Alignment of F0 model parameters with final and non-final accents in Argentinean Spanish

Jorge Gurlekian[1], Hansjörg Mixdorff[2], Diego Evin[1], Humberto Torres[1], Hartmut Pfitzinger[3]

[1] Laboratory of Sensory Research. University of Buenos Aires, Argentina
[2] Department of Informatics and Media, BHT University of Applied Sciences, Berlin, Germany
[3] Institute of Phonetics and Digital Speech Processing, University of Kiel, Germany

jag@fmed.uba.ar,mixdorff@beuth-ochschule.de,diegoevin@gmail.com,hmtorres@fmed.uba.ar
hpt@ipds.uni-kiel.de

Abstract

The goal of this study is to explore the association between tonal accents and fundamental frequency parameters obtained from the Fujisaki model in Buenos Aires Spanish. Results indicate that three-syllable words in final position which are stressed on the third syllable are associated with early peaks. In non-final word accents, late peaks are found for words stressed in the first and second syllables. Alignments were calculated for accent commands relative to the onset of both the accented syllable and its nucleus. Similarly to an earlier study on German, anchoring is supported in favor of tonal transitions than of F0 peaks, considering that alignment of onset and offset times of accent commands appear earlier for final than for non-final accents.

Index Terms: Fujisaki model, prosodic labeling, tonal accents.

1. Introduction

This work replicates a recent study presented by Mixdorff and Pfitzinger [1] for German where different associations were presented between accent labels, parameters of the Fujisaki model [2], intoneme classes according to Stock and Zacharias [3], and focus condition. In the current study the associations to be studied are between the Fujisaki accent command parameters and accents in different word positions within the phrase and in different syllables within the word. Our ultimate purpose is to produce an improved intonation contour from plain text for TTS systems, but also to contribute to the quantitative description of Argentine Spanish intonation.

1.1. Buenos Aires Spanish tonal accents

Two levels of prosodic structure have been found to occur in this Spanish variety: intermediate phrases (ip´s) and intonation phrases (IP). Tonal accents in Buenos Aires Spanish present some typical characteristics which result from the interaction between Spanish and Italian due to immigration in the early twentieth century [4]. Following Sp-ToBi conventions, declarative sentences present a variety of tonal accents in non-final position (H*, L+>H*, L+H*), and H+L* or L* in final position, with a predominance of the latter. A typical configuration for declaratives is L+>H* H- L+H* L* LL%. A distinctive contrast exists between L+H* peaks - where the F0 peak is located in the syllable nucleus - and L*+H or L+>H* (called 'late peaks'), where it is delayed to one or two syllables after the accented one, respectively.

1.2. Speech database

Intonation patterns of Argentine Spanish were collected and analyzed within the international research project AMPER (Multimedial Prosodic Atlas on Romance Areas) [5,6].
The utterances in the corpus have the structure Subject-Predicate-Object (NP + VP + PrepP). The NP contains three-syllable words with pitch accents and lexical stresses on the first (*el* ***trián****gulo* 'the triangle'), second (*la gui****ta****rra* 'the guitar'), or third syllable (*el saxo****fón*** 'the saxophone'), henceforth denoted as types 1:3, 2:3, and 3:3 respectively.
The VP is common to all sentences: *se toca* 'is played'. The PrepP also has pitch accents and lexical stresses in three-syllable words (*con* ***pá****nico* 'with panic', type 1:3), (*con me****su****ra* 'with moderation', type 2:3), (*con obsesión* 'with obsession', type 3:3). The corpus was produced by eight female speakers of Buenos Aires Spanish. The different types of sentences were elicited in a semi-spontaneous way, that is, we presented the instruments and the actions in several contexts to the speakers and they were asked to reply according to each situation.

2. Experiment Design

Nine declarative sentences were created by combining "el triángulo/la guitarra/el saxofón" which are the non-final accent words, with /se toca con /pánico/obsesión/mesura", which are the final accent words. (see Figure 1). Three repetitions of the nine sentences by eight speakers constitute a total of 216 sentences that were employed in this work. All speakers are female and half of them have no college education according to AMPER directives and the other four do.

