TWO TENDENCIES TOWARD ISOCHRONY IN CASTILIAN SPANISH SHORT DECLARATIVE SENTENCES

Takuya Kimura
Tamagawa University, Tokyo, Japan

ABSTRACT

A recording of 96 sentences read by two speakers of Castilian Spanish was analyzed acoustically to investigate the effects of the lexical stress upon two phonetic properties: syllable duration and F0 contours. The results suggest that there are two independent tendencies toward isochrony.

Firstly, unstressed syllables tend to be shorter when there are more unstressed syllables between stressed ones, suggesting a tendency toward isochronous stress in this dialect, traditionally considered to be syllable-timed.

Secondly, Fujisaki's F0 model was applied to the same recording, and a tendency was observed for the interval between two accent commands to be longer when it was too short, and to be shorter when it was too long. This can be seen as a tendency to isochronous occurrence of accent commands.

1. INTRODUCTION

In the last few decades, a number of studies have been done on the nature of rhythm in Spanish. Most of the investigators classify Spanish as a syllable-timed language, while some dialects, such as Buenos Aires Spanish, have been reported to be stress-timed [1, 2]. In this paper I will show that Castilian Spanish also has a tendency toward stress isochrony (Section 3).[1]

Pitch contour is another important problem. While Spanish has traditionally been considered a stress-accent language, recent studies show that the fundamental frequency (F0) is a more crucial acoustic cue for the perception of stress than intensity and duration [3]. However, stressed syllables do not always show higher F0 values than unstressed ones. Spanish F0 contours show complicated variation, caused by intonation, which in turn is determined by such factors as sentence length, syntactic construction and speech style, aside from dialectal varieties [4]. In spite of this, little research has been done so far on the relation between the stress and the F0 contour in Peninsular Spanish. The second purpose of this study is to compare the timings of F0 rise and fall with the timings of onset and end of the stressed syllable in short declarative sentences, a sentence type that arguably has the simplest intonation pattern (Section 4). The definition of a "short declarative sentence" will be given in the next section.

2. MATERIALS USED IN THIS STUDY

A male speaker from Salamanca (Speaker 1) and a female speaker from Madrid (Speaker 2) read 96 different sentences. All the sentences read were short declarative sentences. A "short declarative sentence" means a declarative sentence consisting of only one or two prosodic words. A prosodic word (PW henceforth) is a sentence constituent containing only one lexical stress. For instance, the sentence "Me lo **da** / la mo**re**na." ("The dark-complexioned woman gives it to me." Bold letters indicate stressed syllables.) consists of two PWs, with their boundary indicated by the slash. All the sentences consisted solely of CV-type syllables so that the effects of different syllable types should be eliminated. Length of the PWs ranged from one to four syllables, with the only exception of the sentence "La lla**ma**ba / renova**do**ra." ("He used to call her renovative."), in which the second PW had five syllables. The sentences contained voiced sounds only (with a small number of inevitable exceptions), and all the possible stress patterns were covered. In total, 96(sentences) x 2(speakers) = 192 utterances (1,060 syllables) were recorded in DAT (digitized at 12 kHz with 12 bit precision) and analyzed with KAWAI VOICE ANALYZING SYSTEM Ver. 1.1.

3. SYLLABLE DURATION

3.1. Procedure

Duration of all the syllables contained in the recording was measured by means of visual observation on spectrograms.

3.2. Results

Table 1 shows the mean syllable duration for each speaker. Welch's *t*-test was used to compare the mean values. The table indicates that Speaker 2 read faster than Speaker 1, with a statistically significant difference.

	Mean	SD	*t* (df)	*p*
Speaker 1	153	36	3.74 (1022)	<.01
Speaker 2	145	31		

Table 1. Comparison between the two speakers
(Mean and SD in milliseconds)

Table 2 compares the mean duration of the stressed and unstressed syllables. The stressed syllables were longer than the unstressed ones, and here, too, the differences are statistically significant.

		Mean	SD	*t* (df)	*p*
Speaker 1	stressed	175	37	10.51 (290)	<.01
	unstressed	141	30		
Speaker 2	stressed	163	35	9.32 (259)	<.01
	unstressed	136	24		

Table 2. Comparison between stressed and unstressed syllables
(Mean and SD in milliseconds)

Table 3 shows that the sentence-final syllables were longer than the non-final ones. Here the difference is statistically significant for Speaker 1 but not for Speaker 2. This is probably due to this speaker's habit of weakening the last part of an utterance.

