
The perception of Spanish lexical stress by French speakers: stress identification and time cost

Sandra Schwab$^{1,2}$ and Joaquim Llisterri$^2$

$^1$Université de Genève – $^2$Universitat Autònoma de Barcelona

Sandra.Schwab@unige.ch – Joaquim.Llisterri@uab.cat

ABSTRACT

The perception of lexical stress in Spanish by French speakers with and without knowledge of the language has been studied with a technique allowing the evaluation of the individual or combined effects of the acoustic parameters related to the perception of stress. Results suggest in first place that the exposure to L2 makes the French speakers more sensitive to stress. Secondly, although $F_0$ seems to constitute the crucial cue in the identification of stress position, results point out that, when stress is accurately perceived, the time necessary to detect it is affected by manipulations involving amplitude.

Keywords: L2 perception, lexical stress, stress ‘deafness’, time cost.

1. INTRODUCTION

One of the most salient features in the production of French speakers learning Spanish as an L2 is the tendency to place the lexical stress in the last syllable of a word or phrase. Since French is a fixed-stress language in which stress generally appears in final position, an accentual transfer seems to take place when Francophone speakers attempt to pronounce proparoxytone or paroxytone words in a free-stress language such as Spanish. It has been hypothesized that the origin of this difficulty can be explained with the metaphor of the ‘phonological filter’ (Troubetzkoy 1939) that would be responsible for an insensitivity to perceive, and therefore, an inability to produce, contrastive stress differences. Along the same lines, the notion of ‘stress deafness’ has been put forward by Dupoux and his coworkers (Dupoux, Pallier, Sebastián and Mehler 1997; Peperkamp, Dupoux and Sebastián, 1999; Dupoux, Peperkamp and Sebastián 2001; Dupoux, Sebastián, Navarrete and Peperkamp 2008; Dupoux, Peperkamp and Sebastián 2010).

The results of a series of studies dealing with the perception of lexical stress in Spanish by Francophone speakers carried out by Dupoux, Peperkamp, Sebastián and other collaborators using different experimental procedures seem to suggest that subjects’ sensitivity to stress placement depends on the cognitive charge required by the task and on the phonetic variability and the lexical status of the stimuli. Moreover, the performance in lexical decision or in repetition tasks does not appear to be strongly influenced by the degree of knowledge of Spanish. Taken together, these experiments lead to the conclusion that French speakers are unable to encode contrastive stress in their phonological representations although they might be capable, in certain tasks, to make use of the acoustic cues which are present in the speech signal. The sensitivity to acoustic cues such as fundamental frequency ($F_0$) has also been shown by Mora, Courtois and Cavé (1997) in an experiment on stress placement in Spanish utterances by French speakers. Their results, together with those of Muñoz, Panissal, Bilières and Baqué (2009), mitigate the idea of a complete stress deafness in Francophone subjects.

The present study intends to shed some more light on the perceptual role of the three parameters involved in the phonetic realization of stress in Spanish ($F_0$, amplitude and duration) by considering the effect of their manipulation in the identification of lexical stress in Spanish isolated words by French speakers. The level of competence in L2, the lexical status of the items presented (words vs. pseudowords) and the accentual pattern of the word (proparoxytone, paroxytone or oxytone) are also taken into account, since they appear to be factors influencing the results of similar experiments (Alfano, Llisterri and Savy 2007; Alfano, Savy and Llisterri 2008, 2009; Alfano, Schwab, Savy and Llisterri 2010).

2. METHOD

The experimental procedure adopted has been initially designed for a study with native Spanish speakers (Llisterri, Machuca, de la Mota, Riera and Ríos 2005) and already followed in other works with non-native subjects (Alfano et al. 2007, 2008, 2009, 2010).
2.1. Participants

Two groups of French speaking participants took part in this experiment: a group with advanced knowledge of Spanish and another one with no knowledge of the language. The advanced group was composed of 10 subjects. They were between 21 and 36 years old and were all raised in a French speaking environment with only one language, French. They had been studying Spanish at University of Neuchâtel (Switzerland) during 6-11 years. The group with no knowledge of Spanish was formed by 10 students of the University of Neuchâtel. They were between 19 and 24 years old and were all raised in monolingual environment. None of them reported good knowledge of Italian, which excludes the eventual bias of knowing a free-stress Romance language.