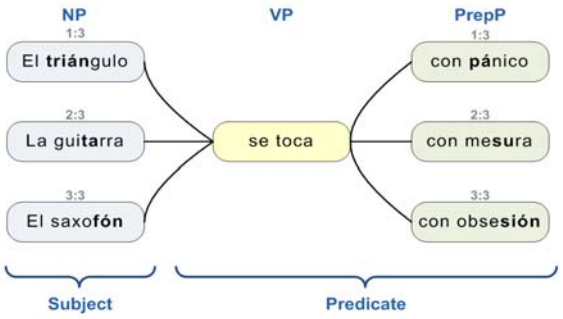

Figure 1. Grammar for the generation of the set of analysis sentences.

The AMPER corpus was labeled according to Sp-ToBI [7] by two trained linguists using a software tool that displays the F0 contour (getF0), and legal phonetic (Arg-SAMPA), graphemic, tonal, break and syllabic symbols, and produces different files for further processing. Figure 2 shows an example of the sentence "El triángulo se toca con pánico." The figure displays from the top to the bottom: The speech waveform, the F0 contour, the intensity contour and the associated segmental and prosodic labels. In this case we find a late F0 peak in the non-final accent for the word "triángulo" and a final accent with an early peak for the word "pánico".

Fujisaki parameters were extracted using the automatic method by Mixdorff [8] and verified with the *FujiParaEditor* [9] tool to obtain the stepwise accent commands associated with accented syllables. Accent commands are described by onset and offset times *T1* and *T2*, amplitude *Aa* and time constant *beta*. Different base frequencies *Fb* were calculated automatically for each of the eight speakers.

Finally a program was developed to find the nearest accent command for each of the accented syllables. We measured and averaged the distances between onset times *T1* and offset times *T2* of accent commands relative to both syllable and vowel onset. If cases with fewer accent commands than accented words, the nearest onset or offset times were selected for the analysis.

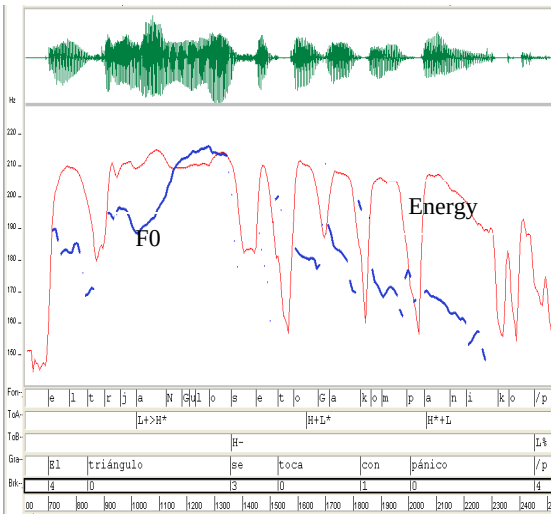

Figure 2. Sentence "El triángulo se toca con pánico". It shows a typical non-final and final accent in a declarative sentence as described in [5]

.

3. Results

Perceptual analysis of the Amper corpus is presented in Table 1. Averaged alignments are presented in Figures 3 (a, b) and 4 (a, b). Alignments according to syllable type and word position can be seen in Figure 5 (a, b).

The main difference between the two groups of speakers (with and without tertiary education) is related to the amplitude of F0 excursion. Educated speakers have more than one octave difference between the lowest and highest F0 values.

Most speakers produce two phrases separated by a ´3´ break. The first phrase indicates a continuation and is therefore an intermediate phrase which can be associated with the non-terminal intoneme, where the non-final accent word is located. The second one is a final phrase that could be associated to the information intoneme and where the final accent word is located [1].