		Mean	SD	t (df)	p
Speaker 1	final	170	37	5.1 (136)	<.01
	non-final	149	35		
Speaker 2	final	149	46	0.93 (106)	>.05
	non-final	144	27		

Table 3. Comparison between sentence-final and non-final syllables
(Mean and SD in milliseconds)

Table 4 compares the syllable duration in PWs of different numbers of syllables. This result clearly indicates that, the more syllables a PW has, the shorter the unstressed syllables are. It suggests a tendency toward stress isochrony in Castilian Spanish, a dialect traditionally considered to be syllable-timed. This, of course, is not enough to argue that Castilian Spanish is a stress-timed language, but it would be fair to say, at least, that it is basically a syllable-timed language with a slight tendency to stress-timing.

	Number of syllables	Mean	SD	t (df)	p
Speaker 1	1	209	37	4.90(21)	<.01
	2	165	18	2.86(87)	<.01
	3	155	17	5.61(104)	<.01
	4	139	13		
Speaker 2	1	201	28	6.25(25)	<.01
	2	155	21	2.95(67)	<.01
	3	145	13	4.15(102)	<.01
	4	136	10		

Table 4. Comparison according to the number of syllables
(Mean and SD in milliseconds)

4. F0 CONTOUR

4.1. Fujisaki's F0 Model

In this section, a quantitative model proposed by Hiroya Fujisaki and his colleagues is used to see the relation between stress and F0 movement. This model simulates the F0 contour as the sum of three components: the base line, the phrase component and the accent component. The base line is constant for each speaker. The phrase component is yielded by one or more impulse-like phrase commands. The accent component is produced by one or more step-like accent commands. The contour of (logarithmic value of) F0 is calculated by the following equation.

$$\log_e F_0(t) = \log_e F_{\min} + \sum_{i=1}^{I} A_{pi} h_1(t - T_{0i}) + \sum_{j=1}^{J} A_{aj} \{h_2(t - T_{1j}) - h_2(t - T_{2j})\}$$

where

$$h_1(t) = \begin{cases} \alpha^2 t \exp(-\alpha t) & (t \ge 0) \\ 0 & (t < 0) \end{cases}$$

$$h_2(t) = \begin{cases} \min[1 - (1 + \beta t)\exp(-\beta t), \theta] & (t \ge 0) \\ 0 & (t < 0) \end{cases}$$

In these equations,
$F_{\min}$ indicates the baseline value of an F0 contour,
I is the number of phrase commands,
J is the number of accent commands,
A_{pi} is the magnitude of the i-th phrase command,
A_{aj} is the amplitude of the j-th accent command,
T_{0i} is the timing of the i-th phrase command,
T_{1j} is the onset of the j-th accent command,
T_{2j} is the end of the j-th accent command,
α is the natural angular frequency of the phrase control mechanism to the phrase command,
β is the natural angular frequency of the accent control mechanism to the accent command, and
θ is a parameter to indicate the ceiling level of the accent component.

To obtain a good prediction, it is essential to assign appropriate values for all the above variables.

This model has been reported to be applicable to Spanish as well as to Japanese, English and many other languages [5]. Figure 1 is an example of application of this model to one of the utterances used in this study. For Spanish, no attempt has been made so far to investigate the relation between the timing of the stressed syllables and that of the accent commands (T_{1j} and T_{2j}), which can be very complicated. In this part of my study, the model was applied to the same material used in the previous section to see how the timings were related.

4.2. Procedure

F0 values were measured at every 10 milliseconds and used as the raw data. To facilitate the application of the model, the following conditions were assumed to be met.

- The values for the variables α, β and θ are fixed to 2.8, 25 and 0.9 respectively [5].
- The $F_{\min}$ value is fixed to 105Hz for Speaker 1, and 122 Hz for Speaker 2.
- Only one phrase command occurs in each utterance. One accent command occurs in a one-PW sentence, and two accent commands occur in a two-PW sentence.
- The polarity of the accent command is restricted to positive.

Based on these assumptions, the values for the variables not fixed were first determined provisionally by visual observation using a graphic edit tool, and then corrected with Analysis-by-Synthesis method [6].

4.3. Results

The following symbols will be used hereafter (j=1, 2). See also Figure 1 for reference.

