

Contribution of prosody to the perception of a foreign *accent*: a study based on Spanish/Italian modified speech

Bianca Vieru-Dimulescu, Philippe Boula de Mareüil

LIMSI-CNRS, BP 133, 91403 Orsay, France

bianca@limsi.fr, mareuil@limsi.fr

Abstract

In this article, we attempt to quantify the contribution of prosody (timing and melody) to the perception of a foreign *accent*, by using a prosodic transplantation paradigm. For this experiment, a dozen sentences, which are spoken in (almost) the same way in Italian and Spanish, were recorded. They were read by 3 Spanish monolinguals, 3 Italian monolinguals and 3 Spanish/Italian bilinguals (once in the Spanish way, once in the Italian way). Prosody crossings were performed, and perceptual tests were run with Italian, Spanish and French listeners. Results suggest that prosody outweighs segmental cues in the perception of Spanish-accented Italian and Italian-accented Spanish.

1. Introduction

This article aims at gaining insight into the role of prosody in the perception of Spanish-accented Italian and Italian-accented Spanish, by using modification of natural speech. What is commonly meant by “foreign *accent*” can be viewed as the effect of a confrontation between two phonological systems in contact, from a native language (L1) and a second language (L2). There is a large body of evidence that a foreign *accent* depends on the total amount, frequency and circumstances of L1/L2 use, on the length of residence in a predominantly L2-speaking environment and on the state of neuro-cognitive development at the time of first exposure to L2 [1] [2]. Why is an *accent* kept in a second language? What causes an immigrant or an L2 learner to retain an *accent* is undoubtedly a crucial and complex question. Another question is what characterises this foreign *accent* physiologically, acoustically and perceptually. It may rely on the phonemic content and on the instantiation of co-articulation phenomena, voice quality, filled pause and hesitation phenomena, melodic *clichés*, pitch range, pitch level at phrase boundaries, durational profiles, stress location errors or other shibboleths, which are interpreted as “foreign”. The combination of fundamental frequency (henceforth, F_0) changes and the pronunciation of certain phonemes, for example, may even result in the perception of a foreign *accent*, while each cue taken separately may not.

New technologies permit us to study the contribution of prosody to the perception of foreign *accent*. Speech synthesis, in particular, allows us to sort out the influence of segmental vs. suprasegmental features (*i.e.* the phoneme string vs. prosody) in what is perceived as a foreign *accent* (e.g. Spanish or Italian). Speech synthesis is a good tool to make allowances, since it enables us to monitor a number of parameters, without presenting as many drawbacks as low-pass filtering: this is the reason why it has been used for delexicalisation and monotonisation purposes [3]. Speech

synthesis has been used, together with simulated or altered speech in research on foreign *accent* [4] [5] [6] [7].

This article carries on with recent experiments reported in [8] and [9], which were made using different languages and different techniques. Using diphone speech synthesis, [8] showed a greater effect of prosody in the perception of Spanish/Italian foreign *accent*, whereas [9], based on modified natural speech, showed a greater effect of segmental cues in the perception of Maghrebien-accented French. The goal is here to verify whether the results obtained originate in the L1-L2 pairing rather than in the employed technique. In the present study, the corpus is the same as in [8]: a dozen sentences which sound alike in Spanish and Italian. These sentences were read by speakers of Spanish and Italian, and their prosody was modified. Perceptual tests were administered to different samples of individuals: Italian, Spanish and French subjects. Results of this experiment suggest that prosody plays a more important role than do segmental cues in the perception of Spanish/Italian foreign *accent*.

2. Experiment

2.1. Material and methodology

To analyse the perception of Spanish/Italian *accent*, we used the same corpus as is in [8]; 14 sentences of 15 syllables on average, which are almost spoken in the same way in Italian and Spanish (see Tab. 1). We tried to minimise the lexical bias which would allow language identification, and we intended to examine what is perceived when we cross the segmental and suprasegmental contents of these two languages.

Table 1: sentences written with Spanish conventions.

1	Al teléfono, Antonio manifestó poca simpatía.
2	La música dura sólo un minuto.
3	Debo arrestarle e identificarle dentro.
4	¿ Ha visto la casa del presidente americano, si o no ?
5	Lentamente, Marina canta “Talla la leña”.
6	La bomba atómica era un problema político.
7	Te lo dirá María que perdono al médico.
8	La mía protesta tristemente cuando bebo tanto vino.
9	La persona que viene sale con un alpinista.
10	Mario compra un piano a crédito.
11	La línea verde señala un itinerario fantástico.
12	Un baño fresco lava naturalmente poco.
13	Un taxi, qué sorpresa! un autobús, qué fenómeno!
14	La polaca prepara la lista.