2.2. Material

The corpus used, taken from Llisterri et al. (2005), was composed of 4 triplets of trisyllabic words (CV.CV.CV) and 4 triplets of trisyllabic analog pseudowords. All words and pseudowords could be proparoxytones (e.g. *número* ‘a number’), paroxytones (e.g. *numero* ‘I number’) and oxytones (e.g. *numeró* ‘he/she numbered’). The corpus was read 10 times by a native Spanish speaker. For each of the three vowels of the target words, the following measure were taken: F0 at the beginning, at the centre and at the end of the segment; amplitude (Ampl) in five equidistant points along the vowel; and, finally, vowel duration (Dur).

The test stimuli were created in the following way: first of all, the original values of F0, amplitude and duration were replaced in each vowel of each stimulus by the values averaged over the 10 repetitions (hereafter, *Base stimuli*); in a second stage, in proparoxytone words, F0, amplitude and duration values for each vowel were replaced by the corresponding F0, amplitude and duration values found in the equivalent paroxytone words (PP>P *Manipulated stimuli*); likewise, in paroxytone words, F0, amplitude and duration values for each vowel were replaced by the corresponding F0, amplitude and duration values found in the equivalent oxytone words (P>O *Manipulated stimuli*). In fact, manipulated stimuli resulted in a shift –to the right– of the accentual information.

The values were modified not only individually, but also simultaneously, obtaining the seven possible combinations of parameters: F0, Ampl, Dur, F0+Dur, F0+Ampl, Dur+Ampl, F0+Dur+Ampl. This strategy has allowed the study of the effects of each acoustic cue both in isolation and in combination with the others. All the manipulations were performed by resynthesis, using the PSOLA algorithm implemented in Praat (Boersma and Weenink 2010).

2.3. Procedure

A total of 136 stimuli (24 base items without manipulation plus 16 x 7 items with manipulations) were presented in the experiment, divided into 4 blocs of 34 items each containing the same number of base and manipulated stimuli and of words and pseudowords.

Subjects performed a stress identification task and were run individually. The stimuli were presented online from a laptop using the DMDX software (Forster 2010), which also recorded the subjects’ responses and their reaction times. Subjects were instructed to listen to each stimulus (e.g. *número* ‘a number’), to make a selection among the three possible choices (e.g. *número* ‘a number’, *numero* ‘I number’, *numeró* ‘he/she numbered’) that appeared in a row on the computer screen, and to press the corresponding button in a response box. The left-to right order of the three choices was always the same across trials: Position 1 corresponded to stimuli with stress on the first syllable, position 2 to stimuli with stress on the second syllable, and position 3 to the stimuli with stress on the third syllable. Each subject received a different randomization of the stimuli.

2.4. Data analysis

Reaction times (RT) were measured from the beginning of the stimuli. RTs inferior to 200 ms were removed (0.18% of the data set) and missing values were not replaced (3.3% of the data set). In order to avoid a possible bias, stimuli duration was subtracted from reaction times. Only RTs on correct responses (n=334) are examined in this paper. It should be noted that ‘correct’ in this context means that the subject has identified the intended position of the stress (i.e. on the second syllable in PP>P manipulations and on the last syllable in P>O manipulations). Reaction times have been analyzed using mixed-effect models (Baayen, Davidson and Bates 2008), since they do not only account for the fixed-effects factors but also for the random-effect factors, such as stimuli and/or participants.
3. RESULTS

The first part of this section is dedicated to base stimuli, and the second part to manipulated stimuli. For both types of stimuli, we summarize the results of correct identification rate reported in Schwab and Llisterri (to appear) and in Alfano et al. (2010) on the same participants described in § 2.1, and we present the results of reaction times for correct responses, i.e. when participants accurately perceived the intended stress position.

