Speech Prosody 2008
Campinas, Brazil
May 6-9, 2008

Prosodic Comparative Study of Mexico City and Madrid Spanish

Eduardo Patricio Velázquez Patiño

Freie Universität Berlin, Germany
Escuela Nacional de Antropología e Historia, Mexico
utka@yahoo.com

Abstract

Intuitively, speakers of the same language, depending on their familiarity with other varieties of their language, perceive dialectal, sociolectal, stylistic and idiosyncratic similarities and differences in speed (syllabic duration and tempo), volume (intensity), register (frequency range), melodic patterns (intonation), or rhythm. In the study which originated this article, an attempt is made to identify the acoustic factors that allow this identification, as well as to provide the necessary tools for a systematic analysis like, in this case, a comparative study of the Spanish varieties of Mexico City and Madrid.

1. Introduction

Spanish prosody has been studied from different theoretical approaches, mainly adapted from models designed for other languages, especially for English. This multitude of approaches, far from impeding the flow of research, has thrown light on different aspects which characterize Spanish prosody (for the specific case of intonation, see Prieto [6]).

As a part of a dissertation project, financed by the National Council of Science and Technology (CONACYT) of Mexico, a research was proposed in order to investigate the acoustic factors having a role in speech characterization and to prove the validity of the popular judgments about the prosodic differences among two standard varieties of the same language. For this study, the Spanish varieties of Madrid and Mexico City were chosen, as representatives of the Iberian and the Mexican varieties.

In order to reach this goal, a review of the theories being currently applied to Spanish was undertaken, and, for each prosodic factor the best approaches were chosen, e.g. the metrical-autosegmental theory and the Sp-ToBI inventory ([5], [7]), as well as the melodic pattern theory (see Cantero [2]), for intonation. Other methodological tools considered suitable for transcription, processing and analysis of spontaneous speech were also adapted or created.

2. Description of the *corpora*

For the acoustic and statistical analysis, two *corpora* with recordings of Madrid and Mexico City speakers were created.

The Madrid *corpus* is composed of 6 recordings of spontaneous dialogues where 7 men and 7 women participate. They belong to the three age groups (18-25, 26-40, and 41-60), although most of them are under 40 years old. Education levels and occupations are very heterogeneous. The quantity of utterances provided by each of them varies from 9 to 58.

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 are also varied. Each of them brings between 11 and 178 utterances to the *corpus*.

3. Transcription and processing in Praat

The recordings of both *corpora* were orthographically transcribed in Praat TextGrids [1]. Some of the prosodic and paralinguistic phenomena were transcribed with norms developed from those used for the C-ORAL-ROM project [3].

In a second selection phase, the utterances were segmented in syllables and transcribed phonetically. In spite of the quality of the recordings, especially those from Mexico City, which was not always optimal, the fundamental frequency was obtained by means of the harmonics (see Font [4]). Moreover, the intonation and rhythmic patterns of each utterance were analyzed and transcribed.

4. Prosodic hierarchical structure in XML

The combination of graphic signs used to represent the different linguistic and paralinguistic phenomena, as well as the boundaries marked with intervals in order to signal pauses and syllable breaks, were the basis from where the prosodic hierarchical structure was mapped (see Velázquez [8]). By means of Praat scripts, the acoustic information of each prosodic unit was extracted and nested into XML-like code. The resulting documents were well-formed and valid XML documents according to a document type definition (DTD) especially designed for the transcription of recorded utterances, enriched with the automatically extracted data.

5. Database creation from XML-Documents

The resulting XML documents are the raw material from which documents in different formats may be obtained. In these new documents, the information may be displayed and combined in different ways, e.g., interactive web pages displaying an image of the acoustic analysis and a link to the recording fragment of each utterance, or tables displaying the acoustic information of all syllables of an utterance, and all utterances in the *corpus*. The data of this latter document may be easily copied into an Excel spread sheet where different formulas and filters may be used to analyze the data.

6. Comparative analysis

Independently of the segmental characteristics of the Mexican and Iberian pronunciation, like *ceceo* vs. *seseo*, or *lleísmo* vs. *yeísmo*, most of the speakers agree that Spaniards speak faster, louder and with a deeper voice than Mexicans. Moreover, the speakers identify, more or less consciously, melodic and rhythmic patterns different from their own.