The sentences were read by 3 Spanish monolinguals, 3 Italian monolinguals and 3 bilinguals. All the talkers were young. The native Spanish bilingual female who was from Madrid and the native Italian bilingual female lived in Pisa; the native

Spanish bilingual male lived in Madrid. The Italian and Spanish monolinguals whose voices were used were also either from Pisa or from Madrid respectively. The recordings were made in the subjects' region of residence, in soundproof booths, under comparable conditions. The data (three repetitions of each utterance per speaker, on average) were transferred with a sampling frequency of 16 kHz and a 16 bit resolution, mono, for further processing (e.g. segmentation into sentences, energy normalisation).

In order to copy the prosodic parameters of an utterance onto another one, a script was written for the PRAAT software, (<http://www.fon.hum.uva.nl/praat/>) which enables speech handling and re-synthesis with the help of the PSOLA algorithm transplantation, phoneme by phoneme, of duration and then pitch. The sentences of the corpus were segmented into phonemes and pauses, by using automatic segmentation and possibly hand-correcting the result. The procedure, whose validity was proved [10], makes use of speech recognition, or more exactly dynamic alignment of continuous-density Hidden Markov Models (HMMs) which are language-dependent. As for F_0 , it was extracted thanks to PRAAT. The transplantation of the duration and F_0 parameters amounts to: (1) checking that the numbers of segments in the two sentences of a given pair are equated; (2) calculating lengthening or shortening coefficients for each phoneme or pause of a speaker with respect to another one; (3) for each phoneme or pause, building new durations which will replace the original durations of either sentence; (4) grafting F_0 from one of the sentences onto the other one, and vice versa. For each pair of sentences in input, two new stimuli are thus obtained: e.g. a Spanish voice speaking with an Italian prosody and conversely.

In addition to cross-language combinations, we crossed each original voice with the prosody of the same language of another speaker, to make sure that the impression of foreign *accent* does not stem from signal manipulation. One repetition of each original was added, to check the responses' consistency. After all, we selected 100 stimuli so as to avoid coverings, to maximise the test diversity: 4 stimuli before counting the results + 2x4x3 originals + 2x4x3 repeated originals + 2x4x3 cross-language crossings + 2x4x3 within-language crossings = 100 stimuli.

2.2. Perceptual tests

The perceptual test is performed via the Internet, through a user-friendly interface (<http://bianca.vieru.free.fr>) which we developed. A short familiarisation phase precedes the test. The instructions follow, written in English, French, Italian and Spanish: through headphones, listen to sentences which are spoken in the same way in Spanish and Italian, read by more or less bilingual speakers, and possibly modified acoustically. Listeners were then requested to focus their attention on the notion of foreign *accent* more than on possible deteriorations of the recording. The subjects, who had to have no known hearing problem, were not paid for this task. They were also asked for their gender, age, place of residence, and familiarity with Spanish and Italian.

After the presentation of each sentence, the task for the natives (Spaniards and Italians) consisted of judging whether the utterance was Spanish, Spanish with an Italian accent (*i.e.* Spanish spoken by an Italian), Italian with a Spanish accent

(*i.e.* Italian spoken by a Spaniard) or Italian. For French listeners, we had to adapt the instructions, as we found it too difficult to ask them to judge whether what they heard was Italian-accented Spanish or Spanish-accented Italian. Nevertheless, we kept a 4-alternative forced-choice by introducing a confidence level: the French subjects were requested to tell if the mother tongue of the speakers they heard was "very probably Spanish", "probably Spanish", "probably Italian" or "very probably Italian". The precision "mother tongue" was brought to the French listeners, who did not understand the meaning of the sentences as the Italian and Spanish listeners could, in case they would have detected a mismatch between the two languages. We expected the French participants' answers to be facilitated when the voice and the type of prosody matched.

The listeners could listen to the sentences as many times as they liked, but could not come back to them after answering. The stimuli were presented in a random order, which changed for each listener. The results of each test were sent directly through the Internet, and were stored in a MySQL database in an XML format, for further processing.

2.3. Acoustic analysis of the corpus

Each test sentences lasted 2 seconds on average: the speech rate was 16 phonemes/second for the Italian sentences and 17.8 phonemes/second for the Spanish sentences. Indeed, the duration ratio between stressed and unstressed syllables is 1.4 for the Italian language, and only 1.1 for the Spanish language, which is coherent with previous work. Mean pitch and pitch range (defined as $12\log_2(F_{0,max}/F_{0,min})$) is comparable across the two languages: 134 Hz, 13 semitones for the male voices; 232 Hz, 16 semitones for the female voices. Accordingly, the temporal phenomena seem to be more salient, more relevant to discriminate Spanish and Italian, which does not exclude that a specific melodic pattern such as a continuation rise on a proparoxytone may be typical of the Spanish language (see Figure 1). The phonetic realisations of a prosodic structure may be specific to each language.

