SPEAKER IDIOSYNCRASY ON PHONETIC REGULARITIES IN FUNCTION OF TEMPORAL PARAMETERS OF VOICE

M.C. Diego, A. Aguilar
University of Barcelona, Dep. Personalitat, E. Y T. P.
Avda. Vall d'Hebrón 171
08035 Barcelona
Mail: mcdiego@psi.ub.es

Abstract

Sex, age, emotional state of the subject, pathological voices... are characteristics than allow us recognise different speakers. The accent difference persons of different regions, indeed their tone establish inequality between analogous voices; intensity and timbre differ voices with similar tones; pauses, tempo, idea's speed... condition an unique and discriminate voice for each speaker.

Previous studies have researched the possibility of studying the speaker idiosyncrasies on phonetic regularities, by mean of temporal parameters like the segmental duration. This kind of characteristics have the property of a very easy measurement, they are very stables by the time and have a minimal change in function of the different environment. All this characteristics allow us, in a practical term, remark the idiosyncrasic pattern of each speaker.

Introduction and aim

Variation characterized speech production. Research has focused on assessing variation that can be attributed to well-determined factors. A large number of factors have been found to affect the duration of segments and sequences of segments; e.g. phrase-final position, differences intrinsic to the segment type, effects of phonetic context, stress... (Hollien, 1990 [1]; Mendoza and Carballo, 1998 [2]). Variation that cannot easily be attributed to identifiable factors is labeled "residual or error variation". Two sources contribute to this kind or variation in speech: intra- and inter-speaker variation. Intra- or within-speaker variation is observed when a single speaker repeats the same utterance several times in the same (paralinguistic) contexts; inter- or between-speaker variation is found when several speakers speak the same utterance in the same context (Heuvel, H. Van den, Rietveld, T. And Cranen, B.,1994 [3]).

Previous studies (Bartkowa, 1988 [4]; Broeders and Rietveld, 1989 [5]; Campbell, 1990 [6]; Campbell and Isard, 1991 [7], addresses a number of issues concerning the effects of speaker idiosyncrasies on phonetic regularities, by means of temporal parameters.

Said regularities provoke that the speech of each speaking will be independent and differentiate, in their characteristic more intimate, of the speech of any speaker of their same language.

In these investigations they observed consistent differences between speakers, also intra-subjects stability for temporal measurements of voice.

Our objective is to establish objective bases that permit us to isolate said intrinsic characteristics to each speaking particular, in individual parameters function. Said parameters could be spectral (tone, pitch, formants), or well temporary parameters as the that occupy us in this case.

In the current study we tried to check if the same temporal parameters provided us similar results in Spanish language. We have check the segmental duration, like the duration of vowels and consonants in each position (Umeda, 1975, 1977 [8,9]; O'Shughnessy, 1981,1984 [10,11]; Crystal and House, 1988ab [12,13]; Johnson, Ladefoged and Lindau, 1993 [14]. We need that the measures of the variables were reliable and constant for each person, if we want to study these difference to know these particularities, but also these variable ought allow us to estimate interindividual and cluster difference. First we want to know the phonetic regularities of the Spanish language. After we need to know the common manner of the phonetics regularities of Spanish people, in the sense of the relative duration of every vowel and consonant, in function of their occupied position in the phrase. Furthermore, we pretended verified that the measures we obtained from different subjects had internal consistence.

Next works will allow us to know how to assess the effect of pathological personality on speech,

Method
Objective
The aim of the study is to call attention to some methodological aspects of speaker variability and its impact on phonetic rule systems. We pretend to analyse and compare temporal patterns that reveal speaker idiosyncrasy. Moreover we study the internal consistence of the different involve variables.

Hypothesis
First: We can observe temporal parameters of voice that provide us information about speakers intrinsical characteristics.
Second: There is internal consistence between obtained measurements for each subject.