However, how many of these characteristics can be acoustically detected and which methods must be used in order to report on the similarities and differences among this two varieties of the same language? The most successful analysis methods will be presented below, as well as the results obtained from the Madrid and Mexico City *corpora*.

10.21437/SpeechProsody.2008-125

6.1. Syllables and utterances

Due to the characteristics of the dialogues or interviews, and also to the number of participants, the speakers produced a very dissimilar number of utterances. Table 1 shows the total number of syllables (**syll**) and utterances (**utt**) produced by each speaker from Madrid or Mexico City, as well as the average of syllables per utterance (**syll/utt**). Besides, according to the feature classification proposed by Cantero [2], ±I (interrogative), ±E (emphatic), ±S (suspended), the utterances are classified into four basic categories: interrogative (/**+I –E –S**/), exclamative (/**–I +E –S**/), suspended (/**–I –E +S**/), and declarative or neutral (/**–I –E –S**/).

	Madrid			*Mexico City*		
	syll	**utt**	**syll/utt**	**syll**	**utt**	**syll/utt**
/+I –E –S/	647	100	6.47	659	104	6.34
/–I +E –S/	59	13	4.54	58	16	3.63
/–I –E +S/	912	123	7.41	3768	499	7.55
/–I –E –S/	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 1: Number of syllables and utterances.

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

6.2. Duration and tempo

Syllabic duration (**dur**) and average variation range of a syllable (**range**) allow us to calculate the average tempo (**syll/sec**), i.e. $1/\overline{dur}(syll)$. Besides, by means of utterance-type and gender filters, we obtain the values respective to the different utterances and speaker groups (see Table 2).

	Madrid			*Mexico City*		
	dur	**syll/sec**	**range**	**dur**	**syll/sec**	**range**
/+I –E –S/**	0.138	7.255	0.141	0.147	6.820	0.147
/–I +E –S/**	0.135	7.419	0.129	0.170	5.887	0.112
/–I –E +S/	0.172	5.810	0.219	0.175	5.730	0.206
/–I –E –S/**	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 2: Syllable duration, average tempo, and variation range.

According to this data, all Madrilenians' productions have shorter syllables than those of the Mexicans (from here onwards, the word *Mexican* will be used to refer only 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.

6.3. Intensity

This section presents the average values corresponding to absolute intensity in dB (**int**), variation range, and standardized intensity (**std int**) (see Table 3), according to formula (1). This formula 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.

$$\overline{Int}_{std} = \frac{\sum Int_i}{n} * 60 / X_1 \quad (1)$$

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	**std int**	**range**	**int**	**std int**	**range**
/+I –E –S/***	60.70	61.36	10.41	73.23	59.74	6.53
/–I +E –S/	61.87	60.88	8.81	73.93	58.97	4.32
/–I –E +S/***	61.94	62.51	11.73	73.46	60.18	7.36
/–I –E –S/***	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 3: Absolute and standard intensity, with variation range.

Standardized intensity values show that, in all cases, the 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.

6.4. Frequency

In order to avoid interpersonal differences in the use of fundamental frequency (**f_0**), formula (2) was used. The idea behind this formula is similar to that of formula (1) with the only difference that the base level is set to 100 Hz.

$$\overline{f0}_{std} = \frac{\sum f0_i}{n} * 100 / X_1 \quad (2)$$

By means of this standardization, departing from an initial common base of 100 Hz (linked to X_1), the values showed in Table 4 were obtained:

	Madrid			*Mexico City*		
	f_0	**std f_0**	**range**	**f_0**	**std f_0**	**range**
/+I –E –S/	179.55	106.55	39.79	191.41	108.40	44.22
/–I +E –S/	212.62	105.26	37.31	197.26	106.86	23.18
/–I –E +S/**	166.11	103.27	37.83	190.95	106.33	38.96
/–I –E –S/***	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 4: Absolute and standardized f_0, with variation range.

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 bases on standardized values, in all cases, Mexicans use a higher frequency, but, except in exclamative utterances, they also use a wider variation range. In both groups of speakers, the interrogative utterances are produced with higher frequencies.

