Acoustic Comparative Study of Spanish Prosody: Mexico City vs. Madrid

Eduardo Velázquez
Escuela Nacional de Antropología e Historia, Mexico

1. Introduction

Speakers intuitively perceive differences in other speakers' pronunciation when using the same language. Of course, this ability depends on their familiarity with other language varieties which could be classified as dialectal, sociolectal, stylistic or idiosyncratic. They identify differences in the way some segments are pronounced, but this identification also occurs in prosodic levels, such as speed (syllabic duration and tempo), volume (intensity), or register (frequency range). All these differences are then typified as inherent features of those language varieties.

Some of the Mexicans who took part in this study were additionally asked to imitate the Spaniards' way of speaking and to name which pronunciation features were different from their own. Independently of the Peninsular segmental articulation features, like *ceceo* vs. *seseo*, or apical vs. predorsal /s/, most of them coincidentally said that Spaniards speak faster, louder and with a deeper voice compared with Mexican Spanish language or, at least, that was the way Mexicans imitated them.

However, how many of these features can be significantly detected acoustically and which methods should be used to report the similarities and differences among these two language varieties? In this article I devote myself to investigating the acoustic factors playing a role in the characterization of language varieties and to verifying the validity of the popular judgments about the prosodic differences among two standard varieties of the same language.

Comparative studies on prosody like the one proposed here have already been conducted by Adell *et al.* (2005), Cunha (2000, 2005), Seong and Hahn (1996), and Ulbrich (2002), coinciding in the examination of oral productions according to almost the same acoustic parameters: pitch (fundamental frequency), duration, and intensity. For this study, the Spanish varieties of Madrid and Mexico City have been chosen as representatives of the Peninsular and the Mexican dialects.

2. Methodology

In order to be able to conduct an acoustic and statistical analysis, I proposed a methodology in Velázquez (2008), which consisted of four steps: 1) choosing from a series of spontaneous recordings collected in Madrid and Mexico City those with the best quality, since, due to the environmental conditions, some of them were relatively noisy –however, they helped to test the analytical methods; 2) transcribing, by means of Praat text grids (Boerma and Weenink, 2008), in an orthographical way and then running a Praat script to obtain a picture and a recording of the transcribed items which would be inserted on an automatically generated web page; 3) selecting the ideal fragments to be transcribed in detail and running another Praat script to extract and insert all acoustic data into a hierarchically structured XML document, and 4) displaying the relevant data in a format compatible with Excel, where the results were obtained by applying different filters and formulas.

 Selected Proceedings of the 4th Conference on Laboratory Approaches to Spanish Phonology, ed. Marta Ortega-Llebaria, 83-90. Somerville, MA: Cascadilla Proceedings Project.

2.1. Transcription and data processing using Praat

As briefly mentioned in the previous section, recordings from Madrid and Mexico City were orthographically transcribed using text grids of the computer program called Praat, specifically designed to perform acoustic analyses. Afterwards, the recordings were also prosodically transcribed following the norms developed from a system used by the C-ORAL-ROM team (Cresti and Moneglia, 2005), which consists of a set of symbols for the segmentation and specification of intonational phrase and utterance types, as well as the signalling of paralinguistic phenomena used in discourse.

Praat has a function that allows the user to automate all internal processes by means of scripts. For this study, the graphical marks were used as anchors to adapt the transcripts to a hierarchical discourse structure, as described in detail in Velázquez (2006). The different display methods allowed a selection of the utterances which had the best acoustic quality to be segmented in syllables and transcribed phonetically.

Some of the utterances, due to their spontaneous nature and the impossibility of avoiding enviromental noises, yielded a poor pitch analysis in Praat. However, average fundamental frequency values were obtained by means of the harmonic frequencies. These corrected values, together with those related to duration, intensity, and other relevant data, were extracted and allocated by means of another Praat script into an XML-based structure which will be explained in the next section.

2.2. Automatic data allocation into an XML-based prosodic hierarchical structure

In the Praat text grids there were two kinds of information which were used to represent the elements of a prosodic hierarchical structure as shown in Figure 1: a) graphical signs, used to represent and organize linguistic and paralinguistic phenomena, such as utterance types, pauses, restarts and interruptions, and b) syllable boundaries, represented by means of intervals in the timeline.

