Pitch scaling and the perception of contrastive focus in L1 and L2 Spanish

Conference Paper · June 2016

DOI: 10.21437/SpeechProsody.2016-121

2 authors:

Covadonga Sánchez Alvarado

Meghan E. Armstrong
University of Massachusetts Amherst

Pitch scaling and the perception of contrastive focus in L1 and L2 Spanish

Covadonga Sánchez-Alvarado[1], *Meghan E. Armstrong*[1]

[1] University of Massachusetts, Amherst (USA)

sancheza@umass.edu, armstrong@umass.edu

Abstract

The pitch accent associated with contrast in English and in Spanish is labeled as L+H* in their respective ToBI labeling systems. Nevertheless, the phonetic implementation of these categories differs, since a wider pitch range is needed in American English to consider a rising tonal movement as L+H*. This study explores the differences between American English and Peninsular Spanish in their use of F0 scaling as a cue to perceiving contrast or lack thereof in nuclear position. Following the predictions of the L2 intonation learning theory (LILt), the hypothesis guiding this study was that Spanish speakers would perceive contrast within a more compressed pitch range than learners of Spanish with American English as their L1. Fourteen native speakers of Spanish and fourteen learners enrolled in a Spanish phonetics class at a U.S. university participated in a forced-choice identification task. The stimuli consisted of manipulations of two utterances produced by native speakers of Peninsular Spanish, expressing contrastive focus; tonal landmarks outside of the focused word were neutralized, and seven-step continua were created manipulating only the height of the peak in the nuclear accent. The results indicate that pitch range is used similarly by native speakers and learners of Spanish.

Index Terms: L2 Prosody, Contrastive Focus, Pitch scaling, Spanish

1. Introduction

The Biological Codes proposed by [1], more precisely the Effort Code, predict that new information will be accompanied by wider pitch excursions (among other prosodic cues) as a way to highlight the importance of the message. Many languages have grammaticalized the Effort Code, showing a rising pitch movement to a high target to intonationally mark contrast. Phonologically, focal pitch accents in English and Spanish are transcribed within the Autosegmental Metrical framework as L+H* [2,3,4,5]. Nonetheless, the phonetic implementation of said category is not necessarily identical; the pitch range required for a tonal movement to be perceived as L+H* in Spanish is not necessarily the same as the one needed for English. In English, contrastive and non-contrastive meanings are encoded by the use of two different pitch accents, L+H* and H*, respectively [2]. Phonetically, the former category is characterized by a wider pitch range excursion while the latter is associated with a simple rise to a maximum in F0 [2]. [6] used a force-choice task to show that the most relevant cue used by American-English speakers to perceive focal prominence is in fact changes in F0, although other prosodic cues such as intensity and duration provide additional information. L+H* is the pitch accent most commonly associated with focal prominence in Spanish [4,5]. [7] explored the realization and perception of contrastive accents in pre-nuclear position. They conclude that, in Spanish, early alignment (as opposed to the late peaks that characterize non-focal accents) and optional longer duration are the main prosodic cues used in production. In perception, however, Spanish speakers appear to make use of as many prosodic features as possible (i.e. alignment, duration, peak height); if not all of them are present, the most systematically used ones are alignment and duration [7]. It is important to note, however, that L+H* can appear in nuclear position in both narrow and broad focus statements, while L* is exclusively used for broad focus.

Apart from differences in phonological categories, languages have been shown to exploit pitch range in different ways [8,9,10]. Little is known about cross-linguistic differences between English and Spanish in this respect, though studies have shown that in general, English shows a wider pitch range when compared to Spanish [10,11,12,13]. These differences may result in the perception of different paralinguistic meanings. [10] shows that a wider pitch range is associated with politeness in English and over-excitement in Spanish; conversely, a narrow pitch range sounds polite in Spanish but rude in English. Regarding the linguistic meanings that may be encoded through prosody, and more precisely the use of pitch scaling for the purpose of conveying narrow focus, [13] shows that British English speakers use a much wider pitch range when compared to speakers of Peninsular Spanish from the Northern region of Spain.