Results of alignments are averaged for the eight speakers given that an analysis of variance of accent command parameters T1 and T2 indicated non-significant differences between the two groups of speakers. For non-final accent words, late peaks are clearly more frequent in all stressed syllables which have a following context syllable. If the accent is in the last syllable before the boundary of the intermediate phrase, all the peaks fall into that syllable. (see Table 1).

Final accent words present a high percentage of occurrences of early peaks if the word has stress in the last syllable. Early peaks also occur for the other types but the percentage distribution show more variation within speakers with a predominance of medial peaks (see Table 2). Besides, most of the peaks have low amplitude.

Table 1. Percentage of non-final tonal accents and boundary tones for each syllable position in the word.

Word type	**Non- final Tonal Accents**	**Peak**	**Phrase Accent**
3:3	**L+H* 100%**	Medial	**H- 92%** **M-** 8%
2:3	**L+>H* 83%** L+H* 17%	Late	**H- 83 %** L- 17%
1:3	**L+>H* 83%** L+H* 17%	Late	**H- 75%** L- 25%

Table 2. Percentage of final tonal accents and boundary tones for each syllable position in the word.

Word type	**Final Tonal Accents**	**Peak position**	**Boundary Tone**
3:3	**H+L* 92%** L+H* 8%	Early Medial	**L-L% 100%**
2:3	**L+H* 62.5%** **L+>H* 21%** H+L* 12.5% H*+L 8%	Medial Late Early Early	**L-L% 100%**
1:3	**L+H* 42%** **H+L* 33%** L+>H*12.5% H*+L 12.5%	Medial Early	**L-L% 100%**

Average measurements of *T1* and *T2* relative to the syllable onset are presented in Figure 3.

For non-final accent words, as shown in Figure 3a, the accent command overlaps with the syllable and the peak appears at the end of the syllable or even after.

For final accent words, as shown in Figure 3b, the accent command begins before the syllable onset and the resulting *F0* contour shows a peak at the beginning or the middle of the syllable.

To see the correspondences more precisely we presented the same figures relative to the syllable nucleus in Figure 4. It can be observed in Figure 4a that the F0 contour decay is overlapping with the vowel -indicated as a shadowed rectangle-, and in Figure 4b, it is the F0 contour onset that overlaps with the vowel. This suggests that the tonal change is better associated with the vowel segment than with the F0 peak itself.