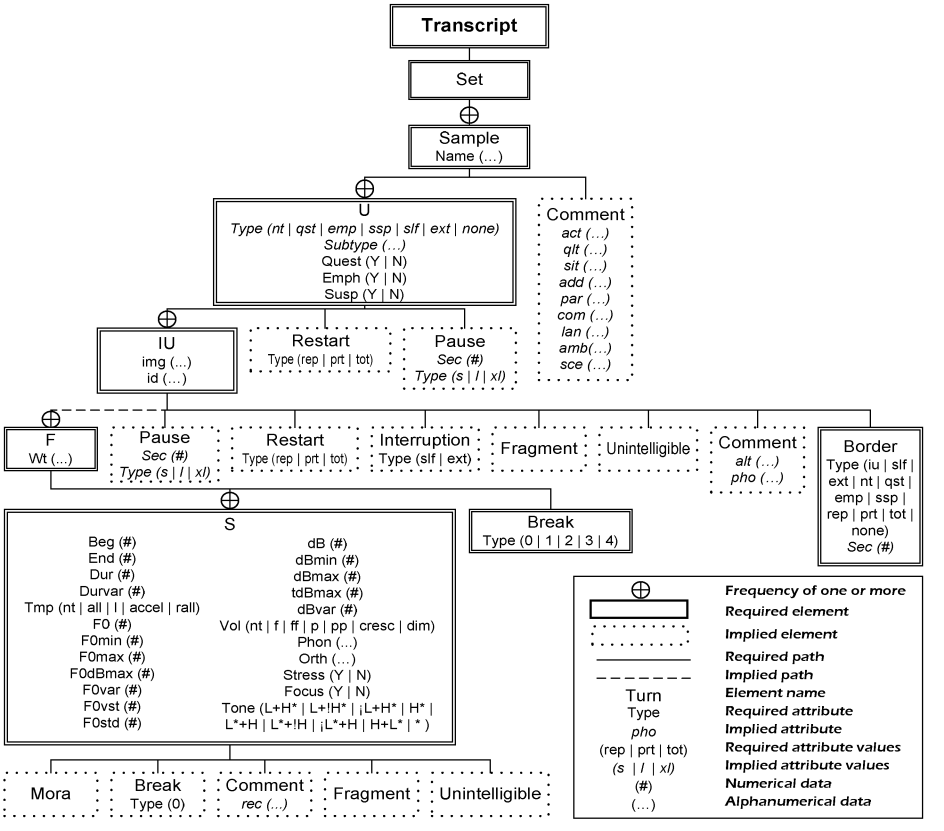

Figure 1: Prosodic hierarchical structure in XML (Extended Markup Language)

The structure presented in Figure 1 is the representation of the document type definition (DTD) of a transcript, according to the nomenclature of XML (Extended Markup Language). According to this DTD, a transcript consists of a set of one or more samples, each one containing one or more utterances (U). The latter must contain one or more intonation units (IU), whereas they may have individual border marks and a rhythmical structure (F, foot), which consist of one or more syllables (S) with a great number of attributes and separated by breaks (see Velázquez 2006, 2008).

Each syllable (i.e. each element called S) has a very important series of attributes: beginning, end, duration, relative variation, and relative tempo of the syllable (Beg, End, Dur, Durvar, Tmp); its fundamental frequency, with minimum, maximum, relative variation in percent and in semitones, and standardized value (F0, F0min, F0max, F0var, F0std), as well as its intensity, with minimum, maximum, relative variation, standardized value, and relative volume (dB, dBmin, dBmax, dBvar, dBstd, Vol), in addition to other phonological attributes. Hence, by means of the Praat script described before, acoustic values corresponding to each of these attributes were automatically extracted and nested into XML-like code. The resulting documents were well-formed and valid XML documents according to the DTD of a transcript.

These latter documents are the raw material from which documents in different formats may be obtained by using user-defined style sheets. For this study, the documents were yielded in HTML format, i.e. web pages, in order to display the acoustic information of all syllables in each utterance from both corpora. The data of these documents were easily copied into an Excel spread sheet where several filters and formulas were used to analyze and test their statistical differences, as shown below.

3. Comparative analysis

The Madrid corpus is composed of 6 recordings of spontaneous dialogues where 7 men and 7 women participated. They belonged to three age groups (18-25, 26-40, and 41-60), although most of them were under 40 years old. As shown in Table 1, education levels and occupations are very heterogeneous. The quantity of utterances provided by each of them varies from 9 to 58.