The similarities and differences between American English and Peninsular Spanish are of interest from an L2 acquisition standpoint. The pitch category L+H* in American English and Peninsular Spanish could very well have quite different phonetic implementations, presenting a challenge for learners. Based on the assumptions of the recently proposed L2 intonation learning theory (LILt) [14], the phonological categories and prosodic parameters present in the L1 constitute the departure point in the development of the L2 intonational grammar. Dissimilarities between the L1 and the L2 may be found along four dimensions: the systemic dimension (if the categories and their distribution differ); the realizational dimension (if what differs is the phonetic implementation); the semantic dimension (if the same category conveys a different meaning); and the frequency dimension (if the same category is used more frequently in one language than in the other). The predictions that can be established from this theory require that a contrastive analysis of the languages under study be established. Given the description about the realization of contrastive focus in English and Spanish presented above, the differences rely on the realizational dimension (the use of a wider pitch range to produce a L+H* in English as compared to Spanish), which in turn results in differences in the semantic dimension (the use of a narrow pitch range will block the perception of a contrastive meaning). While multiple studies have shown how the L1 constraints the perception of prosody in the L2 [15,16,17], fewer studies have looked into the perception of contrastive focus ([18] for Vietnamese learners of English; [19] for Spanish L2 speakers of English]. In these studies, an effect of the L1 was also found, with learners failing to use the prosodic cues that are not available in their native language to convey contrast. To the best of our knowledge, the perception of contrastive focus by L2 learners of Spanish has not been explored yet. Below, we present the results of a perception study investigating how the differences we described above might affect L2 perception of intonational categories.

2. Present study

The purpose of the present study is to examine the role that pitch scaling plays in the perception of contrastive accents in nuclear position in English and in Spanish, particularly for the final rising pitch accent that would be labeled L+H* in either ToBI system. This will provide a greater understanding of the cross-linguistic differences in the use of pitch scaling as well as the influence of the L1 into the L2 in the interpretation of linguistic meanings such as contrast. The research question guiding the study is: do L2 speakers of Spanish (with L1 American English) need the L+H* pitch accent to be realized within a wider pitch range in order to perceive it as a contrastive focus marker in Spanish? Following the premises of the LILt, we expected learners of Spanish to rely on the cues that are most relevant in English. We assume that in order to make the distinction between H* and L+H* in English, the latter category must be produced with a steep enough rise for listeners to perceive a contrast, and thus a wider pitch range must be used for American English speakers to perceive L+H*. Following similar findings regarding the need for wider/narrower pitch ranges in order for learners to perceive specific meanings (namely attitudinal) in their L2 as a result of the phonetic realization in the L1 [10,15], we predicted that differently from American English listeners, listeners whose L1 was Peninsular Spanish would be able to perceive contrast for tokens of L+H* produced in a more compressed pitch range when compared to learners of Spanish with American English as their L1. A forced-choice identification experiment was designed in order to test this hypothesis.

3. Methodology

3.1. Participants

Two groups of participants were included in the study: a group of native speakers of Peninsular Spanish recruited using social media, and a group of Spanish learners enrolled in a Spanish Phonetics class at a U.S. university in the Northeast region of the country. The first group consisted of 14 subjects from different areas of Spain while the second consisted of 14 learners, whose native language was American English. The mean age of the learners was 20.35 (19-21); they had been studying Spanish for an average time of 8.87 years (5-14) at the time of the study. Out of those fourteen students, eleven responded that they had studied abroad at some point (ten of them did so in Spain).

3.2. Materials

The stimuli were created from a set of recordings obtained from two females, both native speakers of Peninsular Spanish, who produced a list of utterances conveying broad focus and narrow contrastive focus elicited through question-answer pairs. All these utterances contained transitive verbs whose objects were paroxytone three-syllable words composed of sonorant segments. The utterances with contrastive focus were presented to ten native speakers, who were asked to judge whether the speakers sounded as if they were making a correction or not. Out of the utterances that were most consistently perceived as contrastive, one was chosen from each speaker and taken as the starting point for the creation of the stimuli: *Mira enanos* ('S/he watches dwarfs'), as produced by speaker S1, and *Escucha baladas* ('S/he listens to ballads'), as produced by speaker S2.