Participants
Subjects are 9 adults of the two sexes (range 15-45 years old) without relevant voice pathology. Five men and four women. Seven of them are bilingual (Spanish and Catalan languages), but they are only demanded to read in Spanish. They were not paid for their services.

Self-report measures
Sex, Age, maternal language, dialect.

Speech data set
192 isolated nonsense disyllabic words of the /CVCV/ -type

Position of each phoneme	Vowels	Consonants
Initial	/a/,/i/,/u/	
Final	/e/	
Inter-vowel		/f/,/s/,/z/,/l/,/m/, /n/, /r/, /j/

Fig.1 Dependent variables and their position

Dependent variables
- Relative duration of the Pre-vowel Consonant – Vowel – Post-vowel Consonant in four letters words with structure Consonant – Vowel – Consonant – Vowel. (Fig.1)
- Temporal effects of coarticulation (final essentially)
- Difference in vowels duration.

to measure the pathological voices, and the progress of its treatment in different people.

Procedure
The nine subjects road the 192 words in Spanish language. The recording will be in natural environment (at work, at home) looking for non noise condition. For the record type we used a Digital Audio Tape Walkman TCD-D8, SONY registered trademark. The speech data were digitised at 48 kHz sampling frequency which allow us to record obtained date. We used an condensing microphone AKG.

Analysis
The analysis will be made by means of the PcVox program and the WISHA target made by the "Departamento de Ingeniería Electrónica E.T.S.I. de Telecomunicaciones" de la Universidad Politécnica de Madrid" [14]. This is an application specifically elaborated for the analysis of voice characteristics. The sample frequency of PcVox is 16 kHz, with a preemphasis coefficient of the 95%. The filter wide of the spectrogram was 45 Hz (355 points by signal). The minimum decibels of the spectrogram was 75, and the maximum was 110 for men and 120 for women.
We presented the spectral information of every 8000 ms. sample and identified the begin and the end of every segment of the word, based on the first formant of the phoneme (Fig.2). In this way we obtained the brute punctuation, then we deducted every initial and final point and we obtained the clear punctuation of the duration of every phonetic segment.
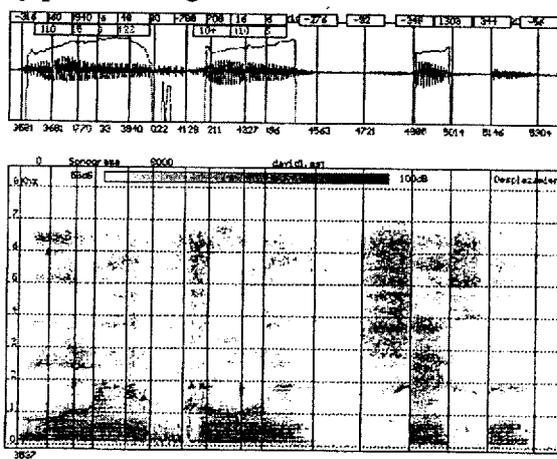

Fig.2 Example of spectrogram window and segments based on first formant. Typical segmentations for / /. From the top to the bottom the panels display the waveform, the RMs-curve (tone) with the intensity, and the spectral information with the formants.

Statistic data

The statistic analysis was made with the computer application SPSS version 8.0 for Windows NT 4.0 of the SPSS Inc.

Results

Eight of the subjects show very homogeneous dates at the reliability analysis. We apply the reliability tests Alpha (Cronbach), Two Halves and Guttman Split-half. We prove the internal soundness of the data with results are the next: α between 0.96 – 0.80; Guttman Split-half 0.97 – 0.75. Only a subject had low reliability punctuation test (α 0.29; Guttman Split-half 0.26 and α part 1 = -11.34 α part 2 = 0.36). Curiously, this subject presents the highest level of introversion.

The first analysis aimed to give an idea of the orders of magnitude of both segment (vowel-consonant) in Spanish language. We analysed the duration in function of the sex of the speaker, and the speaker-related variation in the vowels duration (Fig 3).