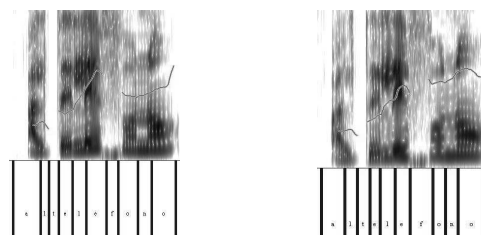


Figure 1: Beginning of sentence 1, read in the Italian manner (left) and the Spanish manner (right).

If we examine the pitch curve of stressed syllables, we can observe certain differences between Italian and Spanish. For some sentences, the pitch curve is rising in Spanish and falling in Italian. In other sentences, the pitch curve of Spanish forms a plateau and the pitch curve of Italian are descending. But these cases are rare, not always systematic nor validated perceptually.

Other clues, of a segmental nature, more punctual and sometimes very subtle, also betray the speakers' origin: (1) the hissing of the /s/ into an apical [s], as well as the lenition of the /d/, especially in an intervocalic position (see Fig. 2), in Spanish speakers; (2) the spirantisation of /k/ in Italian and

the so-called *raddoppiamento fonosintattico*, for instance the redoubling of the [m] in *te lo dirà Maria* (“Maria will tell it to you”) in Tuscan speakers; (3) more open mid vowels in some Italian words than in their Spanish counterparts, which was confirmed by formant measurements.

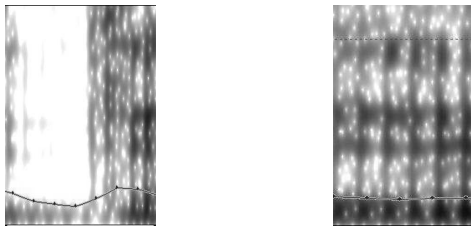


Figure 2: /d/ of *PRESIDENTE* read in the Italian manner (left) and in the Spanish manner (right).

2.4. Listeners

The sentences of the corpus were presented to 60 listeners (32 male, 28 female): 20 Italian, 20 Spanish and 20 French subjects, who were 29 years old on average. They were mainly recruited within newsgroups of Romance language teachers. The Italian subjects were from Rome for half of them; Spanish subjects were either from Madrid or from Barcelona and their surroundings; the French subjects were mainly from the Paris region. Their self-estimated familiarity with Spanish and Italian was checked, but it had no significant influence on their answers according to a χ^2 test.

3. Results

In Tab. 2 and below, Sv:Sp_o and Iv:Ip_o correspond to the original sentences, with Spanish/Italian voice and prosody; Sv:Sp_m and Iv:Ip_m correspond to the modified sentences which result from within-language crossings; Iv:Sp refers to “Italian voice with a Spanish prosody” and Sv:Ip to “Spanish voice with an Italian prosody”. All the results are pooled: we have too few speakers to thoroughly investigate inter-speaker variability.

3.1. Results with Italian, Spanish and French listeners

There is a good correlation between the responses given to the repeated original stimuli (Sv:Sp_o and Iv:Ip_o), at least for the Spanish listeners [$r = .70$] and the Italian listeners [$r = .75$]. Since the task was harder for the French listeners, the weaker (albeit statistically significant) correlation within their responses to the repeated stimuli [$r = .43$] can easily be explained. Even though we were not in a position to control the conditions in which the subjects performed the test, this is a first guarantee that they did it properly.

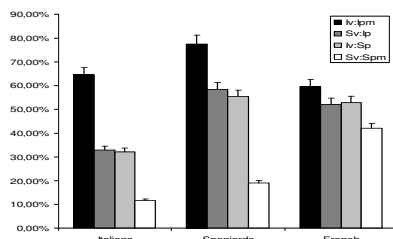


Figure 3. Percentages of responses rated as “Italian” or “Italian-accented Spanish” by Italian and Spanish participants or as “probably Italian” or “very probably Italian”, by French participants, as a function of the stimulus type

Sv:Sp_o and Sv:Sp_m (respectively Iv:Ip_o and Iv:Ip_m) stimuli are identified as Spanish by Spanish listeners (respectively as Italian by Italian listeners) in the majority of cases. The originals and the within-language crossings Sv:Sp_o and Sv:Sp_m (respectively Iv:Ip_o and Iv:Ip_m) are recognised as probably or very probably Spanish (respectively Italian) by French listeners in the majority of cases. Therefore, the deteriorations introduced by the acoustic manipulation of the stimuli do not affect the speakers’ origin identification.