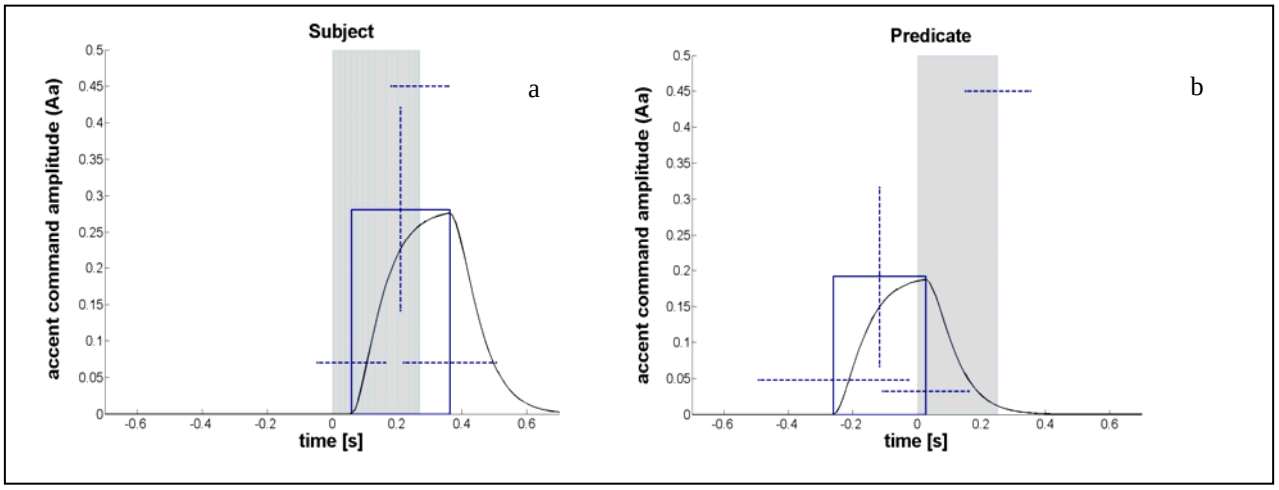

Figure 3. Means and Standard Deviations (whiskers) of accent command alignment relative to the syllable onset, as well as accent accent command amplitude *Aa*. (a) Non-final accents, (b) Final accents, Shadow regions indicates the mean duration of syllables.

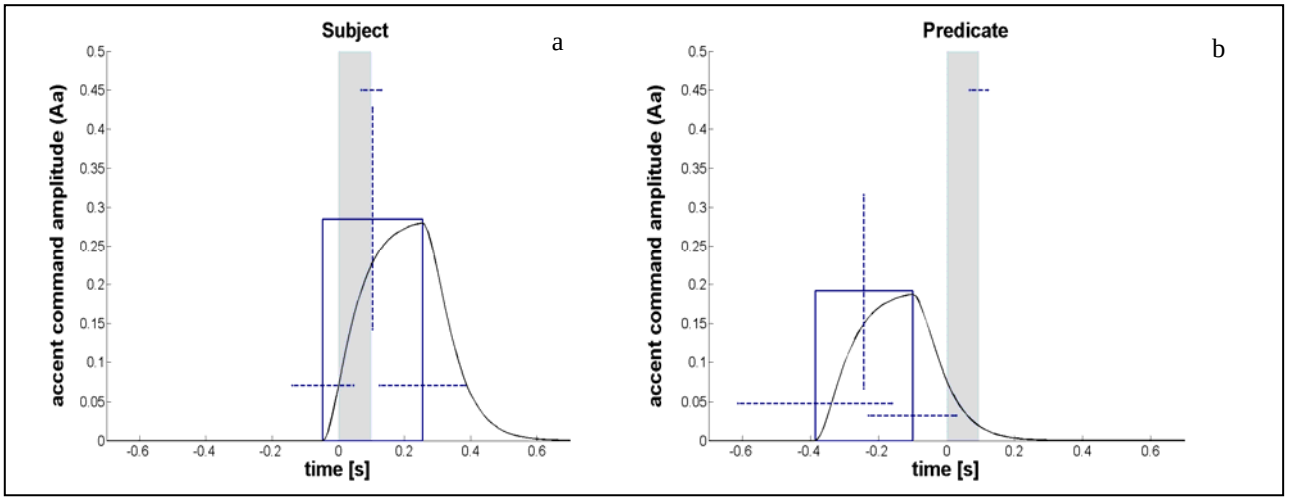

Figure 4. Means and standard deviation (whiskers) of accent command alignment relative to vowel onset, as well as accent command amplitude *Aa*. . (a) Non-final accents, (b) Final accents. Shadow regions indicates the mean duration of vowels