Using the General Linear Model (MANOVA), the results show that unless in the case of / f / initial and / z / between vowel, T of Student does not indicate meaningful differences between sexes.

In the case of the consonants, when we analyse each couple of these in function of their situation in the word (initial or inter-vowel), we observe appreciated differences in the segmental duration for the consonants /z, j, s, l/ in function of the position that they occupied (Fig 4). Furthermore, if we interrelate all the consonants of couple form, we appreciate meaningful correlations at level 0.05 (bilateral) only for the phonemes /f,s,m,j,r,z/ in this way: /f/ in position between vocal interrelates with /s/ (Pearson Corr. = -0.301), /f/ initial with /m, j/ between–vowel (0.314, 0.362) and /r/ initial (0..349); /z/ initial with /l/ between vowel (-0.417) and /r/ initial and between vowel (0.333). We can observe how it does not is increased correlations between same phonemes in different position (beginning of word or position inter - mean).

	Male		Female		Male & Female	
	M	SD	M	SD	M	SD
a	139.4516	26.3152	131.4762	21.4421	135.8053	24.1476
e	124.3600	32.6980	110.5517	31.3911	117.7434	30.4975
o	115.4333	20.9904	110.9286	24.3726	117.5748	93.8463
u	185.7073	50.5724	163.6712	50.9691	167.1935	54.1909
inicial /f/	161.0000	49.0510	176.1533	112.1961	151.7727	70.3603
inter /f/	126.6428	20.6112	137.6000	30.2809	131.6163	20.0205
inicial /z/	119.0000	58.6559	197.0000	56.4712	137.7600	41.5806
inter /z/	133.4286	32.3721	143.5556	27.3438	139.0357	23.3420
inicial /j/	155.4000	42.5655	156.2557	31.9672	148.6563	36.6101
inter /j/	143.0000	33.9293	146.0000	21.6753	135.7556	23.9271
inicial /s/	135.1538	17.9576	156.6000	81.2306	142.8400	49.5413
inter /s/	133.0000	18.9189	143.1429	22.5674	136.1026	17.9069
inicial /m/	118.1667	31.6343	100.3750	29.2468	96.9167	31.6251
inter /m/	87.5455	18.6728	79.8889	12.6337	83.7561	17.2739
inicial /n/	96.6364	39.7776	93.5000	35.4847	95.6596	38.5677
inter /n/	72.6875	26.3887	68.2000	17.4405	71.1000	21.3404
inicial /l/	56.6667	23.7880	74.3333	48.9680	80.0500	30.5967
inter /l/	84.7778	34.7061	75.0000	38.1061	77.1429	32.0393
inicial /r/	149.7657	39.8694	132.8333	56.8472	113.1667	48.0056
inter /r/	105.2222	20.1791	101.1111	22.8580	96.4048	26.5359

inicial = initial position of the consonant in the word; inter = position of the consonant between two vowels. It cant be a, i, u in first case, and e at the end of the word.
Mean segments durations and corresponding standard desviations for the five males and four females speakers. Values are in ms.

Pruebas de los efectos inter-sujetos

Fuente	Variable dependiente	Suma de cuadrados tipo III	gl	Media cuadrática	F	Sig.
Modelo corregido	Vocal A en Castellano	76.514	1	76.514	.194	.666
	Vocal I en Castellano	13.049	1	13.049	.044	.836
	Vocal U en Castellano	183.704	1	183.704	.291	.597
	Vocal E final en Castellano	3661.498	1	3661.498	2.018	.176
	Consonante F inicial en Castellano	12156.302	1	12156.302	6.935	.019
	Consonante F intervocálica en Castellano	310.250	1	310.250	1.069	.318
	Consonante Z inicial en Castellano	12393.441	1	12393.441	7.521	.015
	Consonante Z intervocálica en Castellano	95.886	1	95.886	.146	.707
	Consonante J inicial en Castellano	971.422	1	971.422	1.019	.329
	Consonante J intervocálica en Castellano	1962.796	1	1962.796	3.432	.084
	Consonante S inicial en Castellano	3076.795	1	3076.795	1.419	.252
	Consonante S intervocálica en Castellano	60.250	1	60.250	.208	.656
	Consonante M inicial en Castellano	594.910	1	594.910	.660	.429
	Consonante M intervocálica en Castellano	1118.004	1	1118.004	3.864	.068