Recording	Participants (Gender)	Age*	Education**	Occupation	Items (Duration)
md1	GUI (♂)	A	3	linguist	17 (18.58s)
	MGL (♂)	B	2	hair dresser	20 (22.43s)
	PAT (♀)	B	2	hair dresser	29 (38.76s)
	ROS (♀)	B	3	teacher	23 (29.81s)
md2	FRA (♂)	A	2	AC installer	18 (24.78s)
	PTR (♀)	C	3	telephonist	29 (41.83s)
md3	MIG (♂)	A	2	taxi driver	32 (41.54s)
	PAC (♂)	B	2	AC installer	19 (23.09s)
md4	DAN (♂)	B	3	designer	35 (46.06s)
	LET (♀)	B	3	student	41 (53.33s)
md5	GEM (♀)	A	3	student	35 (46.71s)
	SON (♀)	B	2	telephonist	58 (58.81s)
md6	ABE (♂)	A	3	student	9 (10.13s)
	HEL (♀)	A	3	student	35 (54.78s)
				Total:	400 (8:30.64)

* A = 18-25, B = 26-40, C = 41-60

** 1 = primary or secondary school, 2 = high school, 3 = university student or graduate

Table 1: Participants included in the Madrid corpus

The Mexico City corpus contains 10 recordings of spontaneous interviews and dialogues where 5 women and 6 men from the three age groups participated. Their education levels and occupations were also varied. Each of them brought between 11 and 178 utterances to the corpus (see Table 2).

Recording	Participants (Gender)	Age*	Education**	Occupation	Items (Duration)
df11	ADR (♀)	A	3	receptionist	40 (87.18s)
	MIR (♀)***	A	3	receptionist	23 (32.31s)
df21	MOI (♂)	A	3	receptionist	187 (213.45s)
	ROD (♂)	A	3	receptionist	11 (10.38s)
df22	FRA (♂)	B	2	shop owner	78 (122.71s)
df31	MIR (♀)***	A	3	receptionist	70 (90.98s)
df62	ALB (♂)	B	3	salesman	52 (57.63s)
	CEC (♀)	B	2	saleswoman	28 (26.92s)
df71	FAB (♀)	A	3	student	13 (14.27s)
df81	MAR (♀)	C	1	maid	72 (108.45s)
df82	IVA (♂)***	C	2	plumber	19 (27.56s)
df83	IVA (♂)***	C	2	plumber	63 (97.12s)
df92	HUI (♂)	A	2	electrician	144 (190.14s)
				Total:	800 (17:59.10)

* A = 18-25, B = 26-40, C = 41-60

** 1 = primary or secondary school, 2 = high school, 3 = university student or graduate

*** Same person in different recordings

Table 2: Participants included in the Mexico City corpus

Due to the characteristics of dialogues and interviews, and also to the numbers of participants, speakers produced a very dissimilar number of utterances and syllables, especially per utterance type:

	Madrid			*Mexico City*		
	syll	**Utt**	**syll/utt**	**syll**	**utt**	**syll/utt**
Intr	647	100	6.47	659	104	6.34
Excl	59	13	4.54	58	16	3.63
Susp	912	123	7.41	3768	499	7.55
Decl	1740	239	7.28	1771	251	7.06
Women	2128	300	7.09	2101	277	7.58
Men	1230	175	7.03	4155	593	7.01
Total	3358	475	7.07	6256	870	7.19

Table 3: Number of syllables, number of utterances and average of syllables per utterance