6.5. Intonation

Intonation was transcribed and analyzed according to the tone inventory of the Spanish Tones and Break Indices (Sp-ToBI) [7]. Therefore, following the metrical-autosegmental model, the tonemes are classified into boundary tones (starting and ending), prenuclear tones and nuclear tones.

6.5.1. Starting boundary tones

All initial frequencies above the bottom frequency set individually for each speaker was regarded as starting boundary tones **H%**, and all those at the same height or under the bottom frequency were considered **L%** tones (see Table 5).

	Madrid		*Mexico City*	
	H%	**L%**	**H%**	**L%**
/+I –E –S/	89.0%	11.0%	91.4%	8.7%
/–I +E –S/	92.3%	7.7%	100.0%	0.0%
/–I –E +S/	84.6%	15.5%	88.4%	11.6%
/–I –E –S/	82.0%	18.0%	85.3%	14.7%
Total	84.4%	15.6%	88.0%	12.0%

Table 5: Starting boundary tones for each utterance type.

In general, the preference for H% in both groups of speakers is conclusive, especially in the case of exclamative utterances.

6.5.2. Prenuclear and nuclear tones

Every tone between the first and the penultimate accent are considered prenuclear, whereas the tone around the last accent is called nuclear. According to the Sp-ToBI inventory, there are nine different types of prenuclear or nuclear tones, but, in order to obtain clearer results in this comparative study, they were classified into four "architonemes": those presenting the frequency peak *after* the accented syllable – **S´** (L*+H, L*+!H, ¡L*+H), *during* the accented syllable – **Ś** (L+H*, L+!H*, ¡L+H*, H*), *before* the accented syllable – **´S** (H+L*), or with no perceptible frequency peak – **S** (*).

	Madrid		*Mexico City*	
	Prenucl.	**Nuclear**	**Prenucl.**	**Nuclear**
S´	26.9%	24.6%	28.0%	24.7%
Ś	27.2%	35.6%	37.5%	44.1%
´S	14.5%	15.8%	6.7%	8.2%
S	31.4%	24.0%	27.8%	23.0%

Table 6: Prenuclear and nuclear tones.

In the case of prenuclear tones, the most common ones in the Madrid variety are **S**, whereas in the Mexico City variety they are **Ś**. This latter sort of tones is the most predominant in nuclear position for both groups of speakers. The distribution of tones per utterance type will be shown below.

6.5.3. Ending boundary tones

There are three kinds of ending boundary tones: **H%**, **M%**, **L%**. In Table 7, a clear preference for tone **M%** may be observed, even if part of this high result was possibly due to the fact that only those tones ascending or descending more than ±25% in regard to the last f_0 peak were considered **H%** or **L%**, which might have been too extreme, at least in the case of Madrid.

	Madrid			*Mexico City*		
	H%	**M%**	**L%**	**H%**	**M%**	**L%**
/+I –E –S/	3.0%	92.0%	5.0%	19.2%	75.0%	5.8%
/–I +E –S/	0.0%	61.5%	38.5%	0.0%	93.8%	6.3%
/–I –E +S/	4.1%	91.9%	4.1%	8.6%	84.6%	6.8%
/–I –E –S/	1.3%	93.3%	5.4%	1.6%	90.8%	7.6%
Total	2.3%	91.8%	5.9%	7.7%	85.4%	6.9%

Table 7: Ending boundary tones for each utterance type.

However, in order to report on the most common intonation contours for each utterance type, it is necessary to take into account the combinations of nuclear tones with ending boundary tones, as exemplified in Fig. 1 (the rising and falling dotted lines represent the tones **H%** and **L%** respectively).