Fig 3. Means of duration by sex and their significance with model T of Student.

Prueba de muestras relacionadas

		Diferencias relacionadas							
		Media	Desviación típ.	Error típ. de la media	Intervalo de confianza para la diferencia Inferior	Superior	t	gl	Sig. (bilateral)
Par 1	Consonante F inicial en Castellano - Consonante F intervocálica en Castellano	22.8769	69.8195	19.9040	8.403	44.9158	2.086	40	.042
Par 2	Consonante Z inicial en Castellano - Consonante Z intervocálica en Castellano	-4.0000	45.321	10.1341	-25.2109	17.2109	-.395	19	.697
Par 3	Consonante J inicial en Castellano - Consonante J intervocálica en Castellano	12.3065	48.5393	8.6513	-4.7363	30.5508	1.452	31	.146
Par 4	Consonante S inicial en Castellano - Consonante S intervocálica en Castellano	9.7568	61.2247	10.3653	-10.8568	30.1701	.959	34	.339
Par 5	Consonante M inicial en Castellano - Consonante M intervocálica en Castellano	13.0000	33.9536	5.3027	2.2826	23.7171	2.452	40	.019
Par 6	Consonante N inicial en Castellano - Consonante N intervocálica en Castellano	25.8000	42.5125	6.3374	11.3277	36.6723	3.755	44	.001
Par 7	Consonante L inicial en Castellano - Consonante L intervocálica en Castellano	7.2000	48.545	12.8559	-16.5247	26.921	.685	19	.515
Par 8	Consonante R inicial en Castellano - Consonante R intervocálica en Castellano	15.9024	48.3709	7.5543	.347	31.1702	2.105	40	.042

Fig 4. Related samples test. Meanings for / z, j, s, l / =, 697, 146, 339 and, 515 respectively.

Second term that we analysed was the differences between individuals and the repeated measures of every speaker for any vowel.

Prueba de muestras relacionadas

		t	gl	Sig. (bilateral)
Par 1	Vocal A en Castellano - Vocal I en Castellano	4,381	99	,000
Par 2	Vocal A en Castellano - Vocal U en Castellano	5,340	107	,000
Par 3	Vocal A en Castellano - Vocal E final en Castellano	-7,569	112	,000
Par 4	Vocal I en Castellano - Vocal U en Castellano	-,006	109	,996
Par 5	Vocal I en Castellano - Vocal E final en Castellano	-9,420	112	,000
Par 6	Vocal U en Castellano - Vocal E final en Castellano	-4,957	118	,000

Fig 5. Means of each vowel by subject and their signification with T of Student in a model that compare all the possibles couples.

The results obtained through the repeated measurements analysis show that the duration of the phonemes in the set of the subjects is significantly different, except for the case of the couple /i-u/ (Fig. 5). This last couple are short

vowels highly interrelated between theirself. If we force the extraction of four factors, we found, for every speaker that each factor corresponds to a vowel.

Through then not-parametrical Kruskal-Wallis test, we observe meaningful segment's duration of each vowel (X^2 = /a/; 35.352; /i/ 26.699; /u/ 23.230; /e/ 100.352 with a meaning of 0.000 to 0.003).

Indeed, if we introduced a new variable "word" (Fig. 6), we can observe that do not appear meaningful effects, each subject is more or less constant in the duration of their segments, while the issued word does not has meaningful effects on the segment's duration.