3.1. Base stimuli

As far as percent correct identification of the base stimuli is concerned, Schwab and Llisterri (to appear) and Alfano et al. (2010) reported in the first place that French speakers correctly perceive stress in 71.5% of the cases, suggesting that they might not be so deaf to stress as it has been assumed. Secondly, the advanced learners of Spanish perceive stress more accurately than those with no knowledge of the language, indicating that the exposure to L2 makes French speakers more sensitive to stress. Thirdly, whatever the competence in L2 might be, stress on the first syllable is better perceived than stress on the second syllable, that is in turn better identified than stress on the third syllable. Finally, as for lexical status, no effect and no interactions with other variables were found.

Reaction times were analyzed by means of a mixed-effects model with participant and stimulus as random effects and reaction time as the dependent variable was run on correct responses. The predictors were group (advanced/no knowledge), lexical status (word/pseudoword) and pattern (PP, P, O). Following Baayen (2008), residuals larger than 2.5 times the standard deviation (5 data points out of 334, forming 1.5% of the data) were considered outliers and removed. The model refitted without these data points showed an effect of group, lexical status and pattern. Regarding group (see Fig. 1), advanced participants present shorter reaction times than participants with no knowledge of Spanish ($\beta = -289.37$, $t = 2.568$, $p < 0.05$). As far as pattern is concerned (see Fig. 2), PP pattern shows shorter reaction times in comparison with P pattern ($\beta = -318.42$, $t = -7.425$, $p < 0.001$) and O pattern ($\beta = -411.89$, $t = -8.967$, $p < 0.001$), and reaction times are marginally shorter for P pattern than for O pattern ($\beta = -93.47$, $t = -1.949$, $p = 0.052$). Finally, as for lexical status (see Fig. 3), reaction times are shorter for words than for pseudowords ($\beta = -114.82$, $t = 3.161$, $p < 0.01$). No interaction modulates these effects.

Figures 1, 2 and 3: Reaction times (ms) for the base stimuli as a function of group (in Figure 1, on the left), as a function of pattern (in Figure 2, at the center), and as a function of lexical status (in Figure 3, on the right).

3.2. Manipulated stimuli

The most relevant results reported by Schwab and Llisterri (to appear) and Alfano et al. (2010) concerning the percent correct identification of the manipulated stimuli are the following: firstly, a combined manipulation of $F_0$, duration and amplitude leads to a better perception of the accentual shift than the separate manipulation of each acoustic parameter. This suggests that stress is perceptually not defined by a single parameter, but by the combination of parameters. Secondly, and more interestingly, both groups of French speakers (advanced and with no knowledge) don’t behave in the same way according to the different acoustic manipulations. On the one hand, the advanced group perceives better the accentual shift when the three parameters ($F_0$, amplitude and duration) are jointly manipulated. On the other hand, while both groups are equally sensitive to the manipulation of $F_0$ (in isolation or in combination with amplitude of duration), the group with no knowledge of Spanish is more sensitive to the manipulations of duration or amplitude (in isolation or combined). It appears thus that French speakers with no knowledge of the L2 process stress in a
more acoustic way. Thirdly, $F_0$ (alone or combined with duration or amplitude) seems the most important cue for a syllable to be perceived as stressed by French speakers. Indeed, researches in French (Rigault 1962; Dahan and Bernard 1996) have shown that $F_0$ is the decisive parameter in the perception of prominences in French L1.

As far as reaction times are concerned, a mixed-effects model with participant and stimulus as random effects and reaction time as the dependent variable was run on correct responses (i.e. when the participants perceived the intended stress position on the second syllable in PP>P manipulations and on the last syllable in P>O manipulations). The predictors were group, lexical status, pattern and manipulation. Since manipulation was the only significant fixed effect and no interaction modulated this effect, we ran different mixed-effects models in order to examine in detail the time cost induced by the manipulations in comparison with the base stimuli. In other words, instead of considering the whole set of manipulations, we performed separate analysis on subsets of manipulations.