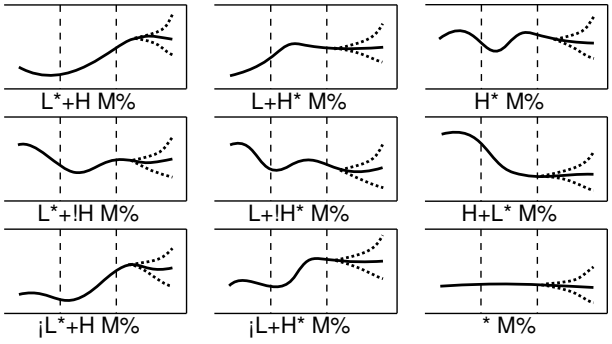

Figure 1: *Stylized f_0 contours of Spanish nuclear tones.*

The most common combinations in the utterance produced by Madrilenians and Mexicans will be counted below.

6.5.4. Tonal combinations in interrogative utterances

Most interrogative utterances with a clearly rising contour (**H%**), or all of them in Madrid, were produced with **S´** tones. Utterances with level contours are mostly produced in both groups with **Ś** tones. However, whereas these latter tones are used more often in Madrid in falling contours, in Mexico City the most common tones are **S´** (see Table 8).

	Madrid			*Mexico City*		
	H%	**M%**	**L%**	**H%**	**M%**	**L%**
S´	100.0%	35.9%	20.0%	70.0%	30.8%	50.0%
Ś	0.0%	43.5%	60.0%	30.0%	47.4%	16.7%
´S	0.0%	7.6%	20.0%	0.0%	2.6%	33.3%
S	0.0%	13.0%	0.0%	0.0%	19.2%	0.0%

Table 8: Tonal combinations in interrogative utterances.

6.5.5. Tonal combinations in exclamative utterances

Cases of exclamative utterances with a rising contour do not seem to exist in any of the two groups of speakers, and as for level contours, the preferred tones in both groups are **S´**. However, whereas tones **S´** are the only ones which appear in falling contours of the Mexican productions, Madrilenians use mostly **S´** y **Ś** (see Table 9).

	Madrid			*Mexico City*		
	H%	**M%**	**L%**	**H%**	**M%**	**L%**
S´	0.0%	50.0%	40.0%	0.0%	46.6%	100.0%
Ś	0.0%	12.5%	40.0%	0.0%	26.7%	0.0%
´S	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
S	0.0%	37.5%	20.0%	0.0%	26.7%	0.0%

Table 9: Tonal combinations in exclamative utterances.

6.5.6. Tonal combinations in suspended utterances

In the case of suspended utterances (see Table 10), the vast majority of rising contours are produced in both groups with **S´** tones, and falling contours with **Ś**. In level contours, Madrilenians lightly prefer **S´** tones, while Mexicans prefer them clearly.

	Madrid			*Mexico City*		
	H%	**M%**	**L%**	**H%**	**M%**	**L%**
S´	60.0%	24.8%	0.0%	67.4%	23.2%	17.6%
Ś	40.0%	26.5%	60.0%	32.6%	40.8%	55.9%
´S	0.0%	23.9%	20.0%	0.0%	11.6%	8.8%
S	0.0%	24.8%	20.0%	0.0%	24.4%	17.7%

Table 10: Tonal combinations in suspended utterances.

6.5.7. Tonal combinations in neutral utterances

Last but not least, in the case of neutral utterances, in Table 11 it may be observed that, to a greater or lesser extent, **Ś** tones are used more often in level and falling contours, whereas in rising contours Madrilenians prefer **S´**, and Mexicans **Ś**.

	Madrid			*Mexico City*		
	H%	**M%**	**L%**	**H%**	**M%**	**L%**
S´	66.7%	17.0%	23.1%	25.0%	13.1%	10.5%
Ś	33.3%	35.4%	61.5%	75.0%	52.2%	47.4%
´S	0.0%	16.6%	15.4%	0.0%	4.4%	26.3%
S	0.0%	31.0%	0.0%	0.0%	30.3%	15.8%

Table 11: Tonal combinations in neutral utterances.

6.6. Rhythm

In the recorded utterances, rhythmic groups were transcribed regarding the stressed syllables as nuclei of up to five subordinated unstressed syllables. From the 21 theoretical combinations, including the monosyllabic **(0)**, the most common rhythmic groups in the Spanish variety of Madrid are **1**, **4**, **0**, **8** and **2**, from which **1**, **4**, and **8** have a stress on the penultimate syllable (62.36%). In the variety of Mexico City, the rhythmic groups which appear more often are **1**, **0**, **4**, **2** and **8**. The groups with a stress on the penultimate syllable represent 57.88% (see Table 12).