We can notice that the Spanish subjects recognise the Spanish stimuli (Sv:Sp_o and Sv:Sp_m) more easily than Italian stimuli (Iv:Ip_o and Iv:Ip_m), and that the Italian subjects recognise the Italian stimuli more easily than the Spanish stimuli. In all the other cases, the Spanish subjects most often answer “Spanish with an Italian accent” and the Italian subjects “Italian with a Spanish accent”. In this respect, the responses of the Spanish and Italian listeners are symmetrical. The artefacts of the manipulated stimuli, which may give an impression of foreignness, do not suffice to explain this result. It had already been observed with diphone synthesis, as soon as listeners understood the meaning of the sentences.

As our main interest lies in the correlates of foreign accent perception rather than language identification, as we are concerned with the origin rather than the target language, we merged the responses “Spanish” and “Spanish-accented Italian” into one “Spanish-accented” category which we opposed to an “Italian-accented” category regrouping “Italian” and “Italian-accented Spanish” responses as in figure 3. For the French group, we defined the “Spanish-accented” and “Italian-accented” categories collapsing “very probably Spanish” and “probably Spanish” on the one hand, “very probably Italian” and “probably Italian” on the other hand.

Table 2: Responses (%) for Italian, Spanish and French listeners (S = “Spanish”, S-acc. I = “Spanish-accented Italian”, I-acc. S = “Italian-accented Spanish”, I = “Italian”; v = voice, p = prosody, o = original, m = modified).

Italian	S	S-acc. I	I-acc. S	I
Sv:Sp _o	21.7	65.8	5.2	7.3
Sv:Sp _m	24.2	64.2	5.0	6.7
Iv:Sp	12.1	55.8	9.6	22.5
Sv:Ip	11.7	55.4	7.9	25.0
Iv:Ip _o	4.6	32.1	10.8	52.5
Iv:Ip _m	2.1	33.3	12.5	52.1
Spanish	S	S-acc. I	I-acc. S	I
Sv:Sp _o	64.4	9.2	20.6	5.8
Sv:Sp _m	69.2	11.7	14.2	5.0
Iv:Sp	29.6	15.0	43.7	11.7
Sv:Ip	20.4	21.3	40.8	17.5
Iv:Ip _o	15.6	13.1	42.5	28.8
Iv:Ip _m	5.4	17.1	46.2	31.3
French listeners	Very probably S	Probably Spanish	Probably Italian	Very probably I
Sv:Sp _o	25.8	37.3	25.8	11.0
Sv:Sp _m	18.3	39.6	36.7	5.4
Iv:Sp	15.0	32.1	40.8	12.1
Sv:Ip	15.0	32.9	37.5	14.6
Iv:Ip _o	15.2	30.0	34.8	20.0
Iv:Ip _m	12.1	28.3	40.8	18.8

For the Italian listeners, the Iv:Sp and Sv:Ip stimuli are respectively perceived as Spanish-accented in 67.9% (12.1 + 55.8) and 67.1% (11.7 + 55.4) of all cases. For the Spanish listeners, Iv:Sp and Sv:Ip stimuli are respectively perceived as Italian-accented in 55.4% (43.7 + 11.7) and 58.3% (40.8 + 17.5) of all cases. But the differences are not highly significant according to a χ^2 test: therefore they do not enable us to deduce if the predominating factor in the perception of Spanish/Italian *accent* is segmental or suprasegmental. By default, we can say that prosody plays an important role, as compared to the articulation of phonemes and voice quality.

In the crossings, both within-language and cross-language, the French subjects were not as sure of their answers as they were with the originals: they answered “probably” more often than “very probably” (Spanish or Italian). The tendency leans towards an importance of segmental and suprasegmental factors on the same footing.

3.2. Generalisation to other listeners

To be able to generalise these results to a wider sample of individuals, and to test if they may also apply to other sentences, it was necessary to map the responses onto two categories. As described above, each response was classified either as “Spanish-accented” or as “Italian-accented”, which allowed us to compute a percentage of say “Italian-accented” response for each subject and each type of stimulus.