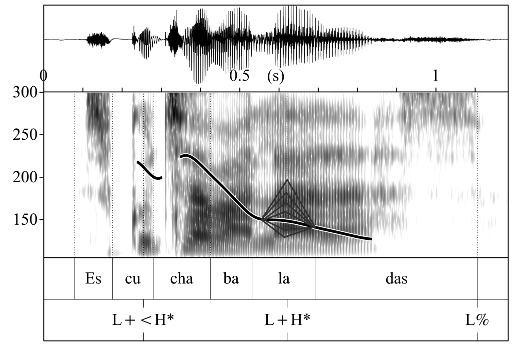

Figure 1: *Stimuli 'She listens to ballads' and schematic representation of the 7-step manipulation.*

In order to reduce the information that could be obtained from prosodic features in parts other than the nuclear accent, three points in the utterance were neutralized: initial F0, the first peak of the first H target, and final L target. The values at these three positions were modified to match the mean values between the utterance with broad focus and the utterance with contrastive focus. Once the neutralization was completed, the height of the peak in the nuclear accent was manipulated. Seven steps were created, with a difference of 12Hz between steps: two steps below the starting point, and four more steps above it, as shown in *Figure 1*. As in previous studies [7], some differences in terms of the phonetic implementation of contrast were found, as a result of individual variation between the two speakers, and they may have had an effect on the results, as discussed below.

3.3. Procedure

Participants were first asked to complete a background questionnaire, which differed based on the group they belonged to (i.e. Spanish native speakers or Spanish learners). They were then asked to read the instructions, which were presented in each group's native language. Participants were presented with two transcribed conversations, one in which there was no contrast (e.g. 'What happened?' 'John saw Bill'), and one in which there was (e.g. 'You saw Bill?' 'I saw JOHN'). They were told that, in the experiment, they would listen to those types of responses out of context, and they would have to determine whether the speaker was correcting the hearer or not. In addition to reading the instructions, the group of Spanish learners was asked to listen to an extract of a conversation in Spanish, taken from a Spanish TV show, so as to set them in "Spanish mode" [20] before they started the experiment. A reminder of the instructions, this time in Spanish, and a practice item, followed the video.

The stimuli were presented using an online web survey platform. Participants were instructed to use headphones and to respond as quickly as possible. The stimuli were divided into different blocks, one for each speaker, and each block included 5 repetitions of each of the seven steps. The items within each repetition were randomized. In total, the experiment consisted of 70 stimuli (7 steps x 5 repetitions x 2 blocks). The experiment took approximately 10 minutes to be completed.

3.4. Analysis

A numerical value was assigned to each response. Non-contrastive responses were coded as "0" and contrastive responses were coded as "1". A three-way analysis of variance (ANOVA) was conducted in order to determine the effect of the different variables. The dependent variable was RESPONSE (or contrast identification score); STEP (with seven levels), GROUP (two levels: native and learner), and SPEAKER (two levels: S1 and S2), were the independent variables and PARTICIPANT was the random effect. For the significant main effects and interactions, pair-wise comparisons were done with t-tests using the Holm adjustment.

4. Results

The results from the ANOVA show a significant main effect of SPEAKER ($F(1,1)= 18.547$, $p<0.001$) and three significant interactions: STEP*GROUP ($F(1,1)= 5.5$, $p<0.05$), STEP*SPEAKER ($F(1,1)=40.3$, $p<0.001$, and STEP*GROUP*SPEAKER ($F(1,1)=7.06$, $P<0.01$). Given the differences between speakers, the results for each utterance will be discussed independently first. Then, a comparison between both sets of stimuli will be established.

Figure 2 below shows the mean contrast identification rates for the stimuli produced by speaker S1. In this case, the scores assigned by the two groups follow a similar trend in that the increase in the scores is parallel and the steps at each end of the continuum are not categorically classified as non-contrastive or contrastive, respectively. Despite this, significant differences between the scores assigned to the different steps were still found, as reported in *Table 2*. For both groups, steps 1, 2 and 3 are significantly different from steps 5, 6 and 7. Step 4 would be almost in between these two clusters, but closer to 1-3, since the mean contrast identification rate is below 50%. Scores for step 5 are slightly above chance, preventing us from concluding that this is the point at which the threshold is. The mean identification rate continues to increase with the stimuli at steps 6 and 7, but not significantly. No significant differences were found between groups at any of the steps, as shown in *Table 4*.