Fig 6. Effects test inter-subject for the variable "word" . Using the F of Snedecor we can observe that the word does not has meaningful effects on the segment's duration. Test M of Box on the equality of the covariance counterfoils: contrast the void hypothesis of that the observed covariance are equal in all the groups. M of Box = 231.74.

Summary and conclusions

In conclusion, each phoneme presents an intrinsic segment's duration, different to other phoneme, but furthermore, each subject presents a duration significantly different to other subject issuing equal phoneme. Its meaning that we observe stability in the intra – subject variability, but also meaningful differences between speakers (the same subject each time that issues a phoneme, will present a duration most or less constant and significantly different of the time than would another speaker take for the emission of the same phoneme). Contrarily we expect, they are not appreciated influences of the variable word in the segment's duration.

We have to consider too the great variability of the data, partially originated for the experimental situation. As well for this cause we should to dispense great number of words, being this number different for each subject. We have observed that all the subjects present meaningful reading mistakes. Some studies have demonstrated that the task of reading induces certain anxiety levels, yet in expert readers[16]. This anxiety is certainly increased because of the recording situation, by the length of the list, thus as by the type of stimulus (words without sense). Observing the great variability of the data distribution, we could propose new studies that consider emotional or personality factor.

Bibliography

[1] Hollien, H. (1990) Psychological Stress and Psychosis. In *The acoustics of crime. The new science of forensic phonetics*. Cap. 12. New York: Plenum Press.

[2] Mendoza, E., Carballo, G. (1998) Acoustics analysis of the induced vocal stress by means of cognitive workload tasks. *Journal of Voice* (in press).

[3]Heuvel, H. Van den, Rietveld, T. And Cranen, B. (1994). Methodological aspects of segment-and speaker-related variability. A study of segmental duration in Dutch. *Journal of Phonetics*, 22, 389-406

[4]Bartkowa, K. (1988). On the use of segmental duration in speaker-independent speech recognition system. In *Proceedings of the 7ᵗʰ FASE Symposium*. Edinburg, 763-770.

[5]Broeders, T. And Rietveld, T. (1989). Segmental marking as a cue in auditory voice identification of telephone speech. In *Proceedings Eurospeech* 89, Paris, 1, 71-74

[6] Campbell, W.N. (1990). Timing invariance in read speech Technology. In *Proceedings Workshop on Speaker Characterization in Speech Technology*, Edinburg, 78-82

[7] Campbell, W.N., Isard, S.D. (1991). Segment duration in a syllable frame. *Journal of Phonetics*, 19, 37-47.

[8] Umeda, N. (1975). Vowel duration in American English. *Journal of the Acoustical Society of America*, 58, 434-445

[9] Umeda, N. (1977). Consonant duration in American English.. *Journal of the Acoustical Society of America*, 61, 846-858

[10] O'Shaughnessy, D. (1981) A study of French vowel and consonant durations, *Journal of Phonetics*, 9, 385-406

[11] O'Shaughnessy, D. (1984) A multispeake analysis of durations in read French paragraphs. *Journal of the acoustical society of America*, 76, 1664-1672

[12] Crystal, Th,. M. and House, A.S. (1988a) The duration of American-English vowels an overview. *Journal of Phonetics*, 16, 263-284

[13] Crystal, Th. M. and House, A.S. (1988b) The duration of American-English stop consonants: an overview. *Journal of Phonetics*, 16, 285-294

[14] Johnson, K., Ladefoged, P., Lindau, M. (1993) Individual differences in vowel production. In *Journal of the Acoustical Society of America*, 94, 701-714

[15] Aguilera, S.; Santos, A. And Muñoz, E. (1991) El visualizador del habla (WISHA). *Infodiac*, 13

[16] Aguilar, A. (1993) *La dislexia como respuesta*. Barcelona P. y Ed. Univ. Barcelona.