An analysis of variance (ANOVA) with the random factor Subject, the between-subject factor Group (Italian vs. Spanish vs. French) and the within-subject factor Type (Iv:Ip_m vs. Sv:Ip vs. Iv:Sp vs. Sv:Sp_m) was performed on our data. In order to test the generalisation to other sentences, we also calculated mean values for each sentence, and a second ANOVA was conducted using the random factor Sentence. It yielded the same pattern of significant results as the analysis by Subject.

The Group effect is significant [$F(2,57) = 40.1; p < .001$]: the Italian subjects answer “Spanish-accented” more often than do the French subjects and the Spanish subjects (respectively in 63%, 51% and 49% of all cases). We can also observe a main effect of Type [$F(3,171) = 27.8; p < .001$]; the interaction Group×Type is not very significant [$F(6,171) = 2.5; p < .1$]. If the analysis is restricted to the cross-language crossings solely, the Sv:Ip stimuli are more perceived as Italian-accented than are the Iv:Sp stimuli [$F(1,57) = 4.6; p < .01$]. However, the results are not statistically significant for the Italian listeners and for the French listeners. We only have a significant difference for the Spanish listeners [$F(1,19) = 13.1; p < .01$].

4. Discussion and conclusion

This experiment, based on natural speech modification, suggests a comparable role of segmental and suprasegmental factors. Subjects are slightly more influenced by prosody than by the voice and possible segmental errors of bilinguals. This result does not allow us to quantify the exact contribution of prosody to the perception of a foreign *accent*, but it underlines the importance of suprasegmental aspects.

In this experiment, we collected perceptual judgements concerning bilinguals’ natural speech. As in a classical language identification experiment, the problem of individual characteristics and of the role they can play is no nearer

solution. Is an *accent* detected or is it the speaker? This may also depend on who judges. In the case of French listeners of Spanish/Italian utterances, what is assessed is the linguistic representation, one’s image of a foreign language (here from the same Romance group). Since the relative influence of prosody may vary as a function of the L1-L2 pairing, the results of this study will be used in future investigations on Spaniards and Italians speaking French.

To study prosody separately, the methodology which is proposed here could apply to the issue of regional or social *accent* (e.g. rural vs. urban). The domain of prosody in dialectal varieties has already been investigated by using speech modification [11]. But when two forms of the same language are mutually intelligible, the need to be understood is not motivated in the same way. We come up against the matter of the prestige attached to this or that variety, this or that dialect. Working on languages of comparable status such as Spanish and Italian enables us to factor out a whole range of stylistic, historical and social parameters.

5. Acknowledgements

We are indebted to M. Adda-Decker and D. Zhu for the automatic segmentation of our corpus of over 300 sentences.

6. References

- [1] Flege, J.E., Bohn, O.-S., & Jang, S., “Effects of experience on non-native speaker’s production and perception of English vowels”, *Journal of Phonetics*, 25: 437-470, 1997.
- [2] Piske, T., MacKay, I.R.A., & Flege, J.E., “Factors affecting degree of foreign accent in an L2: a review”, *Journal of Phonetics*, 29: 191-215, 2001.
- [3] Ramus, F. & Mehler, J., “Language identification with suprasegmental cues: A study based on speech resynthesis”, *Journal of the Acoustical Society of America*, 105 (1): 512-521, 1999.
- [4] Grover, C., Jamieson, D. G., & Dobrovolsky, M.B., “Intonation in English, french and german: perception and production”, *Language and Speech*, 30 (3): 277-295, 1987.
- [5] Munro, J.M., “Non-segmental factors in foreign accent: Ratings of filtered speech”, *Studies in Second Language Acquisition*, 17: 17-34, 1995.
- [6] Magen, H.S., “The perception of foreign-accented speech”, *Journal of Phonetics*, 26: 381-400, 1998.
- [7] Jilka, M., *The Contribution of Intonation to the Perception of Foreign Accent*, PhD thesis: University of Stuttgart, 2000.
- [8] Boula de Mareüil, P., Marotta, G., & Adda-Decker, M., “Contribution of prosody to the perception of Spanish/Italian accents”, *Speech Prosody*, Nara, 2004 (pp. 681-684).
- [9] Boula de Mareüil, P., Brahimi, B. & Gendrot, C., “Role of segmental and suprasegmental cues in the perception of Magrebian-accented French”, *INTERSPEECH-ICSLP*, Jeju, 2004.
- [10] Adda-Decker, M. & Lamel, L., “Pronunciation variants across system configuration, language and speaking style”, *Speech Communication*, 29: 83-98, 1999.
- [11] Van Bezooijen, R. & Gooskens, C., “Identification of language varieties. Contribution of different Linguistic Levels”, *Journal of Language and Social Psychology*, 18 (1): 31-48, 1999.