T_{1j} : the onset of the j-th accent command,

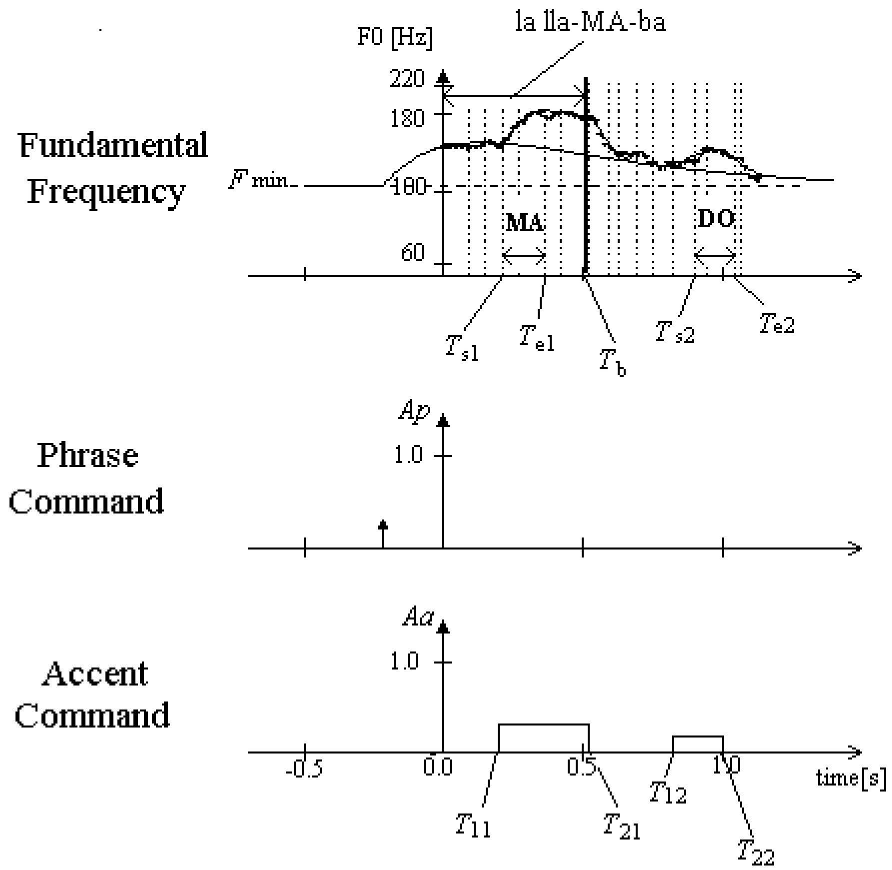

Figure 1. Application of the F0 model to the utterance "La lla-MA-ba / re-no-va-DO-ra." (Speaker 1)

T_{2j} : the end of the *j*-th accent command,
T_{sj}: the onset of the stressed syllable in the *j*-th PW,
T_{ej}: the end of the stressed syllable in the *j*-th PW,
T_b: the PW boundary (only in two-PW sentences).

4.3.1. Overall Tendencies. Table 5 shows the result of the comparison between T_{11} and T_{s1}, and T_{21} and T_{e1} in the one-PW sentences. For example, "$T_{11} - T_{s1} = -37$" means that the accent command occurs 37 milliseconds before the stressed syllable begins. Table 6 compares T_{12} and T_{s2}, and T_{22} and T_{e2} in the second PW of two-PW sentences. The parts of the utterances treated in these two tables are both the last PWs of declarative sentences.

The results shown in these two tables are quite similar to each other. The timing of the accent command and that of the stressed syllable roughly match: the accent command begins approximately 40 milliseconds before the stressed syllable begins, and ends some 20 milliseconds before the stressed syllable ends.

When the last syllable of the sentence is stressed, however, the above observation does not hold. In these cases the accent command ends much earlier than the stressed syllable ends (some 80 milliseconds before). This result agrees with the perception that the pitch falls sharply in the middle of the sentence-final stressed syllable.

	T_{11}-T_{s1}		T_{21}-T_{e1} (last syllable unstressed)		T_{21}-T_{e1} (last syllable stressed)	
	Mean	SD	Mean	SD	Mean	SD
Speaker 1	-37	65	-39	20	-63	73
Speaker 2	-48	53	-11	20	-83	33

Table 5. Time gaps between accent commands and stressed syllables (1-PW sentences) (Mean and SD in milliseconds)

	T_{12}-T_{s1}		T_{22}-T_{e2} (last syllable unstressed)		T_{22}-T_{e2} (last syllable stressed)	
	Mean	SD	Mean	SD	Mean	SD
Speaker 1	-29	49	-43	45	-102	55
Speaker 2	-58	96	+4	52	-76	53

Table 6. Time gaps between accent commands and stressed syllables (second PW in 2-PW sentences) (Mean and SD in milliseconds)

In the first PW of two-PW sentences, a quite different result is obtained, as is shown in Table 7. Here the accent command ends much later than the end of the stressed syllable, approximately at the PW boundary. This result agrees with the perception that, in these circumstances, the high pitch does not end with the stressed syllable but continues until the PW boundary.