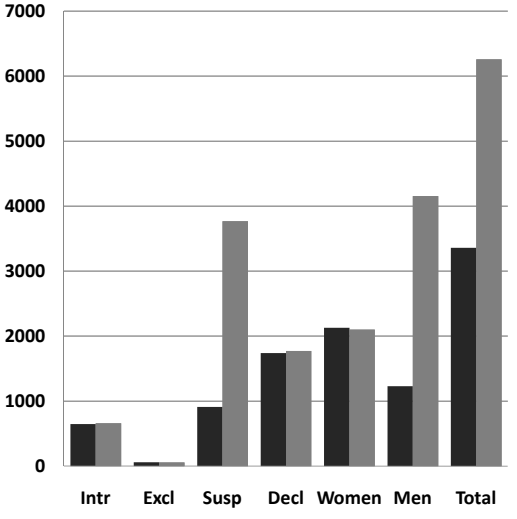

Figure 2: Number of syllables

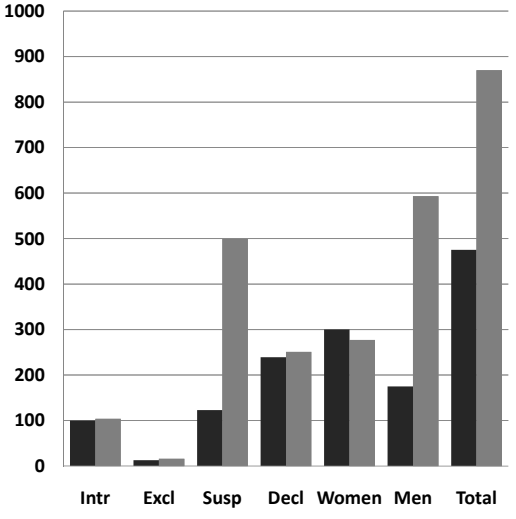

Figure 3: Number of utterances

Table 3 shows the total number of syllables (syll) and utterances (utt) produced by all speakers from Madrid or Mexico City (see also Figures 2 and 3, where dark grey columns correspond to Madrid values and light grey columns to Mexico City values), as well as the average number of syllables per utterance (syll/utt). Besides, utterances are classified into four basic categories: interrogative (Intr), exclamative (Excl), suspended (Susp), and declarative or neutral (Decl).

Therefore, the Madrid corpus included 3358 syllables distributed into 475 utterances, whereas the Mexico City corpus had 6256 syllables into 870 utterances. The least represented utterances in both corpora were the exclamative ones, which were the shortest as well, regarding the average number of syllables per utterance.

3.1. Syllable duration and tempo

Syllable duration (dur) and average variation range of a syllable (var range) allow us to calculate the average tempo, i.e. $1/\overline{dur}$. Besides, by means of utterance type and gender filters, we obtain the values respective to the different utterances and speaker groups (see Table 4, and Figures 4-5).

	Madrid			*Mexico City*		
	dur (sec)	**tempo (syll/sec)**	**var range (sec)**	**dur (sec)**	**tempo (syll/sec)**	**var range (sec)**
Intr**	0.138	7.255	0.141	0.147	6.820	0.147
Excl**	0.135	7.419	0.129	0.170	5.887	0.112
Susp	0.172	5.810	0.219	0.175	5.730	0.206
Decl**	0.144	6.968	0.151	0.151	6.623	0.156
Women***	0.149	6.730	0.159	0.166	6.026	0.182
Men***	0.150	6.650	0.178	0.163	6.123	0.183
Total***	0.149	6.701	0.165	0.164	6.093	0.182

Significance levels of standard t-tests verifying the unequality of the averages are indicated with *, ** and *** for the 10%, 5% and 1% level respectively.

Table 4: Syllable duration, average tempo, and variation range

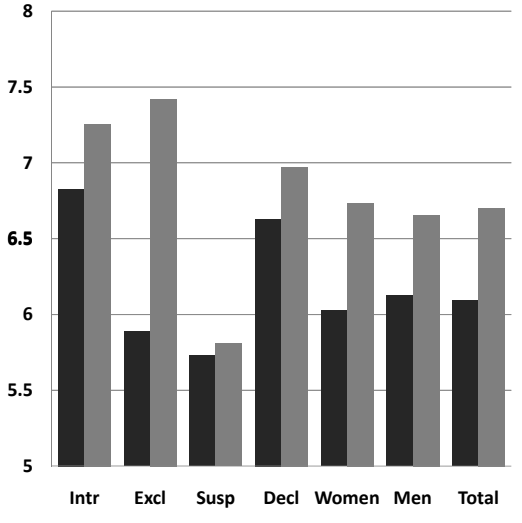

Figure 4: Average tempo (syll/sec)