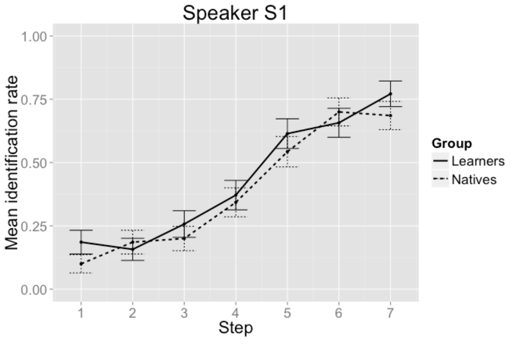

Figure 2: *Mean contrast identification rate for Speaker S1*

	1	2	3	4	5	6
2	1.00					
	0.73					
3	1.00	0.76				
	0.69	1.00				
4	0.10	0.03*	0.76			
	0.01**	0.26	0.31			
5	<0.001***	<0.001***	<0.001***	0.012*		
	<0.001***	<0.001***	<0.001***	0.06		
6	<0.001***	<0.001***	<0.001***	0.0016**	1.00	
	<0.001***	<0.001***	<0.001***	<0.001***	0.26	
7	<0.001***	<0.001***	<0.001***	<0.001***	0.25	0.76
	<0.001***	<0.001***	<0.001***	<0.001***	0.31	1.00

Table 2. *p-values of statistical comparisons of the contrast identification rates between steps for learners (shaded) and native speakers for the stimuli produced by speaker S1.*

Figure 3 presents the mean contrast identification rates for the stimuli produced by speaker S2. In this case, a more clear dissimilarity can be established between the stimuli at both ends of the continuum. As was the case with the stimuli produced by S1, there is a significant difference between steps 1, 2, and 3 on the one hand, and steps 5, 6 and 7, on the other (see *Table 3*). Furthermore, step 4 is significantly different from all others and its scores are at chance for both groups. Despite the fact that the stimuli are grouped similarly for both groups, there were significant differences in the mean identification rates of contrast by each group, as shown in *Table 4* below. Learners were able to more categorically identify the stimuli as non-contrastive or contrastive, and there were significant differences in their judgments at steps 2 ($p<.01$), 6 ($p<.001$) and 7 ($p<.05$). For native speakers, the most categorical distinction is only found with the stimuli at both ends of the continuum.

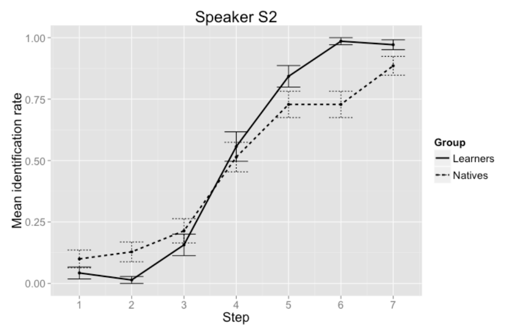

Figure 3: Mean contrast identification rate for Speaker S2

	1	2	3	4	5	6
2	1.00					
	1.00					
3	0.07	0.028*				
	0.37	0.62				
4	<0.001***	<0.001***	<0.001***			
	<0.001***	<0.001***	<0.001***			
5	<0.001***	<0.001***	<0.001***	<0.001***		
	<0.001***	<0.001***	<0.001***	0.014*		
6	<0.001***	<0.001***	<0.001***	<0.001***	0.028*	
	<0.001***	<0.001***	<0.001***	0.014*	1.00	
7	<0.001***	<0.001***	<0.001***	<0.001***	0.043*	1.00
	<0.001***	<0.001***	<0.001***	<0.001***	0.12	0.12

Table 3. *p-values of statistical comparisons of the contrast identification rates between steps for learners (shaded) and native speakers for the stimuli produced by speaker S2.*

The results from the ANOVA had also pointed out the effect of SPEAKER. The differences between the stimuli produced by speaker S1 and speaker S2 resulted in the assignment of significantly different scores by both groups. As shown in *Table 4*, the stimuli that constituted the baseline (step 3) were not judged in a significantly different manner. For most of the remaining stimuli, however, there were differences in the ratings assigned to them. The stimuli produced by S2 seem to have had a facilitative effect for learners, as their rate of identification of a contrastive meaning was significantly lower for the stimuli at steps 1 and 2, and significantly higher for the stimuli at steps 5, 6 and 7. For native speakers, on the other hand, this facilitative effect was only found at steps 4, 5 and 7, while the stimuli at the lower end of the continuum were rated similarly.