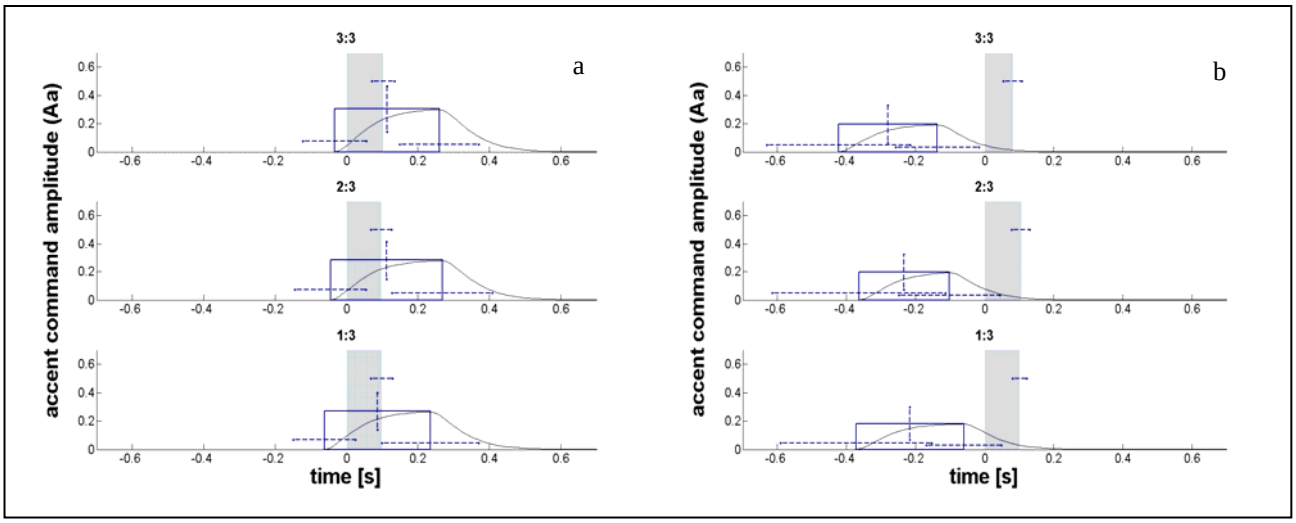

Figure 5. Means and standard deviations of accent command alignment relative to the vowel onset, as well as accent command amplitude *Aa*. From the top to the bottom: accent on third, second, and first syllable respectively. (a) Non-final accents, (b) Final accents Shadow regions indicate the mean duration of vowels.

Furthermore, we separated the results by stressed syllable positions, to explore the relative differences between offset and onset accent command relative to vowel onset. In Table 3 and Figure 5a, T1 is progressively apart from the vowel onset. In the same table and in Figure 5b, it can be seen that T2 is progressively close to the vowel onset for final accent words. These orderings may indicate an apparent one to one correspondence of T1 and T2 associated to syllable type. Interestingly T1 and T2 in those accent positions have smaller standard deviation than *T1* and T2 for the counterpart positions. See Table 3. In order to verify this detailed association we performed an analysis of variance of T1 and *T2* for the three syllable positions. For T2 in final accent words the analysis shows that the alignments are significantly different, (df=2, F=6.38, p=0.002), but not for *T1* in non-final accent words.

Table 3. Mean and standard deviation of accent command parameter values T1 and T2 in ms, for each type of stressed syllable for final and non-final accent words.

Accent	Stress	T1	SD	T2	SD	N
Non-Final	**3:3**	**-33**	89	260	111	72
	2.3	**-45**	100	268	141	72
	1:3	**-62**	87	235	136	72
Final	**3:3**	-423	208	**-137**	120	72
	2.3	-364	250	**-102**	146	72
	1:3	-372	219	**-60**	109	72

4. Discussion and Conclusions

We are interested in the association between the tone and the accented syllable - which is called the primary phonological association - as opposed to the secondary phonological association between the tone and the phrase accent [10]. If the acoustic indicators of tonemes are included they will help to recover the contrast. In summary, tonal accents could carry information of the two processes. Quantitative results indicate an earlier alignment of the accent relative to both the syllable/vowel as opposed to the later alignment for non-final accents. Hence there is a stronger connection between early peak and final word accent, and between late peak and non-final word accent. More precisely, the beginning of the vowel is associated with the offset time T2 of the accent command in final accents, and it is associated with the onset time *T1* of the accent command in non-final accents. Offset time *T2* relative to the vowel onset time distinguish between stressed syllable positions in final accent words. For non-final accent words T1 distinguish between syllable stress positions. Variations due to several factors should still be investigated, such us the presence of low phrase accents, the tendency to produce doubly accented words and to explore other differences between our groups of speakers. Regarding the secondary phonological association, speakers produced two intermediate phrases (ip) separated with a break 3 (see Table 1). Our hypothesis is that in a non-final ip the last pitch accent is influenced by phrase accent and in the final ip the last pitch accent is influenced by phrase accent as well as the boundary tone of the intonational phrase [11]. Present results show that the effect of high phrase accent H- on the non- final H* is to attract it and produce a late peak. The effect of a low boundary tone L% on the final H* is to push it, and make it appear early in the syllable. Similarly to this last result, Shue et al, [12] claimed an early F0 peak for the H* whenever a boundary tone is in the same word supporting the tonal crowding hypothesis. When the nuclear pitch accent is L* and the boundary tone remains L% they reported no such effect. Moreover our alignment results for H* relative to H- in the same word, raise the question if equal tones tend to be attracted and different tones to be separated to enhance tonal contrast.