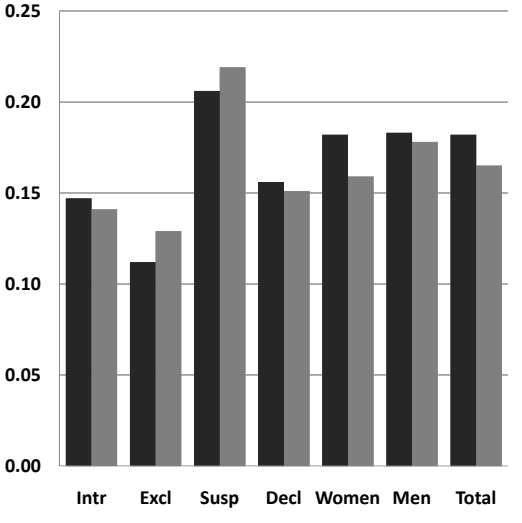

Figure 5: Variation range (duration)

According to these data, all Madrilenians' productions have shorter syllables than those of the Mexicans (from here onwards, the word Mexican will be used to refer to the people of Mexico City). Hence, the average tempo is always higher in the case of Madrid as well, especially in the case of women and exclamative utterances, whereas in Mexico City the highest tempo is reported in men and

in interrogative utterances. Moreover, Mexicans use the widest variation range, except in exclamative and suspended utterances.

3.2. Intensity

This section presents the average values corresponding to absolute intensity in dB (int), variation range, and standardized intensity (std int) (see Table 5, and Figures 6-7), according to formula (1). This formula arbitrarily attaches the value of 60 dB to the first observation (X_1) and adjusts the value of the following syllables according to the relative differences and with respect to the base level of 60 dB to obtain the average of these standardized values.

$$\overline{int}_{std} = \frac{\sum int_i}{n} * 60 / X_1 \qquad (1)$$

Absolute intensity values are normally useless since they depend mainly on the recording equipments and the speakers' distance from the microphone.

	Madrid			*Mexico City*		
	int (dB)	**std int (base 60dB)**	**var range (dB)**	**int (dB)**	**std int (base 60dB)**	**var range (dB)**
Intr* **	60.70	61.36	10.41	73.23	59.74	6.53
Excl	61.87	60.88	8.81	73.93	58.97	4.32
Susp* **	61.94	62.51	11.73	73.46	60.18	7.36
Decl* **	58.79	61.22	10.25	72.71	59.69	6.45
Women**	59.72	60.97	10.40	71.10	59.97	7.72
Men* **	60.68	62.61	10.99	74.17	59.96	6.59
Total* **	60.06	61.56	10.61	73.22	59.96	6.93

Significance levels of standard t-tests verifying the unequality of the standardized averages are indicated with *, ** and *** for the 10%, 5% and 1% level respectively.

Table 5: Absolute intensity, standardized intensity and variation range

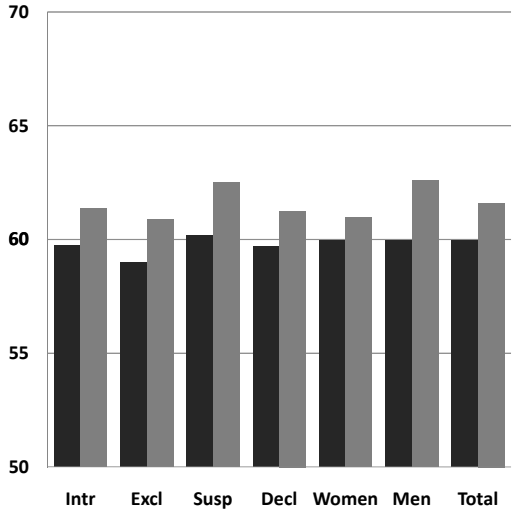

Figure 6: Standardized intensity

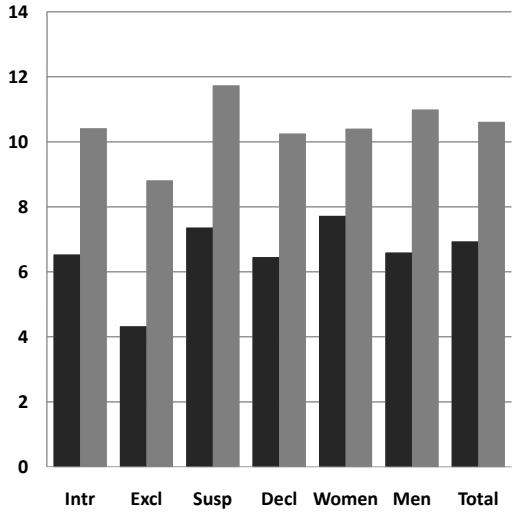

Figure 7: Variation range (intensity)

Standardized intensity values show that, in all cases, Madrilenians' intensity values are higher than those of the Mexicans, especially in the case of men. Besides, the variation range proves that Mexicans use a much narrower intensity spectrum than Madrilenians.

