maximum one of labial stops are very close. This is consistent with the discriminant analysis results, which offered a 100% classification rate just using the slope values.

Table 2. Range slope values across place of articulation in Spanish stops.

	labial	dental	velar
minimum	0.74	0.5	0.93
maximum	0.9	0.69	1.19

However, before raising any conclusion, it must be taken into account the phonetic nature of Spanish stops. Firstly, Spanish voiceless stops show no aspiration, unlike the English ones. Moreover, their VOT values are quite low (6,5 ms for [p]; 10,4 ms for [t] and 25,7 ms for [k], according to [8]), which makes them quite similar to English voiced stops at initial position. Spanish voiced stops are also quite different from their English counterparts: they only occur at initial position and after nasals, and they always have a negative VOT value. Finally, we should also emphasize that the place of articulation of [t] and [d] in Spanish is dento-alveolar —i.e., the tip of the tongue touches the upper teeth, whereas the tongue blade is attached to the alveolar ridge—, as opposed to other languages like English, where they are clearly alveolar. All these differences probably explain why [1]'s method for the establishment of invariant cues for place of articulation of stop consonants failed in Spanish (see [9], [10]). Nevertheless, even though the differences, the resultant locus equations for Spanish stops have proved to be very consistent cross-linguistically, as Table 3 shows.

Table 3. Cross-linguistic comparison of mean slope values.

	labial	dento-alveolar	velar
Spanish	0.83	0.58	1
Thai [4]	0.70	0.30	
English [3]	0.87	0.43	0.66
Swedish [1],[7]	0.63	0.32	0.95
Arabic [4]	0.77	0.25	0.92
Urdu [4]	0.81	0.50	0.97
mean	**0.77**	**0,39**	**0,9**

Both labial and velar Spanish stops are very consistent cross-linguistically. The most important difference appears with dental stops. However, even such a difference can be explained by the above mentioned fact that Spanish [t]-[d] are clearly dento-alveolar and never alveolar, unlike e.g. English. That this might be the reason for the difference in dental slope values is supported by Urdu, a language showing both dental and alveolar stops: dental slope values were higher than the corresponding alveolar values, and hence closer to Spanish values.

One of our primary aims was to find invariant cues which would allow us to carry out automatic speech recognition based on phonetic features. From our statements, we believe that locus equations give us a good basis for the achievement of this aim.

ACKNOWLEDGEMENTS

We are indebted with O. Julià, L. Rallo, and H. Sussman for their help. This work has been supported by grants PB91-0278 from the DGICYT and AP92-46654991 from the Spanish Ministry of Education and Science (MEC).

REFERENCES

[1] Blumstein, S.E. & K.N.Stevens (1979), "Acoustic invariance in speech production", *JASA*, vol. 66, pp. 1001-1017.

[2] Sussman, H. (1991) "The representation of stop consonants in three-dimensional acoustic space", *Phonetica*, vol. 48, pp. 18-31.

[3] Sussman, H., H.A. McCaffrey, & S.A. Matthews (1991), "An investigation of locus equations as a source of relational invariance for stop place categorization", *JASA*, vol. 90, pp. 1309-1325.

[4] Sussman, H., K.A. Hoemeke, & F. S. Ahmed (1993), "A cross-linguistic investigation of locus equations as a phonetic descriptor for place of articulation", *JASA*, vol. 94, pp. 1256-1268.

[5] Lindblom, B. (1963) "On vowel reduction", Speech Transmission Lab. Report #29, The Royal Institute of Technology, Stockholm.

[6] Krull, D. (1988) "Acoustic properties as predictors of perceptual responses", *PERILUS*, vol. VII, pp. 66-70.

[7] Krull, D. (1989) "Second formant locus patterns and coarticulation in spontaneous speech", *PERILUS*, vol. X, pp. 87-108.

[8] Castañeda, M. L. (1986) "El VOT de las oclusivas sordas y sonoras españolas", *Estudios de Fonética Experimental*, vol. II, pp. 91-110.

[9] Núñez-Romero, B. (1995) "Invariación acústica en las oclusivas del español", to appear in *Estudios de Fonética Experimental*, vol. VII.

[10] Villalba, X. (1995) "Los invariantes acusticos y el punto de articulacion de las oclusivas en español", to appear in *Estudios de Fonética Experimental*, vol. VII.