First of all, we considered the subset of base stimuli and the stimuli in which the three parameters ($F_0$, duration and amplitude) were manipulated. A mixed-effects model with participant and stimulus as random effects, reaction time as the dependent variable and manipulation as predictor (base vs. F0_Dur_Ampl) shows no effect of the manipulation, indicating that the modification of the three parameters does not slow down the identification of stress, when stress is accurately identified. This suggests that the result of the acoustic manipulation was natural enough not to be noted by the listeners. Then, we examined different subsets of manipulated stimuli in comparison with base stimuli. Each subset was composed of a pair of complementary manipulations, in the sense that one manipulation concerns only one parameter in isolation (e.g. amplitude), while the other takes simultaneously into account the other two parameters (e.g. $F_0$ and duration). In this way, we can also observe the role of the parameters which were not manipulated (e.g. amplitude in the case of a combined manipulation of duration and $F_0$). The base stimuli were the third component of each subset.

The first subset of complementary manipulations we looked at was composed of stimuli with an isolated manipulation of $F_0$, stimuli with a combined manipulation of duration and amplitude, and base stimuli. A mixed-effects model shows an effect of the manipulation, as can be seen in Fig. 4: reaction times are longer in $F_0$ manipulated stimuli than in base stimuli ($\beta = 119.78, t = 2.310, p < 0.05$) and than in stimuli with a combined manipulation of duration and amplitude ($\beta = 143.2, t = 2.396, p < 0.05$), whereas there is no significant difference between the base stimuli and the stimuli simultaneously manipulated in duration and amplitude ($\beta = 23.46, t = 0.447, p = 0.66$). These results suggest the presence of an inhibitory effect of $F_0$ in the perception of stress, while the combined manipulation of duration and amplitude does not slow down stress perception.

Figures 4 and 5: Reaction times (ms) as a function of manipulation (in Figure 4, on the left: base, isolated manipulation of $F_0$ and combined manipulation of duration and amplitude; in Figure 5, on the right: base, isolated manipulation of amplitude and combined manipulation of $F_0$ and duration).

The second subset of complementary manipulations consisted of stimuli with an isolated manipulation of amplitude, stimuli with a combined manipulation of $F_0$ and duration, and base stimuli. A mixed-effects model again shows an effect of the manipulation: despite the difference we can observe in Fig. 5 between base stimuli and stimuli manipulated in amplitude, reaction times in stimuli manipulated in amplitude don't differ from base stimuli ($\beta = 70.25, t = 1.115, p = 0.27$) nor do they from stimuli manipulated simultaneously in $F_0$ and duration ($\beta = 37.96, t = 0.597, p = 0.55$). On the other hand, reaction times in stimuli manipulated
simultaneously in F₀ and duration are longer than in base stimuli ($\beta = 108.20$, $t = 2.567$, $p < 0.05$). These results seem to indicate the presence of an inhibitory effect of the combined manipulation of F₀ and duration, whereas the isolated manipulation of amplitude does not appear to slow down stress perception.

Finally, the third subset of complementary manipulations was formed of stimuli with the isolated manipulation of duration, stimuli with the combined manipulation of F₀ and amplitude, and base stimuli. A mixed-effects model shows no effect of manipulation, meaning that a manipulation of duration as well as a combined manipulation of F₀ and amplitude does not inhibit stress perception.

4. GENERAL DISCUSSION

We summarize and discuss here the results of stress identification rate (Schwab and Llisterri to appear; Alfano et al. 2010) and the results of reaction times, for base stimuli as well as for manipulated stimuli. As far as base stimuli—which didn't undergo any acoustic changes—are concerned, advanced learners of Spanish do not only perceive stress more accurately than participants with no knowledge of the language, but they are also faster in correctly identifying stress position. This strongly confirms that the exposure to Spanish as an L2 makes the French speakers more sensitive to stress. Secondly, whatever the competence in L2 might be, stress in the first syllable is detected more accurately and quicker than stress in the second syllable, which is, in turn, better and more rapidly identified than stress on the final syllable. Faster reaction times for PP pattern in comparison with P and O patterns are easily accounted for, given the fact that accentual information appears sooner in the PP stimuli than in the P and O stimuli. Thirdly, independently of the competence in L2 and the pattern, stress is perceived as accurately in words as in pseudowords, but when correctly perceived, stress in words is faster identified than in pseudowords. This discrepancy between the similarity in the identification of stress position in words and pseudowords and the difference in the time needed to process stress in words and pseudowords deserves further research.