3.3. Fundamental frequency

Similarly, in order to avoid interpersonal differences in the use of fundamental frequency (f_0), formula (2) was used to get average standardized values (std f_0). The idea behind this formula is similar to that of formula (1) with the only difference being that the base level is set to 100 Hz (linked to X_1).

$$\overline{f_{0\,std}} = \frac{\sum f_{0i}}{n} * 100 / X_1 \qquad (2)$$

By means of this standardization process (see Cantero 2002), the values showed in Table 6 and Figures 8 and 9 were obtained:

	Madrid			*Mexico City*		
	f_0 (Hz)	std f_0 (base 100Hz)	var range (Hz)	f_0 (Hz)	std f_0 (base 100Hz)	var range (Hz)
Intr	179.55	106.55	39.79	191.41	108.40	44.22
Excl	212.62	105.26	37.31	197.26	106.86	23.18
Susp**	166.11	103.27	37.83	190.95	106.33	38.96
Decl***	178.45	100.90	30.65	189.54	104.36	33.41
Women**	208.25	102.19	36.25	247.12	104.33	37.07
Men***	121.86	103.90	31.81	164.37	106.81	37.99
Total***	176.42	102.82	34.61	190.72	106.02	37.70

Significance levels of standard t-tests verifying the unequality of the standardized averages are indicated with *, ** and *** for the 10%, 5% and 1% level respectively.

Table 6: Absolute fundamental frequency, standardized fundamental frequency and variation range

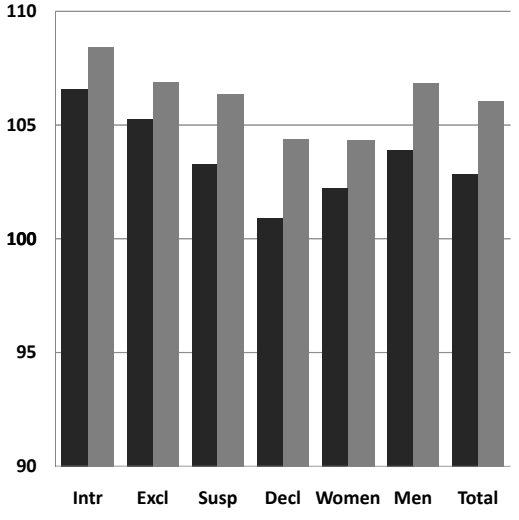

Figure 8: Standardized fundamental frequency

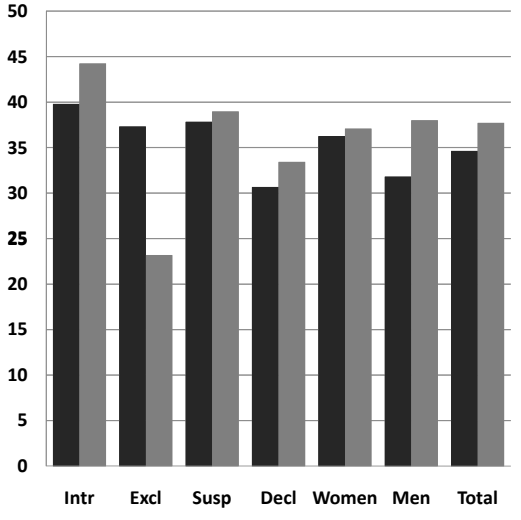

Figure 9: Variation range (fund. frequency)

The absolute values show clearly that, both in the case of men and women, the average fundamental frequency in Madrid is lower than in Mexico City. Even if the comparison is based on standardized values, in all cases, Mexicans use a higher frequency and a wider variation range, except in exclamative utterances. In both groups of speakers, interrogative utterances are produced with higher frequencies.

4. Conclusions

Thanks to the data tables presented in the last sections, it was proven that, by means of this method, compared to Madrilenians, 1) Mexicans produce longer syllables and in a more regular way, which makes their speech seem slower, 2) they speak with a lower and more regular volume or intensity, that is, they do not tend to increase and decrease the intensity much inside an utterance, 3) they use a higher-pitched voice register, which also has an effect on the higher variation of the frequency range, i.e. they speak in a less "monotonous" way.