1	2	3	4	5	6	7
0.007 **↓	0.0023 **↓	0.15	0.028	0.002 **↑	<0.001 ***↑	<0.001 ***↑
1	0.36	0.84	0.041 *↑	0.022 *↑	0.71	0.0037 **↑

Table 4. *p-values of statistical comparisons of contrast identification rates (lower ↓ or higher ↑) for the stimuli of speaker S2 as compared to speaker S1 at each step for learners (shaded) and native speakers.*

5. Discussion

The present study aimed to determine the relevance of pitch scaling in the perception of contrastive focus in American English and Peninsular Spanish by looking at the performance of Spanish learners and Spanish native speakers. The hypothesis guiding the study, which stated that native speakers of Spanish would perceive contrast within a more compressed pitch range than learners of Spanish with American English as their L1, was not confirmed. Overall, native speakers and learners followed similar trends, finding ambiguity in the stimuli located in the middle of the continuum and assigning the higher contrast identification scores to those at the higher ends of the scale. The results will be discussed in light of the individual differences in the phonetic implementation of the stimuli.

Pitch scaling alone only seemed to be used as a determining cue by native speakers and Spanish learners with the stimuli produced by S2. These utterances were characterized as having earlier peaks than those produced by S1 (42 ms earlier), as well as a lower peak in the pre-nuclear accent (281 Hz as opposed to 363 Hz). Such conditions seem to have set up the perfect environment for participants to use pitch scaling as the only cue necessary to distinguish contrastive from non-contrastive statements. The responses provided by native speakers were not as categorical as those provided by learners. It is important to remember, however, that the use of L+H* in Spanish is not exclusively associated with narrow focus readings [5], leading to ambiguity between what could be considered as contrastive and what is not. Spanish speakers may employ other prosodic cues optionally, and other factors, such as the use of discourse context may aid in the disambiguation between both meanings.

There were differences in the realization of focal and pre-focal material between the two stimuli which may explain the differences in perception reported above. In the utterances with contrastive focus produced by the two Spanish speakers, the difference between the pre-nuclear and the nuclear peaks was 167 Hz for S1 and 90.5 Hz for S2 in the utterances with contrastive focus. In the utterances with broad focus, the differences between the height of the pre-nuclear and the nuclear peak was 50 Hz for speaker S1 and 70 Hz for speaker S2, and therefore lower than in the utterance with contrast. It seems that since the pre-nuclear H target was lowered in the stimuli used in the present task, the prosodic realization within the focal accent that coincided with the original one (step 3) was not enough to cue the perception of contrast. If pre-nuclear material did not have an effect on the interpretation of contrast, the stimuli at step 3 should have been perceived as contrastive more than 50% of the time by natives, but this was not the case. These results thus call for the need to explore the role played by other prosodic features associated not only with the focused word, but also with pre-focal material, as well as the relationship between the two. In this regard, Spanish speakers seem to use a wider pitch range in pre-nuclear peaks as a way of "announcing" that the material that is about to come is important. This differs from the strategy used in English, which consists on the use of lower peaks in pre-focal material than in nuclear pitch accents [18]. This use of prosody in the pre-nuclear regions of an utterance could be further tested using gating tasks.

Given the findings reported above, some lingering questions are: would the scores have been the same had the focused words been isolated? What if other prosodic cues (e.g. alignment, intensity or duration) had been modified as well within the focused word? [7] showed that in pre-nuclear position, pitch scaling was not the most relevant cue, while duration and alignment were sufficient to establish contrast. Would these cues play a similar role in nuclear position? And if so, what difference would it make for L2 learners? We also consider the possibility that the group of learners included in this study is already aware of the fact that Spanish uses a narrower pitch range, given that the majority of the learners have had some kind of study-abroad experience in Spain. In this vein, it would be interesting to explore whether learners who lack this experience show a similar behavior. Furthermore, the difference in the role played by pitch scaling in production of L1 and L2 Spanish by learners at different proficiency levels should also reveal whether a cross-linguistic difference between English and Spanish does in fact exist regarding the use of pitch scaling in the expression of contrast.