These results are also in close agreement with the study by Mixdorff and Pfitzinger [1], regarding F0 alignments of non-final and final accents to distinguish between non-terminal and information intonemes. “Non terminal” intoneme is related to a high boundary tone and “information” intoneme is related to a low boundary tone as in declarative sentences [3].

Our findings, that distinguish final and non-final phrases for declarative sentences, could be extended to explore interrogative sentences as well, and to verify the pitch accent interaction between both phrase accents and boundary tones.

5. Acknowledgements

Authors are grateful to R Yanagida and M Tripodi for labeling the corpus. This work was supported by MINCyT and BMBF research agencies of Argentine and Germany respectively.

6. References

[1] Mixdorff, H. and Pfitzinger, H. “A quantitative study of F0 peak alignment and sentence modality”. In Proc. of Interspeech 2009, pp.1003-1006. Brighton, Sept. 2009.

[2] Fujisaki, H. and Hirose, K. “Analysis of voice fundamental frequency contours for declarative sentences of Japanese”, J. Acoust. Soc. Jap. (E)5(4), 233-241. 1984.

[3] Stock E., Zacharias, C., “Deutsche Satzintonation“, VEB Verlag Enzyklopädie, Leipzig, 1982.

[4] Colantoni, L. and Gurlekian, J. “Convergence and intonation: historical evidence from Bs. As. Spanish”. *Bilingualism:Language and Cognition*, 7: 107-119. 2004.

[5] Fernández Planas, A. “Aspectos generales acerca del proyecto Internacional AMPER en España”, Estudios de Fonética experimental XIV: 13-27. 2005.

[6] Gurlekian, J. and Toledo, G. “Datos preliminares del AMPER Argentina: Las oraciones declarativas e interrogativas absolutas sin expansión”. Language Design, J. of Theoretical and Experimental Linguistics Special Issue: Experimental Prosody 2: 213-220. 2008.

[7] Vilaplana, E. and Prieto Vives, P. “La notación prosódica del español: una revisión del sp_ToBI”. *Estudios de Fonética Experimental* XVII: 263-283. 2008.

[8] Mixdorff, H. “A novel approach to the fully automatic extraction of Fujisaki model parameters”. In Proc. of ICASSP 2000, vol 3, pp1281-1294. Istanbul, June 2000.

[9] Mixdorff, H. (1/10/2009). FujiParaEditor. http://public.beuth-hochschule.de/~mixdorff/thesis/fujisaki.html

[10] Toledo, G. “Fonología entonativa: los acentos tonales finales de frase entonativa intermedia (ipT*) frente al tono de frontera (H-) en discursos y texto leídos del español de Buenos Aires”. Language Design, J. of Theoretical and Exp. Lin. 9,2: 129-136. 2008.

[11] Toledo, G. “Fonología autosegmental: contraste entre tonemas ascendentes Intermedios y descendentes finales en el fraseo entonativo del español”. Langues et Linguistique, Vol 32: 149-180. 2008.

[12] Shue, Y-L., Shattuck-Hufnagel, S, Iseli, M, Jun, S-A, Veilleux, N, Alwan,A. “Effects of Intonational Phrase Boundaries on Pitch-Accented syllables in American English. Speech Communication, 52, 2:106-122. 2010.