In spite of the fact that the database has 1345 utterances, there did not appear as many exclamative or interrogative utterances as expected in the analyzed recordings, and, therefore, some results might vary considerably if the same methodology will be applied to a larger database. It is also planned that the sample set will be widened in order to include more regional or local Spanish varieties, or other sets of language varieties, aiming to verify if the model is detailed enough to identify the differences perceived by speakers as the most subtle ones.

References

Adell Mercado, Jordi; Bonafonte Cávez, Antonio and Escudero Mancebo, David (2005): "Analysis of prosodic features: towards modelling of emotional and pragmatic attributes of speech" in *Procesamiento del Lenguaje Natural*, 35, 277-283. [http://www.sepln.org/revistaSEPLN/revista/35/34.pdf]

Boersma, Paul and Weenink, David (2008): "Praat: doing phonetics by computer" (Computer Software, Version 5.1.04). [http://www.praat.org/]

Cantero, Francisco José (2002): *Teoría y análisis de la entonación*. Barcelona: Universitat de Barcelona.

Cresti, Emanuela and Moneglia, Massimo (eds.) (2005): *C-ORAL-ROM. Integrated Reference Corpora for Spoken Romance Languages*. Amsterdam: Benjamins.

Cunha, Cláudia de Souza (2000): "Entoação Regional no Português do Brasil", doctoral dissertation presented at the Federal University of Rio de Janeiro.

Cunha, Cláudia de Souza and Colamarco Cruz Pereira, Manuela (2005): "Do Recife aos Pampas: um experimento prosódico" in *Livro de resumos do IV Congresso Internacional da ABRALIN*. Brasilia: ABRALIN/ UNB. [http://www.letras.ufrj.br/posverna/docentes/72331-1.pdf]

Seong, Cheol-jae and Hahn, Minsoo (1996): "The Prosodic Analysis of Korean Dialogue Speech - Through a Comparative Study with Read Speech" in *Proceedings of International Conference on Spoken Language Processing ICSLP 1996, Philadelphia*. Piscataway, NJ: IEEE. [http://www.asel.udel.edu/icslp/cdrom/vol2/827/a827.pdf]

Ulbrich, Christiane (2002): "A Comparative Study of Intonation in Three Standard Varieties of German" in Bel, Bernard and Marlien, Isabelle (Eds.): *Proceedings of the International Conference on Speech Prosody 2002, Aix-en-Provence*. Aix-en-Provence: Laboratoire Parole et Langage. [http://aune.lpl.univ-aix.fr/sp2002/pdf/ulbrich.pdf]

Velázquez, Eduardo (2006): "Semi-Atomatic Prosodic Transcription of Spoken Spanish in XML" In Hoffmann, Rüdiger and Mixdorff, Hansjörg (Eds.): *Speech Prosody. 3rd International Conference*. Vol. II. Dresden: TUDpress, 623-626. [http://aune.lpl.univ-aix.fr/~sprosig/sp2006/contents/papers/PS5-30_0130.pdf]

Velázquez, Eduardo (2008): "Análisis prosódico comparativo del español oral", doctoral dissertation presented at the Free University of Berlin [http://www.diss.fu-berlin.de/diss/receive/FUDISS_thesis_000000006326]

Selected Proceedings of the 4th Conference on Laboratory Approaches to Spanish Phonology

edited by Marta Ortega-Llebaria

Cascadilla Proceedings Project Somerville, MA 2010

ISBN 978-1-57473-438-6 library binding

Ordering information

Orders for the library binding edition are handled by Cascadilla Press.
To place an order, go to www.lingref.com or contact:

Cascadilla Press, P.O. Box 440355, Somerville, MA 02144, USA
phone: 1-617-776-2370, fax: 1-617-776-2271, e-mail: sales@cascadilla.com

Web access and citation information

This entire proceedings can also be viewed on the web at www.lingref.com. Each paper has a unique document # which can be added to citations to facilitate access. The document # should not replace the full citation.

This paper can be cited as:

Velázquez, Eduardo. 2010. Acoustic Comparative Study of Spanish Prosody: Mexico City vs. Madrid. In *Selected Proceedings of the 4th Conference on Laboratory Approaches to Spanish Phonology*, ed. Marta Ortega-Llebaria, 83-90. Somerville, MA: Cascadilla Proceedings Project. www.lingref.com, document #2369.