		Madrid	**Mexico City**
(0)	**0**	13.7%	19.2%
(sw)	**1**	28.6%	27.8%
(ws)	**2**	10.4%	10.2%
(wsw)	**4**	15.9%	15.4%
(wws)	**5**	4.6%	4.8%
(wwsw)	**8**	11.2%	10.0%
(wwwsw)	**13**	5.5%	3.6%
Others		10.1%	9.0%

Table 12: Rhythmic groups (**s** = strong, **w** = weak).

However, the frequency of certain isolated rhythmic groups gives more information about the preferences on stress placing than on rhythm itself, since only sequences of two or more identical rhythmic groups are perceived by listeners as rhythmical structures (see Table 13).

	Madrid			*Mexico City*		
	2x	**3x**	**4x**	**2x**	**3x**	**4x**
1	28.8%	19.2%	1.9%	36.4%	15.5%	3.6%
2	5.8%			5.5%	2.7%	
4	19.2%			11.8%	7.3%	
5	1.9%			3.6%		
8	11.5%	3.8%		5.5%	1.8%	0.9%
13	3.8%			2.7%	0.9%	
Others	3.8%			1.8%		

Table 13: Number and length of rhythmic patterns.

Both in the productions of the speakers from Madrid and those from Mexico City, the most common rhythms are those created with two, three or four adjacent groups type **1** (trochaic rhythm), type **4** (amphibrachic rhythm), type **8** (*tertius paeon* rhythm), type **2** (iambic rhythm), and type **13** (dibrachic plus amphibrachic rhythm, **ww+wsw**). In the Mexican variety, the anapestic rhythm (formed with groups type **5**) has also a relative importance.

7. Discussion

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.

8. Conclusions

Thanks to the data tables presented in the last sections, it was proven that, by means of this method, from the point of view of 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 rise and lower the intensity much inside an utterance, 3) they use a higher-pitched voice register, which has also an effect on the higher variation of the frequency range, i.e. they speak in a less "monotonous" way, 4) they begin utterances more often with a high pitch; they pronounce more frequently accented syllables with a simultaneous and higher frequency peak, and, comparatively, they finish utterances with rising contours using a higher frequency, and those with falling contours using a per cent lower frequency, 5) in general, although the distribution of nuclear and final tones is very variable, in utterances ended with a level contour (lesser than ±25%), Mexicans use more frequently **Ś** tones, that is, tones whose frequency peak coincides with the accented syllable (L+H*, ¡L+H*, L+!H* or H*), and 6) they tend to form shorter rhythmic structures, preferably with stress on the penultimate syllables (bimetric trochaic rhythms).

9. References

[1] Boersma, P.; Weenink, D., 2007. Praat: doing phonetics by computer (Versión 4.6.36). <http://www.praat.org/>

[2] Cantero Serena, F.J., 2002. *Teoría y análisis de la entonación*. Barcelona: Universitat de Barcelona.

[3] Cresti, E.; Moneglia, M. (eds.), 2005. *C-ORAL-ROM. Integrated Reference Corpora for Spoken Romance Languages*. Amsterdam: Benjamins.

[4] Font Rotchés, D., 2007. *L'entonació del català*. Montserrat: Publicacions de l'Abadia de Montserrat.

[5] Hualde, J.I., 2003. El modelo métrico y autosegmental. In *Teorías de la entonación,* Prieto, P. (ed.). Barcelona: Ariel, 155-184.

[6] Prieto, P. (ed.), 2003. *Teorías de la entonación.* Barcelona: Ariel.

[7] Sosa, J.M., 2003. La notación tonal del español en el modelo Sp-ToBI. In *Teorías de la entonación,* Prieto, P. (ed.). Barcelona: Ariel, 185-208.

[8] Velázquez Patiño, E., 2006. Semi-Atomatic Prosodic Transcription of Spoken Spanish in XML. *Speech Prosody. 3rd Intl. Conference*. Vol. II. TUDpress: Dresden, 623-626. <http:// www.geocities.com/utka/>