6. Conclusions

To conclude, this study has shown that despite the differences between English and Spanish in the use of pitch scaling to signal contrast in nuclear position, native speakers and L2 learners of Spanish use it similarly in perception. As a result of their exposure to Spanish, learners at this proficiency level may be already aware of the differences between English and Spanish at the realizational and semantic dimension when it comes to the perception of contrastive focus [14]. The role of proficiency and study-abroad experience should then be further explored in order to better understand the development of the L2 intonational grammar. Additionally, the results presented in this study also suggest that the role of other prosodic features within focused words (i.e. duration and alignment) as well as in pre-focal material must be taken into consideration and further explored, as these also differ across the two languages under study. Thus, the specific nature of all these prosodic cues should be further explored, not only in perception, but also in production.

7. Acknowledgements

We are thankful to UMass undergraduate researchers Helen Zhang and Alaina Petrullo for their work on this project.

8. References

[1] Gussenhoven, C. (2004). *The Phonology of Tone and Intonation*. Cambridge Univ Press.

[2] Beckman, M. E., & Ayers, E. G. (1997). "Guidelines for ToBI Labelling Preface," *What are the "Guidelines for ToBI Labelling"?* (March), 1–43.

[3] Arvaniti, A., & Garding, G. (2004). "Dialectal variation in the rising accents of American English," *Papers in Laboratory Phonology IX: Change in Phonology,* vol. 9, pp. 547-576

[4] Prieto, P., & Roseano, P. (2010). *Transcription of Intonation of the Spanish Language*. Munich: Lincom Europa.

[5] Hualde, J. I., & Prieto, P. (2014). "Intonational Variation in Spanish: European and American varieties," In S. Frota & P. Prieto (Eds.), *Intonational variation in Romance*. Oxford University Press.

[6] Mahrt, T., & Cole, J. (2014). "The Contribution of Individual Acoustic Cues to the Perception of Focal Prominence," In *Laboratory Phonology 14*.

[7] Vanrell, M. D. M., Stella, A., Fivela, B. G., & Prieto, P. (2013). "Prosodic manifestations of the Effort Code in Catalan, Italian and Spanish contrastive focus," *Journal of the International Phonetic Association*, vol. 43, no. 02, pp. 195–220.

[8] Chen, A. (2005). *Universal and language-specific perception of paralinguistic intonational meaning*. Utrech: LOT.

[9] Mennen, I. (2007). "Phonological and phonetic influences in non-native intonation," *Trends in linguistics studies and monographs*, vol. 186, no. 53.

[10] Estebas-Villaplana, E. (2014) "The evaluation of intonation. Pitch range differences in English and in Spanish," In Thompson, G. & Alba-Juez, L. (Eds.), *Evaluation in Context*. John Benjamins Publishing Company, Amsterdam.

[11] Backman, N. (1987). "Intonation Errors in Second-Language Pronunciation of Eight Spanish-Speaking Adults Learning English," *Interlanguage Studies Bulletin*, vol. 4, no. 2, pp. 239–265.

[12] Fernald, A., Taeschner, T., Dunn, J., Papousek, M., de Boysson-Bardies, B., & Fukui, I. (1989). "A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants," *Journal of child language*, vol. 16, no. 03, pp. 477-501.

[13] García-Lecumberri, M. L. (1995). *Intonational signalling of information structure in English and Spanish: a comparative study*. University of London.

[14] Mennen, I. (2015). Beyond segments: Towards an L2 intonation theory. In Delais-Roussarie, E., M. Avanzi and S. Herment. *Prosody and Languages in Contact*. Singapore: Springer.

[15] Chen, A. (2009). "Perception of paralinguistic intonational meaning in a Second Language," *Language Learning*, vol. 59, no. 2, pp. 367–409.

[16] Cruz-Ferreira, M. (1987). "Non-native interpretive strategies for intonational meaning: an experimental study," *Sound Patterns in Second Language Acquisition*, pp. 10–39.

[17] Zárate-Sández, G. A. (2015). *Perception and Production of Intonation among English-Spanish Bilingual Speakers at Different Proficiency Levels*. Georgetown University.

[18] Nguyễn, T. A. T., Ingram, C. J., & Pensalfini, J. R. (2008). Prosodic transfer in Vietnamese acquisition of English contrastive stress patterns. *Journal of Phonetics*, 36(1), 158-190.

[19] Ortega-Llebaria, M., & Colantoni, L. (2014). L2 English Intonation. *Studies in Second Language Acquisition*, 36(02), 331-353.

[20] Grosjean, F. (1999). *The bilingual's language modes. One Mind, Two Languages: Bilingual Language Processing*, 1–22.