	T_{11}-T_{s1}		T_{21}-T_{e1}		T_{21}-T_{b}	
	Mean	SD	Mean	SD	Mean	SD
Speaker 1	-9	51	+95	116	-27	61
Speaker 2	+13	53	+152	123	+51	104

Table 7. Time gaps between accent commands and stressed syllables / PW boundaries (first PW in 2-PW sentences) (Mean and SD in milliseconds)

4.3.2. Isochronous Tendency. In addition to these basic facts, an interesting rhythmic phenomenon was observed in two-PW sentences. Suppose that K is the number of syllables after the stressed syllable in the first PW, and L the number of syllables before the stressed syllable in the second PW. For instance, in the sentence "La **nó**-ma-da / me lo di-**rá**." ("The nomad will tell me that."), K=2 and L=3. There is a relation between K and T_{s2} (the onset of the second accent command), and between L and T_b (the PW boundary), as shown in Figures 2 and 3, respectively. This reveals a clear tendency for the interval between the two accent commands to be longer when it is too short, and to be shorter when it is too long. This suggests the existence of a tendency to isochronous occurrence of accent commands.

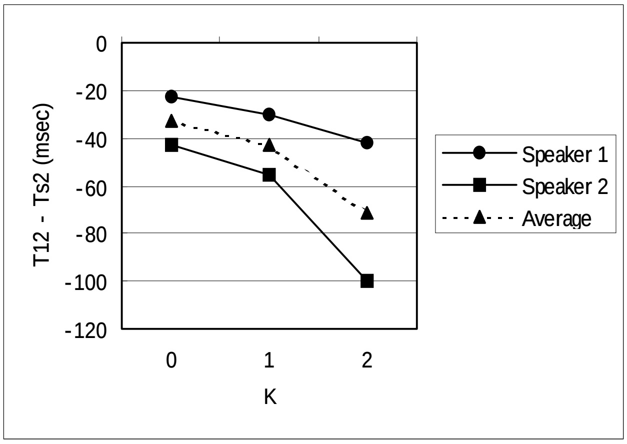

Figure 2. ($T_{12} - T_{s2}$) as a function of K

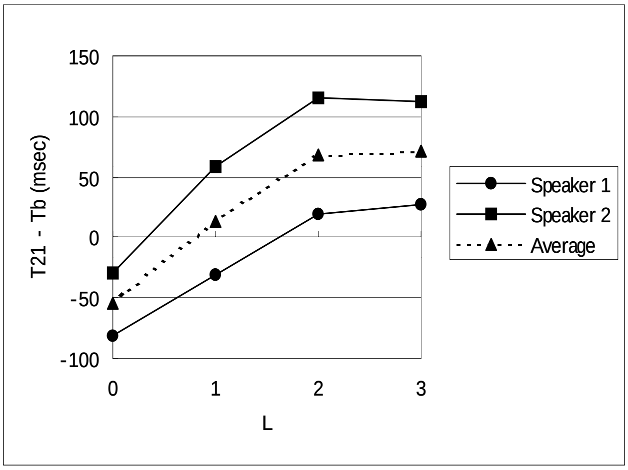

Figure 3. ($T_{21} - T_b$) as a function of L

5. CONCLUSION

This research has demonstrated that there are tendencies toward isochrony in Castilian Spanish, both in terms of stress and pitch movement. It should be kept in mind, however, that the sentences treated here represent the simplest intonation pattern in the language. Investigations on longer and more complicated sentences have yet to be done.

ACKNOWLEDGMENTS

I wish to thank Kathleen B. Riley and Takehiko Makino for reading the draft and making a number of helpful suggestions. I am also indebted to the two Spanish speakers for their collaboration and patience during the recording sessions.

NOTES

1. Castilian Spanish is a dialect of Spanish, widely spoken in central Spain, including Madrid, Salamanca, Burgos and Valladolid. This dialect is usually considered to be Standard Peninsular Spanish.

REFERENCES

[1] Hoequist, C. 1983, Syllable duration in stress-, syllable- and mora-timed languages, *Phonetica*, **40**, 202-237.

[2] Toledo, G. A. 1988, *El ritmo en el español*, Madrid, Gredos.

[3] Quilis, A. 1980, *Fonética acústica de la lengua española.* Madrid, Gredos.

[4] Kimura, T. 1993, Estudio experimental sobre unas configuraciones melódicas del habla española, *Actas del Tercer Congreso de Hispanistas de Asia*, 205-215, Asociación Asiática de Hispanistas.

[5] Fujisaki, H., Ohno, S., Nakamura, K., Guirao, M., Gurlekian, J. 1994. Analysis and of Accent and Intonation in Spanish Based on a Quantitative Model. *Proceedings of 1994 International Conference on Spoken Language Processing*, 355-358.

[6] Furui, S. 1985. *Digital Sound Processing* (In Japanese), Tokyo, Tokai University Press.