Regarding manipulated stimuli, Schwab and Llisterri (to appear) and Alfano et al. (2010) reported that listeners are sensitive to different acoustic cues according to their competence in L2. This conclusion is not supported by results in reaction times, as there is no interaction between competence in L2 and the type of manipulation. In other words, when listeners are able to correctly perceive the stress shift induced by the different acoustic manipulations, advanced learners of Spanish and listeners with no knowledge of the language take the same amount of time to detect stress shift.

Schwab and Llisterri (to appear) and Alfano et al. (2010) also found that a manipulation of F₀ (alone or in combination with other parameters) leads to a more accurate perception of stress shift. When reaction times are considered, the picture is somehow different. We examined pairs of complementary manipulations compared to base stimuli in order to study, in terms of time cost, the role of the accentual information which is shifted to the next syllable (e.g. duration and F₀), and the role of the accentual information which is maintained as in the base stimuli (e.g. amplitude in stimuli manipulated in duration and F₀). We found first that a manipulation of F₀ has an inhibitory effect in the perception of stress, meaning that, when listeners perceive the stress shift induced only by F₀, reaction times are longer than for base stimuli. One possible interpretation is that the unmodified parameters (duration and amplitude) on the originally stressed syllable are strong enough to generate a conflict with the F₀ information on the next syllable. Consequently, the conflict between the original accentual information (duration and amplitude in the originally stressed syllable) and the shifted accentual information (F₀ on the next syllable) slows down the perception of stress in comparison with base stimuli. As far as the combined manipulation of duration and amplitude is concerned, results show no inhibitory effect. In that case, the original accentual information (F₀ in the originally stressed syllable) does not strongly conflict with the shifted accentual information (duration and amplitude on the next syllable), and does not slow down stress perception. This seems to suggest that the combination of duration and amplitude has more weight than F₀ in the perception of stress, given that, when they are maintained in the original syllable, there is a time cost, while there is none when shifted.

Secondly, results showed that a combined manipulation of F₀ and duration has an inhibitory effect in stress perception: when listeners perceive the stress shift induced by F₀ and duration, reaction times are longer than for base stimuli. On the other hand, the isolated manipulation of amplitude has no inhibitory effect. It seems thus that, in the first case, amplitude on the originally stressed syllable is strong enough to conflict with the F₀ and duration information on the next syllable, whereas, in the second case, the original accentual information (F₀ and duration in the originally stressed syllable) does not generate any conflict with the shifted accentual information (amplitude on the next syllable). It appears thus that amplitude has a
stronger weight than the combined parameters of $F_0$ and duration, since, when it is preserved in the original syllable, there is a time cost, whereas there is none when shifted.

Finally, results indicated that neither the isolated manipulation of duration nor the joined manipulation of $F_0$ and amplitude has an inhibitory effect in stress perception: when listeners detect the stress shift, none of the two manipulations slows them down in comparison with base stimuli.

Schwab and Llisterri (to appear) and Alfano et al. (2010) concluded that $F_0$ is the most relevant parameter for the identification of stress position. Considering only cases in which listeners perceive stress shift, we found that the time needed to correctly detect stress is related to amplitude (alone or combined with duration): an unmodified amplitude implies a longer detection time of the shifted intended stress, while a shift in amplitude values does not. Taken together, these results do not minimize the role of $F_0$ in Spanish stress identification by French speakers (i.e. stress position is better identified when a change in $F_0$ is involved), but they draw attention to the effect of amplitude on the time necessary to perceive stress, when its position is correctly identified. Therefore, this research highlights the need to combine identification rate and time cost to get a more coherent picture of the processes involved in the perception of stress in an L2.

5. REFERENCES


NOTES

1. $\beta$ corresponds to the estimated coefficients in a regression model (i.e. the increase per unit change in x).

2. All mixed-effects models carried out on subsets of manipulations were run with participant and stimulus as random effects, reaction time as the dependent variable and manipulation as predictor. Moreover, residuals larger than 2.5 times the standard deviation were considered outliers and removed (Baayen 2008).